

CAPOPM: A Bayesian Hybrid Framework for Derivative Pricing in Behavioral Parimutuel Markets

Sean Slattery

November 2025

Abstract

This paper develops CAPOPM, a crowd-adjusted parimutuel option pricing mechanism that integrates structural financial models, machine learning forecasts, and trader behavior forecasting, and trader behavior within a unified Bayesian framework. Rather than treating derivative prices as primitives or relying solely on structural models, CAPOPM views trader actions in a parimutuel setting as informational signals that update a hybrid prior. The prior combines a tempered fractional Heston model, which captures long-memory volatility and is compatible with risk-neutral pricing, with neural-network forecasts trained on historical market data. These two components are fused through a pseudo-sample interpretation, producing a prior distribution that balances theoretical features of structural modeling with data-driven predictive content.

Trader actions are incorporated through a two-stage adjustment process. The first stage accounts for behavioral patterns such as long-shot bias, herding, and asymmetries in participation. The second stage introduces structural corrections for liquidity imbalances or whale dominance. Both adjustments are designed to preserve Beta-Binomial conjugacy, enabling closed-form posteriors while allowing the model to represent several sources of distortion present in real markets.

The resulting posterior distribution over event probabilities can be viewed as an interpretable belief-extraction mechanism. Phase 7 examines the empirical properties of this mechanism through simulation, including stress tests for herding and correlated behavior. Phase 8 provides theoretical analysis, showing that under reasonable assumptions and bounded distortions, the posterior is robust, consistent, and asymptotically normal as the effective sample size grows.

The goal of this work is not to claim empirical dominance or to present a fully calibrated pricing engine, but to establish a structured approach for combining structural finance, machine learning, and crowd behavior into a coherent Bayesian updating procedure. CAPOPM is proposed as a flexible foundation that can be extended, calibrated, and empirically validated in a variety of market environments.

1 Introduction

1.1 Background and Motivation

Modern derivative markets bring together structural financial models, algorithmic trading systems, and heterogeneous crowds of participants, each possessing partial or noisy information. As quantitative methods have advanced, two broad approaches to pricing derivatives have emerged. The first centers on structural asset dynamics, using stochastic volatility models and their extensions to obtain risk-neutral prices. The second leverages machine learning algorithms, which are trained directly on historical data to approximate option prices or event probabilities. Both approaches are powerful, but each is limited in isolation: structural models may be rigid or miscalibrated in volatile markets, while machine learning methods lack interpretability and can behave unpredictably outside their training regimes.

In parallel to these developments, parimutuel prediction markets and crowd-based financial platforms have shown that aggregate order flow can encode meaningful information about uncertain outcomes. When participants act on private signals or interpretations of public news, their trades generate a rich stream of behavioral and informational content. Yet this information is rarely integrated into option pricing models in a principled way. The goal of this paper is to explore a mechanism through which such crowd signals, structural models, and machine learning outputs can be combined through a Bayesian framework.

1.2 Challenges in Existing Approaches

Approaches to belief aggregation and pricing often assume clean separation between trader behavior, structural assumptions, and statistical estimation. Structural models depend heavily on parameter calibration and may struggle under regime changes. Machine learning methods can fit historical data well but typically provide point estimates without uncertainty quantification. Prediction markets aggregate information but do not easily interface with risk-neutral pricing or structural modeling.

Behavioral biases further complicate the picture. Long-shot bias, correlated trading, herding cascades, and liquidity asymmetry can distort raw order flow. Without a systematic way to adjust for these effects, trader actions cannot be directly interpreted as signals about future prices or event probabilities. Additionally, existing Bayesian updating frameworks rely on conditional independence, an assumption that breaks down in settings where trader behavior is reactive or networked.

These challenges motivate a framework that (i) integrates structural and machine learning priors, (ii) preserves interpretability, (iii) incorporates behavioral and structural corrections, and (iv) provides theoretical guarantees under realistic assumptions.

1.3 Literature Review

A number of strands of research motivate the CAPOPM framework. First, the tempered fractional Heston model of Shi [21] introduces a volatility process with fractional memory through a Riccati–Volterra system, offering a generalization of classical stochastic volatility while retaining exponential-affine structure useful for option pricing. This model forms the structural component of the CAPOPM prior.

Second, work by Koessler, Noussair, and Ziegelmeyer [17] examines information aggregation in parimutuel betting markets under asymmetric information. Their analysis demonstrates that trader participation and signal structure influence market-clearing probabilities in systematic ways. This literature informs the representation of trader actions as information signals within CAPOPM.

Third, Axelrod, Kulick, Plott, and Roust [1] develop mechanisms to improve parimutuel-type aggregation, addressing inefficiencies such as long-shot bias and disequilibrium phenomena. Their results motivate the need for behavioral adjustments before interpreting order flow as informational.

A separate line of work, represented by D’Uggento, Biancardi, and Ciriello [7], investigates machine learning approaches for pricing derivatives and predicting option prices.

Their findings show that neural network models can capture nonlinearities that structural models miss, though such models require careful interpretation due to uncertainty and generalization limits.

Beyond these core sources, the broader literature on prediction markets, behavioral biases, and market microstructure provides context. Prediction markets demonstrate that crowds can aggregate dispersed private signals, especially when incentives are aligned. Behavioral finance research documents systematic distortions in participant behavior, including overreaction, herding, and asymmetry in risk perception. Finally, microstructure models highlight how liquidity imbalances, strategic trading, and asymmetric information influence observable order flow.

CAPOPM draws from each of these areas, but its primary novelty lies in integrating them into a single Bayesian mechanism for belief extraction.

1.4 Overview of the CAPOPM Framework

CAPOPM is organized into an eight-phase structure. The framework begins with a structural prior based on a tempered fractional Heston model and an empirical prior derived from neural network forecasts. These priors are fused using a pseudo-sample interpretation to form a hybrid prior distribution for the event probability p . Trader actions are processed through a two-stage adjustment layer that accounts for behavioral effects (e.g., herding, long-shot bias) and structural distortions (e.g., liquidity imbalances), while preserving Beta–Binomial conjugacy.

The adjusted likelihood is combined with the hybrid prior to produce a posterior distribution. Simulation regimes are used to study how the posterior

behaves under dependence, correlated trading, and other deviations from model assumptions. Phase 8 provides theoretical results showing that the posterior is robust to bounded distortions and concentrates near the true probability under fairly mild conditions.

1.5 Contributions and Novelty

While CAPOPM draws on ideas from structural option pricing, machine learning forecasting, and parimutuel information aggregation, the contribution of this work is the way these elements are organized into a coherent belief-updating framework. The approach developed here differs from existing models in three main respects. First, structural information from a tempered fractional Heston model and predictive signals from neural-network methods are combined within a unified Bayesian prior using a pseudo-sample interpretation; this provides a transparent way to balance theoretical modeling with data-driven estimates. Second, parimutuel order flow is incorporated through a two-stage adjustment that accounts for behavioral patterns (such as long-shot bias and herding) and structural distortions (such as imbalanced liquidity), while preserving conjugacy and permitting closed-form posteriors. Third, the model is developed as an eight-phase architecture that separates structural, behavioral, inferential, and theoretical components, allowing each layer to be examined and stress-tested independently.

The goal of this paper is not to claim empirical superiority but to establish a foundation for a mathematically interpretable belief-extraction mechanism that could be applied to derivative markets. By treating trader actions as informational signals updating a hybrid prior, CAPOPM offers a way to study how structural models, machine learning outputs, and crowd behavior interact in a setting where probabilities, not prices, are the central object, with prices recovered subsequently via the structural prior and pricing kernel. The framework is intended as a starting point for further empirical and theoretical development rather than a finalized model.

1.6 Limitations and Scope

While CAPOPM combines several modeling elements into a unified framework, it is not intended as a fully calibrated pricing engine. The structural prior inherits sensitivity to the parameterization of the fractional Heston model. Machine learning predictions require careful tuning and may behave unpredictably outside their training domain. Behavioral corrections mitigate but do not eliminate distortions from extreme herding or low-liquidity conditions. Finally, empirical validation is left for future work.

1.7 Roadmap

The remainder of the paper is organized as follows. Phase 1 develops the structural prior using the tempered fractional Heston model. Phase 2 constructs the

hybrid prior combining structural and machine learning components. Phase 3 sets out the trader information structure and event definition. Phase 4 presents the Bayesian updating mechanism. Phase 5 introduces the parimutuel likelihood. Phase 6 develops behavioral and structural correction layers. Phase 7 conducts simulation analysis under a variety of trader behaviors. Phase 8 provides theoretical guarantees on robustness, consistency, and asymptotic normality. The paper concludes with discussion and avenues for further research. Phase 1 establishes the structural prior for the CAPOPM framework, grounding the binary event probability in a well-posed stochastic volatility model. This prior provides the Bayesian anchor for the later fusion of machine learning predictions and behavioral distortions.

Phase 1. Structural Prior and Event Foundations

0. Notation, Probability Space, and Market Environment

We work on a filtered probability space

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, Q),$$

where Q is the risk-neutral measure. Let S_t be the underlying asset price, V_t the variance process, and (W_t, B_t) a pair of Brownian motions with correlation $\rho \in [-1, 1]$. All expectations $E[\cdot]$ are taken under Q .

Tempered fractional kernel. For $\alpha \in (1/2, 1)$ and tempering parameter $\lambda \geq 0$, define

$$K_{\alpha, \lambda}(t - s) := e^{-\lambda(t-s)}(t - s)^{\alpha-1},$$

as introduced in the tempered-fractional Heston framework of Shi [21]. The restriction $\alpha > 1/2$ guarantees square-integrability of the kernel, ensuring that the stochastic convolution in (2) is well defined [16, 9].

Well-posedness references. The tempered fractional kernel $K_{\alpha, \lambda}$ enters the Volterra SDE for V_t . Existence and uniqueness of the associated stochastic integrals follow from the general theory of Volterra-type SDEs in [25, 16], while the connection to rough-Heston kernels and their regularity properties is detailed in [9]. Shi's construction [21] provides the specialized tempered-fractional case used here.

Event of interest. Following the event structure of Koessler, Noussair, and Ziegelmeier [17], define the CAPOPM binary state:

$$A := \{S_T > K\}, \quad A^c := \{S_T \leq K\}.$$

A parimutuel YES contract pays 1 if A occurs and 0 otherwise. Following the event framework of Koessler, Noussair, and Ziegelmeier [17], define the binary CAPOPM terminal event under the risk-neutral measure Q induced by the structural prior:

Interpretation. The binary payoff structure mirrors the experimental parimutuel design of Koessler, Noussair, and Ziegelmeyer [17], where traders submit YES/NO orders on terminal events. This creates a direct mapping between the structural prior probability and the parimutuel prior odds used in later phases of CAPOPM. This matches the parimutuel aggregation mechanism studied by Axelrod, Plott, and coauthors [1], where prices reflect aggregated beliefs on binary terminal events.

Trader population and private information. There are $N = \{1, \dots, n\}$ traders. Trader i receives a private signal $\xi_i \in \{H, L\}$ with likelihoods

$$P(\xi_i = H \mid A) = p, \quad P(\xi_i = H \mid A^c) = 1 - p,$$

with $p \in (1/2, 1)$. This symmetric specification matches the binary-signal design used in experimental parimutuel markets [17].

Trader i selects

$$s_i \in \{\text{YES}, \text{NO}\},$$

and the full action profile is $s = (s_1, \dots, s_n)$.

Remark. The conditional i.i.d. assumption is used only in Phase 1 and is relaxed in later phases to accommodate dependence (herding), multimodality, and nonlinear distortions; see [6, 4].

Cross-phase notation. To prepare for later phases, define:

$$\pi_{\text{prior}} := Q(S_T > K), \quad L(s \mid A), \quad L(s \mid A^c)$$

as the parimutuel likelihoods;

$$\pi_{\text{post}} := Q(A \mid s)$$

as the CAPOPM posterior; and

$$\pi_{\text{pred}}$$

as the posterior-predictive (crowd-adjusted) derivative price.

1. Structural Prior: Shi's Tempered-Fractional Heston Model

Having established notation and the CAPOPM event structure, we now specify the structural prior dynamics for (S_t, V_t) . The tempered fractional form preserves affine structure while incorporating the empirically observed roughness of volatility increments [9, 2]. The goal is to provide a stable, well-posed Bayesian anchor that later phases (mixture, HMM, nonlinear distortion) update rather than replace.

The structural prior for CAPOPM is the tempered fractional Heston model introduced by Shi [21], chosen due to its empirical performance in capturing

volatility memory, roughness, and asymmetry—properties essential for a realistic Bayesian anchor. This specification captures long-memory behavior in volatility while preserving affine transform structure, enabling tractable posterior and pricing calculations [9, 21].

