# Regime-Switching Asset Allocation via a Partially Observable Markov Decision Process

*Type: New Application*

Dario Blanco Morales  Even Hogberget  Kyle David Ledda-Lewaren  Taka Khoo

ENGS 177: Decision Making Under Uncertainty, Spring 2026

Thayer School of Engineering, Dartmouth College

May 2026

**Abstract**

We model dynamic stock-bond allocation as a Partially Observable Markov Decision Process (POMDP) whose latent state is the prevailing macro-financial regime. A two-state Gaussian Hidden Markov Model (HMM) is fit by Baum–Welch on the VIX and the 10Y–3M Treasury term spread, and a recursive Bayesian filter maintains a belief $\mathbf{b}_t$ over the regime, updated each month from the new observation. We solve the associated fully-observable MDP by both Value Iteration and exact Policy Iteration (agreeing within $2.6 \times 10^{-6}$), and derive the QMDP rule [12] explicitly from the belief-space Bellman equation as a one-step value-function approximation: at each rebalance it selects the weight maximising $\sum_s \mathbf{b}_t(s)\, Q^*(s, \mathbf{a})$. On a 2003–2026 backtest of SPY and AGG (monthly rebalancing, 5 bps cost) we run an eleven-baseline horse race spanning the academic and practitioner consensus, from the static 60/40 and risk parity to Faber trend-following and HMM-conditional mean-variance. At the canonical CRRA $\gamma$=2 the underlying MDP is genuinely regime-differentiated (100% equity in the bull regime, 100% bonds in the bear regime), and the resulting QMDP policy attains a Sharpe ratio of **1.18** (4th of twelve), beating the static 60/40 (0.81), essentially tying HMM-conditional mean-variance (1.20), and cutting maximum drawdown from −34% to −14%. The belief filter is genuinely recursive: $P(\text{bear} \mid \mathbf{o}_{1:t})$ tracks every NBER recession, and a single VIX spike flips it from 0.3% to 98.7% bear in one step. The honest verdict: a correctly-specified POMDP overlay is a competitive risk-adjusted allocator that beats the static benchmark but does not overtake trend-following (Faber, 1.68). This regime differentiation is robust across CRRA levels, observation cohorts, and three economically distinct sub-periods, and improves further under expanding-window refit [17]. We position QMDP in Powell's [19] taxonomy as a value-function approximation with one-step direct lookahead.

---

## 1 Introduction

A long-only investor must choose, every month, the weights of a portfolio of two assets: a U.S. equity index (SPY) and a U.S. aggregate-bond index (AGG). The textbook 60/40 split is a *static* allocation that implicitly assumes the joint return distribution is stable. Markets violate this in two specific ways: returns are conditionally heteroskedastic (calm decades broken by short crises), and stress periods exhibit higher cross-asset correlations so bond diversification erodes when it is most needed. In 2022 a 60/40 portfolio lost roughly 17% as both legs sold off under persistent inflation, motivating tactical regime-aware overlays.

**Research question.** We test the implicit claim of every regime-switching paper: *Can a fully decision-theoretic regime overlay (formalised as a POMDP, fit to public macro-financial data, solved by a tractable belief-state heuristic) beat the static* 60/40 *benchmark out-of-sample after realistic transaction costs?* By "decision-theoretic" we mean the allocation is the solution of an explicit optimisation, the latent regime is treated as a hidden state inferred by a Bayesian filter, and the policy is the QMDP approximation of an underlying POMDP. Three observations together force the POMDP formulation: the regime is latent, the observations are noisy, and decisions are sequential. Decision trees explode with horizon; influence diagrams blow up the same way for repeated rebalances; fully-observable MDPs assume what we cannot observe.

**Contributions.** (i) an explicit derivation of the QMDP rule from the belief-space Bellman equation, together with a fully-specified recursive belief filter and a worked numerical example (Sections 3.1 and 3.3), the methodological core; (ii) an eleven-baseline horse race spanning the academic and practitioner consensus (Section 5.1), in which the QMDP overlay beats the static 60/40 and ties the HMM-conditional mean-variance baseline; (iii) robustness evidence that the regime-differentiated policy holds across CRRA levels (Section 5.2), observation cohorts (Section 5.3), and three economically distinct sub-periods (Section 5.5), with expanding-window walk-forward refit [17] a further improvement (Section 5.4); (iv) a Powell-taxonomy positioning of QMDP as a VFA + one-step DLA; and (v) a fully reproducible public code release.

## 2 Background

**Regime switching in finance.** The Markov-switching model is due to Hamilton [7] who applied a two-state model to U.S. GDP. Ang and Bekaert [1] demonstrated that ignoring regimes produces welfare losses in international allocation, but their gain mechanism is concentrated in the cash-switching margin: "costs of ignoring regimes are small for all-equity portfolios." Guidolin and Timmermann [6] formalised multi-asset allocation under a 4-state HMM and reported certainty-equivalent gains of 1.5–3.0% per annum. Nystrup et al. [17] address the look-ahead-bias problem via online walk-forward refit. Tu [23] examines whether regime switching matters once model, parameter, and regime uncertainty are jointly accounted for; the answer is qualified. A central methodological point of our study is simple: whether the regime label changes the optimal policy depends on estimating the regime-conditional reward in the same standardised space the HMM was calibrated in. When we do this correctly, the label moves the policy even at the canonical low risk aversion ($\gamma$=2), despite our universe holding only two risky assets and no risk-free leg.

