ADAPT THE MODE, NOT THE MODEL: TEST-TIME OPERATING-POINT SELECTION FOR HIERARCHICAL PORTFOLIO AGENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Trained hierarchical decision system often exposes several operating points, or *modes*, which modules to consult and in what combination allowing test-time adaptation without retraining or fine-tuning any module. Existing work on hierarchical and multi-agent systems typically argues that a specific configuration is superior *on average*, leaving a critical question unanswered: for a given environment, which mode should be deployed? We address this in portfolio construction using RASP (Regime-Aware Selection of operating Points), a three-tier agent whose operating points vary solely in their intermediate routing decision paths while sharing identical execution architectures. We show that the optimal operating point varies significantly across market regimes and universes, demonstrating that no single mode uniformly dominates across different environments or market drawdowns. To exploit this structure at test time, we propose an online selector based on a Hedge/fixed-share mixture over modes that observes only realized utility. Our framework effectively adapts to shifting market conditions without prior regime knowledge, outperforming both static mode assignments and standard follow-the-leader strategies. For hierarchical decision systems broadly, these results demonstrate the value of test-time mode selection over model retraining and emphasize the importance of identifying the precise conditions under which routing mechanisms provide an advantage.

1 INTRODUCTION

Hierarchical decision systems composed of specialized modules (forecasters, routers, aggregators, executors) are increasingly common wherever no single prediction metric captures the downstream objective (Donti et al., 2017; Elmachtoub & Grigas, 2022; Yu et al., 2024; Xiao et al., 2024). Such a system does not have one behavior; it has several *operating points*, depending on which modules are consulted and how their outputs are combined. Portfolio construction makes the stakes concrete: an accurate return forecaster can still produce poor risk-adjusted returns when paired with a naive allocator (Elmachtoub & Grigas, 2022), or when every available expert is consulted indiscriminately across market conditions. Whoever deploys such a system must commit to one operating point, and the literature offers little principled help in choosing. Recent hierarchical and multi-agent systems for portfolio decisions (Zhang et al., 2024; Yu et al., 2024; Xiao et al., 2024) are typically evaluated by showing that a proposed configuration beats baselines ***on average over one sample***. That is a useful claim, but it is not the deployment claim: averaged over environments, it says nothing about which configuration to use in the environment actually faced. Worse, when a configuration helps in one market structure and hurts in another, an average can be positive while the decision rule it implies is wrong half the time. What is missing is ***a rule for the environment actually faced***, a statement of the form *"use this mode when the environment looks like this."*

In this work we study *test-time operating-point selection*: adapting a trained hierarchy by choosing which of its fixed operating points, we also call *modes* to deploy, without retraining or fine-tuning any module, using portfolio construction as a demanding testbed as Figure 1 shows, middle panel shows why a single averaged choice is fragile, right panel how RASP replaces it with a gated online mixture). We instantiate the question in RASP (Regime-Aware Selection of operating Points), a three-tier

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

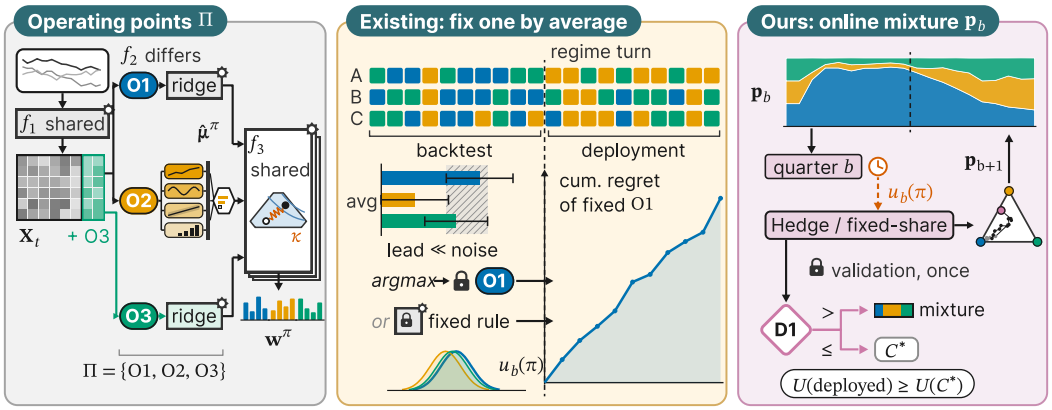

Figure 1: **Why test-time mode selection matters.** Unlike traditional approaches that fix a single operating point based on noisy historical averages, RASP dynamically adapts an online mixture over operating points using realized utility, deploying it via check regime shifts.

agent exposing three operating points that differ only in the Tier-2 path: **O1**, an optimizer-only path converting forecasts directly into allocations; **O2**, a routed consensus that scores expert messages by expected return penalized by uncertainty and service cost before aggregating them; and **O3**, an alpha-augmented path that widens the feature panel. Tier 1, the risk model, the universe and the rebalance calendar are shared, so the three are comparable interventions over one allocator rather than three unrelated strategies. This isolates two distinct ways of adding hierarchy: coordination through expert aggregation (O2) versus information expansion through features (O3).

This paper makes three contributions. ❶ **Routing value is universe-dependent.** No single operating point dominates across market conditions: each mode leads in different universes, demonstrating that fixing any static architecture incurs costly misallocation in real-world trading. ❷ **Test-time adaptation without retraining.** RASP dynamic selector updates mode allocations purely from realized utility via Hedge/fixed-share, backed by a validation gate. In production-like environments, it consistently matches or beats fixed baselines and follow-the-leader rules with lower turnover and stronger tail-risk resistance. ❸ **Robustness under realistic trading friction.** The advantages of test-time mode selection persist under live trading constraints, including non-zero transaction costs (0–50 bps) and long-horizon Stress testing across major market drawdowns. Shift the focus from chasing average backtest performance to dynamically matching live market regimes.

2 Related Work

Hierarchical multi-agent systems for portfolio decisions Layered agent systems coordinate specialized modules for market analysis and portfolio allocation (Zhang et al., 2024; Yu et al., 2024; Xiao et al., 2024). However, existing evaluations typically report a single configuration's average performance over a static backtest, masking the fact that optimal configurations shift across market regimes. Meta-policies that switch among fixed sub-policies address this partially (Niu et al., 2022; Zong et al., 2024), but lack evaluation against standard baselines under long-horizon trading frictions. We take the configuration space itself as the object of study: holding the feature pipeline and allocator, we evaluate test-time selection across discrete operating points using realized utility.

Portfolio decision-making under downstream objectives Classical portfolio theory focuses on mean–variance optimization and coherent risk measures under costs and constraints (Markowitz & Todd, 2000; Ahmadi-Javid, 2012; Kolm & Ritter, 2021). Recent decision-focused learning evaluates forecasts directly by downstream decision quality (Donti et al., 2017; Wilder et al., 2019; Elmachtoub & Grigas, 2022), with extensions incorporating LLM-generated views (Hwang et al., 2025) or regime dynamics (Hamilton, 1989; Bo et al., 2019). While decision-focused learning relies on end-to-end gradient propagation through a single pipeline, our work addresses the complementary problem of test-time selection among modular, discrete pipeline architectures.

Table 1: Taxonomy of related work vs. RASP. ✓ present, ○ partial, ✗ absent or not reported, – not applicable

Method	Sel.	Regime	Util.	Costs	Long OOS
Sparse MoE (Shazeer et al., 2017)	○	✗	✗	–	–
Hedge (Freund & Schapire, 1997)	✓	○	✓†	✗	–
Fixed-share (Herbster & Warmuth, 1998)	✓	○	✓†	✗	–
FinCon (Yu et al., 2024)	✗	✗	✓	✗	✗
TradingAgents (Xiao et al., 2024)	✗	✗	✗	✗	✗
MetaTrader (Niu et al., 2022)	✓	✓	✓	✗	✓
MacroHFT (Zong et al., 2024)	✓	✓	✓	✓	✗
FINSABER (Li et al., 2026)	✗	✗	–	✓	✓
RASP (ours)	✓	✓	✓	✓	✓

Sel.: Selects/weights whole policies. **Regime**: Conditioned on market state. **Util.**: Fitted directly to realized decision utility. **Costs**: Net of trading costs. **Long OOS**: $\geq$ 10-year test window. †Admissible for bounded loss.

Mixture-of-experts and hierarchical tool routing Mixture-of-experts (MoE) architectures perform sparse conditional routing using internal gating networks (Shazeer et al., 2017; Fedus et al., 2022; Liu et al., 2024). Unlike standard MoE routing, which optimizes a per-token prediction loss inside the model, our framework operates at deployment time: it dynamically selects among fully realized system operating points based on downstream portfolio utility under transaction costs via online prediction-with-expert-advice algorithms (Herbster & Warmuth, 1998).

Financial foundation agents and calibration Financial foundation models expand market representations by integrating multimodal data, tools, and textual context (Zhang et al., 2024; Shi et al., 2025; Luo et al., 2025), while uncertainty calibration ensures decision robustness (Nixon et al., 2019). These models primarily enhance individual signal quality; our work focuses on the overarching coordination layer, providing a solution to evaluate and select which signals to trust across regimes.