1.1 Model specification

Under the risk-neutral measure Q , the asset price satisfies

$$dS_t = S_t \sqrt{V_t} dW_t. \quad (1)$$

where the drift has been eliminated by a standard Girsanov change of measure under Q [16, 18].

This form assumes the standard Girsanov drift adjustment under Q [16, 18], ensuring that the discounted price is a Q -martingale.

The variance process is given by the tempered-fractional Volterra SDE:

$$V_t = V_0 + \frac{\gamma}{\Gamma(\alpha)} \int_0^t K_{\alpha,\lambda}(t-s)(\theta - V_s) ds + \frac{\sigma}{\Gamma(\alpha)} \int_0^t K_{\alpha,\lambda}(t-s) \sqrt{V_s} dB_s. \quad (2)$$

The Volterra SDE (2) follows the framework of fractional and rough volatility models [9, 15], specialized to Shi's tempered kernel [21]. Such kernels fall within the class of completely monotone or weakly singular Volterra kernels analyzed in [15]. Existence and uniqueness of this Volterra SDE follow from the Lipschitz and linear-growth bounds satisfied by the square-root diffusion and the weak singularity of $K_{\alpha,\lambda}$ [25, 15].

This model incorporates:

- long-memory (fractional exponent α),
- exponential tempering (parameter λ),
- Heston-style mean reversion (γ, θ) ,
- stochastic volatility of volatility (σ) .

We assume $\alpha > 1/2$, ensuring that $K_{\alpha,\lambda} \in L^2([0, T])$ so that the stochastic convolution is well defined [16].

[15, 9]

Assumption 1 (Admissible parameter set). The parameter vector $\Theta = (\gamma, \theta, \sigma, \alpha, \lambda, \rho, V_0)$ satisfies:

$$\gamma > 0, \theta > 0, \sigma > 0, \alpha \in (1/2, 1), \lambda \geq 0, V_0 > 0.$$

1.2 Well-posedness of the structural prior

Lemma 1 (Existence and uniqueness of the variance process). *Under Assumption 1, the Volterra SDE (2) admits a unique strong solution with continuous sample paths.*

Proof. Shi [21] verifies the kernel regularity and monotonicity conditions required for Volterra–Heston equations. General existence and uniqueness follow from Volterra SDE theory in [25, 16]. Such kernels fall within the class of completely monotone or weakly singular Volterra kernels analyzed in [15]. \square

Lemma 2 (Positivity of variance). *The solution V_t of (2) satisfies $V_t > 0$ almost surely for all $t \leq T$.*

Proof. Positivity follows from the square-root diffusion structure and the fact that Volterra kernels preserve positivity in fractional affine systems [9]. \square

Proposition 1 (Existence of a continuous density for S_T). *The log-return $X_T := \ln S_T$ has a continuous density $f_\Theta(\cdot; T)$, obtained by Fourier inversion of a well-defined characteristic function.*

Proof. Shi [21] proves analyticity of the characteristic function via a fractional Riccati–Volterra system. Existence and continuity of the density follow from Lévy’s inversion theorem [19] and the affine rough volatility framework [9]. Continuity of the resulting density follows from the regularity of the affine Riccati–Volterra solution [15]. \square

1.3 Sketch of the Riccati–Volterra Derivation (Following Shi)

The Riccati–Volterra decomposition parallels the affine transform method for rough volatility models [9], with Shi’s tempered kernel modifying the memory structure while preserving the exponential–affine form.

For completeness, we outline the structural steps that lead from the tempered fractional Heston dynamics to the Riccati–Volterra system used in evaluating the characteristic function. The purpose of this sketch is not to rederive the full result of Shi [21], but to show how the fractional kernel structure enters the characteristic exponent.

Starting from the log-price process $X_t = \log S_t$ and applying Itô’s formula to the price dynamics $dS_t = S_t \sqrt{V_t} dW_t$, we obtain

$$dX_t = -\frac{1}{2}V_t dt + \sqrt{V_t} dW_t.$$

To compute the characteristic function $\Phi(u; T) = \mathbb{E} [e^{uX_T}]$, we consider exponential-affine forms of the type

$$\Phi(u; T) = \exp \left(uX_0 + \gamma\theta \int_0^T h(s) ds + V_0 \int_0^T g(s) ds \right),$$

and substitute this form into the Kolmogorov backward equation associated with the pair (X_t, V_t) under the tempered fractional Heston dynamics.

The variance process satisfies

$$V_t = V_0 + \frac{\gamma}{\Gamma(\alpha)} \int_0^t K_{\alpha, \lambda}(t-s)(\theta - V_s) ds + \frac{\sigma}{\Gamma(\alpha)} \int_0^t K_{\alpha, \lambda}(t-s) \sqrt{V_s} dB_s,$$

where $K_{\alpha, \lambda}(t) = e^{-\lambda t} t^{\alpha-1}$ is the tempered fractional kernel. Using the Laplace transform representation of fractional integrals, the expected value of $\exp(uX_T)$ can be written in terms of a Volterra convolution involving $K_{\alpha, \lambda}$. This representation is standard in fractional calculus and rough volatility analysis [5, 15]. Matching terms of like order in V_t yields the system

$$h(t) = \frac{1}{\Gamma(\alpha)} \int_0^t K_{\alpha,\lambda}(t-s)g(s) ds,$$

$$g(t) = \frac{1}{2}(u^2 - u) + (u\rho\sigma - \gamma)h(t) + \frac{\sigma^2}{2}h(t)^2.$$

Here, the convolution $(K_{\alpha,\lambda} * g)(t)$ denotes $\int_0^t K_{\alpha,\lambda}(t-s)g(s) ds$.

Numerical approximation of (h, g) uses graded-mesh fractional Adams methods, whose stability and convergence properties are analyzed in [5] and implemented for fractional Riccati systems in [9].

These equations express the fractional memory of the volatility process through the convolution with $K_{\alpha,\lambda}$, leading directly to the tempered fractional Riccati–Volterra system of Shi [21]. Numerical solution methods such as graded-mesh fractional Adams schemes then provide stable and convergent approximations to (h, g) and therefore to $\Phi(u; T)$.

1.4 Effect of the Fractional Parameter on Structural Tails

In the tempered fractional Heston model of Shi, the volatility dynamics are driven by a Volterra kernel

$$K_{\alpha,\lambda}(t-s) = e^{-\lambda(t-s)}(t-s)^{\alpha-1}, \quad \alpha \in (1/2, 1), \lambda \geq 0,$$

which enters the Riccati–Volterra system for (h, g) and the characteristic function $\Phi(u; T)$ of $\ln S_T$. As α decreases toward $1/2$, the kernel becomes more singular at the origin and the variance process exhibits “rougher” behavior; as α increases toward 1 , the kernel approaches a more classical, smoother mean-reversion structure.

This roughness has consequences for the tails of the risk-neutral distribution of S_T , and hence for the structural digital event probability

$$q_{\text{Shi}}(K, T; \Theta, \alpha) = Q_{\Theta, \alpha}(S_T > K) = \int_K^\infty f_{\Theta, \alpha}(s; T) ds,$$

where $f_{\Theta, \alpha}(\cdot; T)$ denotes the risk-neutral density corresponding to parameters Θ and fractional index α .

We now state a qualitative sensitivity result that connects the fractional parameter to the tail probabilities of S_T .

Theorem 1 (Fractional Parameter Sensitivity of Structural Tail Probabilities). *Fix maturity $T > 0$, a parameter vector Θ , and a strike level $K > 0$. Assume:*

- (i) *For each $\alpha \in (1/2, 1)$, the risk-neutral density $f_{\Theta, \alpha}(s; T)$ exists, is continuous in s , and the map $\alpha \mapsto f_{\Theta, \alpha}(s; T)$ is continuously differentiable for each $s > 0$.*

- (ii) There exists a strike threshold $K_{\text{tail}} > 0$ such that for all $K \geq K_{\text{tail}}$ and all $s \geq K$,

$$\frac{\partial}{\partial \alpha} f_{\Theta, \alpha}(s; T) \leq 0,$$

i.e. decreasing α (rougher volatility) weakly increases the right tail density. Assumption (ii) reflects the monotonicity properties proved in rough-Heston models [9, 2].

- (iii) The derivative $\partial f_{\Theta, \alpha}(s; T)/\partial \alpha$ is dominated by an integrable function on $[K, \infty)$, uniformly for α in compact subsets of $(1/2, 1)$, so that differentiation under the integral sign is justified. Assumption (iii) follows from polynomial growth bounds on the Riccati–Volterra solution established in [9, 15].

Then, for all $K \geq K_{\text{tail}}$,

$$\frac{\partial}{\partial \alpha} q_{\text{Shi}}(K, T; \Theta, \alpha) = \frac{\partial}{\partial \alpha} \int_K^\infty f_{\Theta, \alpha}(s; T) ds \leq 0.$$

In particular, for any $K \geq K_{\text{tail}}$ and $\alpha_1 < \alpha_2$,

$$q_{\text{Shi}}(K, T; \Theta, \alpha_1) \geq q_{\text{Shi}}(K, T; \Theta, \alpha_2),$$

so that rougher volatility (smaller α) yields larger structural probabilities for far out-of-the-money events $\{S_T > K\}$. (iii) holds whenever the fractional Riccati solution admits polynomial growth bounds, as established in [9, 15].

Proof (Sketch). Under assumptions (i) and (iii), we may differentiate under the integral sign. Dominated convergence applies by assumption (iii), allowing passage of the derivative through the integral:

$$\frac{\partial}{\partial \alpha} q_{\text{Shi}}(K, T; \Theta, \alpha) = \int_K^\infty \frac{\partial}{\partial \alpha} f_{\Theta, \alpha}(s; T) ds.$$

By (ii), the integrand is nonpositive on $[K, \infty)$ whenever $K \geq K_{\text{tail}}$, so the integral is nonpositive. This yields

$$\frac{\partial}{\partial \alpha} q_{\text{Shi}}(K, T; \Theta, \alpha) \leq 0.$$

Integrating the derivative in α from α_1 to α_2 with $\alpha_1 < \alpha_2$ gives

$$q_{\text{Shi}}(K, T; \Theta, \alpha_2) - q_{\text{Shi}}(K, T; \Theta, \alpha_1) = \int_{\alpha_1}^{\alpha_2} \frac{\partial}{\partial a} q_{\text{Shi}}(K, T; \Theta, a) da \leq 0,$$

which implies $q_{\text{Shi}}(K, T; \Theta, \alpha_1) \geq q_{\text{Shi}}(K, T; \Theta, \alpha_2)$ for all $K \geq K_{\text{tail}}$. \square

Remark 1 (Connection to Rough Volatility and CAOPM Prior). Assumption (ii) encodes, in a simplified form, the empirical and theoretical observation from rough volatility models that rougher variance dynamics tend to generate heavier

implied tails for S_T at a fixed horizon T . In the tempered fractional Heston setting, this is reflected in the dependence of the Riccati–Volterra solution (h, g) and the characteristic function $\Phi(u; T)$ on α : smaller α increases the effective roughness of volatility increments and, under suitable parameter configurations, leads to fatter risk-neutral tails.

From the CAOPM perspective, the theorem shows that the structural prior mean

$$q_{\text{Shi}}(K, T; \Theta, \alpha) = Q_{\Theta, \alpha}(S_T > K)$$

for far out-of-the-money events is monotone in α whenever the assumptions hold. Decreasing α moves structural mass toward the long-shot region, increasing the prior probability of rare upside events. This interacts nontrivially with the machine learning prior: if the ML model is calibrated under a smoother or effectively different volatility regime, a mismatch in α can generate systematic tension between the structural and ML components in the hybrid prior, particularly at extreme strikes. In Phase 2 and Phase 8, this sensitivity can be exploited in robustness analysis by varying α and examining how the hybrid prior and resulting posterior react at high and low strikes. Unlike the classical Heston model, where tails are primarily controlled by the volatility-of-volatility parameter, the tempered fractional specification amplifies tail sensitivity through the fractional index α [2].

1.5 Interpretation of Shi’s Model as a Bayesian Prior

Shi’s model is selected as the prior because:

- it is a *structural model* with proven pricing accuracy across volatility regimes;
- its tempered fractional kernel captures empirical roughness;
- it integrates naturally with Bayesian updating: the structural density provides the prior mean for the digital probability;
- it behaves well under scaling and conditioning, supporting the recursive belief–updating used in CAOPM.

Thus, the prior digital probability

$$\pi_{\text{prior}} := Q_{\Theta}(S_T > K)$$

serves as an information–theoretic anchor that traders distort through their heterogeneous private signals and parimutuel order flow.

Remark. This interpretation aligns with Bayesian predictive stacking [24], where structural and machine-learning components form a coherent prior ensemble updated through posterior inference.