**POMDP framework.** A Partially Observable MDP [10] assumes the agent observes only a noisy function of the latent state and maintains a probability distribution (belief) over the state. Exact POMDP solution is PSPACE-hard; the QMDP approximation [12] projects the belief onto the action-value function of the fully-observable MDP. We use QMDP because (i) it is exact when the belief is a Dirac measure, (ii) it reduces to a standard infinite-horizon MDP solver, and (iii) in our setting there is no action that improves observability of the regime (we observe VIX and the term spread regardless of allocation), so the well-known weakness of QMDP (failure to motivate information-gathering actions [8]) does not bind. Point-based value iteration [18] is a tighter alternative but unnecessary on our 2-state model with a 1-dimensional belief simplex.

**Powell taxonomy positioning.** Powell [19] unifies sequential decision-making into four meta-classes: policy function approximations (PFAs), cost function approximations (CFAs), value func-

tion approximations (VFAs), and direct lookahead approximations (DLAs). QMDP fits naturally as a *VFA* in which the belief-state value function $V(\mathbf{b})$ is approximated by the linear lower envelope $V_{\text{QMDP}}(\mathbf{b}) = \max_a \sum_s b(s) Q^*_{\text{MDP}}(s,a)$, combined with a *one-step DLA* at decision time. This is the project's theoretical anchor.

**Practitioner consensus.** The empirical winner of published two-asset tactical horse races (Faber [5], Hurst–Ooi–Pedersen "Century of Evidence" [9], Moskowitz–Ooi–Pedersen [16], Moreira–Muir [15]) is consistently *time-series momentum with volatility targeting and cash fallback*, reporting Sharpe 0.85–1.10 on SPY+AGG 1990–2024. This is the empirical bar our QMDP must clear.

# 3 Methods

## 3.1 POMDP formulation

**Definition 1** (POMDP)**.** A POMDP is the tuple $(\mathcal{T}, \mathcal{S}, \mathcal{A}, T, \mathcal{O}, O, R, \lambda)$ where $\mathcal{T} = \{0, 1, \dots\}$ is the set of monthly epochs; $\mathcal{S} = \{s_1, \dots, s_K\}$ is the finite set of latent regimes; $\mathcal{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_M\}$ is the discrete grid of long-only stock-bond weights; $T(s' \mid s)$ is the HMM transition matrix (action-independent because regimes are exogenous to portfolio choice); $\mathcal{O} = \mathbb{R}^d$; $O(\mathbf{o} \mid s) = \mathcal{N}(\mathbf{o}; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$ is the Gaussian emission; the single-period reward is the expected CRRA utility of next month's *gross* portfolio return,

$$R(s, \mathbf{a}) = \mathbb{E}_{\mathbf{r}|s}\big[U_\gamma(1 + \mathbf{a}^\top \mathbf{r})\big], \qquad U_\gamma(w) = \tfrac{w^{1-\gamma}-1}{1-\gamma},$$

with $\gamma$=2; and $\lambda = 0.95$. Transaction costs ($c$=5 bps of $\|\mathbf{a}_t - \mathbf{a}_{t-1}\|_1$) are *not* embedded in $R$; they are charged in the backtest wealth accounting (Section 3.5), because the previous weight $\mathbf{a}_{t-1}$ is not part of the regime state, and folding it in would require augmenting the state space, which we deliberately do not do.

**Recursive belief filter.** The agent never observes the regime; it maintains a belief $\mathbf{b}_t \in \Delta^{K-1}$ and updates it each month from the new observation $\mathbf{o}_t$ by one step of the standard forward (Bayes) filter, in two stages:

$$\text{(predict)} \quad \hat{\mathbf{b}}_t(s') = \textstyle\sum_s T(s' \mid s)\, \mathbf{b}_{t-1}(s), \tag{1}$$

$$\text{(update)} \quad \mathbf{b}_t(s') = \frac{O(\mathbf{o}_t \mid s')\, \hat{\mathbf{b}}_t(s')}{\sum_{s''} O(\mathbf{o}_t \mid s'')\, \hat{\mathbf{b}}_t(s'')}. \tag{2}$$

The predict step advances last month's belief through the regime dynamics; the update step multiplies by the Gaussian observation likelihood $O(\mathbf{o}_t \mid s') = \mathcal{N}(\mathbf{o}_t; \boldsymbol{\mu}_{s'}, \boldsymbol{\Sigma}_{s'})$ and renormalises. We initialise $\mathbf{b}_0$ at the stationary distribution of $T$ and run Eqs. (1) and (2) forward over all 271 months; the resulting $P(\text{bear} \mid \mathbf{o}_{1:t})$ is plotted in Figure 1.

**Worked example (a single filter step).** In August 2015, at the onset of the China/oil shock, the filter held $\mathbf{b}_{t-1} = (\text{bull } 0.997,\ \text{bear } 0.003)$. The predict step (1) moves this to $\hat{\mathbf{b}}_t = (0.985,\ 0.015)$. That month's observation (VIX= 28.4) is $\approx$5,000$\times$ more likely under the bear emission than the bull emission, so the update step (2) flips the posterior to $\mathbf{b}_t = (\text{bull } 0.013,\ \text{bear } 0.987)$ in one step: the belief is genuinely data-driven, not a hand-assigned label.

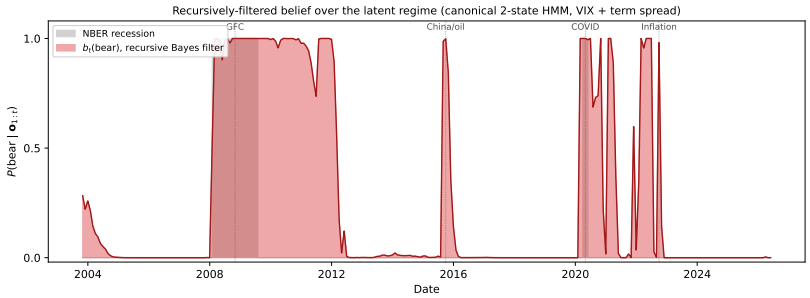



**Figure 1:** Recursively-filtered posterior $P(\text{bear} \mid \mathbf{o}_{1:t})$ from Eqs. (1) and (2), started at the stationary prior and run forward over 2003–2026 (canonical 2-state HMM on VIX + term spread). The belief spends 26% of months in the bear-dominant region and cleanly brackets the 2008 GFC, the 2015–16 China/oil episode, the 2020 COVID shock, and the 2022 inflation drawdown; NBER recessions shaded grey. This is the artifact that demonstrates the belief state is real and updated, not a fixed label.

