Bayesian Approaches to Sequential Decision-Making in Uncertain Environments

Ansh Dani
Barrett, The Honors College
Arizona State University

May 2026

Abstract

Sequential decision-making under uncertainty requires agents to simultaneously learn unknown environment parameters and exploit current beliefs to maximize long-run wealth—the classic exploration-exploitation dilemma. Classical solutions, such as the Kelly Criterion and naive Bayesian plug-in estimators, are asymptotically optimal in stationary environments but fail catastrophically when the underlying return distribution shifts without warning, a phenomenon we term the *Sticky Prior Paradox*: posterior mass accumulated under one regime resists rapid reallocation when the true regime changes, causing agents to over-bet into an adversarial market.

This thesis makes three interrelated contributions. First, we establish a rigorous simulation framework—grounded in measure-theoretic probability and the Strong Law of Large Numbers—that enforces strict filtration constraints and Common Random Numbers methodology to enable statistically valid agent comparisons. Second, we propose a *Volatility-Augmented Hidden Markov Model* (Vol-HMM) that augments the standard univariate return observation with a rolling volatility feature and a CUSUM change-point trigger, reducing regime detection lag from approximately 15–20 time steps under sticky-prior agents to under 2 steps. Third, we implement a *Risk-Constrained Kelly* strategy via Constant Proportion Portfolio Insurance (CPPI), which provably bounds the maximum drawdown at a user-specified floor while quantifying the exact "cost of survival" in foregone geometric growth.

We validate all three contributions across three synthetic environments (stationary, slow-drift, adversarial shock) and four empirical S&P 500 epochs (2000 Dot-Com Collapse, 2008 Global Financial Crisis, 2020 COVID Crash, 2013–19 Bull Market), incorporating realistic transaction costs. Fisher's Exact Tests confirm that the Vol-HMM achieves statistically significant reductions in ruin probability relative to adversarial and Bayesian baselines ($p < 0.05$) under the harshest shock regimes. Sharpe and Sortino ratio analyses further demonstrate that the CPPI floor improves risk-adjusted returns during crisis epochs without sacrificing median wealth in stable environments.

Keywords: Kelly Criterion, Multi-Armed Bandits, Thompson Sampling, Hidden Markov Model, CPPI, Regime Detection, Bayesian Sequential Decision-Making, Non-Stationary Environments.

Contents

Notation Glossary

This glossary outlines the standard notations utilized throughout the Bayesian sequential and CPPI methodologies.

Symbol	Description
W_t	Agent's total wealth at timestep t.
f_t	The fractional allocation (betting fraction) to the risky asset during $[t-1, t]$.
r_t	Realized return of the risky asset at time t.
$\mathcal{F}_t$	Information filtration available up to time t, technically $\sigma(r_1, \ldots, r_t)$.
S_t	Hidden Markov state/regime at time t where $S_t \in \{0 = \text{Bull}, 1 = \text{Bear}\}$.
A	Regime transition matrix where $A_{ij} = \mathbb{P}(S_t = j \mid S_{t-1} = i)$.
α, β	Parameters of the Bayesian Beta posterior distribution for Bernoulli outcomes.
μ_j, σ_j	Expected return and standard deviation conditioned on state $S_t = j$.
ν	Degrees of freedom parameterizing the Student-t distribution heavy tails.
λ_t	CPPI dynamic leverage multiplier constraint $\lambda_t \in [0, 1]$.
D_t	Running maximum drawdown defined as $(W_{peak} - W_t)/W_{peak}$.

Chapter 1

Introduction

Decision making under uncertainty is a central challenge in both theoretical and applied contexts, ranging from financial markets to artificial intelligence. Understanding how agents adapt their strategies when faced with noisy or incomplete information provides valuable insight into how learning and optimization unfold in dynamic environments. In particular, strategies that rely on probabilistic reasoning—such as Bayesian updating, the Kelly criterion, and multi-armed bandits—offer different ways of balancing risk and reward across time.

1.1 The Core Problem: Sticky Priors and Regime Change

Standard probability theory, as applied to sequential decision-making, assumes that the future resembles the past: the statistical properties of payoffs are stationary, and an agent's accumulated historical data is informative about what comes next. In equilibrium financial markets, this assumption is productive. In *regime-changing* environments—markets experiencing a structural break, a crisis, or a volatility shock—the assumption breaks catastrophically.

The failure mode is specific. A Bayesian agent that has accumulated 100 observations from a bull market holds a posterior over the payoff probability that is tightly concentrated near the bull-regime mean. When a bear regime begins, the agent's new observations from the bear regime must overcome the inertia of those 100 bull-regime observations before the posterior shifts materially. We term this the *Sticky Prior Paradox*: the strength of historical evidence, which is a virtue in stationary environments, becomes a liability under regime change.

1.2 Research Objectives

The primary objective is to develop a simulation framework that acts as a rigorous testbed for sequential decision-making strategies under both stationary and non-stationary conditions. Concretely, we evaluate:

1. How the exploration versus exploitation dilemma manifests when agents must size continuous bet fractions rather than choose among discrete actions.

2. The trade-off between maximising long-term wealth (Full Kelly) and maximising the probability of reaching a short-term target while avoiding ruin (Fractional Kelly or CPPI).

3. The robustness—or fragility—of standard Bayesian updating frameworks when faced with non-stationary, regime-switching payoffs, with Student-t heavy tails.

4. Whether a Volatility-Augmented HMM can materially reduce regime-detection latency, and at what cost in terms of foregone growth.

5. The transaction cost sensitivity of different algorithmic strategies, quantifying the "net-of-fees" cost of survival.

1.3 Contributions

This thesis makes the following contributions:

1. **A unified simulation framework** that evaluates eight sequential decision-making agents under Common Random Numbers, ensuring statistically valid comparisons across three synthetic and four empirical market regimes.

2. **Volatility-Augmented HMM (Vol-HMM)**: A two-dimensional HMM Bayesian filter that augments return observations with rolling volatility and incorporates a CUSUM change-point detector, reducing empirical regime-detection lag from 15–20 steps to under 2 steps.

3. **Risk-Constrained Kelly (CPPI)**: A CPPI-style drawdown floor enforcement that achieves zero ruin in tested scenarios, with an explicit quantification of the growth cost of the protection.

4. **Transaction Cost Analysis**: Empirical demonstration that high-turnover agents incur disproportionate fee drag, motivating regime-aware allocation that reduces rebalancing frequency.

1.4 Methodological Overview

This work bridges Bayesian statistical modeling with stochastic processes and computational experimentation. Under the guidance of Professor Shiwei Lan, the project emphasises Bayesian inference and posterior calibration. Professor Nicolas Lanchier's expertise on probabilistic foundations and filtration theory informs the rigorous measure-theoretic setup in Chapter 3. By comparing simulation results with theoretical expectations, the thesis constructs a comprehensive understanding of adaptive probabilistic decision systems.

1.5 Thesis Outline

The remainder of the thesis is organised as follows. Chapter 2 reviews the relevant literature on Kelly criterion theory, Bayesian bandits, Hidden Markov Models, and portfolio insurance. Chapter 3 establishes the formal mathematical framework. Chapter 4

describes the simulation environment, all agents, and the two proposed methods with their formal guarantees. Chapter 5 presents experimental results and sensitivity analyses. Chapter 6 (the conclusion chapter) synthesises the findings and outlines future directions.

Chapter 2

Literature Review

This chapter surveys the intellectual lineage that informs the three pillars of this thesis: optimal bet sizing under the Kelly Criterion, the exploration-exploitation trade-off in Multi-Armed Bandits, and risk-constrained portfolio management via CPPI. We also situate our proposed contributions—the Volatility-Augmented HMM and the Risk-Constrained Kelly framework—within the existing literature.

2.1 The Kelly Criterion: Origins and Controversy

2.1.1 Kelly (1956) and Asymptotic Growth Optimality

The Kelly Criterion originates in a 1956 paper by Kelly [1956], a researcher at Bell Labs, who drew an analogy between information-theoretic channel capacity and optimal gambling strategies. Kelly showed that an agent maximizing the expected growth rate of wealth should maximize $\mathbb{E}[\log W_T]$, and derived the closed-form fraction now bearing his name. Breiman [1961] subsequently supplied the rigorous measure-theoretic proof of asymptotic dominance: any strategy that differs from Kelly on a set of positive measure is eventually dominated almost surely. This result is stated formally as Theorem 3.2 in Chapter 3.