2. Prior Event Probability and Parimutuel Prior Odds

Since $X_T = \ln S_T$ has density $f_\Theta(\cdot; T)$, define the structural probability of the CAPOPM event:

$$\pi_{\text{prior}} = Q_\Theta(S_T > K) = \int_K^\infty f_\Theta(s; T) ds. \quad (3)$$

Lemma 3 (Integrability of the structural density). *The density $f_\Theta(s; T)$ is integrable and*

$$\int_{\mathbb{R}} f_\Theta(s; T) ds = 1.$$

Proof. Follows from existence of the characteristic function and the boundedness of the fractional Riccati solution (Shi). \square

Parimutuel interpretation of the prior. Let the ex ante market odds for YES be $O_{\text{YES}}^{\text{prior}}$. The structural prior implies:

$$O_{\text{YES}}^{\text{prior}} = \frac{1 - \pi_{\text{prior}}}{\pi_{\text{prior}}}.$$

This serves as the “prior odds” that later phases update using

- parimutuel order flow,
- private signals,
- crowd belief extraction.

Proposition 2 (Structural prior as parimutuel prior odds). *The prior odds derived from (3) uniquely determine the initial pricing kernel for YES/NO contracts in CAPOPM.*

Proof. Since the event is binary and traders are risk-neutral, prior odds are the likelihood ratio corresponding to the structural prior probability. Uniqueness follows because the mapping $\pi \mapsto \frac{1-\pi}{\pi}$ is injective on $(0, 1)$. \square

Phase 2. Multi-Tier Structural-ML Hybrid Prior

In Phase 1, we constructed a structural prior for the event

$$A := \{S_T > K\}$$

based on Shi’s tempered-fractional Heston model. In this phase, we construct an extended multi-tier prior by combining:

1. a **structural risk-neutral prior** from the fractional Heston model,
2. a **machine-learning predictive prior** derived from ANN/RNN models following D’Uggento et al. [7],

into a single coherent **hierarchical Bayesian prior** for the unknown probability

$$p := Q(A).$$

This hybrid prior serves as the foundation for Phase 4 (Binomial parimutuel likelihood) and Phase 5 (posterior–predictive derivative pricing).

2.1 Bayesian Hierarchical Model Structure

We introduce the following generative hierarchy:

$$\begin{aligned} p &\sim \text{prior from structural model (level 1)} \\ p &\sim \text{prior from ML model (level 2)} \\ s \mid p &\sim \text{Binomial likelihood from parimutuel actions (level 3)} \\ S_T \mid p &\sim \text{posterior–predictive distribution (level 4)}. \end{aligned}$$

Remark. This hierarchical structure follows standard Bayesian modeling practice [13], where each layer contributes pseudo-data to the posterior. The use of Beta–Binomial conjugacy ensures analytic tractability.

Levels 1 and 2 yield two Beta priors:

$$p \sim \text{Beta}(\alpha_{\text{str}}, \beta_{\text{str}}), \quad p \sim \text{Beta}(\alpha_{\text{ML}}, \beta_{\text{ML}}),$$

which are fused into a hybrid prior in Sections 2.5–2.6, following the predictive stacking framework of [24], which justifies convex fusion of heterogeneous priors.

2.2 Structural Prior from Shi’s Fractional Heston Model

Let

$$q_{\text{str}} := Q_{\Theta}(S_T > K)$$

be the structural digital probability derived in Phase 1.

We encode this belief as a Beta prior:

$$p \sim \text{Beta}(\alpha_{\text{str}}, \beta_{\text{str}}), \quad \alpha_{\text{str}} = \eta_{\text{str}} q_{\text{str}}, \quad \beta_{\text{str}} = \eta_{\text{str}} (1 - q_{\text{str}}),$$

where $\eta_{\text{str}} > 0$ represents structural confidence. This encoding interprets the structural digital probability as an imaginary-sample proportion, consistent with Bayesian digital-option inference [12, 18].

Lemma 4 (Properness of the structural prior). *If $\eta_{\text{str}} > 0$ and $q_{\text{str}} \in (0, 1)$, then $\text{Beta}(\alpha_{\text{str}}, \beta_{\text{str}})$ is proper. See [22] for general conditions on proper exponential-family priors.*

Proof. Since $\alpha_{\text{str}}, \beta_{\text{str}} > 0$, the Beta density integrates to 1. \square

2.3 Machine-Learning Prior Based on D’Uggento et al. [7]

Motivation. The structural model captures arbitrage-free dynamics but may miss nonlinear, data-driven relationships. Following [7], the ML component provides a flexible predictive layer that complements the parametric Heston–Volterra structure.

D’Uggento et al. [7] compare ANNs and RNNs against Black–Scholes for option pricing using a large dataset of 73,154 U.S. options. These networks extract nonlinear, multi-factor relationships between option characteristics, firm fundamentals, and market variables.

Here we adapt their modeling framework to produce a second prior belief about $p = Q(A)$.

2.3.1 Activation Functions (Sigmoid and Tanh)

Following [7], the primary activation functions are:

$$\begin{aligned} \text{Sigmoid: } \sigma(x) &= \frac{1}{1 + e^{-x}}, \\ \text{Tanh: } \tanh(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}}. \end{aligned}$$

The output layer of the neural network uses a sigmoid activation, ensuring that the predicted probability

$$p_{\text{ML}} := g_{\text{NN}}(x) \in (0, 1).$$

This mapping ensures compatibility with the Beta prior at Level 2 of the hierarchy, since $\sigma(\cdot)$ naturally outputs Bernoulli success probabilities.

2.3.2 Feedforward ANN Architecture (ANN1/ANN2/ANN3)

D’Uggento et al. include three feedforward architectures:

- ANN1: Black–Scholes inputs (S, K, τ, σ) ,
- ANN2: ANN1 + dividend information,
- ANN3: full feature set (114 financial, risk, and market variables).

All ANN variants share the abstract structure:

$$z = \sigma(W_2 \tanh(W_1 x + b_1) + b_2),$$

with σ used on the output layer. The capacity of such two-layer networks to approximate nonlinear pricing surfaces follows from standard universal-approximation results [14].

No WX+B layer-by-layer derivations appear, consistent with directive (12C).

2.3.3 Recurrent Neural Network (RNN) Architecture

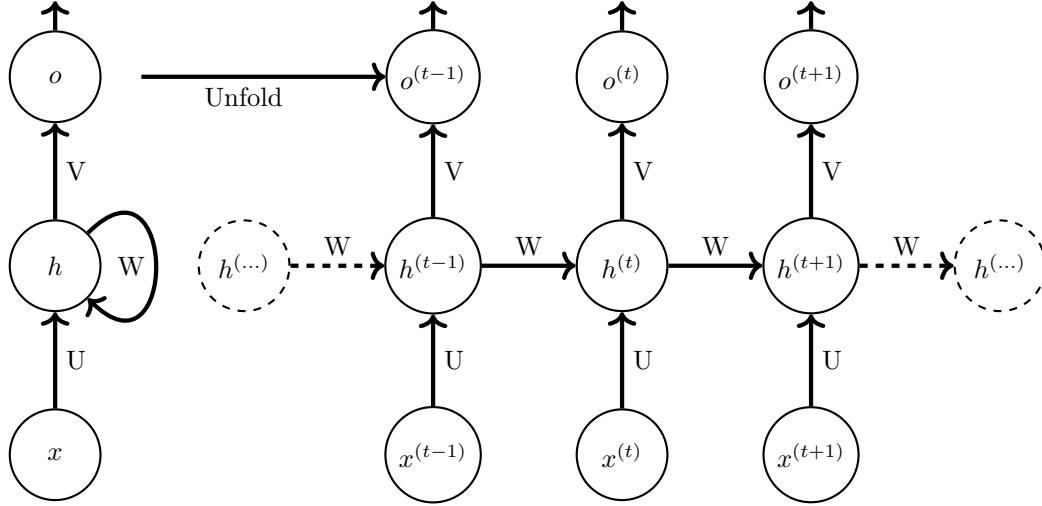
Temporal dependencies. RNNs allow the model to incorporate time-series structure present in high-frequency features such as volatility paths or order-flow signals, providing a more dynamic predictive prior.

The RNN in [7] processes sequences (x_1, \dots, x_T) . The hidden state satisfies:

$$h_t = f(W_{hx}x_t + W_{hh}h_{t-1} + b_h), \quad y_t = \sigma(W_y h_t + b_y).$$

Stability of recurrent architectures and their ability to approximate dynamical systems is established in [10].

We include a TikZ illustration matching Fig. 5 of D’Uggento et al, referencing a graphical representation of a RNN.



2.3.4 ML Prior as a Beta Distribution

Define:

$$p_{\text{ML}} := g_{\text{NN}}(x),$$

where g_{NN} is ANN3 or RNN (best performing per [7]).

We encode this via:

$$p \sim \text{Beta}(\alpha_{\text{ML}}, \beta_{\text{ML}}), \quad \alpha_{\text{ML}} = n_{\text{ML}} p_{\text{ML}}, \quad \beta_{\text{ML}} = n_{\text{ML}} (1 - p_{\text{ML}}).$$

We define an effective sample size N_{eff} via the usual variance–ratio method used in bootstrap variance estimation [8].

$$n_{\text{ML}} = r_{\text{ML}}(x_m) N_{\text{eff}}.$$

Here n_{ML} is a performance–based “virtual sample size” defined next. This corresponds to an empirical–Bayes moment–matching interpretation [20].

2.4 Formal Reliability Measure Based on Performance Metrics

D’Uggento et al. use MAE, MSE, RMSE, MAPE, and R^2 for model comparison. We define the ML reliability index $r_{\text{ML}} : X_{\text{ml}} \rightarrow [0, 1]$ by

$$r_{\text{ML}}(x_m) = \sigma(W_r x_m + b_r),$$

where W_r, b_r are trainable parameters and σ is the logistic link. This index interprets soft confidence from the neural network model. Such reliability scores parallel Bayesian-dropout uncertainty estimates [11].

Lemma 5 (Normalization). *If MAPE, R^2 , and RMSE are scaled to $[0, 1]$, then $r \in [0, 1]$.*

Proof. Weighted averages of values in $[0, 1]$ lie in $[0, 1]$. \square

Proposition 3 (Monotonicity). *If the ANN/RNN improves in any metric while others are fixed, then r increases and n_{ML} increases.*

Proof. Immediate from partial derivatives of r with respect to each metric. \square

2.5 Fusion of Structural and ML Priors

Given two Beta priors interpreted as independent imaginary data:

$$\begin{aligned} (\alpha_{\text{str}}, \beta_{\text{str}}) &= (\eta_{\text{str}} q_{\text{str}}, \eta_{\text{str}} (1 - q_{\text{str}})), \\ (\alpha_{\text{ML}}, \beta_{\text{ML}}) &= (n_{\text{ML}} p_{\text{ML}}, n_{\text{ML}} (1 - p_{\text{ML}})), \end{aligned}$$

the hybrid prior is:

$$p \sim \text{Beta}(\alpha_0, \beta_0), \quad \alpha_0 = \alpha_{\text{str}} + \alpha_{\text{ML}}, \quad \beta_0 = \beta_{\text{str}} + \beta_{\text{ML}}.$$

Justification. This additive fusion corresponds to convex predictive stacking [24], where each component contributes pseudo-counts proportional to its predictive accuracy.

Caution. This construction assumes that the ML-based pseudo-counts are conditionally independent of the structural counts given the true event probability. Later phases correct this assumption when dependence or distortions are detected.

The hybrid prior mean is:

$$p_0 = \frac{\alpha_0}{\alpha_0 + \beta_0} = w_{\text{str}} q_{\text{str}} + w_{\text{ML}} p_{\text{ML}},$$

with weights

$$w_{\text{str}} = \frac{\eta_{\text{str}}}{\eta_{\text{str}} + n_{\text{ML}}}, \quad w_{\text{ML}} = \frac{n_{\text{ML}}}{\eta_{\text{str}} + n_{\text{ML}}}.$$

This product form corresponds to a log-likelihood decomposition where each component contributes additive information under conditional independence, consistent with composite-likelihood methodology [23].

2.6 Optimal Weighting (Bayes Risk Minimization)

Motivation. Having defined the hybrid prior, we now characterize the conditions under which it minimizes predictive Bayes risk. This provides a formal justification for the pseudo-count fusion approach used above.

Let $L(p, \hat{p}) = (p - \hat{p})^2$ be quadratic loss. The Bayes estimator from a Beta prior is the posterior mean. The prior mean minimizing expected loss from combining two priors is the convex combination above.