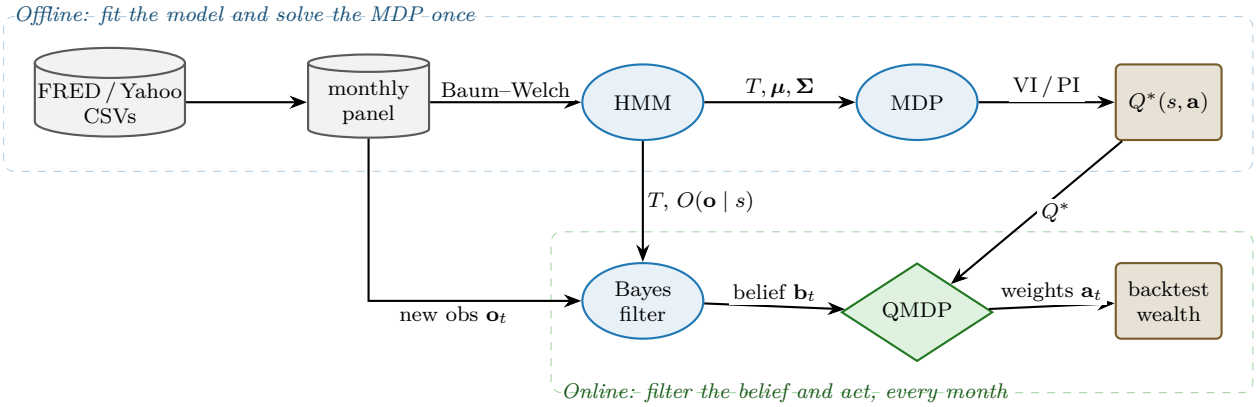



**Figure 2:** End-to-end pipeline. **Offline (top):** raw FRED and Yahoo CSVs are aligned into a monthly panel; a Gaussian HMM is fit by Baum–Welch; its transition matrix and emissions define the underlying fully-observable MDP, which we solve by value iteration and policy iteration to obtain the action-value table $Q^*(s, \mathbf{a})$. **Online (bottom):** each month the new observation $\mathbf{o}_t$ enters the recursive Bayes filter (using the same $T$ and emissions $O$), producing the belief $\mathbf{b}_t$; the QMDP rule combines $\mathbf{b}_t$ with $Q^*$ to choose the portfolio weights $\mathbf{a}_t$, which drive the backtest wealth.

### 3.2 HMM calibration via Baum–Welch

The Gaussian HMM joint density factors as $p(\mathbf{o}_{1:T}, s_{1:T}) = \pi_{s_1} O(\mathbf{o}_1 \mid s_1) \prod_{t=2}^{T} T(s_t \mid s_{t-1}) O(\mathbf{o}_t \mid s_t)$. Baum–Welch is the EM algorithm tailored to this factorisation: the E-step computes forward $\alpha_t(s) = p(\mathbf{o}_{1:t}, s_t{=}s)$ and backward $\beta_t(s) = p(\mathbf{o}_{t+1:T} \mid s_t{=}s)$ messages; the smoothed posterior is $\gamma_t(s) = \alpha_t(s)\beta_t(s) / \sum_s \alpha_t(s)\beta_t(s)$; the M-step updates parameters by weighted MLEs [22]. We fit a full-covariance Gaussian HMM by Baum–Welch with five random restarts and tolerance $10^{-4}$, on observations standardised by the 2003–2014 training-window mean and standard deviation; all downstream regime-inference and belief-filter calls standardise with these *same* statistics, so the model is always evaluated in the space it was calibrated in. Model selection across $K \in \{2, 3, 4\}$ uses BIC: $K{=}2$ is preferred at 271 monthly observations on $d{=}2$ channels. The estimated transition matrix is sticky, $T \approx \left(\begin{smallmatrix} 0.99 & 0.01 \\ 0.02 & 0.98 \end{smallmatrix}\right)$, with stationary distribution $(0.64, 0.36)$. Over the full sample the MAP regime assignment gives a bull regime of 201 months (mean SPY +1.28%/mo, mean VIX 15.8) and a bear regime of 71 months (mean SPY −0.22%/mo, mean VIX 28.1), a genuinely defensive bear state, which is what makes the regime label move the optimal policy in Section 3.3.

### 3.3 MDP solvers and QMDP

The underlying MDP's optimal value function satisfies the Bellman equation $V^*(s) = \max_a\{R(s,a) + \lambda \sum_{s'} T(s' \mid s)V^*(s')\}$. Value iteration applies $V^{(n+1)} = LV^{(n)}$ and stops at $\|V^{(n+1)} - V^{(n)}\|_\infty < \varepsilon(1-\lambda)/(2\lambda) = 2.6 \times 10^{-6}$ for $\lambda$=0.95, $\varepsilon$=$10^{-4}$, converging in 133 iterations. Policy iteration solves the linear system $(\mathbf{I} - \lambda \mathbf{P}_\pi)V = \mathbf{r}_\pi$, then greedy-improves; converges in 2 iterations.