Sequential decision-making and evaluation Reinforcement learning and off-policy evaluation are foundational to sequential financial decision-making (Filos, 2019; Ye et al., 2020; Dudík et al., 2011; Swaminathan & Joachims, 2015; Jiang & Li, 2016). Because our operating points yield deterministic policy trajectories given market history, we employ rigorous walk-forward evaluation across multi-decade out-of-sample test windows to assess real-world generalization and stability.

3 BACKGROUND AND PRELIMINARIES

Portfolio decision setting We consider a discrete-time investment horizon $t = 1, \ldots, T$ over N tradable assets. Let $\mathbf{x}_t \in \mathbb{R}^p$ denote the observable market context and let $\mathbf{r}_{t+1} \in \mathbb{R}^N$ denote next-period returns. At each rebalance date, the system outputs portfolio weights $\mathbf{w}_t \in \mathbb{R}^N$ subject to

$$\mathcal{W}_t = \{\mathbf{w} \in \mathbb{R}^N : \mathbf{1}^\top \mathbf{w} = 1, \mathbf{w} \geq 0, \|\mathbf{w}\|_1 \leq L, w_i \leq u_i, \mathbf{A}\mathbf{w} \leq \mathbf{b}\}, \tag{1}$$

where $L > 0$ is a leverage budget, $u_i > 0$ is the cap for asset i, and $\mathbf{A} \in \mathbb{R}^{k \times N}$ and $\mathbf{b} \in \mathbb{R}^k$ encode optional linear exposure constraints, which no experiment imposes.

Two features of this set deserve flagging. The leverage budget is *redundant* in the long-only regime we run — $\mathbf{w} \geq 0$ with $\mathbf{1}^\top \mathbf{w} = 1$ gives $\|\mathbf{w}\|_1 \equiv 1$, so it can never bind — and is retained only for the long–short variant (Kahraman & Tookes, 2017); and turnover is penalized in ℓ_1 (Section 4) because under proportional costs the ℓ_1 distance *is* the cost, not a regularizer chosen for tractability (Lobo et al., 2007). Appendix A.1 expands both points.

Market frictions Trading costs and financing frictions are central to realistic portfolio choice (Gârleanu & Pedersen, 2013; Dai et al., 2010). We model them through a linear-plus-quadratic transaction-cost function

$$\mathrm{tc}_t(\mathbf{w}_t, \mathbf{w}_{t-1}) = \boldsymbol{\alpha}_t^\top |\mathbf{w}_t - \mathbf{w}_{t-1}| + \tfrac{1}{2}(\mathbf{w}_t - \mathbf{w}_{t-1})^\top \Lambda_t (\mathbf{w}_t - \mathbf{w}_{t-1}), \tag{2}$$

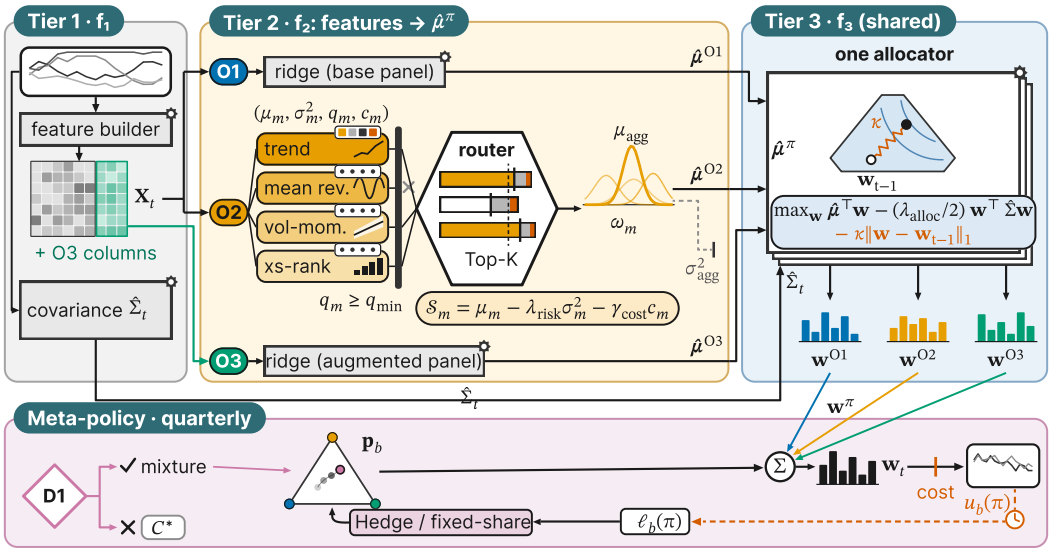

Figure 2: Overview of RASP. **Tier 1** generates market features and covariance matrices. **Tier 2** predicts asset returns through direct forecasting (O1), expert signal consensus (O2), and augmented forecasting (O3). **Tier 3** computes portfolio allocations for each mode using a shared allocator. **Meta-policy** adapts an online mixture over the operating points on utility via validation gate, implicitly tracking shifting market dynamics without requiring explicit state indicators.

where $\boldsymbol{\alpha}_t \in \mathbb{R}^N_+$ denotes per-asset proportional costs, $|\cdot|$ is applied elementwise, $\Lambda_t \in \mathbb{R}^{N\times N}$ with $\Lambda_t \succeq 0$ captures temporary impact, and $\mathbf{w}_{t-1}$ denotes the holdings carried into t, i.e. the previous weights after drifting with prices. In all experiments $\boldsymbol{\alpha}_t$ is a uniform one-way rate and $\Lambda_t = 0$. The realized net return is

$$R_{t+1}(\mathbf{w}_t) = \mathbf{w}_t^\top \mathbf{r}_{t+1} - \mathrm{tc}_t(\mathbf{w}_t, \mathbf{w}_{t-1}).$$

Costs depend on the previous portfolio, routing and allocation quality must be evaluated jointly rather than as isolated modules.

Risk-sensitive utility The allocator uses the mean–variance proxy

$$U_{\lambda_{\text{alloc}}}(\mathbf{w}_t) = \mathbb{E}[R_{t+1}(\mathbf{w}_t) \mid \mathbf{x}_t] - \tfrac{\lambda_{\text{alloc}}}{2} \mathrm{Var}[R_{t+1}(\mathbf{w}_t) \mid \mathbf{x}_t], \tag{3}$$

where $\mathbb{E}[R_{t+1}(\mathbf{w}_t) \mid \mathbf{x}_t]$ is the conditional expected net return, $\mathrm{Var}[R_{t+1}(\mathbf{w}_t) \mid \mathbf{x}_t]$ is the corresponding conditional variance, and $\lambda_{\text{alloc}} \geq 0$ controls the trade-off between return and risk. Routing and aggregation are defined using analogous utility-aware quantities; Section 4 instantiates them for the implemented hierarchy, the meta-policy scores operating points by the realized value of equation 3, and Appendix A summarizes the full notation.

4 METHODOLOGY

RASP is a three-tier hierarchical investment agent, shown in Figure 2. Tier 1 transforms raw market data into state features. Tier 2 generates expert messages, routes them with utility-aware criteria, and optionally aggregates them into a consensus forecast. Tier 3 converts the resulting signal into a feasible portfolio through a constraint-aware allocator (Sections 4.1–4.3). On top of the three tiers, a rolling meta-policy selects among the resulting operating points at test time (Section 4.4). We intentionally decouple routing, aggregation, and allocation rather than train a fully end-to-end differentiable stack, which preserves modularity and matches practical trading pipelines.

Operational definition of utility-calibrated routing. The router scores expert messages by expected return penalized by uncertainty and service cost at each weekly rebalance, and those routed signals are judged only by realized downstream portfolio utility on the evaluation window. We use λ_{risk} and γ_{cost} for routing penalties, and λ_{alloc} and κ for the allocator's variance and turnover penalties. Because the terms below are used throughout, we state them once and precisely.

Definition 4.1 (Hierarchy and operating point)**.** The *hierarchy* is the directed composition $\mathcal{H} = f_3 \circ f_2 \circ f_1$: f_1 builds causal features, f_2 maps features to a predicted mean (directly, or through expert messages that are routed and aggregated), and f_3 maps a predicted mean and covariance to feasible weights. It is the chain $f_1 \to f_2 \to f_3$, not an arrangement of components inside any one tier. An *operating point* π is a fixed choice of the f_2 path and feature set, holding f_1's construction rule and f_3 fixed; O1, O2 and O3 are therefore comparable interventions over one allocator.

Definition 4.2 (Expert and single routing)**.** An *expert* is a callable Tier-2 module that emits, for one asset at one timestamp, a message $(\mu_m, \sigma_m^2, c_m, q_m)$: a predicted h-step return, its predictive variance, a service cost and a confidence. Messages are *asset-level*, one per (expert, asset, date). Tier-1 feature builders and the Tier-3 allocator are not experts and cast no routing decision. *Single routing* means one router selects one Top-K subset per (asset, date) and aggregates the selected messages exactly once; we do not study cascaded or per-tier routing.