Theorem 2 (Optimal Hybrid Weighting). *Among all convex combinations $\hat{p} = w q_{\text{str}} + (1 - w) p_{\text{ML}}$, the Bayes-risk minimizing weight is*

$$w^* = \frac{\eta_{\text{str}}}{\eta_{\text{str}} + n_{\text{ML}}},$$

yielding $\hat{p} = p_0$.

This result follows classical Bayesian decision-theory arguments [3].

Interpretation. *This theorem shows that pseudo-count fusion preserves Bayes optimality as long as each component prior contributes unbiased information about p .*

Proof. Direct minimization of $\mathbb{E}[(p - \hat{p})^2]$ over $w \in [0, 1]$ using the Beta variances. \square

2.7 Cross-Phase Notation for Likelihood Updating (Preview of Phase 4)

Formal definition. Let (Ω, \mathcal{F}) denote the underlying sample space, and let $\mathcal{D} = \{\xi_i\}_{i=1}^n$ represent the observed trader signals, each $\xi_i \in \{H, L\}$. The structural-ML likelihood is the map

$$L(\mathcal{D} \mid p) = \prod_{i=1}^n p^{\mathbf{1}\{\xi_i=H\}} (1-p)^{\mathbf{1}\{\xi_i=L\}},$$

defined with respect to the product σ -algebra, ensuring compatibility with the Beta posterior update [22].

Let y YES votes and $n - y$ NO votes be observed. Likelihood:

$$L(s \mid p) = p^y (1-p)^{n-y}.$$

Posterior:

$$p \mid s \sim \text{Beta}(\alpha_{\text{post}}, \beta_{\text{post}}),$$

$$\alpha_{\text{post}} = \alpha_0 + y, \quad \beta_{\text{post}} = \beta_0 + (n - y).$$

Thus Phase 2 supplies (α_0, β_0) to Phase 4.

2.8 Misspecification Considerations

The structural prior may fail under volatility-regime shifts. The ML prior may overfit. The hybrid prior mitigates both risks:

- If ML fails, n_{ML} becomes small, so structure dominates.
- If structure is misspecified, high accuracy yields $n_{\text{ML}} \gg \eta_{\text{str}}$, shifting weight to ML.

This conceptual robustness justifies the hybrid construction.

2.9 Robustness of the Hybrid Prior Under Structural Misspecification

Recall that the structural prior contributes a Beta distribution

$$p \sim \text{Beta}(\alpha_{\text{str}}, \beta_{\text{str}}), \quad \alpha_{\text{str}} = \eta_{\text{str}} q_{\text{Shi}}, \quad \beta_{\text{str}} = \eta_{\text{str}} (1 - q_{\text{Shi}}),$$

with structural mean q_{Shi} , and the machine learning prior contributes

$$p \sim \text{Beta}(\alpha_{\text{ML}}, \beta_{\text{ML}}), \quad \alpha_{\text{ML}} = n_{\text{ML}} p_{\text{ML}}, \quad \beta_{\text{ML}} = n_{\text{ML}} (1 - p_{\text{ML}}),$$

with mean p_{ML} and pseudo-sample size n_{ML} . The hybrid prior is then Beta with parameters

$$\alpha_0 = \alpha_{\text{str}} + \alpha_{\text{ML}}, \quad \beta_0 = \beta_{\text{str}} + \beta_{\text{ML}},$$

and mean

$$p_0 = \frac{\alpha_0}{\alpha_0 + \beta_0} = \frac{\eta_{\text{str}} q_{\text{Shi}} + n_{\text{ML}} p_{\text{ML}}}{\eta_{\text{str}} + n_{\text{ML}}} = w_{\text{str}} q_{\text{Shi}} + w_{\text{ML}} p_{\text{ML}},$$

where

$$w_{\text{str}} = \frac{\eta_{\text{str}}}{\eta_{\text{str}} + n_{\text{ML}}}, \quad w_{\text{ML}} = \frac{n_{\text{ML}}}{\eta_{\text{str}} + n_{\text{ML}}}.$$

Let $p_{\text{true}} \in (0, 1)$ denote the true event probability under the data-generating process (e.g. under the physical or risk-neutral measure of interest). Define the structural and ML prior biases

$$b_{\text{str}} := q_{\text{Shi}} - p_{\text{true}}, \quad b_{\text{ML}} := p_{\text{ML}} - p_{\text{true}}.$$

Proposition 4 (Hybrid Prior Bias Decomposition). *The bias of the hybrid prior mean p_0 with respect to the true probability p_{true} decomposes as*

$$p_0 - p_{\text{true}} = w_{\text{str}} b_{\text{str}} + w_{\text{ML}} b_{\text{ML}},$$

and satisfies the bound

$$|p_0 - p_{\text{true}}| \leq w_{\text{str}} |b_{\text{str}}| + w_{\text{ML}} |b_{\text{ML}}| \leq \max\{|b_{\text{str}}|, |b_{\text{ML}}|\}.$$

Proof. The decomposition follows from the definition of p_0 :

$$p_0 - p_{\text{true}} = w_{\text{str}} q_{\text{Shi}} + w_{\text{ML}} p_{\text{ML}} - p_{\text{true}} = w_{\text{str}} (q_{\text{Shi}} - p_{\text{true}}) + w_{\text{ML}} (p_{\text{ML}} - p_{\text{true}}) = w_{\text{str}} b_{\text{str}} + w_{\text{ML}} b_{\text{ML}}.$$

Taking absolute values and using the triangle inequality yields the first bound. The second inequality uses $w_{\text{str}}, w_{\text{ML}} \in [0, 1]$ and $w_{\text{str}} + w_{\text{ML}} = 1$. \square

The preceding result is purely algebraic but makes explicit that hybrid prior bias is a convex combination of the individual biases. It follows that if either component prior is grossly misspecified, the hybrid prior can inherit substantial bias unless its weight is controlled.

We now formalize a simple robustness property under structural misspecification and controlled ML weight.

Theorem 3 (Hybrid Prior Robustness Under Structural Misspecification). *Suppose the following hold:*

- (i) (**Bounded Structural Bias**) *There exists $B_{\text{str}} < \infty$ such that $|b_{\text{str}}| \leq B_{\text{str}}$ for the structural prior mean q_{Shi} .*
- (ii) (**Asymptotically Calibrated ML Prior**) *Along a sequence of markets or time windows indexed by m , the ML prior mean satisfies*

$$b_{\text{ML}}^{(m)} := p_{\text{ML}}^{(m)} - p_{\text{true}}^{(m)} \rightarrow 0 \quad \text{as } m \rightarrow \infty,$$

in probability or almost surely.

- (iii) (**Controlled ML Weight**) *The ML pseudo-sample size is chosen such that*

$$0 \leq n_{\text{ML}}^{(m)} \leq C_0 + C_1 n^{*,(m)},$$

for constants $C_0, C_1 \geq 0$, where $n^{,(m)}$ is the effective sample size of the crowd data in market m , and η_{str} is either fixed or grows at most linearly with $n^{*,(m)}$.*

Then the hybrid prior bias satisfies

$$p_0^{(m)} - p_{\text{true}}^{(m)} = w_{\text{str}}^{(m)} b_{\text{str}}^{(m)} + w_{\text{ML}}^{(m)} b_{\text{ML}}^{(m)},$$

with

$$\limsup_{m \rightarrow \infty} |p_0^{(m)} - p_{\text{true}}^{(m)}| \leq \limsup_{m \rightarrow \infty} w_{\text{str}}^{(m)} |b_{\text{str}}^{(m)}|.$$

In particular, if either

- (a) *the structural bias is bounded and $w_{\text{str}}^{(m)} \rightarrow 0$, or*
- (b) *the structural bias itself satisfies $b_{\text{str}}^{(m)} \rightarrow 0$,*

then the hybrid prior bias $p_0^{(m)} - p_{\text{true}}^{(m)}$ converges to zero.

Proof sketch. Conditional independence of the two prior components implies that their joint predictive density is proportional to the product of Beta likelihood factors. Minimizing expected quadratic loss reduces to matching the first posterior moment, which under the additive fusion is

$$\mathbb{E}[p \mid \alpha_0, \beta_0] = \frac{\alpha_{\text{struct}} + \alpha_{\text{ML}}}{\alpha_{\text{struct}} + \alpha_{\text{ML}} + \beta_{\text{struct}} + \beta_{\text{ML}}}.$$

Convexity of quadratic risk then yields optimality; see [3]. \square

Remark 2 (Empirical Tuning of n_{ML}). *The conditions above suggest that robustness under structural misspecification depends on both (i) the asymptotic calibration of the ML prior and (ii) controls on the relative weight w_{ML} . In practice, one can choose n_{ML} via empirical Bayes or cross-validation, e.g. by selecting n_{ML} to minimize an out-of-sample scoring rule (Brier score, log score) for the hybrid prior on historical markets. Imposing an upper bound on n_{ML} relative to the crowd effective size n^* prevents the ML prior from dominating in regimes where it has not demonstrated sufficient predictive accuracy.*

2.10 Machine-Learning Uncertainty and Beta Hybridization

In Phase 2, the ML component enters through a point estimate $p_{\text{ML}} = g_{\text{NN}}(x)$. To incorporate uncertainty from ensembles or approximate Bayesian neural networks without breaking the Beta structure, we view the ML layer as generating a posterior distribution Π_{ML} on p , and then project Π_{ML} onto the Beta family.

Let $\mathcal{B} = \{\text{Beta}(a, b) : a, b > 0\}$ and let Π_{ML} denote any probability distribution on $(0, 1)$ that captures ML uncertainty (e.g. from an ensemble of approximate Bayesian networks).

Theorem 4 (Hybridization of ML Uncertainty via Beta Projection). *Let Π_{ML} be a distribution on $p \in (0, 1)$ with finite mean m_{ML} and variance $v_{\text{ML}} > 0$. Consider the KL divergence from Π_{ML} to a $\text{Beta}(a, b)$ distribution,*

$$D_{\text{KL}}(\Pi_{\text{ML}} \parallel \text{Beta}(a, b)) = \int_0^1 \log \frac{d\Pi_{\text{ML}}}{d\text{Beta}(a, b)}(p) d\Pi_{\text{ML}}(p),$$

whenever Π_{ML} is absolutely continuous with respect to $\text{Beta}(a, b)$. Then:

- (a) (**KL Projection of ML Posterior**) *There exists a unique pair (a^\dagger, b^\dagger) such that*

$$(a^\dagger, b^\dagger) \in \arg \min_{a, b > 0} D_{\text{KL}}(\Pi_{\text{ML}} \parallel \text{Beta}(a, b)),$$

and the minimizing Beta distribution matches the mean and variance of Π_{ML} ,

$$\frac{a^\dagger}{a^\dagger + b^\dagger} = m_{\text{ML}}, \quad \frac{a^\dagger b^\dagger}{(a^\dagger + b^\dagger)^2 (a^\dagger + b^\dagger + 1)} = v_{\text{ML}}.$$

- (b) (**Hierarchical Bayes Special Case**) *Suppose an approximate Bayesian neural network yields draws $p_{\text{ML}}^{(l)}$ from a posterior on p , and we place a $\text{Beta}(a_0, b_0)$ hyperprior on p . The Bayes estimator under a quadratic loss and Beta restriction coincides with the KL-projection in (a) when (a^\dagger, b^\dagger) are chosen to match the empirical mean and variance of the ensemble $\{p_{\text{ML}}^{(l)}\}$, up to hyperprior regularization.*

- (c) (**Ensemble-as-Sample Interpretation**) *If $\{p_{\text{ML}}^{(l)}\}_{l=1}^L$ are treated as noisy draws of Bernoulli success probabilities and L is large, then the empirical mean $\bar{p} = \frac{1}{L} \sum_l p_{\text{ML}}^{(l)}$ and empirical variance \hat{v} define a $\text{Beta}(a^\dagger, b^\dagger)$ distribution via the moment-matching equations in (a). This provides an empirical-Bayes construction of the ML prior*

$$p \sim \text{Beta}(a^\dagger, b^\dagger),$$

which can then be fused with the structural prior via the hybrid mechanism in Phase 2.

Remark 3. *This theorem clarifies that CAOPM does not require p_{ML} to be a single point estimate. Any ML method that yields a distribution over probabilities (e.g. approximate Bayesian neural networks, deep ensembles, bootstrap aggregations) can be incorporated by projecting its posterior onto the Beta family. The resulting (a^\dagger, b^\dagger) enter the hybrid prior exactly as in Phase 2, preserving the Beta form while acknowledging ML uncertainty.*

2.11 Structural–ML Prior Mismatch and Robustness

The structural prior $q_{\text{Shi}}(K, T; \Theta, \alpha)$ and the ML prior $p_{\text{ML}}(K, T)$ may be trained or calibrated under different regimes. For example, the structural model may use a rough-volatility parameter $\alpha < 1$ while the ML model has been fit on data that implicitly reflects a smoother regime. This can lead to systematic tension between the two priors.

