Learning to Switch:
A Weekly RL Meta-Controller over HRP/HERC Portfolio Experts

Payam Taebi [1]

[1]Sharif University of Technology

Motivation

- Market regimes shift; covariance estimates are noisy; naïve optimization produces unstable weights.
- Practical portfolios must respect **costs** and **limits** (turnover, per-asset caps).
- Objective: a **transparent**, **constraint-aware** allocator that adapts to regimes without overfitting.

Methodology Overview

Expert set. Construct long-only HRP and HERC portfolios over multiple daily lookbacks (e.g., 60, 120, 252, 504, 756, 1008 days).

- **HRP** (Hierarchical Risk Parity): (i) build correlation distance tree, (ii) quasi-diagonalize covariance, (iii) allocate risk top-down by cluster variance; no Σ^{-1} inversion.
- **HERC** (Hierarchical Equal Risk Contribution): on the same tree, solve for weights that equalize risk contribution at each split (downside-aware sizing variants allowed).

Meta-controller. Weekly (Friday) a PPO policy selects one *expert* or *HOLD*. The selection is discrete (13 actions: 12 experts + HOLD), making behavior interpretable.

State s_t. Compact features using information up to t only:

- For each expert: realized performance over past {1, 4, 12} weeks (36 dims).
- Regime cues (4 dims): EWMA volatility proxy; trend flags for SPY and TLT; a simple weekly stress flag (e.g., SPY down and IEF up).

Execution. Map action $\to$ target weights $w_t^\star$; enforce:

- L1 turnover cap $\leq$ **20%** per week; per-asset cap **35%**; up to **5%** cash if caps bind.
- Trading cost: **2 bps** per unit turnover.

Reward (net). Decide at t, realize at $t+1$:

$$r_{t+1} = w_t^\top r_{t+1} - c_{\text{bps}} \|w_t - w_{t-1}\|_1 - \kappa \|w_t - w_{t-1}\|_1 + b_{\text{hold}} \cdot \mathbf{1}\{\text{stress \& HOLD}\}.$$

Here w_t are executed weights after caps; c_{bps} is the cost rate (2 bps); κ shapes turnover; b_{hold} is a small stress-only nudge.

Data & Protocol

- **Universe:** 10 ETFs (SPY, IEFA, EEM, IEI, IEF, TLT, LQD, HYG, GLD, DBC).
- **Cadence:** Weekly decisions; returns from adjusted close.
- **Splits:** Train 2010–2018, Validate 2019, **Out-of-Sample (OOS)** 2020–2025.
- **Training:** PPO with small MLP; select learning rate by validation Sharpe; fixed seed; identical frictions across methods.

OOS Equity (Net of Costs)