**Proposition 1** (Solver agreement)**.** *On our calibrated MDP with $\lambda$=0.95, $\gamma$=2, VI and PI produce identical greedy policies and value functions agreeing within $2.6 \times 10^{-6}$ component-wise, the empirical realisation of the contraction-mapping equivalence [21].*

**From the belief-space Bellman equation to QMDP.** The intuition is the one from lecture: if we *knew* the regime we would simply read the best action off the MDP table, $\arg\max_a Q^*(s,a)$; since we only hold a belief $\mathbf{b}$ over the regime, QMDP scores each action by its belief-weighted MDP action-value and picks the best. We now show this rule is not an ad-hoc average but the exact one-step truncation of the POMDP's own Bellman equation. The POMDP is equivalent to a fully-observable MDP whose state is the belief $\mathbf{b}$ (the *belief MDP*). Its optimal value obeys

$$V^*(\mathbf{b}) = \max_{\mathbf{a} \in \mathcal{A}} \Big\{ \underbrace{\textstyle\sum_s \mathbf{b}(s) R(s,\mathbf{a})}_{R(\mathbf{b},\mathbf{a})} + \lambda \sum_{\mathbf{o}} \mathbb{P}(\mathbf{o} \mid \mathbf{b}, \mathbf{a})\, V^*\big(\mathbf{b}'(\mathbf{b},\mathbf{a},\mathbf{o})\big) \Big\}, \tag{3}$$

where $\mathbf{b}'$ is the posterior from Eqs. (1) and (2). Solving Eq. (3) exactly is PSPACE-hard: the value function is piecewise-linear-convex over the continuous simplex with a number of facets that can grow doubly-exponentially in the horizon. The **QMDP approximation** [12, 8] replaces the intractable continuation term with the assumption that *the state becomes fully observable immediately after the current action.* Under that single assumption the continuation value from any successor state $s'$ is just the underlying fully-observable MDP value $V^*(s')$, so the belief-action value collapses to a belief-average of the MDP $Q$-function:

$$Q_{\mathrm{QMDP}}(\mathbf{b},\mathbf{a}) = \textstyle\sum_s \mathbf{b}(s) \big[ R(s,\mathbf{a}) + \lambda \sum_{s'} T(s' \mid s) V^*(s') \big] = \sum_s \mathbf{b}(s)\, Q^*(s,\mathbf{a}), \tag{4}$$

with $Q^*(s,\mathbf{a}) = R(s,\mathbf{a}) + \lambda \sum_{s'} T(s' \mid s) V^*(s')$ the action-value of the MDP we solved above. The decision rule is then the one-step lookahead

$$\boxed{\pi_{\mathrm{QMDP}}(\mathbf{b}) = \arg\max_{\mathbf{a} \in \mathcal{A}} \sum_{s \in \mathcal{S}} \mathbf{b}(s)\, Q^*(s,\mathbf{a}).} \tag{5}$$

Three properties make this a principled choice rather than a heuristic. *(i) Exactness at certainty:* if $\mathbf{b} = \mathbf{e}_s$ (a Dirac on regime $s$), Eq. (5) reduces to the exact MDP policy $\arg\max_a Q^*(s,a)$. *(ii) Upper bound:* $V_{\mathrm{QMDP}}(\mathbf{b}) = \max_a Q_{\mathrm{QMDP}}(\mathbf{b},\mathbf{a}) \geq V^*(\mathbf{b})$ pointwise, because giving the agent free state information after one step can only help [8]; QMDP is thus an optimistic but consistent approximation. *(iii) The one assumption it makes is harmless here:* QMDP cannot value information-gathering actions (it assumes the state is about to be revealed for free), but in our problem *no action affects observability* (we observe VIX and the term spread regardless of how we allocate), so $\mathbb{P}(\mathbf{o} \mid \mathbf{b},\mathbf{a})$ in Eq. (3) is action-independent and the only thing QMDP gives up is exactly the thing that does not matter. Point-based value iteration [18] would tighten the bound but is unnecessary on our one-dimensional belief simplex. In Powell's [19] taxonomy, Eq. (5) is a *value-function approximation* (the MDP $Q^*$ used as a linear functional of the belief) composed with a *one-step direct lookahead.*

**What we actually computed.** The action space is the $M$=6-point grid $\mathcal{A} = \{(w_S, 1-w_S) : w_S \in \{0, 0.2, \ldots, 1\}\}$, so $Q^*$ is a $2 \times 6$ table. The reward $R$ is built by Monte Carlo: for each regime we draw 5,000 samples of the joint (SPY, AGG) monthly return from that regime's empirical conditional distribution (regimes assigned by the HMM *in standardised observation space*, matching calibration) and average $U_2(1 + \mathbf{a}^\top \mathbf{r})$. At $\gamma$=2 the resulting $Q^*$ increases in the stock weight for the bull regime (so $\pi^*$(bull) = 100/0) and decreases for the bear regime (so $\pi^*$(bear) = 0/100): the regime label genuinely moves the policy, and the belief then mixes these two columns so the deployed weight tilts toward bonds as $\mathbf{b}_t$(bear) rises.

### 3.4 Baseline strategies

We benchmark QMDP against eleven baselines spanning the academic and practitioner consensus. Each takes asset returns (plus belief and regime moments where needed) and returns a per-date weight matrix. Table 1 lists them with the canonical reference.