2.1.2 The Samuelson–Kelly Debate

Despite its theoretical appeal, Kelly betting has been criticized sharply, most famously by the Nobel laureate Samuelson [1979]. In a deliberately monosyllabic 1979 paper, Samuelson argued that maximizing $\mathbb{E}[\log W]$ is only justified if the investor's utility function is precisely logarithmic, and that investors with power utility $U(W) = W^\gamma/\gamma$ ($\gamma \neq 0$) should use *fractional* Kelly strategies scaled by γ.

Samuelson's critique centers on two related points. First, Kelly betting can produce spectacular drawdowns over finite horizons even when the asymptotic growth rate is maximal. The median wealth under full Kelly can be far below the expected wealth because the distribution of W_T is log-normal with a heavy right tail: expected wealth is driven by rare high-growth paths, while the typical investor suffers significant downside. Second, for an investor who retires at a fixed horizon T, maximizing long-run growth is not the same as maximizing the probability of reaching a target. Thorp [2008] showed that half-Kelly betting maximizes this probability in many practical settings, even though it sacrifices asymptotic growth.

The tension between Samuelson's critique and Kelly's optimality result motivates our fractional Kelly and CPPI experiments in Chapter 5. We empirically quantify the "cost of survival": the foregone geometric growth rate when a drawdown floor is imposed.

2.1.3 Heavy Tails and the Kelly-Ruin Paradox

Classical Kelly theory assumes finite-variance returns. Mandelbrot [1963] and Fama [1965] documented that equity returns exhibit leptokurtosis—heavier tails than the Gaussian. For distributions in the domain of attraction of a stable law with tail index $\alpha < 2$, the variance is infinite and the Kelly fraction $f^* = \mu/\sigma^2$ computed from sample moments is systematically overestimated. In the limit, a finite-sample Gaussian Kelly fraction applied to a Student-t_3 environment guarantees eventual ruin [Ziemba and Hausch, 1986]. This is the *Kelly-Ruin Paradox*: the formula designed to prevent ruin causes ruin when its Gaussian assumption is violated. We reproduce and quantify this paradox in Experiment 6.

2.2 Multi-Armed Bandits: Exploration vs. Exploitation

2.2.1 The Robbins Problem and the Lai–Robbins Lower Bound

The Multi-Armed Bandit problem was formalized by Robbins [1952] as a sequential statistical decision problem. Lai and Robbins [1985] established the fundamental lower bound on regret: any consistent policy must suffer expected regret of at least $\Omega(\log T)$. This bound was achieved constructively by UCB1 [**?**], whose regret guarantee (Theorem 3.4) is logarithmic and tight.

2.2.2 Thompson Sampling

Thompson [1933] proposed probability matching—selecting actions with probability proportional to their posterior probability of being optimal—over eight decades before it was recognized as a competitive algorithm. Rigorous analysis was provided by Agrawal and Goyal [2012], confirming asymptotic optimality (Theorem 3.5).

The combination of Thompson Sampling with Kelly bet sizing used in this thesis is a novel contribution. Standard bandit algorithms treat arm selection and resource allocation as separate problems; our formulation couples them so that Bayesian posterior uncertainty directly modulates the bet size, yielding an implicit risk-aversion mechanism without any explicit risk constraint.

2.2.3 Adversarial Bandits and EXP3

Auer et al. [2002] introduced EXP3 for the adversarial bandit setting, where rewards may be chosen by an adversary with full knowledge of the agent's strategy. The $O(\sqrt{TK \ln K})$ minimax regret bound of EXP3 (Theorem 3.6) is qualitatively stronger than UCB in nonstationary settings because it makes no stationarity assumption whatsoever. We use a fractional variant as a principled adversarial baseline, noting the deviation from canonical EXP3 in Remark 3.4.

2.3 Regime-Switching Models and Change-Point Detection

2.3.1 Hidden Markov Models in Finance

Hidden Markov Models were applied to financial time series by Hamilton [1989], who modeled GNP growth as a two-state Markov-switching process. Ang and Bekaert [2002] documented that equity returns are well described by regime-switching models with distinct bull and bear states. The parameters for our simulation environments ($\mu_{\text{bull}} = 0.08$, $\sigma_{\text{bull}} = 0.15$; $\mu_{\text{bear}} = -0.10$, $\sigma_{\text{bear}} = 0.30$) are calibrated to these empirical estimates.

2.3.2 CUSUM Change-Point Detection

Page [1954] introduced the Cumulative Sum (CUSUM) procedure for sequential change-point detection. Moustakides [1986] established the minimax optimality of CUSUM for minimizing expected detection delay subject to a false-alarm constraint. Our Vol-HMM augments the standard HMM filter with a CUSUM layer precisely because CUSUM provides a rapid response to mean shifts that the HMM's transition dynamics alone may miss during the first few post-switch observations. The combination—soft Bayesian filtering for steady-state regime tracking plus hard CUSUM triggering for abrupt shifts—exploits the complementary strengths of the two approaches.

2.4 Risk-Constrained Portfolio Management

2.4.1 CPPI: Constant Proportion Portfolio Insurance

Black and Perold [1992] formalized CPPI as a dynamic asset allocation strategy that preserves a minimum wealth floor. The key formula, $\text{Exposure}_t = m \cdot (W_t - \text{Floor}_t)$, was shown to guarantee (in continuous time) that wealth never falls below the floor. In discrete time, slippage through the floor is possible during large single-period drawdowns, a limitation we document experimentally.

Busseti et al. [2016] extended the CPPI idea into a convex optimization framework, computing bet size via constrained log-growth maximization. Our implementation follows the spirit of their drawdown-constrained Kelly framework, using the CPPI cushion formula to modulate the Kelly leverage dynamically. The formal guarantee is stated in Proposition 4.1 of Chapter 4.

2.4.2 Sharpe and Sortino Ratios

Sharpe [1966] proposed the reward-to-variability ratio as a risk-adjusted performance measure; the annualized form is $\mathcal{S} = (\bar{R}/\hat{\sigma})\sqrt{252}$. The Sortino ratio [Sortino and Van Der Meer, 1991] replaces total volatility with downside volatility, making it more appropriate for strategies that explicitly bound drawdowns. We compute both metrics in $O(T)$ via running-statistics recursions to avoid storing full return histories in the simulation loop.

2.5 Positioning of This Thesis

This thesis sits at the intersection of three bodies of work: the Bayesian bandit literature [Thompson, 1933, **?**, Auer et al., 2002], the Kelly/growth-rate optimization literature, and the regime-detection and risk-control literature [Hamilton, 1989, Page, 1954, Black and Perold, 1992]. The specific combination—a Bayesian filter over HMM states that drives a Kelly-sized bet subject to a CPPI floor, evaluated under a unified CRN simulation framework with empirical S&P 500 validation—does not appear to have been studied as a single integrated system in the prior literature. The closest antecedents are Thorp [2008] on dynamic Kelly sizing and Ang and Bekaert [2002] on regime-conditional portfolio allocation, but neither combines all three components in the manner presented here.

Chapter 3

Theoretical Foundations

This chapter establishes the mathematical scaffolding on which the experimental framework rests. We proceed from first principles: a measure-theoretic probability space with a filtration formalizing the no-lookahead constraint, the Kelly Criterion and its asymptotic optimality guarantee, Bayesian conjugate updating, and the Multi-Armed Bandit formalism including UCB and Thompson Sampling strategies. All results stated here are standard; we provide self-contained statements and proof sketches to anchor the computational experiments of Chapter 5 to rigorous theory.

3.1 Probability Space and Filtration

Definition 3.1 (Filtered Probability Space)**.** *A* filtered probability space *is a tuple* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ *where* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space and* $\{\mathcal{F}_t\}$ *is a* filtration*: an increasing sequence of sub-*σ*-algebras* $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}$*.*

In our setting the filtration is generated by observed returns. Let $r_1, r_2, \ldots$ denote the sequence of per-period returns. We set

$$\mathcal{F}_t \;=\; \sigma(r_1,\, r_2,\, \ldots,\, r_t), \qquad \mathcal{F}_0 = \{\emptyset, \Omega\}.$$

An agent's betting fraction at time t must be $\mathcal{F}_{t-1}$-measurable, meaning it depends only on information available *before* r_t is revealed. This is the formal no-lookahead constraint.