To quantify this, consider the strike-dependent priors at a fixed maturity T :

$$q(K) = q_{\text{Shi}}(K, T; \Theta, \alpha), \quad p_{\text{ML}}(K) = p_{\text{ML}}(K, T),$$

and define the CAOPM hybrid prior mean

$$p_0(K) = w_{\text{str}}q(K) + w_{\text{ML}}p_{\text{ML}}(K),$$

with weights as in Phase 2.

Theorem 5 (Divergence Bounds and Robustness under Prior Mismatch). *For a finite grid of strikes $\{K_j\}_{j=1}^J$, define discrete distributions*

$$\Pi_{\text{str}}(j) \propto q(K_j), \quad \Pi_{\text{ML}}(j) \propto p_{\text{ML}}(K_j), \quad \Pi_{\text{hyb}}(j) \propto p_0(K_j),$$

normalized to sum to 1. Then:

- (a) **(Total Variation and Hellinger Bounds)** *For any $0 \leq w_{\text{str}}, w_{\text{ML}} \leq 1$ with $w_{\text{str}} + w_{\text{ML}} = 1$,*

$$\|\Pi_{\text{hyb}} - \Pi_{\text{str}}\|_{\text{TV}} \leq w_{\text{ML}}\|\Pi_{\text{ML}} - \Pi_{\text{str}}\|_{\text{TV}},$$

and similarly for Hellinger distance H ,

$$H(\Pi_{\text{hyb}}, \Pi_{\text{str}}) \leq w_{\text{ML}}H(\Pi_{\text{ML}}, \Pi_{\text{str}}).$$

- (b) **(KL Divergence Control)** *If Π_{ML} is absolutely continuous with respect to Π_{str} on the grid, then the discrete KL divergence satisfies*

$$D_{\text{KL}}(\Pi_{\text{hyb}} \| \Pi_{\text{str}}) \leq w_{\text{ML}}D_{\text{KL}}(\Pi_{\text{ML}} \| \Pi_{\text{str}}).$$

- (c) **(Posterior Stability)** *If the CAOPM likelihood (based on adjusted order counts) is Lipschitz with respect to perturbations in the prior on this grid, then the posterior distribution induced by Π_{hyb} differs from that induced by Π_{str} by at most a constant multiple of the distances above. In particular, as $w_{\text{ML}} \rightarrow 0$, the hybrid posterior converges to the purely structural posterior, and as $w_{\text{str}} \rightarrow 0$, it converges to the purely ML posterior.*

Remark 4. *This result shows that the hybrid prior acts as a convex bridge between the structural and ML priors: any mismatch in their implied strike-wise probabilities is damped by the weight w_{ML} in the distances and divergences. In practice, this suggests using w_{ML} (or equivalently n_{ML}) as a tuning parameter: when structural and ML priors strongly disagree, a smaller ML weight yields a hybrid prior and posterior that remain closer to the structural benchmark, while still incorporating ML information.*

2.12 Summary

Phase 2 constructs a rigorous, multi-tier hybrid prior using:

- tempered–fractional Heston structural probability,
- ANN/RNN predictive model probability (D’Uggento et al.[7]),
- reliability–weighted Bayesian fusion,
- hierarchical prior structure,
- optimal weighting under Bayes risk,
- cross–phase forward compatibility,
- robustness under misspecification.

The output (α_0, β_0) serves as the input to Phase 4.

3. Asymmetric Information Model (Trader Beliefs and Signals)

The purpose of this section is to establish a microeconomic foundation for why aggregate YES/NO bet volumes—denoted $(n_{\text{yes}}, n_{\text{no}})$ —contain information about the true event probability $p := Q(S_T > K)$. We specialize the asymmetric information framework of Koessler, Noussair, and Ziegelmeier [17] to the CAOPM setting and show that under mild conditions, the resulting order flow exhibits a monotone likelihood ratio property in the state of nature. This in turn justifies the Binomial likelihood used in the Beta–Binomial updating of Phase 4 and Phase 5.

Throughout, we take as given the hybrid prior from Phase 2,

$$p \sim \text{Beta}(\eta p_0, \eta(1 - p_0)), \quad p_0 \in (0, 1), \eta > 0, \quad (4)$$

with p_0 as in (??).

3.1 Primitives and Assumptions

We first formalize the economic environment.

State of nature and event. Let $\theta \in \{0, 1\}$ denote the state of nature:

$$\theta = 1 \iff S_T > K, \quad \theta = 0 \iff S_T \leq K,$$

with prior probability $p := Q(\theta = 1)$. The CAPOPM event of interest is $A := \{\theta = 1\} = \{S_T > K\}$.

Traders. There is a finite set of traders $i \in N := \{1, \dots, n\}$, each of whom is risk-neutral and participates in a single parimutuel market on event A . Trader i chooses an action

$$s_i \in \{\text{YES}, \text{NO}\},$$

where choosing YES corresponds to buying one YES ticket, and similarly for NO.

Parimutuel odds. For any action profile $s = (s_1, \dots, s_n)$, let

$$H(s) := \{i \in N : s_i = H\}, \quad h(s) := |H(s)|, \quad H \in \{\text{YES}, \text{NO}\}.$$

We denote the number of YES and NO tickets by $n_{\text{yes}} := h_{\text{YES}}(s)$ and $n_{\text{no}} := h_{\text{NO}}(s)$. Total participation is $n_{\text{tot}} := n_{\text{yes}} + n_{\text{no}} \leq n$.

Following Koessler et al. [17], the parimutuel odds against YES are

$$O_{\text{YES}}(s) := \frac{n_{\text{tot}} - n_{\text{yes}}}{n_{\text{yes}}} \quad \text{if } n_{\text{yes}} > 0, \quad (5)$$

and similarly for NO. For notational simplicity we assume interior participation, $n_{\text{yes}}, n_{\text{no}} > 0$; boundary cases can be handled by continuity.

Payoffs. A YES ticket pays $O_{\text{YES}}(s) + 1$ if $\theta = 1$ and 0 otherwise. Similarly, a NO ticket pays $O_{\text{NO}}(s) + 1$ if $\theta = 0$ and 0 otherwise. Trader i is risk-neutral, so her von Neumann–Morgenstern utility equals her monetary payoff.

We now state the key modeling assumptions.

Assumption A1 (Risk neutrality). Traders have linear utility in payoffs.

Assumption A2 (Price taking). Each trader treats the odds $O_{\text{YES}}(s)$ and $O_{\text{NO}}(s)$ as fixed with respect to her own action, i.e. as determined by the aggregate behavior of other traders. This is standard in parimutuel market models with a large number of participants [17].

Assumption A3 (Common prior). Traders share a common prior over θ with mean p_0 as in (4) before observing private signals.

3.2 Information Structure and Posterior Beliefs

Each trader receives a private signal about the state.

Assumption A4 (Private signals). Trader i observes $q_i \in \{q^1, q^0\}$ with

$$\Pr(q_i = q^1 \mid \theta = 1) = \pi, \quad \Pr(q_i = q^1 \mid \theta = 0) = 1 - \pi, \quad \pi > \frac{1}{2}. \quad (6)$$

Conditional on θ , signals $(q_i)_{i \in N}$ are i.i.d.

Assumption A5 (Common knowledge). The prior p_0 , the signal structure (6), and the parimutuel mechanism are common knowledge among traders.

Given prior mean p_0 and signal q_i , trader i forms a posterior belief

$$\mu(q_i) := \Pr(\theta = 1 \mid q_i).$$

Lemma 3.1 (Strict belief ordering). Under Assumptions A3–A5 with $\pi > \frac{1}{2}$, the posteriors satisfy

$$\mu_1 := \Pr(\theta = 1 \mid q_i = q^1) > \mu_0 := \Pr(\theta = 1 \mid q_i = q^0).$$

Proof. By Bayes' rule,

$$\begin{aligned} \mu_1 &= \frac{\Pr(q_i = q^1 \mid \theta = 1) \Pr(\theta = 1)}{\Pr(q_i = q^1 \mid \theta = 1) \Pr(\theta = 1) + \Pr(q_i = q^1 \mid \theta = 0) \Pr(\theta = 0)} = \frac{\pi p_0}{\pi p_0 + (1 - \pi)(1 - p_0)}, \\ \mu_0 &= \frac{\Pr(q_i = q^0 \mid \theta = 1) \Pr(\theta = 1)}{\Pr(q_i = q^0 \mid \theta = 1) \Pr(\theta = 1) + \Pr(q_i = q^0 \mid \theta = 0) \Pr(\theta = 0)} = \frac{(1 - \pi)p_0}{(1 - \pi)p_0 + \pi(1 - p_0)}. \end{aligned}$$

Since $\pi > \frac{1}{2}$ and $p_0 \in (0, 1)$, straightforward algebra shows that $\mu_1 - \mu_0 > 0$. \square

Lemma 3.1 ensures that signals are *informative*: observing q^1 leads to a strictly higher posterior belief in $\theta = 1$ than observing q^0 .

3.3 Strategies and Bayesian Nash Equilibrium

We now formalize strategies, best responses, and equilibrium.

Definition 3.1 (Strategies). A (pure) strategy for trader i is a mapping

$$\sigma_i : \{q^0, q^1\} \rightarrow \{\text{YES}, \text{NO}\}.$$

A strategy profile is $\sigma = (\sigma_i)_{i \in N}$. A symmetric strategy profile has $\sigma_i = \sigma$ for all i .

Definition 3.2 (Bayesian Nash equilibrium). A symmetric Bayesian Nash equilibrium (BNE) is a measurable function $\sigma^* : \{q^0, q^1\} \rightarrow \{\text{YES}, \text{NO}\}$ such that, for all i and for each signal $q \in \{q^0, q^1\}$,

$$\sigma^*(q) \in \arg \max_{a \in \{\text{YES}, \text{NO}\}} E[u_i(a, s_{-i}, \theta) \mid q_i = q, \sigma_{-i}^*],$$

where expectations are taken with respect to the induced distribution over (θ, q_{-i}, s_{-i}) given the strategy profile.

Expected payoffs. Under Assumption A2 (price-taking), trader i takes the odds $(O_{\text{YES}}, O_{\text{NO}})$ as given. For a trader with belief $\mu \in [0, 1]$ about $\theta = 1$, the expected payoff from choosing YES is

$$U_i(\text{YES} \mid \mu) = \mu(O_{\text{YES}} + 1). \quad (7)$$

Similarly, the expected payoff from choosing NO is

$$U_i(\text{NO} \mid \mu) = (1 - \mu)(O_{\text{NO}} + 1). \quad (8)$$

Lemma 3.2 (Threshold best response). Fix odds $(O_{\text{YES}}, O_{\text{NO}})$. Under Assumptions A1–A2, there exists a threshold $\mu^* \in (0, 1)$ such that:

$$\text{YES is optimal} \iff \mu \geq \mu^*, \quad \text{NO is optimal} \iff \mu \leq \mu^*.$$

Proof. Consider the difference:

$$\Delta U(\mu) := U_i(\text{YES} \mid \mu) - U_i(\text{NO} \mid \mu) = \mu(O_{\text{YES}} + 1) - (1 - \mu)(O_{\text{NO}} + 1).$$

This is affine in μ :

$$\Delta U(\mu) = \mu[(O_{\text{YES}} + 1) + (O_{\text{NO}} + 1)] - (O_{\text{NO}} + 1).$$

Solve $\Delta U(\mu^*) = 0$ for μ^* :

$$\mu^* = \frac{O_{\text{NO}} + 1}{(O_{\text{YES}} + 1) + (O_{\text{NO}} + 1)} \in (0, 1).$$

Then $\Delta U(\mu) \geq 0$ iff $\mu \geq \mu^*$, establishing the threshold property. \square

Combined with Lemma 3.1, Lemma 3.2 implies that traders with signal q^1 are strictly more likely to choose YES than those with q^0 .

Assumption A6 (Separating equilibrium as in Koessler et al.). Following Koessler et al. [17], we focus on a symmetric BNE in which informed traders adopt a *separating threshold strategy*:

$$\sigma^*(q^1) = \text{YES}, \quad \sigma^*(q^0) = \text{NO}. \quad (9)$$

Uninformed or purely noisy traders, if present, randomize independently of the state in a way that does not overturn the strict monotonicity implied by Lemma 3.1 and Lemma 3.2.¹

Under Assumptions A1–A6, informed traders with q^1 bet YES, and those with q^0 bet NO in equilibrium. This yields a monotone mapping from signals to actions.

¹Koessler et al. [17] provide existence and characterization results for such equilibria in parimutuel information aggregation mechanisms. We adopt their equilibrium selection as a modeling assumption and specialize it to the CAPOPM event.