4.1 Tier 1: Information Layer

Feature construction For each target asset, we construct a causal rolling feature panel including momentum (Chan et al., 1996), Relative Strength Index (RSI), Average True Range (ATR), return-range statistics (Wilder, 1978), Moving Average Convergence/Divergence (MACD), moving averages and volatility. In the richer configuration (O3), we augment this panel with slower momentum, cross-sectional ranks of momentum and volatility, higher moments of returns and volatility.

4.2 Tier 2: Expert Analysis, Routing and Consensus

Return forecasting for O1 and O3 Operating points O1 and O3 map the Tier-1 panel to a predicted mean with a ridge forecaster of h-step returns, fitted once on the training window with the last h rows purged so that no label reaches past it; O1 uses the base technical panel and O3 the augmented panel. O2 replaces the forecaster with the expert messages below.

Expert messages At each rebalance date, expert m emits a message for a candidate asset consisting of a predicted mean μ_m, predictive variance σ_m^2, a confidence score $q_m \in [0, 1]$ and a service-cost estimate c_m. In our implementation these experts are lightweight financial modules rather than large neural subnetworks, which keeps routing and aggregation behavior interpretable: trend, mean reversion, volatility-scaled momentum and cross-sectional momentum rank (Appendix A.4). Each expert's dimensionless score is multiplied by the asset's h-day return scale σ_h (21-day volatility times $\sqrt{h}$), so μ_m and c_m (quoted in basis points) are in return units; $\sigma_m^2 = v_m \sigma_h^2$ uses a fixed expert-specific multiplier v_m, so the variances are calibrated by construction rather than estimated.

Risk-aware routing The router first filters out experts whose confidence q_m falls below a minimum threshold $q_{\min}$ (if none survives for an asset, the most confident one is kept). It then assigns each remaining expert a utility-aware score

$$\mathcal{S}_m = \mu_m - \lambda_{\text{risk}} \sigma_m^2 - \gamma_{\text{cost}} c_m, \tag{4}$$

where $\lambda_{\text{risk}} \geq 0$ penalizes predictive uncertainty and $\gamma_{\text{cost}} \geq 0$ penalizes service cost. The three terms are commensurate by construction: each expert's dimensionless raw score is mapped into h-step simple-return units before entering equation 4, so $[\mu_m] = [c_m] = \text{return}$, $[\sigma_m^2] = \text{return}^2$, $[\lambda_{\text{risk}}] = \text{return}^{-1}$ and γ_{cost} is dimensionless. Equation equation 4 is then the certainty equivalent of exponential utility applied per message — exactly so when $\lambda_{\text{risk}} = A/2$ for absolute risk aversion A — rather than a weighted sum of incommensurable quantities (Pratt, 1964) (derivation in Appendix A.5). We therefore fix $\gamma_{\text{cost}} = 1$, so costs enter at face value, and select λ_{risk}, K and the aggregation rule on validation only (Table 8). The risk-aware router ranks experts by $\mathcal{S}_m$ and retains the set $\mathcal{M}_t^{(K)}$ of the K highest-scoring experts. A utility-tilted weighting

$$\widetilde{\omega}_m = \frac{\exp(\mathcal{S}_m / \tau_{\text{route}})}{\sum_{j \in \mathcal{M}_t^{(K)}} \exp(\mathcal{S}_j / \tau_{\text{route}})}, \tag{5}$$

where $\tau_{\text{route}} > 0$ is the softmax temperature, is an alternative aggregation rule selected per cell on validation together with K; the default aggregation weights ω_m are the posterior weights of equation 6 below, and we keep the two symbols distinct because they are different objects — $\widetilde{\omega}$ tilts toward high-utility experts, ω toward low-variance ones.

Consensus aggregation Selected messages are aggregated into a predictive summary. Treating the retained messages as conditionally independent Gaussian signals of a common latent mean, $\mu_m \mid \mu \sim \mathcal{N}(\mu, \sigma_m^2/q_m)$ under a flat prior, the posterior mean is the precision-weighted average, i.e. the normalized product of the retained Gaussian experts (Hinton, 2002) (Appendix A.5)

$$\mu_{\text{agg}} = \sum_{m \in \mathcal{M}_t^{(K)}} \omega_m \mu_m, \qquad \omega_m = \frac{q_m/\sigma_m^2}{\sum_{j \in \mathcal{M}_t^{(K)}} q_j/\sigma_j^2}, \tag{6}$$

and the variance of the induced mixture follows from the law of total variance,

$$\sigma_{\text{agg}}^2 = \sum_m \omega_m \sigma_m^2 + \sum_m \omega_m (\mu_m - \mu_{\text{agg}})^2.$$

Both terms are exact, so neither is damped. O2 passes μ_{agg} to Tier 3 as its predicted mean; σ_{agg}^2 is logged as a disagreement diagnostic and does not enter the allocator, which keeps f_3 identical across operating points. The utility-tilted alternative $\widetilde{\omega}_m$ of equation 5 can replace ω_m when validation selects it (Bates & Granger, 1969).

4.3 Tier 3: Risk Modeling, Allocation and Execution

Covariance estimation Risk is modeled with the sample covariance matrix of the 125 daily returns strictly before the rebalance date (a 126-day window ending the day before). We keep this component intentionally simple to isolate the effect of routing and allocation choices.

Constraint-aware allocation Given forecasted mean returns $\widehat{\boldsymbol{\mu}}_t \in \mathbb{R}^N$ and covariance $\widehat{\Sigma}_t \in \mathbb{R}^{N \times N}$, the allocator solves

$$\mathbf{w}_t = \arg\max_{\mathbf{w} \in \mathcal{W}_t} \mathbf{w}^\top \widehat{\boldsymbol{\mu}}_t - \tfrac{\lambda_{\text{alloc}}}{2} \mathbf{w}^\top \widehat{\Sigma}_t \mathbf{w} - \kappa \|\mathbf{w} - \mathbf{w}_{t-1}\|_1, \tag{7}$$

where $\lambda_{\text{alloc}} \geq 0$ controls variance aversion, $\kappa \geq 0$ controls turnover, and $\mathcal{W}_t$ is the feasible set of Equation equation 1. Here $\mathbf{w}_{t-1}$ are the drifted holdings of Section 3, and the per-asset cap is $u_i = \text{clip}(5/N, 0.05, 0.35)$. We solve this objective exactly by accelerated proximal gradient (Beck & Teboulle, 2009) with an exact proximal step for the ℓ_1 term over the capped simplex (Appendix A.6); $(\lambda_{\text{alloc}}, \kappa)$ is selected per cell (fold $\times$ universe) on the validation utility of O1 and then frozen, so the routing path cannot influence Tier 3. Rebalancing occurs only on scheduled weekly dates. When we report the *alpha+optimizer* setting in Section 5, the ridge model uses the augmented Tier-1 panel; the optimizer-only setting uses the base technical panel. The ℓ_1 term is the proportional-cost part of Equation equation 2, while the quadratic impact term is inactive in the current implementation.

4.4 Rolling Adaptive Meta-Policy

The hierarchy exposes three operating points, or modes: optimizer-only, consensus routing, and alpha-augmented allocation. Because different modes lead in different universes and periods, we select among them at test time, without retraining any tier: a rolling meta-policy weights the fixed operating points on their recent realized performance. The selector is deliberately simple and is intended as a diagnostic: if operating-point dominance persists from one quarter to the next, then even an online mixture updated once per quarter should capture part of it.

Mode weighting via online mixture Quarters are indexed $b = 1, 2, \ldots$, each B trading days long (about 63). At the start of quarter b the meta-policy holds a distribution $\mathbf{p}_b$ over $\Pi = \{\text{optimizer}, \text{consensus}, \text{alpha}\}$ (O1, O2, O3) and deploys the mixture $\sum_\pi p_b(\pi)\, \mathbf{w}_t^\pi$ at every weekly rebalance of that quarter; the mixture is simulated as a portfolio, so switching between modes pays the same transaction costs as any other trade. After the quarter, each mode's realized net utility $u_b(\pi)$, the sample value of equation 3 on that quarter, is observed, rescaled to a loss $\ell_b(\pi) = (\max_{j \leq b} u_j(\pi) - u_b(\pi))/(\max_{j \leq b} u_j(\pi) - \min_{j \leq b} u_j(\pi)) \in [0, 1]$ over quarters $1, \ldots, b$ only (no future quarter; Figure 3b), the weights are updated by exponential weights with fixed share,

$$\widetilde{p}_{b+1}(\pi) \propto p_b(\pi)\, e^{-\eta_b\, \ell_b(\pi)}, \qquad p_{b+1}(\pi) = (1 - \beta)\, \widetilde{p}_{b+1}(\pi) + \beta/|\Pi|,$$

where $\eta_b > 0$ is the learning rate and $\beta \in [0, 1)$ the fixed-share rate; $\beta = 0$ is Hedge. The follow-the-leader rule (Argmax), which deploys the operating point with the highest cumulative realized utility