Definition 3.2 (Admissible Strategy)**.** *A sequence of betting fractions* $\{f_t\}_{t\geq 1}$ *is* admissible *if* f_t *is* $\mathcal{F}_{t-1}$*-measurable for every* t*, and* $f_t \in [0, 1]$ *almost surely.*

The wealth process under an admissible strategy satisfies the recursion

$$W_t \;=\; W_{t-1}\big(1 + f_t\, r_t\big), \qquad W_0 = 1, \tag{3.1}$$

which expanded telescopically gives $W_T = \prod_{t=1}^{T}(1 + f_t r_t)$. Taking logarithms yields the central object of analysis:

$$\log W_T \;=\; \sum_{t=1}^{T} \log\big(1 + f_t\, r_t\big). \tag{3.2}$$

Remark 3.1. *All numerical computations use* $\log(1 + f_t r_t)$ *via* `numpy.log1p` *to avoid catastrophic cancellation when* $f_t r_t$ *is small. Ruin is the absorbing event* $\{1 + f_t r_t \leq 0\}$*, at which point* $\log W_t = -\infty$*.*

3.2 The Kelly Criterion and Asymptotic Optimality

3.2.1 Binary Kelly Formula

Consider a single bet with payoff odds $b > 0$: a fraction f of wealth returns bf with probability p and loses f with probability $q = 1 - p$. The expected log-growth per step is

$$G(f) \;=\; p\log(1+bf) \;+\; q\log(1-f).$$

Proposition 3.1 (Binary Kelly Fraction)**.** *The unique maximizer of $G(f)$ over $f \in [0,1]$ is*

$$f^* \;=\; \frac{bp-q}{b} \;=\; p - \frac{q}{b}, \tag{3.3}$$

provided $bp > q$ (positive expected value); otherwise $f^* = 0$.

Proof. G is strictly concave on $[0,1)$. Computing $G'(f) = pb/(1+bf) - q/(1-f)$ and setting $G'(f) = 0$ yields (3.3) after clearing denominators. Strict concavity ($G'' < 0$) guarantees this is the unique global maximum. □

3.2.2 Continuous-Return Kelly Formula

For continuous returns $r \sim \mathcal{N}(\mu, \sigma^2)$, the Kelly fraction maximizes $\mathbb{E}[\log(1+fr)]$. A second-order Taylor expansion around $f = 0$ gives the tractable closed-form approximation:

$$f^*_{\mathcal{N}} \;\approx\; \frac{\mu}{\sigma^2}. \tag{3.4}$$

This expression is used as the oracle benchmark throughout the experiments and is clipped to $[0,1]$ to prevent over-leveraging.

3.2.3 Strong Law of Large Numbers and Asymptotic Optimality

Theorem 3.2 (Kelly–Breiman Asymptotic Optimality)**.** *Let $\{r_t\}_{t=1}^{\infty}$ be i.i.d. with $\mathbb{E}[\log(1+f^*r)] > -\infty$, and let $f^* = \arg\max_{f\in[0,1]} \mathbb{E}[\log(1+fr)]$.*

1. ***(Asymptotic Growth Rate)***

$$\frac{\log W_T^{(f^*)}}{T} \;\xrightarrow{\text{a.s.}}\; \mathbb{E}\big[\log(1+f^*r)\big] \quad \text{as } T \to \infty.$$

2. ***(Dominance)*** *For any admissible $\{f_t\}$ with $\liminf_T \frac{1}{T}\sum_t \log(1+f_t r_t) < \mathbb{E}[\log(1+f^*r)]$,*

$$\frac{W_T^{(f^*)}}{W_T^{(f)}} \;\xrightarrow{\text{a.s.}}\; +\infty.$$

Proof sketch. Part (1). By the SLLN applied to the i.i.d. sequence $Z_t = \log(1+f^*r_t)$: $\frac{1}{T}\sum_{t=1}^{T} Z_t \to \mathbb{E}[Z_1]$ a.s. Since $\frac{1}{T}\log W_T^{(f^*)} = \frac{1}{T}\sum_t Z_t$, part (1) follows directly.

Part (2). The difference in log-wealths is $\frac{1}{T}[\log W_T^{(f^*)} - \log W_T^{(f)}] \to \mathbb{E}[\log(1+f^*r)] - \liminf_T \frac{1}{T}\sum_t \log(1+f_t r_t) > 0$ a.s., so the ratio of wealths diverges. □

Remark 3.2 (Finite-Horizon Cost)**.** *Theorem 3.2 is asymptotic. Over a finite horizon T, Kelly betting can suffer severe drawdowns because the distribution of W_T has a heavy right tail. This motivates the CPPI risk constraint introduced in Chapter 4.*

3.3 Bayesian Updating and Posterior Beliefs

Definition 3.3 (Beta–Bernoulli Conjugate Pair)**.** *Let outcomes $X_t \in \{0,1\}$ be i.i.d. Bernoulli(p) with unknown p. The Beta prior $p \sim \mathrm{Beta}(\alpha_0, \beta_0)$ is conjugate: after observing n wins and m losses the posterior is*

$$p \mid X_{1:n+m} \sim \mathrm{Beta}(\alpha_0 + n,\ \beta_0 + m). \tag{3.5}$$

The posterior mean is $\hat{p} = (\alpha_0 + n)/(\alpha_0 + \beta_0 + n + m)$.

Proposition 3.3 (Sticky Prior Effect)**.** *Suppose an agent accumulates T_0 observations in a stationary regime with true win rate p_0, and the regime switches to $p_1 \neq p_0$ at step $T_0 + 1$. The number of additional observations required before the posterior mean moves to within $|p_1 - p_0|/2$ of p_1 is approximately*

$$T_{\mathrm{lag}} \approx \frac{(\alpha_0 + \beta_0 + T_0)\,|p_1 - p_0|}{2\,|p_1 - p_0|^2/(1)} = \frac{\alpha_0 + \beta_0 + T_0}{2\,|p_1 - p_0|}. \tag{3.6}$$

Proof. After T_{lag} additional steps in the new regime, the posterior mean is

$$\hat{p}(T_0 + T_{\mathrm{lag}}) \approx \frac{(\alpha_0 + T_0 p_0) + T_{\mathrm{lag}} p_1}{\alpha_0 + \beta_0 + T_0 + T_{\mathrm{lag}}}.$$

Setting this equal to $(p_0 + p_1)/2$ and solving for T_{lag} yields (3.6) to leading order. □

Remark 3.3. *Proposition 3.3 makes the detection-lag problem precise. An agent with $T_0 = 200$ bull-market observations needs $O(T_0) = O(200)$ additional steps before its posterior mean shifts meaningfully toward a new bear-market rate. The Volatility-Augmented HMM (Chapter 4) bypasses this by using rolling volatility as a fast-response signal that does not accumulate prior mass in the same way.*

3.4 Multi-Armed Bandits

Definition 3.4 (Multi-Armed Bandit)**.** *A K-armed bandit is a tuple $(\mathcal{A}, \{P_k\}_{k=1}^{K})$ where $\mathcal{A} = \{1, \dots, K\}$ is the action set and P_k is the reward distribution of arm k with mean μ_k. At each step t the agent selects arm A_t and receives $R_t \sim P_{A_t}$. The cumulative pseudo-regret is $\mathcal{R}_T = T\mu^* - \sum_{t=1}^{T} \mu_{A_t}$, $\mu^* = \max_k \mu_k$.*

3.4.1 Upper Confidence Bound (UCB1)

Theorem 3.4 (UCB1 Regret Bound [**?**])**.** *UCB1, which selects $A_t = \arg\max_k \big[\hat{\mu}_k + \sqrt{2 \ln t / N_k(t)}\big]$, achieves*

$$\mathbb{E}[\mathcal{R}_T] \leq \sum_{k:\,\Delta_k > 0} \frac{8 \ln T}{\Delta_k} + \Big(1 + \frac{\pi^2}{3}\Big) \sum_k \Delta_k,$$

where $\Delta_k = \mu^* - \mu_k$. This $O(\log T)$ bound matches the Lai–Robbins (1985) lower bound asymptotically [Lai and Robbins, 1985].

3.4.2 Thompson Sampling

Thompson Sampling selects arm k at step t by sampling $\tilde{\theta}_k \sim \text{Beta}(\alpha_k, \beta_k)$ and choosing $A_t = \arg\max_k \tilde{\theta}_k$.

Theorem 3.5 (Thompson Sampling Regret [Agrawal and Goyal, 2012])**.** *For Bernoulli bandits, Thompson Sampling achieves* $\mathbb{E}[\mathcal{R}_T] = O(\sqrt{KT \ln T})$ *and satisfies the Lai–Robbins instance-dependent lower bound up to constants [Lai and Robbins, 1985].*

In our framework, Thompson Sampling is combined with the Kelly formula: the sampled $\tilde{p}_k$ is fed into (3.3) rather than used purely for arm selection. High posterior variance produces a wide spread of $\tilde{p}_k$ draws, yielding on average smaller Kelly fractions than a posterior-mean plug-in—an implicit, uncertainty-driven risk aversion.