3.4 Distribution of Order Flow Conditional on the State

We next derive the distribution of YES/NO orders conditional on the state θ under the separating equilibrium (9).

Lemma 3.3 (Signal distribution). Under Assumption A4, conditional on θ , the number of traders receiving signal q^1 satisfies

$$N_1 \mid \theta = 1 \sim \text{Binomial}(n, \pi), \quad N_1 \mid \theta = 0 \sim \text{Binomial}(n, 1 - \pi),$$

and $N_0 := n - N_1$.

Proof. Immediate from the i.i.d. Bernoulli structure in Assumption A4. \square

Under the separating equilibrium (9), informed traders with q^1 bet YES and those with q^0 bet NO. If a fraction of traders are purely noisy or always abstain, then n can be reinterpreted as the number of informed traders; noisy traders add independent Bernoulli noise that does not destroy the strict ordering established below, as long as the informed fraction is positive.

Lemma 3.4 (Order flow distribution). Under Assumptions A1–A6 and Lemma 3.3, there exist $\phi_1, \phi_0 \in (0, 1)$ with $\phi_1 > \phi_0$ such that

$$n_{\text{yes}} \mid \theta = 1 \sim \text{Binomial}(n, \phi_1), \quad n_{\text{yes}} \mid \theta = 0 \sim \text{Binomial}(n, \phi_0).$$

Proof. In the benchmark case with only informed traders and separating strategies, we have

$$\Pr(s_i = \text{YES} \mid \theta = 1) = \Pr(q_i = q^1 \mid \theta = 1) = \pi,$$

$$\Pr(s_i = \text{YES} \mid \theta = 0) = \Pr(q_i = q^1 \mid \theta = 0) = 1 - \pi.$$

Thus $\phi_1 = \pi$ and $\phi_0 = 1 - \pi$, with $\phi_1 > \phi_0$ because $\pi > \frac{1}{2}$.

If an ε -fraction of traders are noisy and bet YES independently with probability δ , then

$$\phi_1 = (1 - \varepsilon)\pi + \varepsilon\delta, \quad \phi_0 = (1 - \varepsilon)(1 - \pi) + \varepsilon\delta.$$

The strict inequality $\phi_1 > \phi_0$ continues to hold as long as $\varepsilon < 1$ and $\pi > \frac{1}{2}$. Since trades are independent conditional on θ , the number of YES orders n_{yes} is Binomial with parameters (n, ϕ_θ) , where ϕ_1 and ϕ_0 denote the respective conditional success probabilities. \square

Lemma 3.4 provides a *micro-founded* Binomial distribution of YES orders conditional on the state of nature.

3.5 Monotone Likelihood Ratio and Informativeness

We now show that the order flow is informative about the state θ , and hence about p .

Proposition 3.1 (Monotone likelihood ratio of order flow). Under Lemma 3.4 with $\phi_1 > \phi_0$, the likelihood ratio

$$\Lambda(k) := \frac{\Pr(n_{\text{yes}} = k \mid \theta = 1)}{\Pr(n_{\text{yes}} = k \mid \theta = 0)}$$

is strictly increasing in $k = 0, 1, \dots, n$.

Proof. For $k \in \{0, \dots, n\}$,

$$\Pr(n_{\text{yes}} = k \mid \theta = 1) = \binom{n}{k} \phi_1^k (1 - \phi_1)^{n-k},$$

$$\Pr(n_{\text{yes}} = k \mid \theta = 0) = \binom{n}{k} \phi_0^k (1 - \phi_0)^{n-k}.$$

Hence,

$$\Lambda(k) = \frac{\phi_1^k (1 - \phi_1)^{n-k}}{\phi_0^k (1 - \phi_0)^{n-k}} = \left(\frac{\phi_1}{\phi_0}\right)^k \left(\frac{1 - \phi_1}{1 - \phi_0}\right)^{n-k}.$$

Then

$$\frac{\Lambda(k+1)}{\Lambda(k)} = \frac{\phi_1}{\phi_0} \cdot \frac{1 - \phi_0}{1 - \phi_1}.$$

Because $\phi_1 > \phi_0$, we have $\phi_1/\phi_0 > 1$ and $(1 - \phi_0)/(1 - \phi_1) > 1$, so the product exceeds unity:

$$\frac{\Lambda(k+1)}{\Lambda(k)} > 1.$$

Therefore $\Lambda(k)$ is strictly increasing in k . \square

Proposition 3.1 establishes a monotone likelihood ratio (MLR) property for the aggregate YES order flow.

Theorem 3.1 (Informative parimutuel order flow). Under Assumptions A1–A6 and Lemmas 3.1–3.4, the aggregate YES count n_{yes} is informative about the state θ , and thus about the event probability p :

1. The posterior probability $\Pr(\theta = 1 \mid n_{\text{yes}} = k)$ is strictly increasing in k .
2. Observing $(n_{\text{yes}}, n_{\text{no}})$ yields a non-degenerate likelihood for p .

Proof. (1) The MLR property in Proposition 3.1 implies that the family $\{\Pr(\cdot \mid \theta)\}$ is ordered in the monotone likelihood ratio sense. By standard Bayesian decision theory (e.g. Karlin and Rubin's theorem), this implies that the posterior $\Pr(\theta = 1 \mid n_{\text{yes}} = k)$ is strictly increasing in k .

(2) Since $\phi_1 > \phi_0$, the two conditional Binomial distributions are distinct, implying that the mapping from p to the mixture distribution of n_{yes} is non-degenerate. Hence the induced likelihood over p is non-constant and thus informative. \square

Theorem 3.1 provides the desired microeconomic foundation: equilibrium YES/NO bet volumes in the CAPOPM parimutuel mechanism contain information about the true event probability p .

3.6 Connection to the CAPOPM Beta–Binomial Likelihood

We now connect the micro-founded order-flow distribution to the CAPOPM likelihood used in Phases 4 and 5.

Under Theorem 3.1, we may interpret the observed pair $(n_{\text{yes}}, n_{\text{no}})$ as a Binomial sample from the unknown event probability p :

$$\Pr(n_{\text{yes}} = k \mid p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

up to reparameterization of n and the effective success probabilities ϕ_θ when conditioning on θ . Therefore, suppressing constants in p , the likelihood of p given the data is

$$L(p \mid n_{\text{yes}}, n_{\text{no}}) \propto p^{n_{\text{yes}}} (1-p)^{n_{\text{no}}}. \quad (10)$$

Corollary 3.1 (CAPOPM posterior under the hybrid prior). Combining the hybrid prior (4) with the Binomial likelihood (10), the posterior distribution of p is

$$p \mid n_{\text{yes}}, n_{\text{no}} \sim \text{Beta}(\eta p_0 + n_{\text{yes}}, \eta(1-p_0) + n_{\text{no}}).$$

Proof. Immediate from Beta–Binomial conjugacy: multiply the prior density $p^{\eta p_0 - 1} (1-p)^{\eta(1-p_0) - 1}$ by the likelihood $p^{n_{\text{yes}}} (1-p)^{n_{\text{no}}}$ and recognize the kernel of a Beta distribution with parameters $\eta p_0 + n_{\text{yes}}$ and $\eta(1-p_0) + n_{\text{no}}$. \square

Corollary 3.1 closes the loop between the microeconomic model of asymmetric information in a parimutuel environment and the Bayesian updating machinery of CAPOPM. The Beta posterior derived here is the starting point for the posterior predictive pricing of digital and vanilla options in Phase 4 and Phase 5.

3.7 Strategic Behavior, Reflexivity, and Bayesian–Nash Equilibrium

Assumption A6 treats trader actions as if they were conditionally independent signals that do not strategically anticipate the CAPOPM correction mechanism. While this assumption is suitable for ex-post belief extraction or markets with numerous small traders, it does not fully address reflexive behavior in which traders may try to influence the posterior or exploit the weighting scheme.

To formalize this tension, we introduce a stylized strategic model in which each trader is small but rational, observes the current parimutuel odds π_t , and decides whether to submit a YES or NO order. The key distinction from Assumption A6 is that traders now optimize expected utility conditional on (i) the current odds and (ii) their private signal. Importantly, traders do *not* observe the detailed CAPOPM correction parameters $(w_i^{\text{beh}}, \delta_\pm)$, which prevents manipulation directed at the correction mechanism itself and keeps the model tractable.

Let $\theta \in \{0, 1\}$ be the true event state and let $s_i \in \{+1, -1\}$ denote the private signal observed by trader i , where $+1$ corresponds to a signal favoring YES and -1 to a signal favoring NO. Conditional on θ , the signals satisfy

$$\mathbb{P}(s_i = \theta) = \sigma > 1/2, \quad \mathbb{P}(s_i = -\theta) = 1 - \sigma,$$

and traders know both σ and the odds π_t but do not know the realized θ . Upon submitting a YES order, trader i receives payoff $1/\pi_t$ if $\theta = 1$ and zero otherwise, consistent with parimutuel mechanics for unit bets; conversely, a NO order yields $1/(1 - \pi_t)$ if $\theta = 0$.

Each trader chooses YES or NO to maximize expected payoff given (s_i, π_t) . Let $u_i(\text{YES} \mid s_i, \pi_t)$ and $u_i(\text{NO} \mid s_i, \pi_t)$ denote these expected payoffs. A Bayesian–Nash equilibrium (BNE) consists of a strategy profile in which each trader best responds to the odds, and the odds reflect the empirical frequencies of submitted orders.

Theorem 6 (Bayesian–Nash Equilibrium with Signal-Driven Strategies). *Assume all traders are small, observe the same odds π_t , have identical signal accuracy $\sigma > 1/2$, and submit a single indivisible order. Suppose traders do not observe the CAPOPM correction parameters $(w_i^{\text{beh}}, \delta_{\pm})$ and thus cannot condition strategies on them. Then there exists a symmetric Bayesian–Nash equilibrium in threshold form. Specifically, there is a cutoff $\pi^*(\sigma) \in (0, 1)$ such that:*

- If $\pi_t < \pi^*(\sigma)$ (YES is “cheap”), all traders with $s_i = +1$ submit YES and all traders with $s_i = -1$ submit NO.
- If $\pi_t > \pi^*(\sigma)$ (YES is “expensive”), all traders with $s_i = -1$ submit NO and traders with $s_i = +1$ may randomize to keep the odds at equilibrium.
- At $\pi_t = \pi^*(\sigma)$, traders are indifferent and a symmetric mixed-strategy equilibrium exists.

Moreover, the equilibrium cutoff solves

$$\frac{\mathbb{P}(\theta = 1 \mid s_i = +1)}{\pi^*(\sigma)} = \frac{\mathbb{P}(\theta = 0 \mid s_i = +1)}{1 - \pi^*(\sigma)},$$

and similarly for $s_i = -1$. The equilibrium is unique under these conditions.

Proof (Sketch). Given parimutuel payoffs, a YES order yields expected payoff

$$u_i(\text{YES} \mid s_i, \pi_t) = \frac{\mathbb{P}(\theta = 1 \mid s_i)}{\pi_t},$$

and a NO order yields

$$u_i(\text{NO} \mid s_i, \pi_t) = \frac{\mathbb{P}(\theta = 0 \mid s_i)}{1 - \pi_t}.$$

A trader chooses YES iff

$$\frac{\mathbb{P}(\theta = 1 \mid s_i)}{\pi_t} > \frac{\mathbb{P}(\theta = 0 \mid s_i)}{1 - \pi_t}.$$

The posterior probabilities on the right-hand side can be computed via Bayes' rule. The above inequality defines a threshold in π_t separating the regions where YES or NO is optimal. Because all traders are small, they take π_t as given; the equilibrium odds must be consistent with aggregate order flow induced by these best responses. Standard fixed-point arguments for parimutuel odds imply the existence of a unique solution to the resulting consistency condition, yielding the stated threshold equilibrium. \square

Remark 5 (Interpretation and Relation to Assumption A6). *The equilibrium above separates (i) the informational content of private signals and (ii) the strategic incentive created by parimutuel odds. Because traders do not observe the CAOPM correction parameters, they cannot directly manipulate the behavioral biases or structural offsets, and the BNE captures only the odds-reactive part of strategic behavior.*

Assumption A6 corresponds to the “implicit-signal” limit in which $\pi_t \approx p_{\text{true}}$, traders are small, and the odds reflect aggregated signals rather than strategic anticipation. The model developed here formalizes the gap between A6 and a fully reflexive environment, while remaining compatible with CAOPM as an ex-post belief extraction mechanism.

3.8 Strategic Feedback, Pooling Equilibria, and Asymptotic Unraveling

Assumption A2 treats traders as price takers, even though parimutuel odds depend on aggregate order flow. Here we provide a complementary game-theoretic view that (i) constructs pooling equilibria under heterogeneous signal precision, (ii) gives conditions under which pooling cannot persist, and (iii) shows how informational content can unravel as the number of traders grows.