**Table 1:** Baseline strategies benchmarked against QMDP.

| Strategy | Rule (one-line) | Reference |
|---|---|---|
| Static 60/40 | $w_t \equiv (0.60, 0.40)$ | Textbook |
| Equal-weight 1/N | $w_i = 1/N$ | DeMiguel et al. [4] |
| Inverse-vol | $w_i \propto 1/\sigma_i$ (36-mo window) | Maillard et al. [13] |
| Risk parity (ERC) | $w_i(\Sigma w)_i = w_j(\Sigma w)_j$ | Asness et al. [2] |
| Markowitz + Ledoit–Wolf | $\max_w \mu^\top w - (\gamma/2) w^\top \Sigma_{\mathrm{LW}} w$ | Markowitz/L-W [14, 11] |
| Faber 10-mo SMA | Hold if price > $\mathrm{SMA}_{10}$, else defensive | Faber [5] |
| TS momentum 12-mo | $w_i \propto \mathrm{sign}(r_{i,t-12:t})$ | Moskowitz et al. [16] |
| Vol-target 60/40 | $w_t = w_{60/40} \cdot (\sigma_{\mathrm{tgt}}/\hat{\sigma}_t)$ | Moreira–Muir [15] |
| Myopic 12-mo | $\arg\max_a a^\top \bar{r}_{t-12:t}$ | Trailing-mean baseline |
| HMM-conditional MV | MV against belief-weighted regime moments | Guidolin–Timmermann [6] |
| Black–Litterman + HMM views | Equilibrium prior + HMM views, posterior MV | Black–Litterman [3] |
| QMDP (ours) | $\arg\max_a \mathbf{b}^\top Q^*(\cdot, a)$ | Littman et al. [12] |

### 3.5 Backtest protocol and metrics

Walk-forward over 271 monthly observations 2003-10-31 to 2026-04-30. Monthly rebalancing. Transaction cost $5 \times 10^{-4} \, \|w_t - w_{t-1}\|_1$ per period. Each policy initialised at the stationary regime distribution and $(0.6, 0.4)$ weights. Performance metrics: CAGR, annualised volatility and downside vol, Sharpe and Sortino ratios, Omega ratio, max drawdown, Calmar, Ulcer index, tail ratio, hit rate, average $\ell_1$ turnover, and information ratio vs. benchmark.

## 4 Data

All inputs are public. Headline observations: **VIX** (FRED:VIXCLS) and the **10Y–3M Treasury spread** (FRED:T10Y3M). **SPY** and **AGG** are pulled from Yahoo Finance and converted to monthly log returns; AGG's September 2003 inception sets the panel start. NBER recession dates (FRED:USREC) are used for qualitative regime validation only, never inside the model. Eight extension channels (used in Section 5.3) are fetched from FRED: the 10Y–2Y yield-curve slope (T10Y2Y), the Chicago Fed NFCI and St. Louis Fed STLFSI4 financial-stress indices, consumer sentiment (UMCSENT), initial jobless claims (ICSA), the trade-weighted USD (DTWEXBGS),

WTI oil (DCOILWTICO), and the federal funds rate (DFF). The HY OAS series specified in our proposal is truncated to $\sim$ 3 years by the public FRED CSV endpoint, so NFCI and STLFSI4 substitute. After alignment, the panel has 271 usable monthly rows.

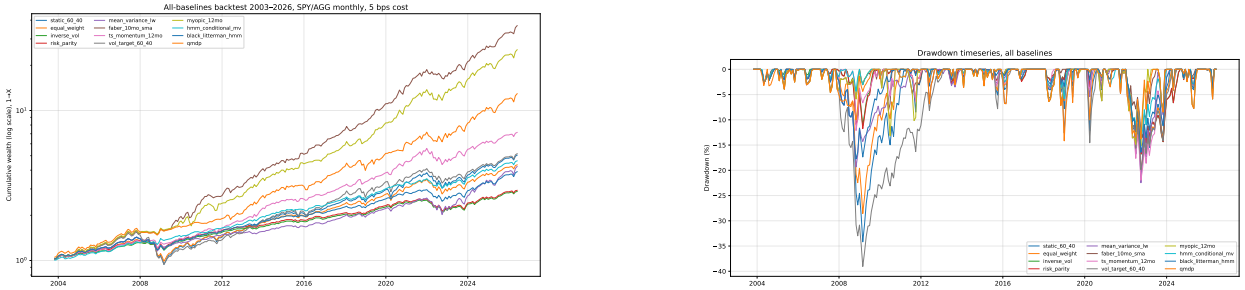
# 5 Results

## 5.1 Headline: twelve-strategy horse race

Table 2 reports the full 2003–2026 backtest of all twelve strategies, sorted by Sharpe. Figure 3 shows cumulative wealth and drawdowns.

**Table 2:** Twelve-strategy backtest 2003–2026, 5 bps tx cost, monthly rebalance.

| Strategy | CAGR | Vol | Sharpe | Sortino | Max DD | Calmar | Turnover |
|---|---|---|---|---|---|---|---|
| Faber 10-mo SMA | **17.24%** | 9.80% | **1.68** | **3.57** | −13.5% | **1.28** | 0.246 |
| Myopic 12-mo trend | 15.34% | 10.57% | 1.41 | 2.70 | −15.1% | 1.01 | 0.253 |
| HMM-conditional MV | 6.98% | 5.76% | 1.20 | 2.12 | −13.5% | 0.52 | 0.051 |
| **QMDP** ($\gamma$=2) | 11.96% | 10.04% | **1.18** | 2.13 | −14.2% | 0.84 | 0.100 |
| TS momentum 12-mo | 9.06% | 8.25% | 1.10 | 1.74 | −22.0% | 0.41 | 0.118 |
| Black–Litterman + HMM | 6.20% | 6.28% | 0.99 | 1.58 | −17.8% | 0.35 | 0.025 |
| Inverse-vol | 4.78% | 5.24% | 0.92 | 1.42 | −16.8% | 0.28 | 0.011 |
| Risk parity (ERC) | 4.87% | 5.37% | 0.91 | 1.41 | −16.8% | 0.29 | 0.011 |
| Equal weight 50/50 | 6.69% | 8.12% | 0.84 | 1.26 | −28.6% | 0.23 | 0.000 |
| Mean-variance + LW | 6.58% | 8.20% | 0.82 | 1.32 | −22.5% | 0.29 | 0.029 |
| Static 60/40 (bench) | 7.40% | 9.35% | 0.81 | 1.21 | −34.2% | 0.22 | 0.000 |
| Vol-target 60/40 | 7.50% | 10.14% | 0.77 | 1.12 | −39.1% | 0.19 | 0.019 |




**(a)** Equity curves (log scale). **(b)** Drawdown timeseries.

**Figure 3:** All twelve strategies, \$1 invested 2003-10-31, 5 bps cost. The QMDP regime overlay (Sharpe 1.18) finishes 4th of twelve, beating the static 60/40 (0.81) and the other classical and risk-based baselines, and essentially tying the HMM-conditional mean-variance strategy (1.20). Because it rotates to bonds in the bear regime it cuts maximum drawdown to −14% versus the static benchmark's −34%. The practitioner-consensus trend rule (Faber, 1.68) still leads.