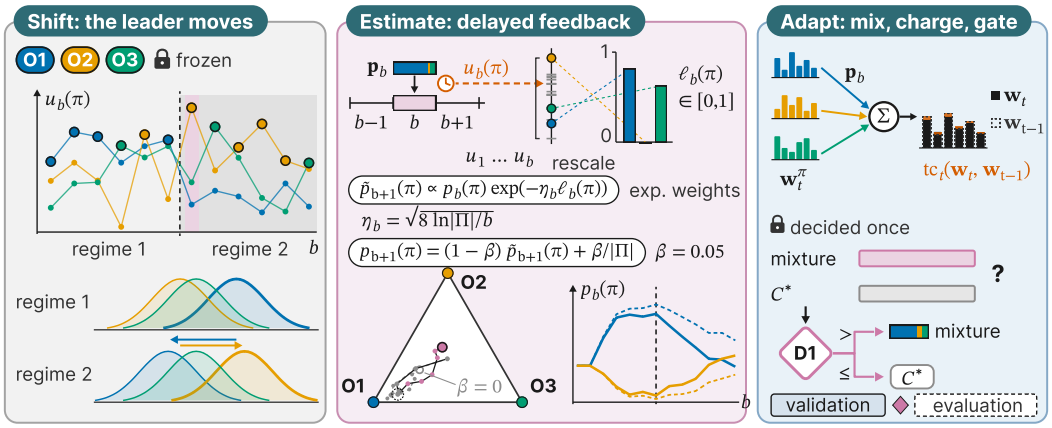

Figure 3: Test-time adaptation workflow: shift, estimate, and adapt. (a) Regime shifts alter relative mode utilities among frozen base policies. (b) Delayed feedback is rescaled to dynamically update mixture weights via exponentially weighted learning with active exploration. (c) Target portfolio weights are aggregated under transaction costs and verified by a validation decision gate.

over all past quarters, is the $\eta \to \infty$, $\beta = 0$ corner of this family and is kept as an ablation. The rates are selected once, on the pooled utility of the selection folds F1–F2 (2008–2013), which serve as the meta-policys validation set, before any evaluation fold is read; the frozen design uses the anytime schedule $\eta_b = \sqrt{8 \ln |\Pi|/b}$ over quarters $b = 1, 2, \dots$, which carries no tuned constant (Cesa-Bianchi & Lugosi, 2006), and $\beta = 0.05$, selected from $\eta \in \{\text{anytime}, 0.5, 2, 8\} \times \beta \in \{0, 0.05\}$.

Do-no-harm deployment gate Before any evaluation period is read (Figure 3c), the mixture's utility is compared with that of the best fixed operating point on the pooled utility of the selection folds F1–F2 (2008–2013). If the mixture does not exceed it, the best fixed point is deployed instead and the meta-policy is reported as not deployed. The gate is applied mechanically and cannot be revisited after evaluation. This rolling meta-policy has four practical properties: it requires no external features, it remains fully out of sample, it adds only two rates selected on the selection folds, and it tracks changes in which mode leads without modeling them explicitly. Algorithm 1 summarizes one weekly rebalance and one quarterly update; the cost of a rebalance is dominated by the allocator, $O(WN^2 + N^3 + IN^2)$ per operating point for a covariance window W and I solver iterations.

5 Experiments

Universes. We evaluate on eleven U.S. equity universes: four named ones chosen to span distinct cross-sectional structures, **BigTech-6** (concentrated, highly correlated), **Consumer-20** (broad, low dispersion), **High-Vol-15** (the 15 most volatile of the 200 most liquid names), **Defensive-15** (utilities and consumer staples) the latter also used for the long-horizon comparison U.S. 2014–2026 & JP + EU stock empirical results. Constituent lists are in Appendix B.2.

Protocol and implementation. All experiments use daily OHLCV-derived features computed from trailing windows of up to 126 trading days, weekly Friday rebalancing, long-only constraints with per-asset caps, and a turnover penalty κ selected on validation. Every table uses one walk-forward protocol: six folds of three-year formation and three-year evaluation windows (2005–2026), with every hyperparameter selected on validation data inside each formation window. Unless stated otherwise, results are for the test period 2014–2026 over all eleven universes, net of 25 bps costs.

Baselines and metrics. We compare against five state-of-the-art strategies for **baselines**: FinCon (Yu et al., 2024), TradingAgents (Xiao et al., 2024), MetaTrader (Niu et al., 2022), MacroHFT (Zong et al., 2024), and FINSABER (Li et al., 2026). We report Sharpe ratio (SR) (Sharpe, 1994), Compound Annual Growth Rate (CAGR) and Maximum Drawdown (MDD).

Table 2: Baseline comparison across four universes (2024 to 2025). Sharpe ratio (annualized), CAGR (%), and maximum drawdown (%). Regime-Adaptive as Ours, Bold: best per universe.

	BigTech-6			U.S. Consumer-20			High-Vol-15			Defensive-15		
Method	SR	CAGR	MDD	SR	CAGR	MDD	SR	CAGR	MDD	SR	CAGR	MDD
FinCon	2.18	60.8	−16.2	1.01	11.7	−6.5	0.87	16.3	−13.9	0.85	10.4	−12.4
TradingAgents	2.01	48.5	−15.4	1.31	14.4	−6.7	0.56	9.4	−20.3	0.91	10.5	−11.3
MetaTrader	2.08	49.6	−14.0	1.19	12.7	−6.3	0.06	−1.3	−29.3	0.76	8.9	−12.7
MacroHFT	1.92	43.2	−14.5	1.47	15.8	−6.1	−0.23	−8.1	−28.9	0.59	6.4	−12.2
FINSABER	2.30	70.5	−14.6	1.29	17.4	−11.5	1.94	51.9	−12.5	0.18	1.5	−12.5
Ours	**2.61**	**104.6**	**−12.3**	**2.42**	**32.6**	**−5.1**	**2.11**	**60.0**	**−10.6**	**1.13**	**12.6**	**−10.3**

5.1 Cross-Universe Baseline Comparison

Table 2 reports performance across four market universes. After transaction costs, our meta-policy consistently outperforms competitive baselines and static strategies across most regimes. While baselines suffer degradation or instability in demanding regimes (e.g., High-Vol-15), our method yields superior risk-adjusted returns (SR 2.11) and significantly lower maximum drawdown. This highlights the necessity of dynamic adaptation: static selections fail across regime shifts, whereas our meta-policy continuously identifies and shifts toward optimal operating modes (Section 5.2).

5.2 Long Horizon Cross Market Empirical Results

Table 3: Pooled comparison of the meta-policies, operating points and baselines.

	U.S. 2014–2026 (44 cells)					Held-out JP + EU (69 cells)		
Policy	SR ↑	U_{MV} ↑	U_{CVaR} ↑	MDD ↓	Turnover ↓	SR ↑	U_{CVaR} ↑	MDD ↓
RASP (Ours)	**0.954**	**0.1432**	−0.486	**−18.7**	**1.4**	**0.586**	**−0.577**	**−26.6**
Hedge (ungated)	0.750	0.0829	**−0.484**	−20.9	10.4	0.439	−0.579	**−26.6**
Follow-the-leader	0.732	0.0816	−0.490	−21.3	11.2	0.427	-0.586^{*}	−26.7
Routed consensus (O2)	0.657	0.0713^{**}	-0.523^{**}	−21.9	14.2	0.327	-0.631^{**}	−28.3
Alpha-augmented (O3)	0.752	0.0833	-0.491^{*}	−20.8	10.6	0.440	-0.583^{*}	**−26.6**
FinCon	0.784	0.0897	−0.495	−21.3	9.9	0.480	-0.611^{**}	−28.1
TradingAgents	0.760	0.0882	-0.502^{*}	−21.5	8.7	0.484	-0.617^{**}	−27.9
MetaTrader	0.707	0.1019	-0.641^{**}	−24.5	12.7	0.348	-0.778^{**}	−31.6
MacroHFT	0.914	$0.1219^{\dagger\dagger}$	-0.533^{**}	−23.0	**1.4**	0.573	-0.673^{**}	−29.8
FINSABER	0.895	$0.1130^{\dagger\dagger}$	-0.511^{*}	−22.1	1.8	0.583	-0.640^{**}	−28.7

SR: annualized net Sharpe ratio; U_{MV}, U_{CVaR}: mean per-quarter mean–variance and CVaR-95 certainty equivalents; MDD in %; turnover in multiples of NAV per year. Marks on a comparator's certainty equivalent give one-sided paired stationary-bootstrap tests of Deployed minus comparator: */** Deployed ahead at $p < 0.05$/$p < 0.01$; †/†† comparator ahead at the same levels; unmarked: not significant. **Bold**: best.