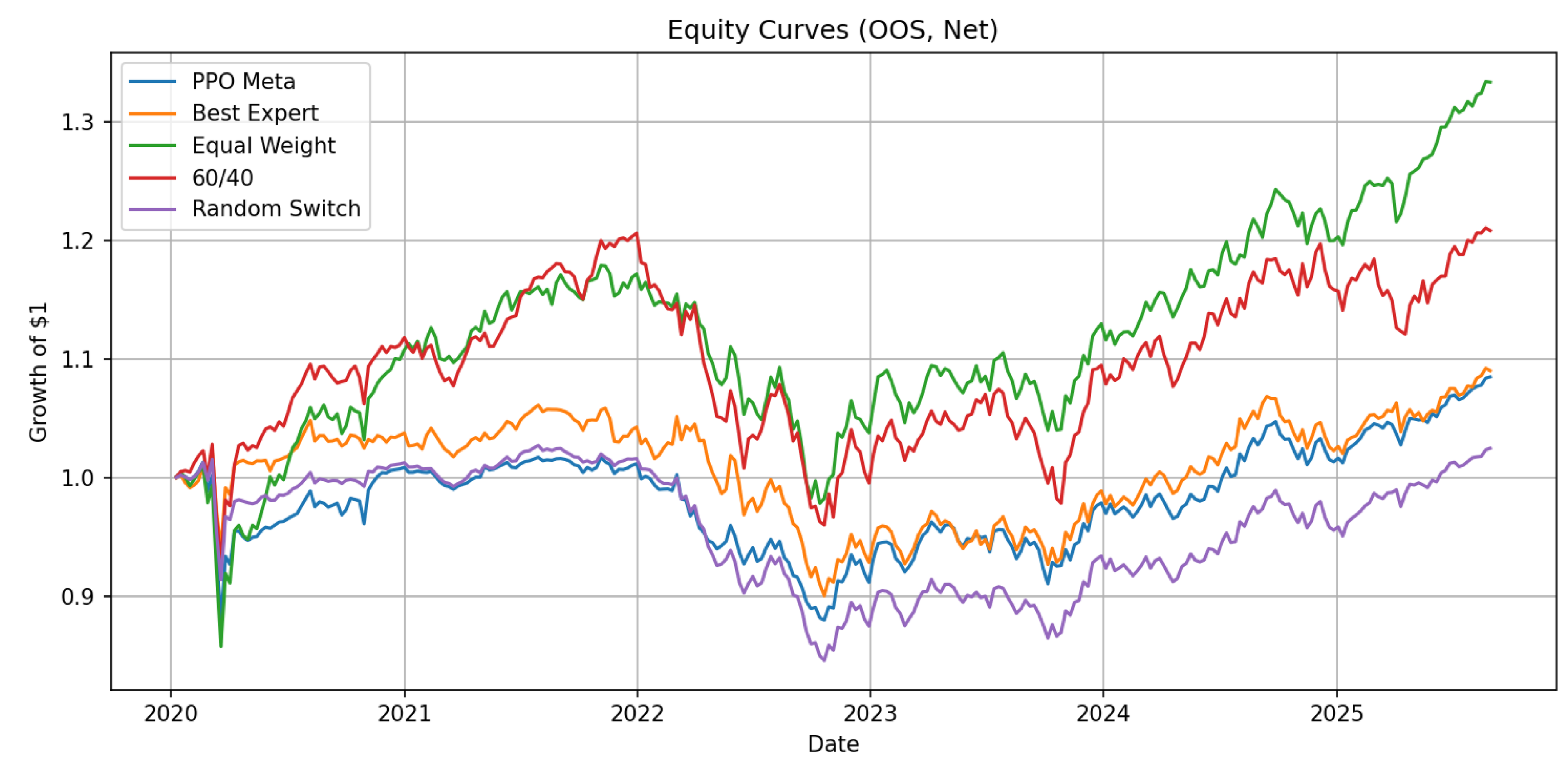

Figure 1. Growth of $1 (2020–2025), net of costs. All strategies share identical caps and cost model.

Quantitative Results (OOS)

Strategy	Ann.%	Sharpe	CVaR5%	MaxDD	Avg TO	Weeks
PPO_Meta	1.45	0.236	−2.28%	–	8.62%	295
BestExpert (HERC_CDaR5_L60)	1.54	0.251	−2.18%	−15.14%	12.13%	295
EqualWeight	5.21	0.591	−2.99%	−17.02%	0.00%	295
60/40	3.39	0.420	−2.62%	−20.40%	0.43%	295
RandomSwitch	0.44	0.099	−2.05%	−17.65%	16.67%	295

CVaR5%: mean of the worst 5% weekly returns (lower is better). MaxDD: maximum peak-to-trough drawdown on net equity. Avg TO: mean weekly L1 turnover.