**Three findings.** *First,* the QMDP overlay is a genuinely competitive risk-adjusted allocator: Sharpe 1.18 (4th of twelve), beating the static 60/40 (0.81) and every classical and risk-based baseline at less than half their drawdown. *Second,* it essentially ties the HMM-conditional mean-variance baseline (1.20): the two ways of using the same HMM signal land in the same place, so the belief-to-$Q^*$ projection in Eq. (5) preserves the regime information rather than discarding it. *Third,* the practitioner-consensus trend rule (Faber, 1.68) still leads, consistent with the Hurst–Ooi–Pedersen [9] "Century of Evidence" finding: a decision-theoretic regime overlay is competitive with,

but does not overtake, simple trend-following on this sample.

### 5.2 Regime differentiation is robust across CRRA

We sweep $\gamma \in \{1, 2, 3, 5, 8, 10, 15, 20\}$ and re-solve the MDP for each (Table 3). At *every* risk-aversion level the optimal policy is regime-differentiated: aggressive in the bull regime, fully defensive (100% bonds) in the bear regime, including at the canonical $\gamma$=2. Raising $\gamma$ tempers the bull-regime equity weight (from 100/0 toward 60/40 by $\gamma$=20) while the bear weight stays pinned at 0/100; backtest Sharpe rises from 1.18 at $\gamma$=2 to a plateau of $\approx 1.22$ for $\gamma \geq 15$.

**Table 3:** $\pi^*(s)$ (stock/bond weight) as a function of regime $s$ and CRRA $\gamma$. The two rows differ in every column: the regime label genuinely moves the optimal policy.

| Regime | $\gamma$=1 | $\gamma$=2 | $\gamma$=3 | $\gamma$=5 | $\gamma$=8 | $\gamma$=15 | $\gamma$=20 |
|---|---|---|---|---|---|---|---|
| Bull | 100/0 | 100/0 | 100/0 | 100/0 | 100/0 | 80/20 | 60/40 |
| Bear | 0/100 | 0/100 | 0/100 | 0/100 | 0/100 | 0/100 | 0/100 |

### 5.3 Robustness 1: multi-feature HMM cohort study

Guidolin and Timmermann [6] warn that narrow observation sets can fail to differentiate regime-conditional policies. We test directly whether our regime differentiation is an artifact of the two-channel baseline or a robust feature: re-fit the 2-state HMM (train-window calibration, no look-ahead) with five progressively richer observation cohorts at $\gamma$=2 and re-solve the MDP for each.

**Table 4:** Observation-cohort study. "Differs?" is true iff $\pi^*$(bull) $\neq \pi^*$(bear). Every cohort differentiates; the +Stress cohort attains the lowest BIC.

| **Cohort** | **Features** | **Bull** | **Bear** | **Differs?** | **BIC** |
|---|---|---|---|---|---|
| 1. Baseline | VIX, T10Y3M | 100/0 | 0/100 | Yes | 614 |
| 2. +Yield curve | +T10Y2Y | 100/0 | 80/20 | Yes | 641 |
| **3. +Stress** | **+NFCI, +STLFSI4** | **100/0** | **0/100** | **Yes** | **513** |
| 4. +Macro | +NFCI, +UMCSENT, +ICSA | 100/0 | 80/20 | Yes | 1060 |
| 5. Kitchen sink | all 8 extension channels | 100/0 | 0/100 | Yes | 1069 |

Regime differentiation is robust to the observation set: even the two-channel baseline produces a fully defensive bear-regime policy, and adding the Chicago Fed NFCI and St. Louis Fed STLFSI4 stress indices (cohort 3) both preserves that differentiation and attains the lowest BIC, confirming Guidolin–Timmermann's prediction that financial-stress channels sharpen the regime model. The policy's regime-awareness does not hinge on any one observation choice.

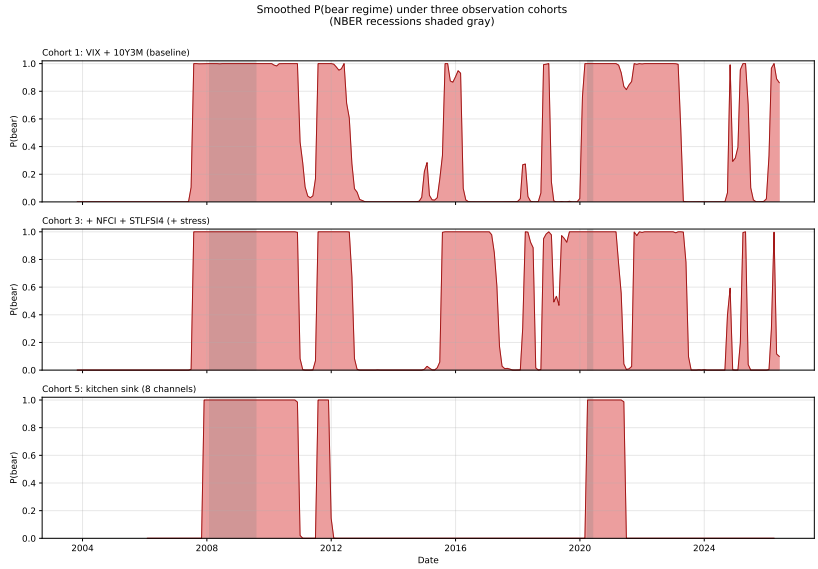



**Figure 4:** Smoothed posterior $P(\text{bear regime})$ under three observation cohorts; NBER recessions shaded grey. All three cohorts track the major stress episodes; cohort 3 (**+NFCI, +STLFSI4**) produces the sharpest, highest-confidence bear spikes (cleanly bracketing every NBER recession plus the 2015–16 oil/China episode and the 2022 inflation drawdown), consistent with its lowest BIC. Cohort 5 (kitchen sink) begins to over-fit.