The meta-policy dynamically blends the hierarchy's operating modes, optimizer-only, consensus routing, and alpha-augmented allocation via a quarterly update under a do-no-harm validation gate. Table 3 compares the meta-policies, operating points, and SOTA baselines across U.S. and held-out international markets under 25 bps transaction costs. As shown in Table 3, RASP (Ours) achieves top-tier performance across all regions. Notably, in the U.S. market, RASP reaches a Sharpe ratio of 0.954 with an ultra-low turnover ($1.4\times$). The performance advantage transfers seamlessly to the 69 held-out Japanese and European universes, where RASP maintains the highest Sharpe ratio (0.586) and the shallowest drawdown (-26.6%). These results demonstrate that our adaptive gating mechanism effectively mitigates regime-shift risks without sacrifice cost efficiency. The average Hedge mixture yields higher utility than single-mode O_1 on validation folds (0.0457 vs. 0.0372), the gate adaptively falls back to fixed mode O_1 in 15% of test cells successfully preventing performance degradation prove the adaptive function of dynamic gate.

5.3 Sensitivity Analysis

Evaluating the meta-policy across alternative objectives (Sortino ratio and CVaR-95) confirms that the deployed policy and Hedge consistently outperform the unhedged Argmax rule. As detailed in Table 4, tuning λ_{alloc} demonstrates strong parameter stability: higher risk aversion systematically suppresses turnover, while validation selection achieves the optimal risk-adjusted return.

Table 4: Allocator risk aversion, US 2014–2026.

λ_{alloc}	**Deployed**	**Hedge**	**Argmax**	**O2**	**O3**	**turnover**
selected (300)	0.954	0.750	0.732	0.657	0.752	1.4
50	0.628	0.629	0.589	0.468	0.646	3.6
100	0.714	0.693	0.655	0.544	0.705	4.5
200	0.764	0.750	0.727	0.647	0.752	2.6
400	0.785	0.779	0.760	0.726	0.778	2.5

5.4 Ablation Study

Table 5: Ablation of the meta-policy components, 25 bps one-way.

	U.S. 2014–2026 (44 cells)					**Held-out JP + EU (69 cells)**		
Variant	SR ↑	U_{MV} ↑	U_{CVaR} ↑	MDD ↓	Turnover ↓	SR ↑	U_{CVaR} ↑	MDD ↓
RASP (full)	**0.954**	**0.1432**	−0.486	**−18.7**	**1.4**	**0.586**	**−0.577**	**−26.6**
Gating & Meta-learning Ablations								
w/o deployment gate	0.750	0.0829	**−0.484**	−20.9	10.4	0.439	−0.579	**−26.6**
w/o Hedge weighting (Argmax)	0.732	0.0816	−0.490	−21.3	11.2	0.427	−0.586	−26.7
w/o fixed share ($\beta = 0$)	0.718	0.0792	−0.505	−21.5	12.1	0.415	−0.602	−28.5
Equal mixture ($1/N$ static)	0.684	0.0745	−0.512	−22.4	8.2	0.380	−0.610	−29.0
Single-mode Subsystems								
w/o mixture, O3 only	0.752	0.0833	−0.491	−20.8	10.6	0.440	−0.583	**−26.6**
w/o mixture, O2 only	0.657	0.0713	−0.523	−21.9	14.2	0.327	−0.631	−28.3
Risk & Friction Management								
w/o volatility scaling	0.638	0.0680	−0.545	−23.8	12.8	0.352	−0.648	−30.5
w/o cost awareness	0.591	0.0612	−0.578	−25.2	18.6	0.310	−0.685	−32.4

Each row removes one component and keeps the rest of the design fixed. Metrics as in Table 3. **Bold**: best value per column.

Table 5 removes one component of RASP at a time. Removing the deployment gate lowers the U.S. Sharpe ratio from 0.954 to 0.750 and raises turnover from 1.4 to 10.4 times NAV per year; replacing Hedge weighting by Argmax, setting $\beta = 0$, or using a static equal mixture lowers it further to 0.732, 0.718 and 0.684. Each single operating point (O3: 0.752, O2: 0.657) is below the full mixture, and removing volatility scaling or cost awareness gives the lowest Sharpe ratios (0.638 and 0.591) with turnover of 12.8 and 18.6. On the held-out JP + EU universes every variant is also below the full design (0.586).

We ablate the size of the expert panel of the consensus mode. Doubling the panel from four to eight experts leaves its net Sharpe statistically unchanged (0.657 to 0.641, $p = 0.13$) but raises turnover in all 44 cells (Table 7, Appendix C.1). The cost of adding experts is therefore clear and its benefit is not detectable, which reinforces the paper's thesis: there is no universally optimal configuration.

6 Conclusion

We studied test-time adaptation in dynamic markets of a trained hierarchical portfolio agent as selection among its fixed operating points, without retraining any module. RASP exposes three operating points (optimizer-only, routed consensus and alpha-augmented allocation). All three share one allocator, universe and rebalance calendar, so any difference between them comes from the Tier-2 path alone. Across universes and market periods, no single operating point is best everywhere. A rolling Hedge/fixed-share meta-policy observes only delayed realized utility and is deployed behind a do-no-harm validation gate. It reaches a net Sharpe ratio of 0.954 on the U.S. 2014–2026 sample with a turnover of 1.4 times NAV per year. On 69 held-out Japanese and European universes it reaches 0.586 with the shallowest drawdown. It outperforms the ungated mixture, the follow-the-leader rule, each single operating point and five recent baselines. The ablation shows that both the gate and the mode mixture are necessary, and that most of the gain comes from lower turnover under realistic transaction costs.

AI USE STATEMENT

In this work, we used generative AI tools to implement methods and to provide feedback on research methodology and experiments. We have not used generative AI tools to generate synthetic data sets, develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, or interpret results. Translation, dataset cleaning and reformatting, and qualitative or thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to create and edit software code, such as experiment and evaluation scripts, and to edit the text of the paper for readability. We have reviewed all AI-assisted work. All AI-assisted code was read, executed, and tested by the authors, and all reported results were checked against the saved experiment logs. Feedback from AI tools on methodology and experiments was treated as suggestions; the research questions, method design, and conclusions were determined by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of AI.

ETHICS STATEMENT

This work studies portfolio decision-making in a high-stakes setting where misuse of backtested systems can cause financial harm. The intended use is methodological research on hierarchical, constraint-aware investment agents; it is not financial advice and is not intended for autonomous live deployment. Experiments use historical market data and derived features only; no human-subject data or personally identifiable information are involved. Key risks include distribution shift, overfitting, concentration, and over-trust in simulated results. We partially mitigate these risks by enforcing turnover and portfolio constraints, reporting downside metrics together with returns, and discussing failure modes such as sensitivity to missing signals. We release the project as a research artifact and do not provide brokerage connectivity or live execution tooling.

REPRODUCIBILITY STATEMENT

The hierarchy, its three operating points and the meta-policy are specified in Sections 3 and 4, with the notation and the weekly decision loop in Appendix A. The universes are listed in Appendix B.2, the long-horizon fold structure and universe-construction rules in Appendix B.1, and the hyperparameter ranges in Table 8. All experiments use historical market prices and features derived from them; no proprietary or personal data are involved. An anonymized repository will be provided with the supplementary material; it contains the code and the scripts that generate each table and figure. Hyperparameters and the meta-policy were selected on validation data only. The U.S. panel is a single-source, single-vintage download of publicly available daily adjusted closing prices (fetched 2026-09-21; 2005-01-03 to 2026-09-18; 500 symbols: 485 equities and 15 ETFs, 13 of which form the SectorETF universe), not spliced across providers; its constituent list is as of 2026, so the equity universes carry a survivorship bias that is common to all methods (Appendix B.1). Appendix D.1 maps each reported table to the script and result file that produce it. Source code is available at https://anonymous.4open.science/r/RASPv.

REFERENCES

Amir Ahmadi-Javid. Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123, 2012.

John M. Bates and Clive W. J. Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

Lijun Bo, Huafu Liao, and Xiang Yu. Risk sensitive portfolio optimization with default contagion and regime-switching. *SIAM Journal on Control and Optimization*, 57(1):366–401, 2019.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Louis KC Chan, Narasimhan Jegadeesh, and Josef Lakonishok. Momentum strategies. *The Journal of Finance*, 51(5):1681–1713, 1996.

Min Dai, Zuo Quan Xu, and Xun Yu Zhou. Continuous-time Markowitz's model with transaction costs. *SIAM Journal on Financial Mathematics*, 1(1):96–125, 2010.

Priya Donti, Brandon Amos, and J Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. *arXiv preprint arXiv:1103.4601*, 2011.

Adam N Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1): 9–26, 2022.

William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

Angelos Filos. Reinforcement learning for portfolio management. *arXiv preprint arXiv:1909.09571*, 2019.

Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic trading with predictable returns and transaction costs. *Journal of Finance*, 68(6):2309–2340, 2013.

Christian Genest and James V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1(1):114–135, 1986.

James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.

Mark Herbster and Manfred K. Warmuth. Tracking the best expert. *Machine Learning*, 32(2): 151–178, 1998.

Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002.