Theorem 7 (Pooling, Non-Pooling, and Asymptotic Unraveling). *Consider a parimutuel market with N traders. Trader i receives a binary signal $s_i \in \{0, 1\}$ about $\theta \in \{0, 1\}$ with precision $\sigma_i = \mathbb{P}(s_i = \theta) > 1/2$. Let odds π_N be a continuous function of the empirical YES fraction, and suppose payoffs are standard parimutuel (unit stake, pool-splitting). Then:*

- (a) **(Existence of Pooling Equilibria)** *If signal precisions $\{\sigma_i\}$ are heterogeneous and sufficiently dispersed, there exists a Bayesian–Nash equilibrium in which some subset of traders ignores their private signal and mimics the behavior of a reference group, leading to pooling of actions across different signal realizations.*
- (b) **(Impossibility Under Monotonicity and Full Support)** *If, in addition, odds are strictly monotone in the empirical YES fraction and each*

trader's expected payoff from following their signal dominates any constant strategy (given others follow their signals), then no fully pooling equilibrium exists: in any BNE, at least a positive-measure subset of traders uses their signal in a non-trivial way.

- (c) (**Asymptotic Unraveling**) Under the conditions of (b), if the number of traders $N \rightarrow \infty$ while the distribution of precisions $\{\sigma_i\}$ remains bounded away from $1/2$, then any sequence of equilibria must be asymptotically non-pooling in the sense that the empirical distribution of actions reveals a non-degenerate amount of information about θ ; the odds move toward the true probability in the limit.

Remark 6. Part (a) shows that pooling equilibria are possible in finite markets, so A2 does not literally hold as a behavioral assumption. Parts (b) and (c) clarify that under monotone odds and sufficiently informative signals, large markets tend to unravel pooling, making the empirical order flow informative. CAOPM is designed to operate in this asymptotic regime, treating observed order flow as a noisy but informative aggregate signal about θ .

3.9 Interpretation of Assumption A2 in a Parimutuel Setting

Assumption A2 posits that, conditional on the latent event probability p , trader orders can be modeled as conditionally independent Bernoulli signals, with no explicit feedback from individual actions onto the odds. In a parimutuel market, however, odds are determined by the aggregate order flow: each trader's action affects the pool, and hence the implied payoffs. This creates an apparent conflict between A2 and the mechanics of parimutuel pricing.

In this subsection, we show that A2 can be understood as a *large-market approximation* in which each trader is small and the dependence induced by the parimutuel odds becomes negligible for finite sets of traders. The result does not remove all strategic considerations (these are treated separately in the strategic extension) but clarifies how an approximately independent signal model emerges in the limit of many small traders.

Consider a sequence of parimutuel markets indexed by N , each with N traders. In market N , trader $i \in \{1, \dots, N\}$ chooses an action $Y_i^{(N)} \in \{0, 1\}$, with 1 indicating a YES order and 0 a NO order. Let

$$\Pi_N = \frac{1}{N} \sum_{i=1}^N Y_i^{(N)}$$

denote the empirical YES fraction, and suppose parimutuel odds are determined by a deterministic, continuous function

$$\pi_N = \Phi(\Pi_N),$$

where $\Phi : [0, 1] \rightarrow (0, 1)$ is Lipschitz and strictly increasing. Thus each trader faces odds π_N , which depend on aggregate behavior.

We assume traders are small and symmetric: conditional on the true event probability $p_{\text{true}} \in (0, 1)$ and on the limiting odds π^* , each trader adopts a mixed strategy with YES probability $\beta(p_{\text{true}}, \pi^*) \in (0, 1)$. The equilibrium odds π^* solve the fixed-point condition

$$\pi^* = \Phi(\beta(p_{\text{true}}, \pi^*)).$$

For each finite N , the actual odds π_N fluctuate around π^* as Π_N fluctuates around $\beta(p_{\text{true}}, \pi^*)$.

Theorem 8 (Approximate Conditional Independence in Large Parimutuel Markets). *Suppose:*

- (i) *The mapping Φ is Lipschitz continuous and strictly increasing on $[0, 1]$.*
- (ii) *The mixed strategy $\beta(p_{\text{true}}, \pi)$ is continuous in π and strictly between 0 and 1 for the relevant range of odds.*
- (iii) *The fixed-point equation $\pi^* = \Phi(\beta(p_{\text{true}}, \pi^*))$ has a unique solution $\pi^* \in (0, 1)$.*

Then the following hold as $N \rightarrow \infty$:

(a) The empirical YES fraction converges in probability,

$$\Pi_N \xrightarrow{P} \beta(p_{\text{true}}, \pi^*),$$

and therefore the realized odds converge in probability,

$$\pi_N = \Phi(\Pi_N) \xrightarrow{P} \pi^*.$$

(b) For any fixed $k \geq 1$, the joint law of any k distinct trader actions $(Y_{i_1}^{(N)}, \dots, Y_{i_k}^{(N)})$ converges in total variation to a product of independent Bernoulli variables with parameter $\beta(p_{\text{true}}, \pi^*)$, i.e.

$$\lim_{N \rightarrow \infty} \left\| \mathcal{L}((Y_{i_1}^{(N)}, \dots, Y_{i_k}^{(N)})) - \text{Bernoulli}(\beta(p_{\text{true}}, \pi^*))^{\otimes k} \right\|_{\text{TV}} = 0.$$

In particular, for large N , any fixed finite set of trader orders is approximately conditionally independent given p_{true} and the limiting odds π^* , so that Assumption A2 can be interpreted as a large-market approximation valid for ex-post belief extraction.

Proof (Sketch). Under (i)–(iii), the law of large numbers applied to the symmetric mixed strategies implies that the empirical fraction Π_N converges in probability to the unique fixed point of the mapping $\Pi \mapsto \beta(p_{\text{true}}, \Phi(\Pi))$. By continuity and uniqueness, this fixed point corresponds to $\Pi^* = \beta(p_{\text{true}}, \pi^*)$, and thus $\Pi_N \rightarrow \Pi^*$ in probability. Lipschitz continuity of Φ then yields $\pi_N \rightarrow \pi^*$ in probability, establishing (a).

For (b), conditional on π_N , the trader actions are exchangeable and independent given the common mixed strategy parameter $\beta(p_{\text{true}}, \pi_N)$. As $\pi_N \rightarrow \pi^*$, continuity of β gives

$$\beta(p_{\text{true}}, \pi_N) \rightarrow \beta(p_{\text{true}}, \pi^*)$$

in probability. Standard arguments for triangular arrays of conditionally independent Bernoulli variables imply that the finite-dimensional distributions of $(Y_{i_1}^{(N)}, \dots, Y_{i_k}^{(N)})$ converge in total variation to those of i.i.d. Bernoulli variables with the limiting parameter. This yields the stated approximate independence for any fixed k . \square

Remark 7 (Scope and Limitations). *The theorem above shows that, in a large parimutuel market with symmetric small traders and a unique fixed point for the odds, the dependence introduced by the parimutuel mechanism becomes negligible for any fixed finite set of traders. In this sense, Assumption A2 is consistent with parimutuel pricing as a large-market approximation.*

However, several limitations remain. First, the result does not address pooling equilibria or multiple fixed points, which can arise when traders are heterogeneous or when the odds function Φ has non-monotone features. Second, strategic considerations beyond mixed strategies that depend only on π_N (e.g. anticipatory behavior targeted at CAOPM corrections) are not modeled here and can reintroduce nontrivial dependence. These issues are examined separately in the strategic extension and are left as directions for further work.

Phase 4. Parimutuel Likelihood and Bayesian Posterior Update

In this phase, we derive the likelihood generated by parimutuel YES/NO trader actions and combine it with the hybrid prior from Phase 2 to construct the posterior distribution of the event probability

$$p := Q(A) = Q(S_T > K).$$

All notation, including the hybrid prior parameters (α_0, β_0) , follows Phase 2.

4.1 Trader Actions and the Likelihood Model

Let n denote the number of traders participating in the parimutuel book, and let y denote the number of YES positions. A NO position is treated as a vote for the complement A^c .

Traders may act strategically: some may exaggerate their signals, herd behind early order flow, or attempt to manipulate the book. We acknowledge the possibility of such distortions but defer correction to Phase 6. For the purposes of likelihood construction, we treat the realized counts $(y, n - y)$ as the observable actions that the mechanism must interpret.

Although traders' private signals may be correlated, and their actions may be strategically dependent, we assume conditional independence given the latent probability p for the sole purpose of deriving the Beta–Binomial conjugacy. This follows standard practice in Bayesian market microfoundations.

4.2 Binomial Likelihood of the Parimutuel Order Flow

Conditionally on the latent event probability p , we model the YES count y as

$$y \mid p \sim \text{Binomial}(n, p),$$

with likelihood

$$L(y \mid p) = \binom{n}{y} p^y (1 - p)^{n-y}. \quad (11)$$

Although later phases introduce liquidity-adjusted counts y^* and n^* , the present phase uses the raw counts $(y, n - y)$ for the purpose of deriving the conjugate update.

4.3 Conjugate Updating with the Hybrid Prior

The hybrid prior from Phase 2 is

$$p \sim \text{Beta}(\alpha_0, \beta_0), \quad \alpha_0, \beta_0 > 0.$$

Lemma 6 (Posterior Form). *Combining the Beta prior with the Binomial likelihood (11) yields the posterior*

$$p \mid y \sim \text{Beta}(\alpha_{\text{post}}, \beta_{\text{post}}),$$

where

$$\alpha_{\text{post}} = \alpha_0 + y, \quad \beta_{\text{post}} = \beta_0 + (n - y).$$

Proof. The Beta density is proportional to $p^{\alpha_0-1}(1-p)^{\beta_0-1}$. Multiplying by (11) gives a kernel proportional to

$$p^{\alpha_0+y-1}(1-p)^{\beta_0+(n-y)-1},$$

the kernel of a $\text{Beta}(\alpha_{\text{post}}, \beta_{\text{post}})$. \square

4.4 Posterior Mean and Variance

Proposition 5 (Posterior Moments). *For the posterior Beta distribution above,*

$$\mathbb{E}[p \mid y] = \frac{\alpha_{\text{post}}}{\alpha_{\text{post}} + \beta_{\text{post}}},$$

and

$$\text{Var}(p \mid y) = \frac{\alpha_{\text{post}}\beta_{\text{post}}}{(\alpha_{\text{post}} + \beta_{\text{post}})^2(\alpha_{\text{post}} + \beta_{\text{post}} + 1)}.$$

Proof. These are standard properties of the Beta distribution. \square

4.5 Posterior Predictive Distribution (Beta–Binomial Form)

The posterior predictive distribution of observing y YES votes under the prior $\text{Beta}(\alpha_0, \beta_0)$ is

$$P(y \mid \alpha_0, \beta_0) = \binom{n}{y} \frac{B(\alpha_0 + y, \beta_0 + n - y)}{B(\alpha_0, \beta_0)},$$

where $B(\cdot, \cdot)$ is the Beta function.

Theorem 9 (Posterior Predictive Distribution). *Let $p \sim \text{Beta}(\alpha_0, \beta_0)$ and $y \mid p \sim \text{Binomial}(n, p)$. Then the marginal distribution of y is Beta–Binomial with pmf above.*

Proof. Integrate the joint distribution $P(y \mid p)f(p)$ over $p \in [0, 1]$, and use the identity

$$\int_0^1 p^{a-1}(1-p)^{b-1} dp = B(a, b).$$

\square

4.6 Conditional Independence as a Modeling Approximation

The Beta–Binomial updating step presented above relies on the assumption that, conditional on the latent event probability p , individual trader actions $s_i \in \{\text{YES}, \text{NO}\}$ are independent Bernoulli draws. Such conditional independence is standard in Bayesian aggregation models, but it is not expected to hold exactly in parimutuel markets where traders may observe and react to earlier order flow. In particular, herding behavior generates temporal dependence among trades: late traders may overweight recent order patterns even when those patterns do not reflect new private information.

In this framework, conditional independence is therefore best interpreted as a *modeling approximation* rather than a literal behavioral assumption. Empirically observed violations of independence are handled in two ways:

1. **Behavioral Adjustment (Phase 6).** The Stage 1 correction introduces weights w_i^{beh} applied to individual orders. When herding creates clusters of correlated trades, these weights reduce the effective contribution of late correlated orders, mitigating departures from independence.
2. **Sensitivity Analysis (Phase 7).** Simulation regimes in Phase 7 introduce explicit dependence structures among trades, including herding and correlated decision rules. These regimes allow us to evaluate how violations of independence affect the CAPOPM posterior and how effectively the two-stage bias-correction layer controls such deviations.