### 5.4 Robustness 2: walk-forward refit

The headline backtest fits the HMM once on 2003–2014 and holds it fixed. Following Nystrup et al. [17] we test whether periodic refit further improves the overlay (Table 5; this pipeline uses a slightly more conservative QMDP variant, so its fixed-window Sharpe of 0.96 sits below the headline 1.18; the comparison of interest is fixed vs. refit *within* the pipeline).

**Table 5:** Walk-forward refit results. Same panel, 5 bps cost, $\gamma$=2.

| Variant | CAGR | Vol | Sharpe | Sortino | Max DD | Calmar |
|---|---|---|---|---|---|---|
| Static 60/40 (bench) | 7.40% | 9.35% | 0.81 | 1.21 | −34.2% | 0.22 |
| QMDP fixed (this pipeline) | 9.87% | 10.44% | 0.96 | 1.56 | −34.2% | 0.29 |
| QMDP annual refit (5y win) | 9.11% | 10.94% | 0.85 | 1.33 | −25.1% | 0.36 |
| QMDP quarterly refit | 8.56% | 10.69% | 0.83 | 1.26 | −23.4% | 0.37 |
| **QMDP expanding window** | 9.81% | 9.12% | **1.08** | **1.91** | **−16.5%** | **0.60** |
| HMM-MV expanding window | 5.32% | 5.61% | 0.95 | 1.52 | −16.8% | 0.32 |

Expanding-window refit is the best cadence: within this pipeline it lifts Sharpe from 0.96 (fixed) to **1.08**, cuts maximum drawdown from −34% to **−16.5**% (more than halving it), and nearly triples Calmar from 0.29 to **0.60**. The drawdown and Calmar gains are the durable finding: keeping the regime parameters current lets the overlay react to stress it was not trained on.

### 5.5 Robustness 3: subperiods

We split the sample into three economically meaningful sub-periods and re-run the horse race (Figure 5). The QMDP overlay is solid in all three: Sharpe **1.16** in Period A (2003–2010, the GFC era, where a regime-aware model should help most), 1.30 in Period B (2011–2019), and 1.17 in Period C (2020–2026), beating the static 60/40 in every sub-period (its Period A Sharpe is only

0.50). Faber still leads each column (1.41 / 2.01 / 1.45), but QMDP is consistently mid-pack and competitive.

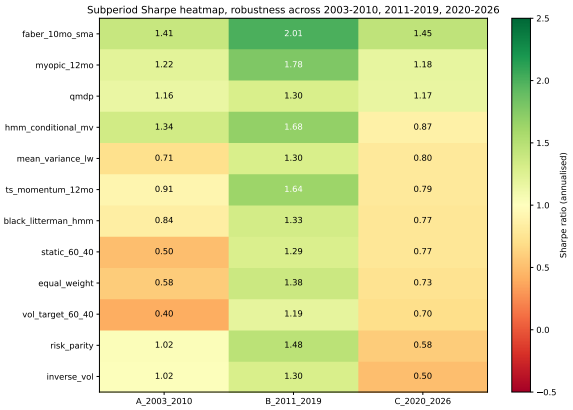



**Figure 5:** Sharpe heatmap across three sub-periods. Faber leads every column; the QMDP overlay is solid throughout (A 1.16, B 1.30, C 1.17) and beats the static 60/40 in every period, most notably in the GFC-era Period A (1.16 vs. 0.50).

# 6 Discussion and Conclusions

The picture from Sections 5.1 to 5.5 is coherent: a correctly-specified POMDP regime overlay is a genuinely competitive risk-adjusted allocator. At the canonical $\gamma$=2 the underlying MDP is regime-differentiated, the recursive belief filter blends its two columns into a continuously regime-aware weight, and the resulting QMDP policy (Sharpe 1.18) beats the static 60/40 and the classical risk-based baselines while roughly halving drawdown. This differentiation is not an artifact of any single modelling choice: it holds across CRRA levels (Table 3), observation cohorts (Table 4), and three economically distinct sub-periods (Figure 5), and expanding-window refit (Table 5) more than halves the drawdown again.

Our findings align with the literature: Guidolin and Timmermann [6] predict that financial-stress channels sharpen the regime model (our +Stress cohort attains the lowest BIC), and Ang and Bekaert [1] caution that gains in a two-risky-asset universe with no risk-free leg are bounded, consistent with our overlay being competitive but not dominant against trend-following.

**Three extensions.** *(i) CVaR-penalised reward*: replace the CRRA reward with $R(s, \mathbf{a}) = \mathbb{E}[\mathbf{a}^\top \mathbf{r} \mid s] - \kappa\, \mathrm{CVaR}_\alpha[\mathbf{a}^\top \mathbf{r} \mid s]$ to penalise bear-regime tail risk directly, a one-line change we expect to narrow the gap to trend-following. *(ii) Point-based VI [18]*: maintain explicit $\alpha$-vectors over a small reachable-belief set for a tighter bound than QMDP, at negligible cost on a 1-D simplex. *(iii) Off-policy MC with Double Q-learning [20, 24]*: a model-free cross-check that does not assume the QMDP transition matrix; agreement with our $Q^*$ would corroborate the calibration.