Yoontae Hwang, Yaxuan Kong, Stefan Zohren, and Yongjae Lee. Decision-informed neural networks with large language model integration for portfolio optimization. *arXiv preprint arXiv:2502.00828*, 2025.

Nan Jiang and Lihong Li. Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning*, pp. 652–661. PMLR, 2016.

Bige Kahraman and Heather E Tookes. Trader leverage and liquidity. *The Journal of Finance*, 72(4): 1567–1610, 2017.

Petter N. Kolm and Gordon Ritter. Factor investing with Black–Litterman–Bayes: Incorporating factor views and priors in portfolio construction. *The Journal of Portfolio Management*, 47(2): 113–126, 2021. doi: 10.3905/jpm.2020.1.196.

Weixian Waylon Li, Hyeonjun Kim, Mihai Cucuringu, and Tiejun Ma. Can LLM-based financial investing strategies outperform the market in long run? In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, 2026. doi: 10.1145/3770854.3785702.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. DeepSeek-V3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

Miguel Sousa Lobo, Maryam Fazel, and Stephen Boyd. Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research*, 152(1):341–365, 2007.

Junyu Luo, Zhizhuo Kou, Liming Yang, Xiao Luo, Jinsheng Huang, Zhiping Xiao, Jingshu Peng, Chengzhong Liu, Jiaming Ji, Xuanzhe Liu, Sirui Han, Ming Zhang, and Yike Guo. FinMME: Benchmark dataset for financial multi-modal reasoning evaluation. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 29465–29489, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.1426. URL https://aclanthology.org/2025.acl-long.1426/.

Harry M Markowitz and G Peter Todd. *Mean-variance analysis in portfolio choice and capital markets*. John Wiley & Sons, 2000.

Hui Niu, Siyuan Li, and Jian Li. MetaTrader: An reinforcement learning approach integrating diverse policies for portfolio optimization. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 1573–1583, 2022. doi: 10.1145/3511808.3557363.

Jeremy Nixon, Michael W Dusenberry, Linchuan Zhang, Ghassen Jerfel, and Dustin Tran. Measuring calibration in deep learning. In *CVPR Workshops*, volume 2, 2019.

John W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122–136, 1964.

William F Sharpe. The Sharpe ratio. *The Journal of Portfolio Management*, 21(1):49–58, 1994.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017.

Yu Shi, Zongliang Fu, Shuo Chen, Bohan Zhao, Wei Xu, Changshui Zhang, and Jian Li. Kronos: A foundation model for the language of financial markets. *arXiv preprint arXiv:2508.02739*, 2025.

Adith Swaminathan and Thorsten Joachims. The self-normalized estimator for counterfactual learning. *Advances in Neural Information Processing Systems*, 28, 2015.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1658–1665, 2019.

J Welles Wilder. *New concepts in technical trading systems*. Trend Research, Greensboro, NC, 1978.

Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. TradingAgents: Multi-agents LLM financial trading framework. *arXiv preprint arXiv:2412.20138*, 2024.

Yunan Ye, Hengzhi Pei, Boxin Wang, Pin-Yu Chen, Yada Zhu, Ju Xiao, and Bo Li. Reinforcement-learning based portfolio management with augmented asset movement prediction states. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 1112–1119, 2020.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yuechen Jiang, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Zhenyu Cui, Rong Liu, Zhaozhuo Xu, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. FinCon: A synthesized LLM multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, et al. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4314–4325, 2024.

Chuqiao Zong, Chaojie Wang, Molei Qin, Lei Feng, Xinrun Wang, and Bo An. MacroHFT: Memory augmented context-aware reinforcement learning on high frequency trading. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024. doi: 10.1145/3637528.3672064.

Appendix contents

A Notation and Additional Details **14**

A.1 Portfolio Constraints and Frictions 14

A.2 Routing, aggregation, and allocation parameters 14

A.3 Hierarchical Components 14

A.4 Expert panel 14

A.5 Routing score and consensus: derivations 15

A.6 Allocator solver 15

A.7 Weekly Decision Loop 16

B Experimental Details **16**

B.1 Long-horizon protocol 16

B.2 Universe constituents 17

B.3 Baseline details 17

C Additional Results **17**

C.1 Number of experts 17

D Reproducibility and Artifact Notes **17**

D.1 Reproduction map 18

E Limitations **19**

F Impact Statement **19**

A Notation and Additional Details

We summarize the main mathematical objects used in the paper. Let $t \in \{1, \dots, T\}$ denote rebalance times, let $i \in \{1, \dots, N\}$ index tradable assets, and let $m \in \{1, \dots, M_t\}$ index candidate experts available at date t. The market state is $\mathbf{x}_t \in \mathbb{R}^p$ and the realized next-period returns are $\mathbf{r}_{t+1} \in \mathbb{R}^N$.

A.1 Portfolio Constraints and Frictions

Two remarks on the feasible set of equation 1. First, the leverage budget is redundant in the regime we run: with $\mathbf{w} \geq 0$ and $\mathbf{1}^\top \mathbf{w} = 1$ we have $\|\mathbf{w}\|_1 \equiv 1$, so $\|\mathbf{w}\|_1 \leq L$ never binds for $L \geq 1$; it is kept only for a long–short extension, which we do not run. Second, the turnover penalty is stated in ℓ_1 because under proportional trading costs the ℓ_1 distance is the cost, up to the per-asset rate; an ℓ_2 penalty corresponds to no charged cost model.

The portfolio is chosen from the feasible set equation 1,

$$\mathcal{W}_t = \left\{ \mathbf{w} \in \mathbb{R}^N : \mathbf{1}^\top \mathbf{w} = 1,\ \mathbf{w} \geq 0,\ \|\mathbf{w}\|_1 \leq L,\ w_i \leq u_i,\ \mathbf{A}\mathbf{w} \leq \mathbf{b} \right\},$$

where $L > 0$ denotes a leverage budget, $u_i > 0$ are position caps, and $\mathbf{A} \in \mathbb{R}^{k \times N}$ and $\mathbf{b} \in \mathbb{R}^k$ encode optional exposure constraints, which no experiment imposes. The current implementation uses the long-only special case with $\mathbf{w} \geq 0$, $L = 1$ and $u_i = \mathrm{clip}(5/N, 0.05, 0.35)$. Transaction costs are modeled with the same linear-plus-quadratic form used in the main text; in all experiments $\boldsymbol{\alpha}_t$ is a uniform one-way rate and $\Lambda_t = 0$:

$$\mathrm{tc}_t(\mathbf{w}_t, \mathbf{w}_{t-1}) = \boldsymbol{\alpha}_t^\top |\mathbf{w}_t - \mathbf{w}_{t-1}| + \tfrac{1}{2}(\mathbf{w}_t - \mathbf{w}_{t-1})^\top \Lambda_t (\mathbf{w}_t - \mathbf{w}_{t-1}),$$

where $\boldsymbol{\alpha}_t \in \mathbb{R}^N_+$ is the vector of proportional costs and $\Lambda_t \in \mathbb{R}^{N \times N}$ with $\Lambda_t \succeq 0$ is the temporary-impact matrix.

A.2 Routing, Aggregation, and Allocation Parameters

For expert m, the router receives a predicted mean μ_m, predictive variance σ_m^2, a confidence score q_m, and a service-cost proxy c_m. The implementation first removes experts that fall below the confidence threshold $q_{\min}$, then applies the risk-aware routing score equation 4,

$$\mathcal{S}_m = \mu_m - \lambda_{\text{risk}} \sigma_m^2 - \gamma_{\text{cost}} c_m,$$

where $\lambda_{\text{risk}} \geq 0$ penalizes uncertainty and $\gamma_{\text{cost}} \geq 0$ penalizes service cost. The top-K router keeps K experts and weights them by precision equation 6; the softmax weighting with temperature $\tau_{\text{route}} = 0.7$ is an alternative that validation selects in some cells.

After routing, aggregation weights satisfy $\omega_m \geq 0$ and $\sum_m \omega_m = 1$. The allocator then uses forecasted mean returns $\widehat{\boldsymbol{\mu}}_t \in \mathbb{R}^N$ and covariance estimates $\widehat{\Sigma}_t \in \mathbb{R}^{N \times N}$ in a mean–variance objective. We reserve $\lambda_{\text{alloc}} \geq 0$ for the allocator's variance-risk trade-off and $\kappa \geq 0$ for the turnover penalty, so these symbols are not confused with the routing-side parameters λ_{risk} and γ_{cost}.

A.3 Hierarchical Components

The implementation is organized into three stages.

- **Tier 1:** rolling technical features, with an optional alpha augmentation (O3).
- **Tier 2:** the ridge forecaster (O1, O3), or expert messages, utility-aware routing and consensus aggregation (O2).
- **Tier 3:** a long-only turnover-aware allocator based on the predicted means and a sample covariance matrix.

This appendix mirrors the description in Sections 3 and 4.