Policy Behavior

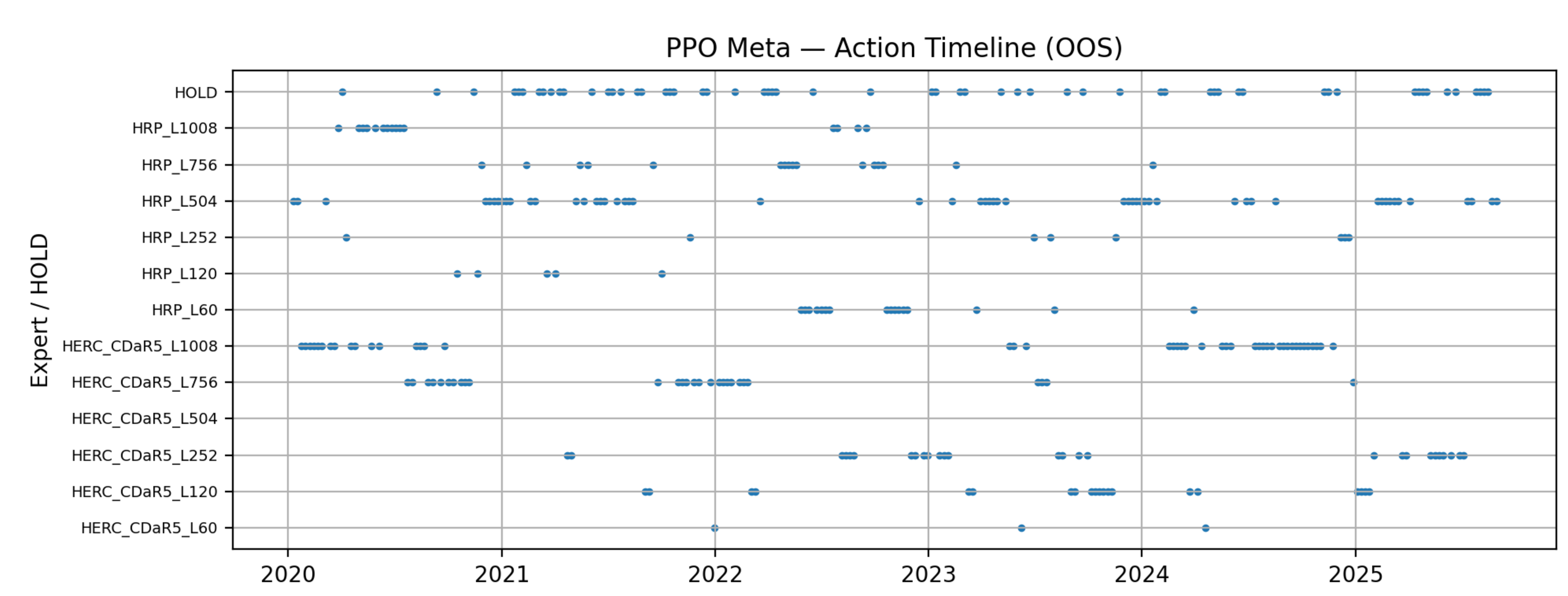

Figure 2. Action timeline: expert choice or HOLD by week. Blocks indicate regime persistence; HOLD appears in choppy periods.

Risk & Trading Characteristics

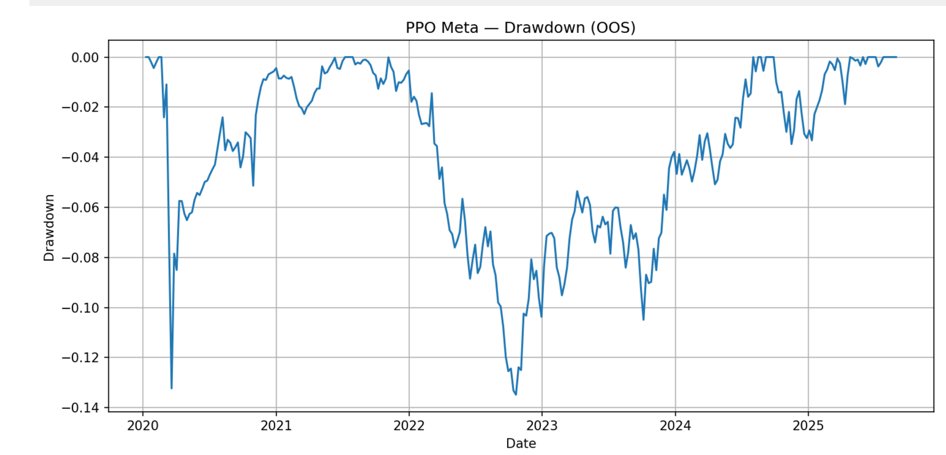

Figure 3. PPO drawdown over OOS. Shallower declines and steady recoveries support the smoother equity path.

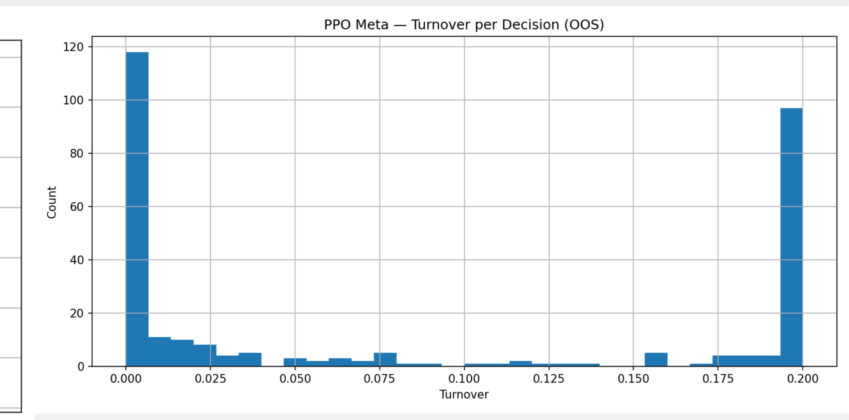

Figure 4. PPO weekly turnover. Most weeks are low; occasional cap hits occur after regime transitions.