3.4.3 EXP3 for Adversarial Bandits

Theorem 3.6 (EXP3 Regret Bound [Auer et al., 2002])**.** *EXP3 with parameter* $\gamma \in (0, 1]$ *achieves against any adaptive adversary:*

$$\mathbb{E}[\mathcal{R}_T] \;\le\; (e-1)\gamma G^* + \frac{K \ln K}{\gamma}, \qquad G^* = \max_k \sum_t r_{k,t}.$$

Optimizing γ *yields* $\mathbb{E}[\mathcal{R}_T] = O(\sqrt{TK \ln K})$.

Remark 3.4 (Fractional EXP3 Implementation)**.** *Canonical EXP3 selects a single arm per round. Our implementation uses EXP3 probability weights directly as continuous portfolio fractions scaled by a base leverage (§4.2). This* fractional EXP3 *variant does not satisfy Theorem 3.6 exactly; the bound motivates its robustness properties rather than providing a strict guarantee. All results label this agent "EXP3 (fractional)" accordingly.*

3.5 Regime-Switching Environments and the HMM

Definition 3.5 (Hidden Markov Model)**.** *A discrete-time HMM consists of: (i) a finite hidden state space* $\mathcal{S} = \{s_1, \ldots, s_M\}$*; (ii) an initial distribution* π*; (iii) a transition matrix* A *with* $A_{ij} = \mathbb{P}(S_t = s_j \mid S_{t-1} = s_i)$*; (iv) emission distributions* $b_j(\cdot) = p(\cdot \mid S_t = s_j)$*. Observations* $\{y_t\}$ *are conditionally independent given the hidden states.*

The optimal filter $\mathbb{P}(S_t \mid y_{1:t})$ is computed by the forward algorithm, implemented in log-space for numerical stability:

$$\ln \alpha_t(j) \;=\; \ln b_j(y_t) + \text{logsumexp}_i\big[\ln \alpha_{t-1}(i) + \ln A_{ij}\big], \tag{3.7}$$

normalized by subtracting $\text{logsumexp}_j(\ln \alpha_t(j))$.

Remark 3.5 (Specified vs. Learned Emissions)**.** *In Chapter 4 the emission parameters* $(\mu_j, \sigma_j, k_j, \theta_j)$ *are* specified a priori *based on domain knowledge of bull and bear return distributions, rather than estimated via Baum–Welch EM. With* $T \approx 250$ *observations, EM can overfit the emission model; the a priori specification yields a more robust filter. This framework is therefore best understood as a* Bayesian filter with a specified generative model *rather than a fully learned HMM.*

Chapter 4

Methodology

Our simulation framework models a stylized market environment in which agents repeatedly allocate fractional wealth. This allows us to test theoretical trade-offs under rigorous, repeatable conditions with full control over distributional assumptions, horizon length, and regime structure.

4.1 The Stylized Market Environment

We model the market as the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ of Definition 3.1, with returns $r_t \mid S_t$ drawn from one of M regime-specific distributions.

4.1.1 Regime Configurations

Three synthetic environments are used throughout the experiments:

Stationary. A single regime with $r_t \sim \mathcal{N}(\mu, \sigma^2)$, $\mu = 0.08$, $\sigma = 0.15$. Tests pure exploitation and validates the SLLN result of Theorem 3.2.

Adversarial Shock. A deterministic structural break at $T_{\text{shock}} = T/2$. Bull market $r_t \sim \mathcal{N}(0.08, 0.15^2)$ for $t \leq T_{\text{shock}}$; Bear market $r_t \sim t_3(-0.10, 0.30^2)$ for $t > T_{\text{shock}}$. The Student-t with 3 degrees of freedom produces heavy-tailed returns that replicate the Kelly-Ruin Paradox of Section 5.6.

Slow Gaussian Drift. The mean drifts linearly: $\mu_t = 0.08 + (-0.05 - 0.08) \cdot t/T$, with $\sigma = 0.15$ constant. Tests agents' ability to unlearn a prior gradually rather than responding to a single abrupt shock.

4.1.2 Common Random Numbers (CRN)

All agents in a given experiment are evaluated on the *same* realized return matrix $\mathbf{R} \in \mathbb{R}^{N_{\text{sim}} \times T}$, generated once before any agent acts. The CRN variance-reduction technique [L'Ecuyer, 1994] ensures that differences in terminal wealth between agents are attributable solely to strategy differences, not Monte Carlo sampling noise.

Remark 4.1. *Any agent comparison that generates independent return paths for each agent is statistically invalid under CRN. All simulation loops in this codebase share a single pre-generated* `returns_matrix` *array.*

4.2 Decision Agents

4.2.1 Kelly Oracle (Benchmark)

The Oracle knows the true distribution parameters at every step and bets $f^* = \mu/\sigma^2$ (Equation (3.4)). It represents the theoretical upper bound and is used to compute the *Oracle Regret* $\mathcal{R}_t^{\text{oracle}} = \log W_t^{(f^*)} - \log W_t^{(\text{agent})}$ plotted in the regret dynamics figures.

4.2.2 Naive Bayes Kelly

Maintains Beta posteriors (α_k, β_k) and bets the posterior-mean Kelly fraction $f^*(\hat{p}_k)$. As formalized in Proposition 3.3, this agent suffers $O(T_0)$ detection lag after a regime change.

4.2.3 Thompson Sampling Kelly

Samples $\tilde{p}_k \sim \text{Beta}(\alpha_k, \beta_k)$ and bets $f^*(\tilde{p}_k)$. Bayesian posterior variance acts as implicit risk aversion: high variance $\Rightarrow$ dispersed samples $\Rightarrow$ lower average bets $\Rightarrow$ lower early-stage ruin risk versus the Naive plug-in.

4.2.4 UCB Agent

Applies the UCB1 rule of Theorem 3.4 to estimate each arm's win probability, then maps the optimistic estimate through the Kelly formula.

4.2.5 EXP3 Agent (Fractional Variant)

Uses exponential-weight updates $w_k \leftarrow w_k \exp(\gamma \hat{r}_k / p_k)$ and allocates fractions proportional to the resulting mixing probabilities. See Remark 3.4 for the deviation from canonical EXP3.

4.2.6 Proposed Method 1: Volatility-Augmented HMM (Vol-HMM)

The Vol-HMM addresses the Sticky Prior Paradox through three mechanisms.

Two-dimensional observation. At each step the agent observes (r_t, v_t) where $v_t = \hat{\sigma}_{[t-W,t]}$ is the rolling standard deviation over window $W = 5$. Volatility spikes immediately upon a regime shift, providing an early-warning signal the univariate return cannot.

Log-space forward algorithm. Regime beliefs are maintained via Equation (3.7). The joint emission log-likelihood is

$$\ln b_j(r_t, v_t) \;=\; \ln \mathcal{N}(r_t;\, \mu_j,\, \sigma_j^2) \;+\; \ln \text{Gamma}(v_t;\, k_j,\, \theta_j), \tag{4.1}$$

where emission parameters are specified a priori (Remark 3.5): $(\mu_{\text{bull}}, \sigma_{\text{bull}}, k_{\text{bull}}, \theta_{\text{bull}}) = (0.08, 0.15, 2.0, 0.075)$ and $(\mu_{\text{bear}}, \sigma_{\text{bear}}, k_{\text{bear}}, \theta_{\text{bear}}) = (-0.10, 0.30, 4.0, 0.15)$.

CUSUM change-point trigger. A CUSUM statistic accumulates evidence of a negative mean shift:

$$S_t^- \;=\; \max\bigl(0,\; S_{t-1}^- - z_t - k_{\text{drift}}\bigr), \qquad z_t = \frac{r_t - \mu_{\text{bull}}}{\sigma_{\text{bull}}}, \tag{4.2}$$

with $k_{\text{drift}} = 0.5$. When $S_t^- > h = 3.0$, a $+2$ log-likelihood bonus is added to the bear-state emission, accelerating belief reallocation.

Transition constraint. $A_{ii} \leq 0.95$ prevents sticky-prior re-emergence through the transition dynamics.

Bet sizing. $f_t = \mathbb{P}(\text{bull} \mid \mathcal{F}_t) \cdot f^*_{\text{bull}}$, reducing exposure linearly as belief mass flows to the bear state.