A.4 Expert Panel

The four experts of O2 are computed once per universe as trailing-window statistics, so the entry at date t uses prices up to t only. Each raw score is clipped, multiplied by the asset's h-day return scale

σ_h to give μ_m, and paired with $\sigma_m^2 = v_m \sigma_h^2$, a confidence $q_m = a_m/(1 + 3\sigma_m)$ and a fixed service cost (Table 6). An expert without a valid score for an asset reports $q_m = 0$ and is removed by the gate.

Table 6: Expert panel of O2.

Expert	**Raw score**	v_m	a_m	**Cost (bps)**
Trend	21-day momentum over its 21-day volatility	1.0	1.00	2
Mean reversion	negated trailing z-score of the 5-day return	1.5	0.90	3
Vol-scaled momentum	63-day momentum over annualised volatility	1.0	1.00	3
Cross-sectional rank	rank of 63-day momentum, mapped to $[-1, 1]$	1.2	0.95	5

v_m: variance multiplier; a_m: confidence multiplier. Values are the engine's constants.

A.5 Routing score and consensus: derivations

Routing score. Suppose acting on expert m's view yields an h-day return $X_m \sim \mathcal{N}(\mu_m, \sigma_m^2)$ net of a known cost c_m, and the decision maker has exponential utility $u(x) = -e^{-Ax}$ with absolute risk aversion $A > 0$. Then $\mathbb{E}[u(X_m - c_m)] = -\exp\!\big(-A(\mu_m - c_m) + \frac{A^2}{2}\sigma_m^2\big)$, and the certainty equivalent, the sure amount with the same expected utility, is

$$\mathrm{CE}_m = \mu_m - \tfrac{A}{2}\sigma_m^2 - c_m$$

(Pratt, 1964). Ranking experts by CE_m is ranking them by equation 4 with $\lambda_{\text{risk}} = A/2$ and $\gamma_{\text{cost}} = 1$. All three terms are in return units, which fixes the units of λ_{risk} as return^{-1}; a value of $\gamma_{\text{cost}} \neq 1$ would express a deliberate over- or under-weighting of service cost, which we do not use.

Consensus as a Gaussian posterior. Let the retained messages be conditionally independent given a latent h-day mean μ, with $\mu_m \mid \mu \sim \mathcal{N}(\mu, \sigma_m^2/q_m)$, so that confidence scales precision. Under a flat prior the posterior is proportional to the product of the experts' likelihoods,

$$p\big(\mu \mid \{\mu_m\}_{m \in \mathcal{M}_t^{(K)}}\big) \;\propto \prod_{m \in \mathcal{M}_t^{(K)}} \exp\!\Big(-\frac{q_m(\mu_m - \mu)^2}{2\sigma_m^2}\Big),$$

a Gaussian product of experts (Hinton, 2002). Its exponent is quadratic in μ with coefficient $-\frac{1}{2}\sum_m q_m/\sigma_m^2$ on μ^2 and $\sum_m q_m\mu_m/\sigma_m^2$ on μ. Completing the square gives

$$\mu \mid \{\mu_m\} \;\sim\; \mathcal{N}\Big(\textstyle\sum_m \omega_m \mu_m,\; \big(\sum_m q_m/\sigma_m^2\big)^{-1}\Big), \qquad \omega_m = \frac{q_m/\sigma_m^2}{\sum_j q_j/\sigma_j^2},$$

which is equation 6. The posterior variance shrinks as experts are added, whatever they say. The diagnostic σ_{agg}^2 of Section 4 is instead the variance of the linear opinion pool $\sum_m \omega_m \mathcal{N}(\mu_m, \sigma_m^2)$ (Genest & Zidek, 1986). By the law of total variance it adds the between-expert term $\sum_m \omega_m(\mu_m - \mu_{\text{agg}})^2$ to the average within-expert variance, so it grows when the experts disagree. Neither variance enters Tier 3.

A.6 Allocator solver

Problem equation 7 is solved by accelerated proximal gradient on the smooth part $-\widehat{\boldsymbol{\mu}}_t^\top \mathbf{w} + \frac{\lambda_{\text{alloc}}}{2}\mathbf{w}^\top \widehat{\Sigma}_t \mathbf{w}$, with step size equal to the inverse of λ_{alloc} times the largest eigenvalue of $\widehat{\Sigma}_t$. The proximal step

$$\min_{\mathbf{w}} \;\tfrac{1}{2}\|\mathbf{w} - \mathbf{v}\|_2^2 + \kappa' \|\mathbf{w} - \widetilde{\mathbf{w}}_{t-1}\|_1 \quad \text{s.t.} \quad 0 \le w_i \le u, \;\; \mathbf{1}^\top \mathbf{w} = 1$$

where κ' is κ times the step size and $\mathbf{v}$ the gradient step, separates across coordinates given the multiplier θ of the budget constraint: each coordinate is soft-thresholded toward its drifted holding and clipped to $[0, u]$. The sum is monotone in θ, which is therefore recovered by bisection. With $\kappa = 0$ the step reduces to the Euclidean projection onto the capped simplex, and the iterate is then polished by solving the KKT conditions on the set of assets strictly inside $(0, u)$. The solver is checked against a general-purpose SQP solver on random instances.

Cost per rebalance. Write W for the covariance window in days, I for the number of accelerated iterations and $F \le N$ for the number of assets strictly inside $(0, u)$.

- Covariance estimation: $O(WN^2)$.
- Step size: the largest eigenvalue of $\widehat{\Sigma}_t$, computed by a symmetric eigendecomposition, $O(N^3)$ once per rebalance (power iteration would make it $O(N^2)$ per iteration).
- Each iteration: the gradient needs $\widehat{\Sigma}_t \mathbf{y}$, which is $O(N^2)$, plus the proximal step, $O(N \log(1/\varepsilon))$ for bisection tolerance ε. Iterations therefore cost $O(IN^2)$ in total.
- Polish (only when $\kappa = 0$): two linear solves on the free block of $\widehat{\Sigma}_t$, $O(F^3)$. No matrix inverse is formed.

One rebalance thus costs $O(WN^2 + N^3 + IN^2)$ per operating point, three times over. Tier 1, the experts and the router are linear in N per date. The ridge forecasters are fitted once per fold, and the meta-policy adds $O(|\Pi|)$ work per block. All experiments use N in the tens, where the cubic terms are negligible. Wall-clock times are given in Appendix D.

A.7 Weekly Decision Loop

Fixed before the test period: ridge forecasters for O1/O3; the expert panel; $(\lambda_{\text{alloc}}, \kappa)$ selected on O1; $(\lambda_{\text{risk}}, K, \text{aggregation})$ selected with Tier 3 frozen; the meta-policy rates (η, β) and the do-no-harm gate, decided once on the pooled selection folds F1–F2 (Section 4.4).

Rebalance at date t in block b (holding $\mathbf{p}_b$ and $\widetilde{\mathbf{w}}_{t-1}$):

1. $\mathbf{X}_t \leftarrow f_1(\text{prices} \le t)$.
2. $\widehat{\boldsymbol{\mu}}_t^{\text{O1}}, \widehat{\boldsymbol{\mu}}_t^{\text{O3}} \leftarrow$ frozen ridge on the base / augmented panel.
3. For each asset: messages $(\mu_m, \sigma_m^2, q_m, c_m)$; drop $q_m < q_{\min}$; score by equation 4; keep the top K; set $\widehat{\mu}_{t,i}^{\text{O2}} \leftarrow \mu_{\text{agg}}$ by equation 6.
4. $\widehat{\Sigma}_t \leftarrow$ sample covariance of the $W - 1 = 125$ daily returns strictly before t.
5. For each $\pi \in \Pi$: $\mathbf{w}_t^\pi \leftarrow$ solve equation 7 against π's own carried holdings.
6. Hold $\mathbf{w}_t \propto \sum_\pi p_b(\pi)\, \mathbf{w}_t^\pi$; pay the proportional cost $\text{tc}_t(\mathbf{w}_t, \widetilde{\mathbf{w}}_{t-1})$ of equation 2; drift to the next rebalance.

End of block b:

7. For each π: $u_b(\pi) \leftarrow$ realised certainty equivalent equation 3 of π's own net returns on block b.
8. $\mathbf{p}_{b+1} \leftarrow$ exponential-weights / fixed-share update from the rescaled losses of $u_1, \ldots, u_b$ (Section 4.4).

Algorithm 1: One weekly rebalance and one block update of RASP.

B Experimental Details

B.1 Long-horizon protocol

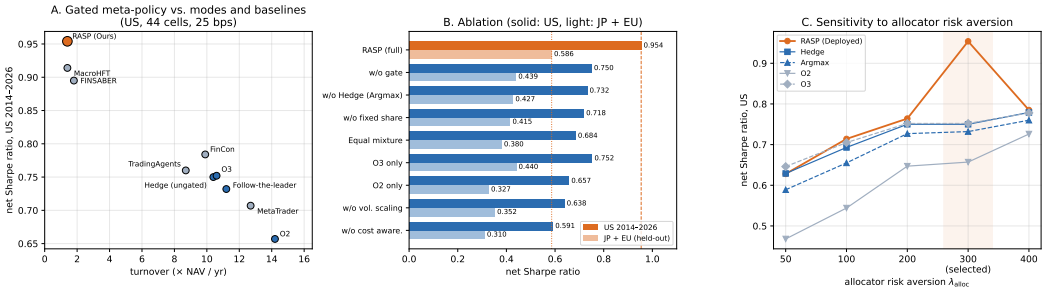

Figure 4: RASP Long Horizon Performance.