Portfolio Weights Over Time

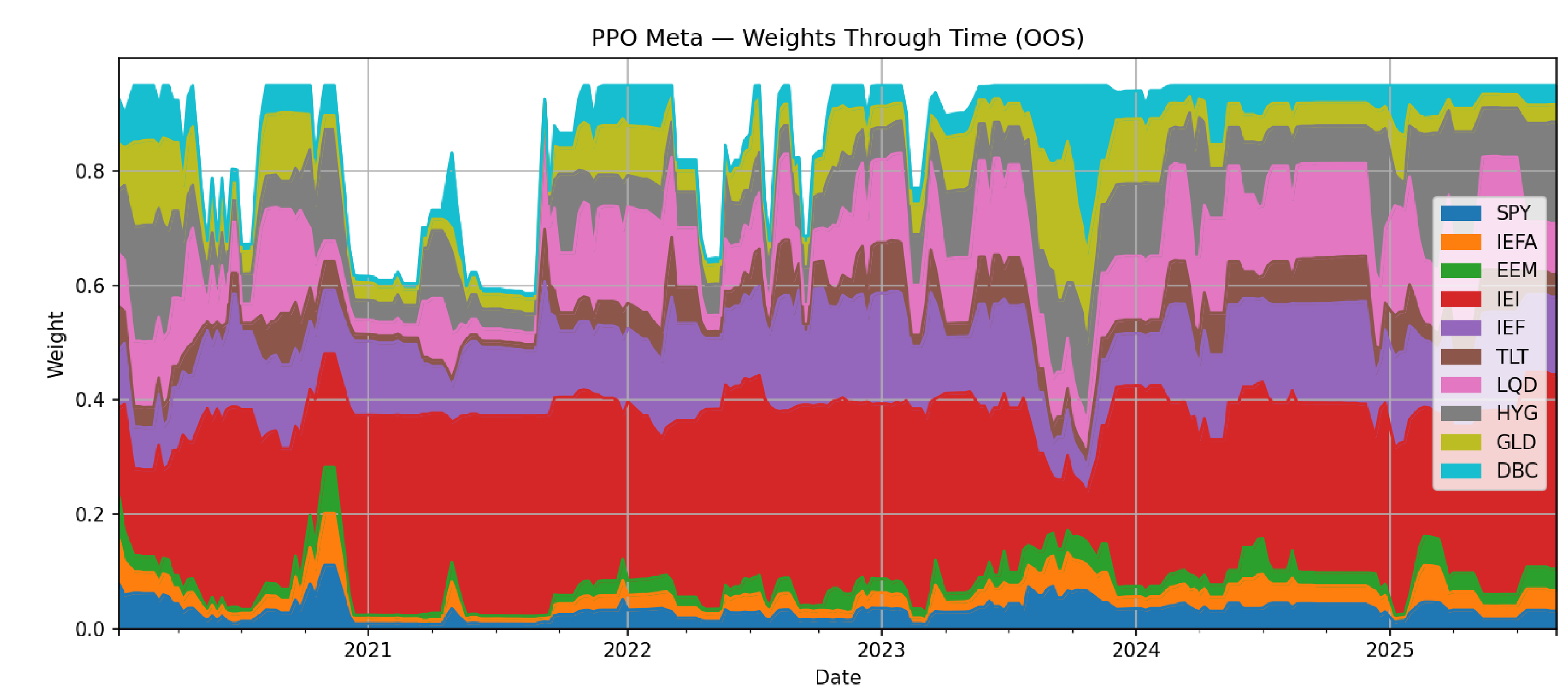

Figure 5. Executed weights heatmap (OOS). Rotations across equity, duration, credit, and diversifiers with caps respected.

Interpretation

- Discrete switching among robust experts yields **interpretable** decisions and **controlled** trading.
- Compared with static baselines, the meta-controller emphasizes **drawdown and tail control** with modest cost drag.
- Equal-Weight attains higher Sharpe in this window; the RL approach prioritizes path stability under identical frictions.

Limitations & Future Work

- Explore mild downside-aware shaping in reward; small action-change penalty; expand expert set (include 1/N).
- Walk-forward validation and capacity analysis for deployment.