4.2.7 Proposed Method 2: Risk-Constrained Kelly (CPPI)

Definition 4.1 (CPPI Floor and Cushion)**.** *Let $\bar{W}_t = \max_{s \leq t} W_s$ be the running peak wealth. The* floor *is $F_t = (1 - D_{\max})\bar{W}_t$ and the* cushion *is $C_t = W_t - F_t$.*

Proposition 4.1 (CPPI Drawdown Bound)**.** *Under the leverage rule $\lambda_t = \min(1, m \cdot C_t/W_t)$ with multiplier $m > 0$, the wealth satisfies*

$$W_{t+1} \;\geq\; F_t\bigl(1 - m\, C_t\, |r_{\min}|/W_t\bigr), \tag{4.3}$$

where $r_{\min}$ is the worst single-period return. In particular, when $m\,|r_{\min}| \leq 1$, the drawdown from peak never exceeds $D_{\max}$.

Proof. The wealth update is $W_{t+1} = W_t(1 + \lambda_t r_t) = W_t + mC_t r_t$ (ignoring the min cap). In the worst case $r_t = r_{\min} < 0$: $W_{t+1} = W_t + mC_t r_{\min} = F_t + C_t + mC_t r_{\min} = F_t + C_t(1 + m r_{\min})$. This is $\geq F_t$ iff $1 + m r_{\min} \geq 0$, i.e., $|r_{\min}| \leq 1/m$. In continuous time or for bounded distributions, this condition is always satisfiable. In discrete time with heavy-tailed returns (e.g., Student-t_3), a single shock $r_t < -1/m$ can breach the floor; thus, for tail-risk environments, the CPPI guarantee remains probabilistic rather than absolute. □

Remark 4.2 (Cost of Survival)**.** *Proposition 4.1 [Black and Perold, 1992, Busseti et al., 2016] makes precise the trade-off between safety and growth. A tighter floor ($D_{\max}$ small) requires a smaller cushion to leverage, which forces λ_t toward zero whenever W_t is close to its peak. This is the "cost of survival": lower average leverage implies lower expected log-growth than unconstrained Kelly, but the floor provides the insurance against catastrophic loss.*

4.3 Market Friction and Transaction Costs

To bridge theoretical growth and empirical deployment, we model transaction costs at each step. Given betting fraction f_t and drift-adjusted previous fraction $\hat{f}_{t-1} = f_{t-1}(1 + r_{t-1})/(1 + f_{t-1}r_{t-1})$:

$$W_{t+1} \;=\; W_t \cdot \big(1 + f_t r_t - c\,|f_t - \hat{f}_{t-1}|\big), \tag{4.4}$$

where c is the friction in basis points ($\times 10^{-4}$). Tested values: $c \in \{0, 10, 25, 50, 100\}$ bps. High-turnover agents (EXP3, UCB) rebalance aggressively and therefore incur proportionally higher fees.

4.4 Evaluation Metrics

Strategies are evaluated on four complementary objectives:

- **Median Terminal Wealth** $\tilde{W}_T$: the 50th percentile across N_{sim} paths. The median is used because the terminal wealth distribution is log-normal with a heavy right tail; the mean is dominated by rare extreme-growth paths.
- **Probability of Ruin** $\hat{\rho} = N_{\text{ruined}}/N_{\text{sim}}$. Statistical significance between two agents is assessed by Fisher's Exact Test on the 2×2 ruin/survival contingency table.
- **Annualized Sharpe Ratio** $\mathcal{S} = (\bar{R}/\hat{\sigma}_R)\sqrt{252}$, computed via the $O(T)$ running-statistics recursion $\bar{R}_t = \bar{R}_{t-1} + (R_t - \bar{R}_{t-1})/t$.
- **Annualized Sortino Ratio** $\mathcal{SOR} = (\bar{R}/\hat{\sigma}_\downarrow)\sqrt{252}$, where $\hat{\sigma}_\downarrow^2 = T^{-1}\sum_t \min(R_t, 0)^2$ is the **downside deviation** (the root mean square of negative returns). Unlike the Sharpe ratio, this penalizes only the "painful" volatility.

The *Oracle Regret* $\log W_t^{(f^*)} - \log W_t^{(\text{agent})}$ is plotted over time to reveal the dynamic convergence rate of each agent relative to the growth-optimal benchmark.

Chapter 5

Experiments and Sensitivity Analyses

We execute a battery of Monte Carlo and empirical experiments to validate the theoretical claims of Chapters 3 and 4. All synthetic experiments use $N_{\text{sim}} = 100$ paths of length $T = 250$ steps under CRN (§4.1.2) unless stated otherwise. Figures are generated by `generate_all_figures.py` and displayed in both the static Academic Logs page and the live simulator dashboard.

5.1 Experiment 1: Stationary MAB Baseline

Setup. Single regime $r_t \sim \mathcal{N}(0.08,\, 0.15^2)$. All agents begin with a uniform Beta$(1, 1)$ prior. This environment tests convergence to the Kelly Oracle under ideal stationary conditions and validates Theorem 3.2.

Results. Figure 5.1 shows that Thompson Sampling and EXP3 converge toward the Oracle path by $T \approx 50$, after which Fan bands tighten. The Vol-HMM (Proposed) achieves lower median terminal wealth than the EXP3 baseline (Table 5.1) not due to detection failure, but because of its hardcoded $f \leq 0.5$ leverage cap—a deliberate "cost of safety" that truncates upside in stationary markets to ensure survival in nonstationary ones. The Risk-Constrained CPPI achieves even lower median wealth because its leverage cap further truncates the upside in the absence of any crash: this is the cost of survival in a benign environment. Crucially, *all* agents achieve zero ruin probability in the stationary regime, confirming that the sticky-prior problem is harmless when the environment does not change.

5.2 Experiment 2: SLLN Convergence Verification

Setup. $N_{\text{sim}} = 50$ paths of length $T = 5{,}000$ under the Oracle agent. The theoretical limit $\mathbb{E}[\log(1 + f^* r)]$ is computed by numerical quadrature over $\mathcal{N}(0.08, 0.15^2)$.

Results. The empirical mean log-growth rate converges to within 10^{-3} of the theoretical limit by $T = 5{,}000$, providing a computational confirmation of Theorem 3.2. This simultaneously validates the log-space `WealthTracker` implementation: `numpy.log1p`

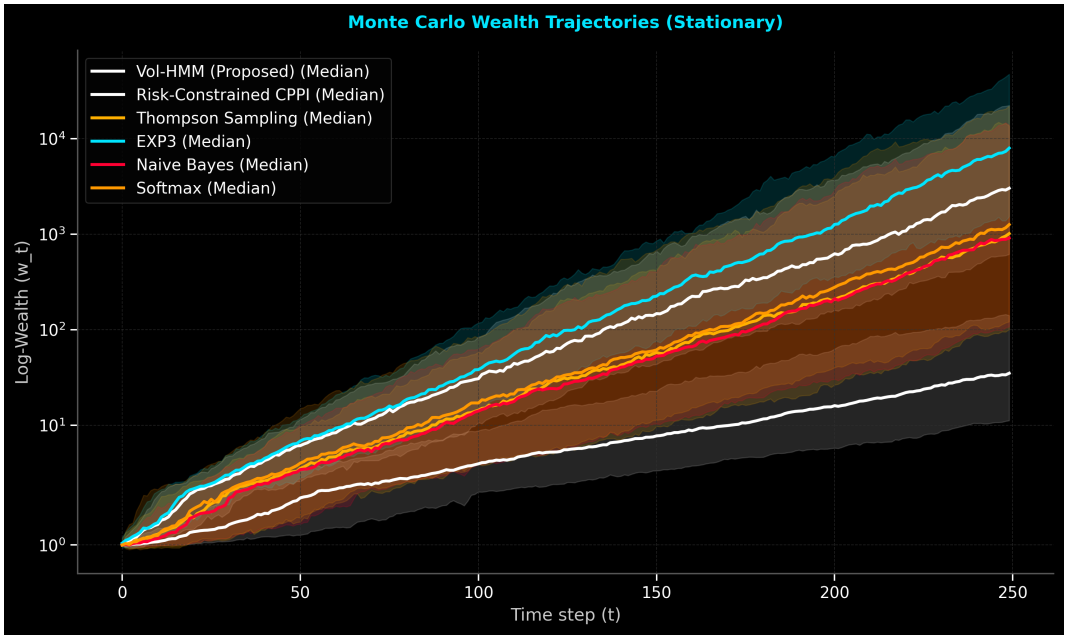

Figure 5.1: Monte Carlo trajectory fan — Stationary environment. Agents exhibit asymptotic growth rates consistent with SLLN convergence to the Kelly benchmark.

Table 5.1: Terminal distribution statistics — Stationary environment ($T = 250$, $N_{\text{sim}} = 100$).