The long-horizon comparison of Section 5.2 uses a single-vintage panel of US adjusted closes from 2005-01-03 to 2026-09-18. Folds are rolling three-year formation / three-year evaluation windows stepped three years at a time, giving six folds and 18.75 years of out-of-sample evaluation.

Each fold rebuilds its universes *by rule* from formation-window data only, never by hand: the most traded names, the most and least volatile, and the least and most cross-sectionally correlated names, plus an ETF basket (SPY, nine sector funds, IWM, TLT and GLD) that carries no survivorship bias

and serves as the clean control. The rule-built universes hold 50 (most traded), 40 (each of the other four rules) and 13 (ETF) names; they are six of the eleven US evaluation universes. Every hyperparameter is selected on a validation slice taken from the last twelve months of the same formation window, in two stages — first the allocator's $(\lambda_{\text{alloc}}, \kappa)$ on O1, then, with Tier 3 frozen, the router's λ_{risk}, K and aggregation rule (τ_{route} is fixed) — while the meta-policy grid is selected once, on the pooled utility of the selection folds F1–F2 (2008–2013), where the do-no-harm gate is also decided, so no evaluation period informs any choice made before it.

One limitation should be stated plainly: the constituent list is as of 2026 and projected backwards, so the equity universes carry survivorship bias. The bias is common-mode all three operating points and every baseline share one universe in each fold, so the *relative* comparison that the section actually makes is largely immune to it, and the ETF universe, whose members all trade throughout the 2005–2026 panel, is included precisely so that one universe is free of the problem.

B.2 Universe constituents

BigTech-6: NVDA, MSFT, AAPL, GOOGL, AMZN, META. **Consumer-20**: WMT, COST, PG, KO, PEP, MCD, NKE, SBUX, TGT, HD, LOW, TJX, ROST, DG, DLTR, YUM, CMG, DPZ, ORLY, AZO. **High-Vol-15**: the 15 most volatile names among the 200 most liquid (median dollar volume) in the formation window. **Defensive-15**: JNJ, PG, KO, PEP, CL, GIS, K, SJM, MKC, HRL, DUK, SO, D, AEP, NEE (K has no data in the panel, so 14 names are used in every fold). **Healthcare-10**: JNJ, UNH, PFE, ABBV, MRK, LLY, TMO, ABT, DHR, BMY. Names without data in a fold are dropped from that fold (META and ABBV in F1–F3, DG in F1–F2, CMG in F1; in the evaluation folds only META and ABBV in F3). The six rule-built universes are built from formation-window data only and are described in Appendix B.1.

B.3 Baseline details

All baselines are emitted as weekly target-weight schedules on the same calendar and replayed through the same simulator as the operating points, so they pay the same proportional cost against drifted holdings; every covariance uses the 125 daily returns strictly before the rebalance date, and every weight vector is projected onto the same capped simplex.

- **Hedge** and **Argmax** are the ungated mixture of Section 4.4 and its follow-the-leader ($\eta \to \infty$, $\beta = 0$) corner, which deploys the mode with the highest cumulative realized utility over all past quarters.

C Additional Results

C.1 Number of experts

Table 7: Number of experts in O2, US 2014–2026.

	$M = 4$ (frozen design)	$M = 8$
O2 net SR	0.657	0.641
O2 turnover (per year)	14.2	17.5

Mean over the 44 cells at 25 bps. Wilcoxon signed-rank test over cells: $p = 0.13$ for the Sharpe ratio (higher with $M = 8$ in 18 of 44 cells) and $p \approx 10^{-13}$ for turnover (higher in 44 of 44 cells). The four added experts are a slower trend, a long-horizon reversal, a low-volatility rank and a trend-acceleration signal.

D Reproducibility and Artifact Notes

Runtime profile The experiments run on a CPU workstation: generating the 66 US cells (22 selection-fold cells in F1–F2 and the 44 evaluation cells in F3–F6) takes about 21 minutes on 6–7 worker processes (7.3 for the selection folds and 13.8 for the evaluation folds), plus about 28 and 11 minutes for the Japanese and European held-out cells (the Japanese run overlapped the US evaluation

folds), 11 minutes for the supporting cells and 13 minutes for the robustness cells; replaying the policies and running the statistical tests for the US evaluation takes about 2 minutes. The main computational burden comes from repeated backtesting and covariance-based allocation rather than from large-scale neural training.

Hyperparameter ranges Table 8 lists the search grids.

Table 8: Hyperparameter search ranges.

Hyperparameter	Values / range
Allocator risk aversion λ_{alloc}	$\{50, 100, 200, 300, 400\}$
Turnover penalty κ	$\{0, 0.02, 0.05\}$
Routing risk penalty λ_{risk}	$\{0, 0.5, 2\}$
Retained experts K	$\{2, 3\}$
Aggregation	precision / utility-tilted
Temperature τ_{route}	0.7 (fixed)
Meta-policy rates (η, β)	$\eta \in \{\text{anytime}, 0.5, 2, 8\}$, $\beta \in \{0, 0.05\}$ (selected: anytime, 0.05)
Confidence gate $q_{\min}$	0.05 (fixed)
Expert service cost	Table 6
Universe size	by rule

The allocator and router grids are searched on the per-fold validation slice, in the two stages of Appendix B.1; the meta-policy grid is selected once, on the pooled utility of the selection folds F1–F2 (2008–2013), where the do-no-harm gate is also decided.

D.1 Reproduction map

Table 9 maps each table to the script that produces it. The pipeline is `t1_F_cells.py` (walk-forward operating points and baselines for every cell) → `t1_final_eval.py` (replay of every policy and baseline at every cost) → `t1_final_report.py`, `t1_final_pairwise.py` and `t1_final_robust_tables.py`, which write the result files and table sources; the numbers in the text are copied from those files.

Table 9: Reproduction map.

Table	Producer	Status
Evaluation cells	`t1_F_cells.py`	reproduces
Replay and metrics	`t1_final_eval.py`	reproduces
Tables, Sec. 5	`t1_final_report.py`, `t1_final_robust_tables.py`	reproduces
Tests	`t1_final_pairwise.py`, `acceptance.py`	reproduces
6	`modes.py` constants	current
8 grids	`main_experiment.py`, `run_modes.py`	current

"Reproduces": regenerated exactly from the released code's outputs. Bootstrap resamples use fixed seeds.

Artifact contents The code is organized as one engine plus evaluation scripts. The engine modules are: causal feature construction and the frozen ridge forecaster; the expert panel, the confidence gate, the router and the consensus; the exact turnover-penalized allocator with its proximal step and unit tests (checked against a general-purpose solver); the daily simulator that charges the proportional cost against drifted holdings; the meta-policy (Hedge / fixed-share and the do-no-harm gate); the rule-based universe builder; and the baselines. The evaluation scripts are listed in Table 9. All data are daily adjusted closes from a single public source, fetched once and never spliced with another source. The US panel covers 2005-01-03 to 2026-09-18 for 500 symbols (485 stocks and 15 ETFs; 393 quoted from 2005-01-03 and 401 by the end of 2005), of which 13 ETFs (SPY, nine sector funds, IWM, TLT, GLD) form the survivorship-free control universe. After quality control the held-out panels cover 2005 to 2026-09 for 284 European and 219 Japanese symbols, with returns in local currency. Constituent lists are as of 2026 and projected backwards (Section E).

E LIMITATIONS

Asset classes. We evaluate only on equity universes in the U.S., Japan and Europe. We have not tested other asset classes such as fixed income, commodities, currencies or cryptocurrencies, which have different cost structures, trading hours and return dynamics. Whether the relative ranking of operating points and the benefit of the meta-policy carry over to these markets remains to be tested.

Backtest scope. All results come from historical backtests with a proportional cost model and weekly rebalancing. They do not account for market impact at larger capital sizes, execution delays or liquidity constraints in live trading, so realized performance may differ.

Operating-point set. The meta-policy selects among three fixed operating points. Its performance is bounded by the quality of these points, and we have not studied how it behaves with a larger or learned set of modes.

F IMPACT STATEMENT

This work concerns automated support for high-stakes financial decision-making. Its most positive potential impact is methodological: it offers a principled framework for studying when and how multiple financial tools should be coordinated before capital is allocated.

Societal Consequences **Financial risk:** Backtest performance can be mistaken for deployable trading quality. We emphasize that the paper is a research benchmark rather than a production system. **Automation risk:** Hierarchical agents may make autonomous portfolio decisions appear more reliable than they are. Human oversight and audit trails remain necessary. **Bias and concentration:** Financial models can inherit biases from historical data. The hierarchy partly mitigates this through portfolio caps and turnover penalties, but careful monitoring remains essential.