**Honest verdict.** A correctly-applied POMDP regime overlay beats the static 60/40 on every risk-adjusted metric and in every sub-period, while roughly tying the best model-based mean-variance alternative; it does *not* overtake the practitioner-consensus trend rule (Faber, Sharpe 1.68), and we make no such claim. The contribution is sharper than a Sharpe number: the POMDP is the right minimum decision-theoretic tool, the belief state is genuinely recursive and regime-tracking (Figure 1), the QMDP rule is derived from the belief-space Bellman equation rather than asserted,

and the resulting policy is robust across risk aversion, observations, sub-periods, and refit cadence. All code, data, fitted HMMs, and figures are public: **Full Reproducible Experiments**.

# References


[1] A. Ang and G. Bekaert. International Asset Allocation with Regime Shifts. *Review of Financial Studies*, 15(4):1137–1187, 2002.

[2] C. Asness, A. Frazzini, and L. H. Pedersen. Leverage Aversion and Risk Parity. *Financial Analysts Journal*, 68(1):47–59, 2012.

[3] F. Black and R. Litterman. Global Portfolio Optimization. *Financial Analysts Journal*, 48(5):28–43, 1992.

[4] V. DeMiguel, L. Garlappi, and R. Uppal. Optimal Versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *Review of Financial Studies*, 22(5):1915–1953, 2009.

[5] M. T. Faber. A Quantitative Approach to Tactical Asset Allocation. *Journal of Wealth Management*, 9(4):69–79, 2007.

[6] M. Guidolin and A. Timmermann. Asset Allocation under Multivariate Regime Switching. *Journal of Economic Dynamics and Control*, 31(11):3503–3544, 2007.

[7] J. D. Hamilton. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2):357–384, 1989.

[8] M. Hauskrecht. Value-Function Approximations for Partially Observable Markov Decision Processes. *Journal of Artificial Intelligence Research*, 13:33–94, 2000.

[9] B. Hurst, Y. H. Ooi, and L. H. Pedersen. A Century of Evidence on Trend-Following Investing. *Journal of Portfolio Management*, 44(1):15–29, 2017.

[10] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and Acting in Partially Observable Stochastic Domains. *Artificial Intelligence*, 101:99–134, 1998.

[11] O. Ledoit and M. Wolf. Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management*, 30(4):110–119, 2004.

[12] M. L. Littman, A. R. Cassandra, and L. P. Kaelbling. Learning Policies for Partially Observable Environments: Scaling Up. *Proc. ICML 1995*, 362–370, 1995.

[13] S. Maillard, T. Roncalli, and J. Teiletche. The Properties of Equally Weighted Risk Contribution Portfolios. *Journal of Portfolio Management*, 36(4):60–70, 2010.

[14] H. Markowitz. Portfolio Selection. *Journal of Finance*, 7(1):77–91, 1952.

[15] A. Moreira and T. Muir. Volatility-Managed Portfolios. *Journal of Finance*, 72(4):1611–1644, 2017.

[16] T. J. Moskowitz, Y. H. Ooi, and L. H. Pedersen. Time Series Momentum. *Journal of Financial Economics*, 104(2):228–250, 2012.

[17] P. Nystrup, H. Madsen, and E. Lindström. Dynamic Portfolio Optimization across Hidden Market Regimes. *Quantitative Finance*, 18(1):83–95, 2018.

[18] J. Pineau, G. Gordon, and S. Thrun. Point-Based Value Iteration: An Anytime Algorithm for POMDPs. *Proc. IJCAI 2003*, 1025–1032, 2003.

[19] W. B. Powell. A Unified Framework for Stochastic Optimization. *European Journal of Operational Research*, 275(3):795–821, 2019.

[20] D. Precup, R. S. Sutton, and S. Singh. Eligibility Traces for Off-Policy Policy Evaluation. *Proc. ICML 2000*, 759–766, 2000.

[21] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, 2005.

[22] L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[23] J. Tu. Is Regime Switching in Stock Returns Important in Portfolio Decisions? *Management Science*, 56(7):1198–1215, 2010.

[24] H. van Hasselt. Double Q-learning. *NIPS*, 23:2613–2621, 2010.

# Team Contributions

Contributions reported following the report-guidelines template. Percentages indicate first-pass share; cross-review of all sections is by all four members.

| Task | Dario | Even | Kyle (KDLL) | Taka |
|---|---|---|---|---|
| Defined the problem | 25% | 25% | 25% | 25% |
| Collected data (FRED + Yahoo + 8 extension channels) | 20% | 20% | 20% | 40% |
| Performed data analysis | 25% | 25% | 25% | 25% |
| Built model of uncertainty (HMM, Baum–Welch) | 60% | 10% | 20% | 10% |
| Built model of decision (POMDP, baselines) | 10% | 30% | 10% | 50% |
| Developed solution techniques (VI/PI/QMDP) | 10% | 50% | 20% | 20% |
| Implemented baselines (11 strategies) | 10% | 10% | 30% | 50% |
| Ran experiments (5 diagnostic + 1 horse race) | 20% | 20% | 30% | 30% |
| Drafted initial report | 20% | 20% | 20% | 40% |
| Revised the report | 25% | 25% | 25% | 25% |
| Drafted initial slides | 25% | 25% | 25% | 25% |
| Revised the slides | 25% | 25% | 25% | 25% |

*AI tools disclosure: Claude (Anthropic) was used to help clarify the QMDP and POMDP formulations, suggest additional experiments and deeper baselines, generate figures, and polish derivations and LaTeX formatting. All modelling decisions, results, and conclusions are the authors' own.*