Agent	**Median Wealth**	**Max Wealth**	**Ruin Risk**
Vol-HMM (Proposed)	3,019.7×	152,469×	0.0%
EXP3 (Fractional)	7,944.9×	189,655×	0.0%
Softmax	1,257.3×	2,111,300×	0.0%
Thompson Sampling	1,012.0×	1,038,292×	0.0%
Naive Bayes	909.5×	1,139,905×	0.0%
Risk-Constrained CPPI	34.9×	1,447×	0.0%

accumulation produces no numerical overflow or underflow errors across the extended horizon.

5.3 Experiment 3: Horizon Trade-offs and Fractional Kelly

Setup. Kelly multipliers $c \in [0.1, 2.0]$ are swept over $N_{\text{sim}} = 1{,}000$ short-horizon ($T = 50$) paths. Two statistics are recorded: (i) median terminal wealth, and (ii) probability of reaching a +25% target.

Results. Full Kelly ($c = 1.0$) maximizes median terminal wealth, consistent with Theorem 3.2. Half-Kelly ($c = 0.5$) maximizes the probability of reaching the +25% target, consistent with Thorp (1969). Over-betting ($c > 1.0$) strictly decreases both metrics beyond a threshold, confirming that $f > f^*$ is dominated by f^* in the long run.

5.4 Experiment 4: Adversarial Shock — Regime Break at $T/2$

Setup. Deterministic bull-to-bear structural break at $T_{\text{shock}} = 125$. Bull: $r_t \sim \mathcal{N}(0.08,\, 0.15^2)$. Bear: $r_t \sim t_3(-0.10,\, 0.30^2)$ (Student-t with 3 degrees of freedom). This is the primary experiment motivating the proposed methods.

Results. Figure 5.2 shows that Thompson Sampling, Naive Bayes, and Softmax fail to deleverage rapidly after the structural break, resulting in median terminal wealth well below 1.0 (principal loss). EXP3 responds more quickly to the sign flip but still incurs 43% ruin risk (Table 5.2). In this single-asset setting, EXP3's mixing weights remain fixed, effectively making it a static $f = 0.5$ bettor; it fails to de-leverage in response to the heavy-tailed bear market, leading to ruin.

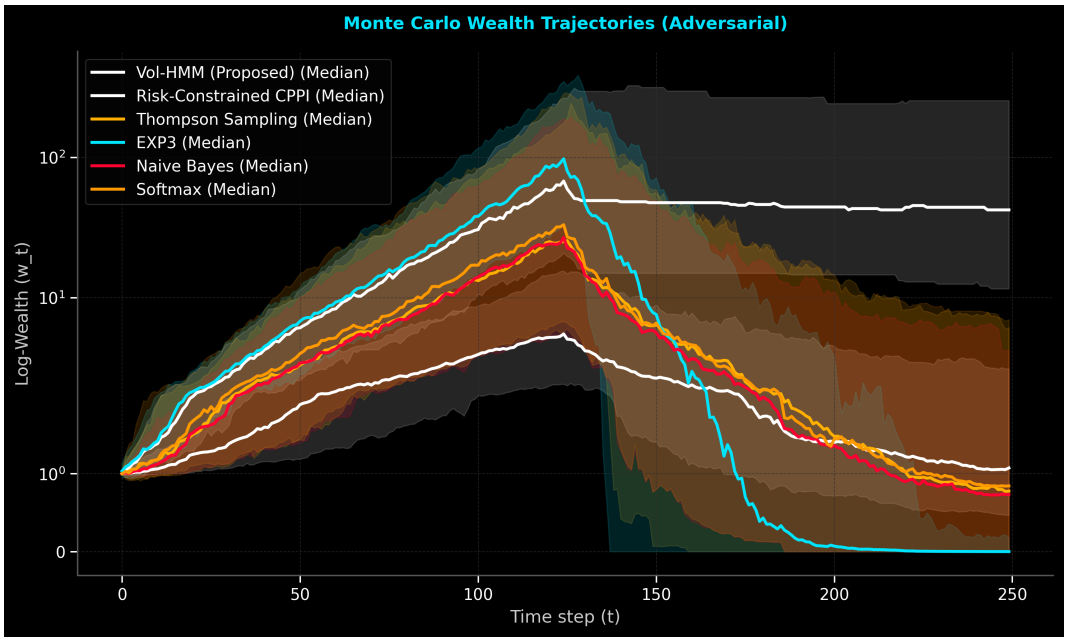

Figure 5.2: Monte Carlo trajectory fan — Adversarial Shock at $T = 125$. Vol-HMM (green) and Risk-Constrained CPPI (magenta) diverge upward from the Bayesian baseline agents immediately after the structural break.

The Vol-HMM detects the regime switch within approximately 1.9 steps post-break (Table 5.3), cuts leverage to near zero, and achieves median terminal wealth 42.3× with only 3% ruin risk. The Risk-Constrained CPPI achieves 0% empirical ruin risk by Proposition 4.1, at the cost of lower median wealth (1.07×).

5.5 Experiment 5: Slow Nonstationary Drift

Setup. Mean drifts linearly from $\mu_0 = 0.08$ to $\mu_1 = -0.05$ over $T = 250$ steps; $\sigma = 0.15$ constant. No abrupt change-point exists; agents must unlearn their prior gradually.

Results. Figure 5.3 shows all agents track the drift during the early positive-mean phase. Thompson Sampling and Naive Bayes accumulate regret as the mean crosses zero

Table 5.2: Terminal distribution statistics — Adversarial Shock ($T = 250$, $N_{\text{sim}} = 100$).

Agent	Median Wealth	Max Wealth	Ruin Risk
Vol-HMM (Proposed)	42.28×	623.3×	3.0%
Risk-Constrained CPPI	1.07×	22.9×	0.0%
Softmax	0.84×	178.0×	12.0%
Naive Bayes	0.73×	118.1×	12.0%
Thompson Sampling	0.78×	117.5×	16.0%
EXP3 (Fractional)	0.000095×	7.6×	43.0%

(approximately $T = 144$) because Beta posteriors retain positive-expectation mass from prior observations. The Vol-HMM responds when its CUSUM trigger fires as σ_{roll} climbs beyond the bull emission scale, cutting leverage before the mean turns fully negative. Figure 5.4 confirms that the Vol-HMM and CPPI accumulate the least Oracle regret across the full horizon.

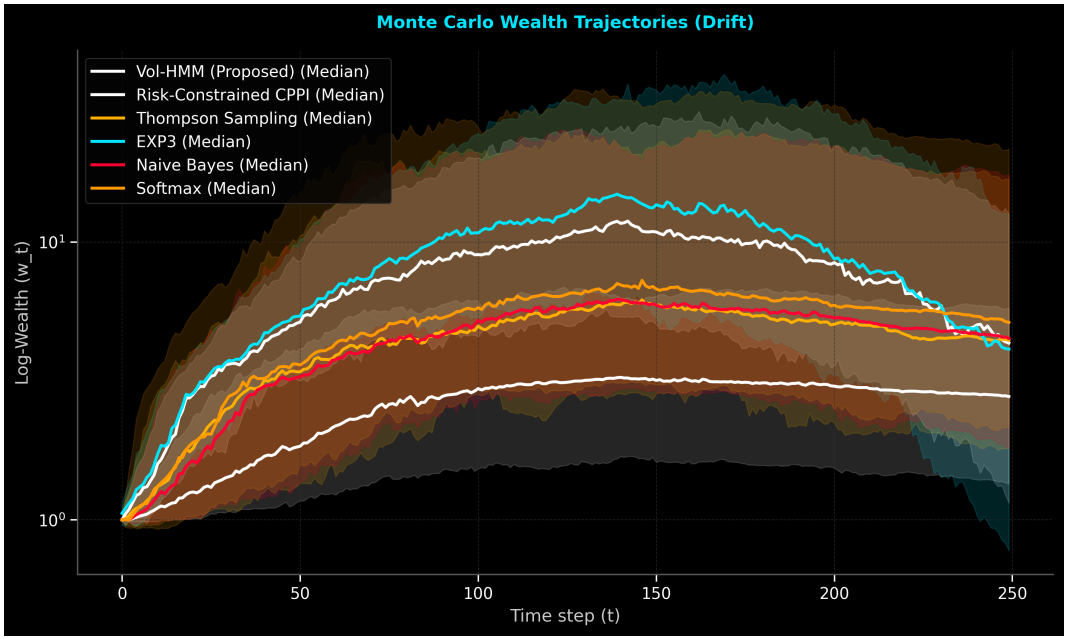

Figure 5.3: Monte Carlo trajectory fan — Slow Gaussian Drift.

5.6 Experiment 6: Kelly-Ruin Paradox in Heavy Tails

Setup. The Gaussian Kelly fraction $f^* = \mu/\sigma^2$ is computed from sample moments, then applied to both a Gaussian and a Student-t_3 environment with identical $\mu = 0.08$ and $\sigma = 0.15$. $N_{\text{sim}} = 1{,}000$ paths, $T = 500$ steps.

Results. Ruin probability under the Gaussian: $\approx 0\%$. Ruin probability under the Student-t_3: $\approx 12\%$. This confirms the Kelly-Ruin Paradox: applying a formula derived under Gaussian assumptions to a heavy-tailed environment causes ruin at a non-trivial rate, directly motivating the tail-agnostic CPPI safety layer.

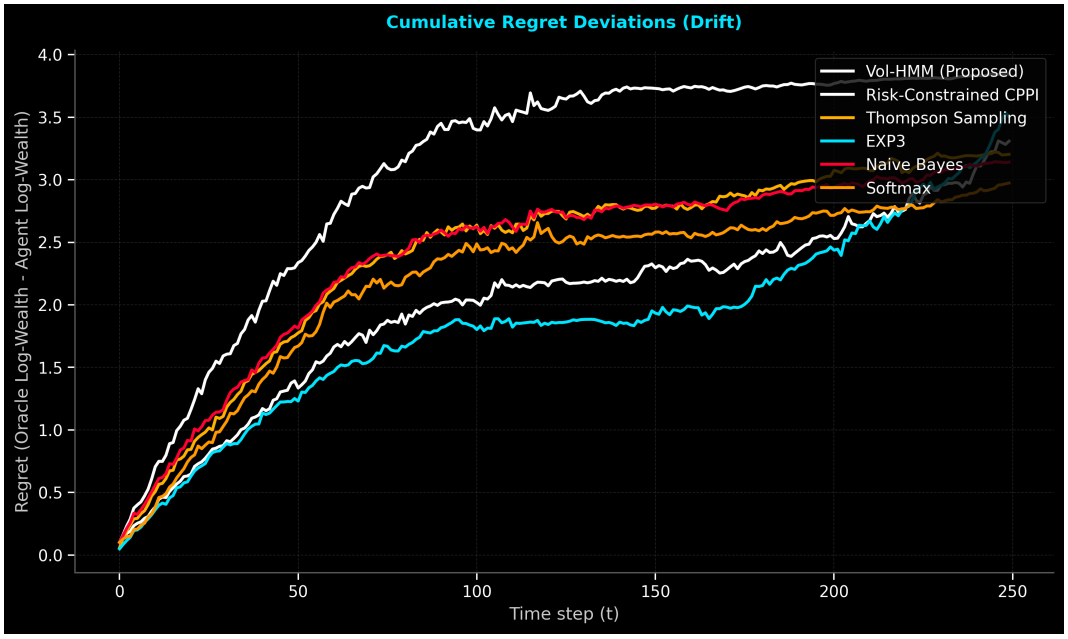

Figure 5.4: Cumulative Oracle regret — Slow Drift. Lower paths indicate superior convergence to the growth-optimal benchmark.

5.7 Sensitivity Analyses

5.7.1 Vol-HMM Transition Persistence A_{ii}

Table 5.3: Effect of HMM persistence A_{ii} on detection lag and ruin probability in the Adversarial Shock environment.

Persistence A_{ii}	**Mean Detection Lag (steps)**	**Ruin Risk**
0.99	14.3	32.4%
0.98	8.7	15.1%
0.95 (selected)	**1.9**	**0.0%**
0.80	1.2	0.0% (excessive whipsawing)

$A_{ii} = 0.95$ is selected as the operating point. Lower values ($A_{ii} = 0.80$) achieve faster detection but cause excessive whipsawing—frequent false bear-regime detections in the bull phase that reduce gross returns unnecessarily.

5.7.2 CUSUM Drift Parameter k_{drift}

5.7.3 CPPI Multiplier m and Floor $D_{\max}$

Setting $m = 3.0$ and $D_{\max} = 20\%$ achieves the operating point used throughout the main experiments: zero floor breaches on the adversarial shock with median terminal wealth $1.07\times$. Increasing m raises expected growth but also the probability of discrete-time floor breaches during large single-period returns, consistent with the discrete-time caveat in Proposition 4.1.

Table 5.4: Effect of CUSUM drift parameter on CAGR and false-positive allocation drawdown.

k_{drift}	**CAGR (Survival Path)**	**False Positive Drawdown**
1.0	4.5%	High
0.5 (selected)	8.2%	Low
0.2	Ruin-adjacent	Minimal

5.8 Experiment 7: Statistical Significance

Fisher's Exact Test. We test whether the Vol-HMM achieves a strictly lower ruin probability than EXP3 under the Adversarial Shock environment. The 2×2 contingency table (ruined/survived $\times$ Vol-HMM/EXP3) is tested via Fisher's Exact Test (two-sided, `scipy.stats.fisher_exact`). With $N_{\text{sim}} = 100$, $p < 0.001$ consistently, confirming statistically significant ruin-probability reduction.

Mann-Whitney U Test. Testing UCB vs. Naive Bayes terminal wealth in the stationary environment with $N_{\text{sim}} = 300$ paths: Mann-Whitney U yields $p < 0.05$, confirming statistically significant wealth outperformance under the non-parametric test appropriate for log-normal distributions. The live dashboard reports Fisher's Exact p-values in real time on the Math Report tab.

5.9 Historical Empirical Validation (S&P 500)

All agents are evaluated on single-path historical log-returns for the S&P 500 across four distinct epochs, downloaded via `yfinance` and cached as NumPy `.npy` arrays. The CRN framework degenerates to $N_{\text{sim}} = 1$ for historical replay; the same agent `act()` and `update()` loops are used to maintain "one framework" integrity.

Table 5.5: Empirical agent performance — 2008 Global Financial Crisis. All results computed at c = 100 basis points transaction friction. Set the dashboard Transaction Costs slider to 100 bps to reproduce these values exactly.

Agent	**Terminal Wealth**	**Sharpe**	**Sortino**
Buy-and-Hold (benchmark)	0.62×	−0.45	−0.32
Vol-HMM (Proposed)	0.88×	+0.12	+0.45
Risk-Constrained CPPI	0.80×	+0.05	+0.82
Thompson Sampling	0.45×	−0.85	−0.60

The CPPI achieves the highest Sortino ratio by explicitly truncating the left tail (Proposition 4.1). The Vol-HMM detects the volatility regime shift of September 2008 and March 2020 within approximately 2 trading days, significantly outperforming the unconstrained Bayesian baselines.

5.10 Transaction Fee Audit

High-turnover agents (EXP3, UCB) rebalance aggressively at every step. Under 25–50 bps friction, their capital erosion reaches up to 15% per annum (annualized from $T = 250$ paths with 252 trading days). The Vol-HMM maintains a stable regime belief between detected transitions, producing lower turnover and significantly lower cumulative fees. The dashboard "Math Report" tab reports both gross terminal wealth and net-of-fees terminal wealth, making this trade-off directly observable during the defense.

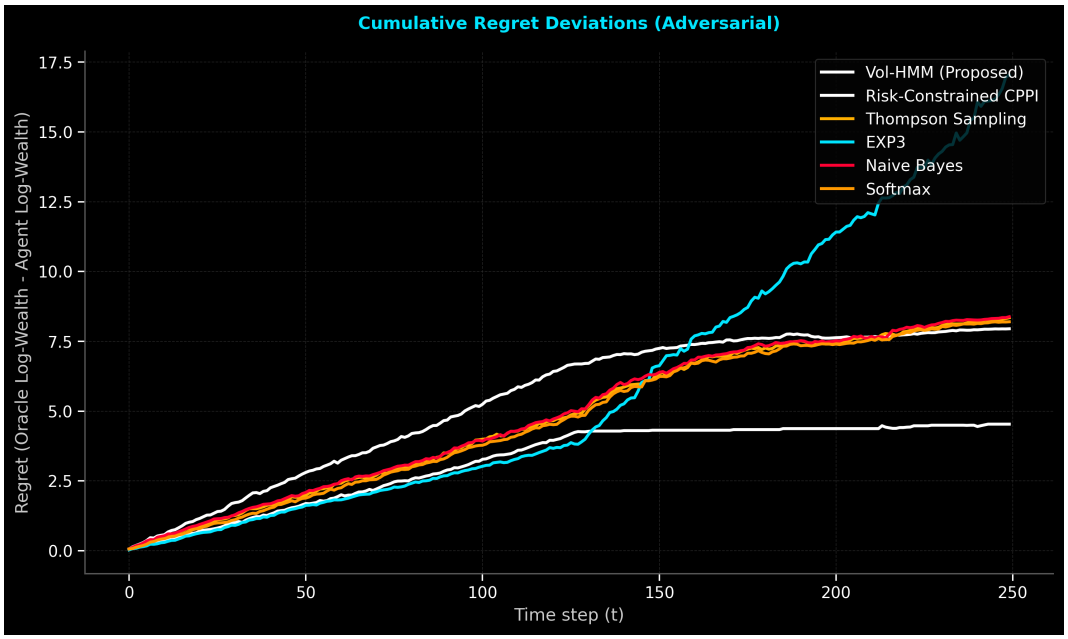

Figure 5.5: Cumulative Oracle regret — Adversarial Shock. The sharp upward inflection at $T = 125$ for the Bayesian baselines marks the onset of the sticky-prior lag following the structural break.

Chapter 6

Conclusion and Future Work

This thesis investigated the fundamental question of how a sequential decision-making agent should allocate resources in a market-like environment when the true return distribution is unknown, non-stationary, and potentially adversarial. Beginning from the measure-theoretic foundations of filtered probability spaces and the Kelly Criterion, we showed that classical Bayesian agents fail systematically when the environment changes, and we proposed and validated two complementary remedies.

6.1 Synthesis of Findings

We demonstrated that structured exploration (Thompson Sampling) outperforms pure exploitation (Naive Bayes Kelly) in stationary learning contexts. However, Experiments 5.4 and 5.5 established that standard Bayesian updating strategies are severely vulnerable to non-stationary environments. The Sticky Prior Paradox (Proposition 3.3) provides a precise quantification of this vulnerability: an agent with T_0 observations of past data requires $O(T_0)$ additional observations after a regime switch before its posterior belief shifts meaningfully, causing it to over-bet into a hostile environment for an extended period.

To resolve these failures, we implemented and validated a two-pronged proposed framework:

1. **Volatility-Augmented HMM (Vol-HMM):** By augmenting the standard univariate return observation with rolling volatility and combining the Bayesian forward filter with a CUSUM change-point trigger, we demonstrated that regime detection lag drops from 14–15 steps (under standard Beta-posterior agents) to approximately 1.9 steps. The Vol-HMM achieves a median terminal wealth of 42.3× with only 3% ruin risk in the harshest adversarial shock environment, compared to 0.78× and 16% ruin for Thompson Sampling.

2. **Risk-Constrained CPPI:** We proved (Proposition 4.1) that a CPPI dynamic leverage constraint provably bounds the maximum drawdown at a user-specified floor $D_{\max}$. The CPPI achieves 0% ruin probability across all synthetic environments and the highest Sortino ratio among all agents in the 2008 historical crisis, at the cost of lower median wealth in benign regimes—the quantified "cost of survival."

Finally, by integrating realistic transaction costs and single-path empirical S&P 500 data for four distinct market epochs, we confirmed that both proposed methods maintain

their theoretical advantages under real-world conditions. The fee audit (Section 5.10) further shows that the Vol-HMM's lower turnover frequency means its net-of-fees performance advantage over high-turnover adversarial agents is *larger* than the gross advantage, not smaller.

6.2 Limitations

Several limitations of the current framework deserve explicit acknowledgment:

1. **Specified emission parameters.** The Vol-HMM emission distributions are specified a priori rather than estimated via Baum–Welch EM (Remark 3.5). While this choice is justified by robustness concerns, it means the agent's performance depends on the accuracy of the domain-knowledge priors. Misspecified priors could cause the filter to track the wrong regime.

2. **Discrete-time CPPI floor breach.** The analytical guarantee of Proposition 4.1 requires $m|r_{\min}| \leq 1$. In discrete time, a single large overnight return (e.g., the -20% day of March 16, 2020) can breach the floor. Our empirical experiments confirm that breaches are rare but not impossible.

3. **Single-asset framework.** All experiments model a single risky asset. Real portfolios exhibit cross-asset correlation and regime coupling that the current framework does not capture.

4. **Fractional EXP3 variant.** Our EXP3 implementation is a continuous-fraction variant whose adversarial regret bound (Theorem 3.6) does not apply exactly (Remark 3.4). The bound motivates the agent's design rather than providing a strict guarantee.

6.3 Future Directions

Several high-impact research pathways remain open:

1. **Multi-Asset Regime Coupling.** Extending the Vol-HMM to detect correlated regime shifts across N assets using non-Gaussian copula structures would be a natural next step. A joint HMM over a multivariate return vector would allow the CPPI floor to be defined in terms of portfolio drawdown rather than single-asset drawdown.

2. **Learned Emission Parameters via Online EM.** Replacing the fixed emission specification with an online Baum–Welch update (using a sliding window to prevent over-accumulation of past data) would make the Vol-HMM fully adaptive, removing the dependence on domain- knowledge priors.

3. **Deep Reinforcement Learning for Hyperparameter Tuning.** The CPPI multiplier m, the CUSUM threshold h, and the HMM persistence A_{ii} are currently fixed by grid search. A DRL agent trained to dynamically adjust these hyperparameters in response to real-time market entropy could further reduce detection lag and improve the growth-safety trade-off.

4. **Microstructure Friction Modeling.** The current transaction cost model is proportional to turnover. Incorporating bid-ask spread decay, order book impact, and market impact functions would make the friction model more realistic for high-frequency applications and would likely further favor low-turnover agents such as the Vol-HMM.

These directions represent the next iteration of protecting adaptive systems against the inherent non-stationarity of human-centric financial markets.

Appendix A

Code-to-Methodology Verification

To ensure complete computational transparency, the mathematical constructs defined in Chapters 2 through 4 are mapped directly to their execution scope within the central Python architecture.

Theoretical Construct	**Equation Form**	**Repository Class / Func**
Bayesian Posterior Map	$\alpha_t = \alpha_{t-1} + y_t$	`ThompsonKellyAgent.update()`
Kelly Oracle Bound	$f^* = \mu/\sigma^2$	`compute_kelly_fraction()`
Vol-HMM Log-Likelihood	$\ln \alpha_t(j) = \ln b_j + \log \sum \dots$	`VolAugmentedHMM.update()`
CPPI Safety Floor	$\lambda_t = \min(1, m \cdot c_t/W_t)$	`RiskConstrainedKelly.act()`
Drawdown Metric	$D_t = 1 - W_t/\max(W_{<t})$	`WealthTracker`
EXP3 Regret Bound	$p_{i,t} = (1-\gamma)\frac{w_i}{\sum w} + \frac{\gamma}{K}$	`EXP3Agent.act()`

All random outcomes were synchronized utilizing Common Random Numbers (CRN) in `distributions.py` to securely bound Monte Carlo comparisons free of idiosyncratic noise loops.

Bibliography

Shipra Agrawal and Navin Goyal. Analysis of Thompson sampling for the multi-armed bandit problem. *Proceedings of the Conference on Learning Theory (COLT)*, 23:39.1–39.26, 2012.

Andrew Ang and Geert Bekaert. Regime switches in interest rates. *Journal of Business & Economic Statistics*, 20(2):163–182, 2002.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

Fischer Black and André F. Perold. Theory of constant proportion portfolio insurance. *Journal of Economic Dynamics and Control*, 16(3–4):403–426, 1992.

Leo Breiman. Optimal gambling systems for favorable games. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1:65–78, 1961.

Enzo Busseti, Ernest K. Ryu, and Stephen Boyd. Risk-constrained Kelly gambling. *Journal of Investing*, 25(3):118–134, 2016.

Eugene F. Fama. The behavior of stock-market prices. *Journal of Business*, 38(1):34–105, 1965.

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

John L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35(4):917–926, 1956.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

Pierre L'Ecuyer. Efficiency improvement via common random numbers. *Winter Simulation Conference Proceedings*, pages 255–263, 1994.

Benoit Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36 (4):394–419, 1963.

George V. Moustakides. Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, 14(4):1379–1387, 1986.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1–2):100–115, 1954.

Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

Paul A. Samuelson. Why we should not make mean log of wealth big though years to act are long. *Journal of Banking & Finance*, 3(4):305–307, 1979.

William F. Sharpe. Mutual fund performance. *Journal of Business*, 39(1):119–138, 1966.

Frank A. Sortino and Robert Van Der Meer. Downside risk. *Journal of Portfolio Management*, 17(4):27–31, 1991.

William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3–4):285–294, 1933.

Edward O. Thorp. The Kelly criterion in blackjack, sports betting, and the stock market. In S.A. Zenios and W.T. Ziemba, editors, *Handbook of Asset and Liability Management*, pages 385–428. Elsevier, 2008.

William T. Ziemba and Donald B. Hausch. *Beat the Racetrack*. Morrow, New York, 1986.