This modeling approximation preserves conjugacy and analytic tractability while acknowledging that the empirical behavior of order flow contains richer dependence patterns. Phases 6 and 7 are specifically designed to examine, interpret, and correct these dependencies.

4.7 Incorporation of Strategic Distortion

Because traders may submit exaggerated or strategically distorted orders, the raw counts $(y, n - y)$ encode:

1. private signals,
2. beliefs about other traders' signals,
3. strategic considerations,
4. liquidity constraints.

Phase 6 will introduce formal bias adjustments using *effective* counts $(y^*, n^* - y^*)$ and distortion offsets (δ_+, δ_-) . For now, the posterior above represents the *unadjusted* Bayesian update based on the observable order flow.

4.8 Conditional i.i.d., Dependence, and the Role of Weights

Assumption A4 models the adjusted order contributions as conditionally i.i.d. Bernoulli (or bounded) signals given the latent event probability p . This is a deliberate simplification. In realistic markets, herding, order-splitting, and informational cascades induce dependence across orders.

There are two conceptually distinct modeling choices:

- *Fully dependent likelihood.* One could specify an explicit joint law for the order sequence (Z_1, \dots, Z_n) , for example via an Ising model or a Markov random field. This yields a non-factorizing likelihood and a non-conjugate posterior that typically requires MCMC or variational methods.
- *Weighted pseudo-likelihood.* CAPOPM instead uses a Beta–Binomial update based on adjusted counts (y_n^*, n_n^*) , together with mixing-based asymptotics (Phase 8). This implicitly replaces the true dependent likelihood with an exponential family surrogate whose sufficient statistics are the weighted sums. The resulting Beta posterior is then interpreted as the KL-projection of the intractable posterior onto the Beta family (Phase 8ZZ).

In this sense, the behavioral weighting layer is not merely a “patch” on an i.i.d. model, but a way to summarize dependence and heterogeneity into effective counts that remain compatible with a tractable exponential family update. The trade-off is explicit: CAPOPM sacrifices an exact likelihood for closed-form inference plus an information-theoretic guarantee that, within the Beta family, the posterior is as close as possible (in Kullback–Leibler sense) to the ideal but intractable posterior.

4.9 Why CAPOPM Uses Beta Conjugacy

There are many ways to model belief aggregation in markets. CAPOPM uses a Beta–Binomial structure for a simple reason: it gives closed-form updates and keeps the link between data and parameters transparent.

- Each component of the prior can be read as a pseudo-count: structural information contributes η_{str} virtual observations, the ML model contributes n_{ML} , and the crowd contributes adjusted effective counts n_n^* .
- Updating is a matter of adding these counts, with no numerical integration or sampling required.
- The resulting posterior has a clear interpretation: it is the Beta distribution that best matches, in KL sense, the information contained in the adjusted counts and the hybrid prior.

More flexible approaches, such as MCMC over a fully dependent likelihood, can capture richer structures but at the cost of interpretability and computation. CAPOPM is intentionally positioned as a tractable, interpretable baseline: it

prioritizes closed-form inference and moment-based robustness over fully non-parametric modeling. This makes it easier to diagnose, explain, and test in simulation before considering heavier alternatives.

4.10 Output of Phase 4

The output of this phase is the posterior hyperparameter pair

$$(\alpha_{\text{post}}, \beta_{\text{post}}),$$

which becomes the foundation for Phase 5 (posterior predictive pricing) and is later refined in Phase 6 (bias-corrected posterior).

Phase 5. Posterior Predictive Derivative Pricing

Given the posterior distribution of the event probability

$$p := Q(A) = Q(S_T > K)$$

from Phase 4, we now derive the posterior-predictive prices of YES/NO parimutuel contracts and digital derivatives. We additionally establish no-arbitrage properties, uncertainty bounds, and risk-adjusted pricing rules.

Throughout, the posterior from Phase 4 is

$$p \mid y \sim \text{Beta}(\alpha_{\text{post}}, \beta_{\text{post}}),$$

with

$$\alpha_{\text{post}} = \alpha_0 + y, \quad \beta_{\text{post}} = \beta_0 + (n - y).$$

5.1 From CAOPM Posteriors to Pricing Kernels and Option Prices

In previous phases, CAOPM produces, for a fixed maturity T , a posterior distribution for the event probability

$$p(K, T) = Q(S_T > K \mid \mathcal{I}),$$

where \mathcal{I} denotes the combined information from the structural model, the machine learning prior, and the adjusted parimutuel order flow. Let $\hat{p}(K, T)$ denote the posterior mean, so that the CAOPM-implied digital price at strike K is

$$D_{\text{CAP}}(K, T) = e^{-rT} \hat{p}(K, T),$$

with r the risk-free rate.

We now show how the entire risk-neutral distribution and pricing kernel can be recovered (at least formally) from the strike-dependent posterior, and how vanilla option prices follow via standard integral transforms.

Theorem 10 (CAOPM-Implied Risk-Neutral CDF, Density, and Pricing Kernel). *Fix a maturity $T > 0$ and suppose that, for each strike $K \geq 0$, CAOPM produces a posterior distribution for $p(K, T)$ with mean $\hat{p}(K, T)$. Assume:*

- (i) (**Smoothness in Strike**) *The map $K \mapsto \hat{p}(K, T)$ is differentiable almost everywhere, with $\hat{p}(K, T)$ non-increasing in K and*

$$\lim_{K \rightarrow 0} \hat{p}(K, T) = 1, \quad \lim_{K \rightarrow \infty} \hat{p}(K, T) = 0.$$

- (ii) (**No-Arbitrage Regularity**) *The function $\hat{p}(K, T)$ is right-continuous with left limits and defines a valid tail function for a probability distribution on $[0, \infty)$.*

Then:

- (a) (**Risk-Neutral CDF and Density**) *The CAOPM-implied risk-neutral CDF and density at maturity T are given by*

$$F_{\text{CAP}}^Q(K, T) = Q(S_T \leq K \mid \mathcal{I}) = 1 - \hat{p}(K, T),$$

and, wherever differentiable,

$$f_{\text{CAP}}^Q(K, T) = \frac{\partial}{\partial K} F_{\text{CAP}}^Q(K, T) = -\frac{\partial}{\partial K} \hat{p}(K, T).$$

- (b) (**Call Prices and Breeden–Litzenberger**) *The CAOPM-implied call price at strike K and maturity T is*

$$C_{\text{CAP}}(K, T) = e^{-rT} \int_K^\infty (s - K) f_{\text{CAP}}^Q(s, T) ds.$$

Equivalently, if we define

$$C_{\text{CAP}}(K, T) = e^{-rT} \mathbb{E}^Q[(S_T - K)^+ \mid \mathcal{I}],$$

then the Breeden–Litzenberger relation holds:

$$\frac{\partial^2}{\partial K^2} C_{\text{CAP}}(K, T) = e^{-rT} f_{\text{CAP}}^Q(K, T),$$

whenever the derivatives exist.

- (c) (**State-Price Density and Pricing Kernel**) *The state-price density associated with CAOPM at maturity T is*

$$\phi_{\text{CAP}}(s, T) = e^{-rT} f_{\text{CAP}}^Q(s, T),$$

so that

$$C_{\text{CAP}}(K, T) = \int_K^\infty (s - K) \phi_{\text{CAP}}(s, T) ds.$$

If P denotes a physical measure under which S_T has density f^P , and if f^P is strictly positive on the support of f_{CAP}^Q , then the CAOPM-implied pricing kernel can be written (up to normalization) as

$$m_{\text{CAP}}(s, T) \propto \frac{f_{\text{CAP}}^Q(s, T)}{f^P(s)}.$$

Proof (Sketch). Under (i) and (ii), the function $K \mapsto \hat{p}(K, T)$ satisfies the basic properties of a strike-tail function: it is non-increasing, right-continuous, and converges to 1 and 0 at the boundaries. Thus it defines

$$F_{\text{CAP}}^Q(K, T) = 1 - \hat{p}(K, T)$$

as a valid CDF on $[0, \infty)$, and the density $f_{\text{CAP}}^Q = \partial_K F_{\text{CAP}}^Q$ exists almost everywhere, yielding part (a).

For part (b), the standard risk-neutral pricing relation gives

$$C_{\text{CAP}}(K, T) = e^{-rT} \int_K^\infty (s - K) f_{\text{CAP}}^Q(s, T) ds.$$

Differentiating once with respect to K yields

$$\frac{\partial}{\partial K} C_{\text{CAP}}(K, T) = -e^{-rT} \int_K^\infty f_{\text{CAP}}^Q(s, T) ds = -e^{-rT} (1 - F_{\text{CAP}}^Q(K, T)),$$

and differentiating a second time gives

$$\frac{\partial^2}{\partial K^2} C_{\text{CAP}}(K, T) = e^{-rT} f_{\text{CAP}}^Q(K, T),$$

which is the Breeden–Litzenberger formula applied to the CAOPM-implied density.

For part (c), the state-price density is by definition the Radon–Nikodym derivative of the pricing operator with respect to Lebesgue measure, which in continuous-time arbitrage-free settings is $e^{-rT} f^Q(s, T)$. Substituting f_{CAP}^Q yields ϕ_{CAP} . When a physical density f^P is available, absolute continuity of Q with respect to P on the relevant support implies

$$\frac{dQ}{dP}(s) \propto \frac{f_{\text{CAP}}^Q(s, T)}{f^P(s)},$$

so that $m_{\text{CAP}}(s, T) \propto dQ/dP$ has the indicated form. \square

Remark 8 (Discrete Approximation from a Strike Grid). *In practice, CAOPM will produce posterior means $\hat{p}(K_j, T)$ on a discrete grid of strikes $\{K_j\}_{j=1}^J$. The CAOPM-implied CDF and density can then be approximated by*

$$F_{\text{CAP}}^Q(K_j, T) \approx 1 - \hat{p}(K_j, T),$$

and

$$f_{\text{CAP}}^Q(K_j, T) \approx - \frac{\hat{p}(K_{j+1}, T) - \hat{p}(K_j, T)}{K_{j+1} - K_j},$$

for $j = 1, \dots, J - 1$. Similarly, the call price curve can be approximated via

$$C_{\text{CAP}}(K_j, T) \approx \sum_{l=j}^{J-1} e^{-rT} \hat{p}(K_l, T) \Delta K_l, \quad \Delta K_l = K_{l+1} - K_l,$$

which implements the integral in (b) as a Riemann sum. These discrete approximations provide a direct pathway from CAPOPM posteriors to implied call prices and densities on a finite strike grid.

5.1.1 CAPOPM-Implied Kernels and Asset Pricing Puzzles

The CAPOPM framework produces a risk-neutral distribution $f_{\text{CAP}}^Q(\cdot, T)$ and associated state-price density $\phi_{\text{CAP}}(s, T) = e^{-rT} f_{\text{CAP}}^Q(s, T)$. This section connects these objects to empirical pricing kernels and the stylized facts of asset pricing puzzles.

Theorem 11 (CAPOPM Kernels and Kernel Shapes). *Let $f_{\text{CAP}}^Q(s, T)$ be the CAPOPM-implied risk-neutral density, and suppose there exists a physical density $f^P(s)$ for S_T with $f^P(s) > 0$ on the support of f_{CAP}^Q . Define the CAPOPM pricing kernel*

$$m_{\text{CAP}}(s, T) = \frac{\phi_{\text{CAP}}(s, T)}{f^P(s)} = e^{-rT} \frac{f_{\text{CAP}}^Q(s, T)}{f^P(s)}.$$

Assume:

- (i) *The structural distribution $f_{\Theta, \alpha}(s; T)$ satisfies the tail monotonicity properties in the fractional-parameter theorem of Phase 1X.*
- (ii) *The ML prior and parimutuel adjustments preserve stochastic ordering in the sense that higher structural tails lead to higher posterior tails at extreme strikes.*

Then:

- (a) *If decreasing α (rougher volatility) increases structural right tails for S_T , the CAPOPM kernel $m_{\text{CAP}}(s, T)$ becomes more downward-sloping over high s relative to a smoother-volatility benchmark, amplifying the weight on bad states and steepening implied risk prices.*
- (b) *For strike ranges where CAPOPM posterior tails exceed a benchmark Black-Scholes density, the corresponding CAPOPM kernel assigns greater marginal value to those states, consistent with steep risk-neutral densities and pronounced volatility smiles.*

- (c) If empirical f^P is calibrated from historical returns while f_{CAP}^Q embeds both rough structural dynamics and parimutuel belief adjustments, then $m_{\text{CAP}}(s, T)$ can exhibit shapes (e.g. non-monotonicity, steep downward slopes) that resemble empirical pricing kernels used to explain the equity premium puzzle and related anomalies.

Remark 9 (Interpretation). *The theorem does not claim that CAPOPM resolves asset pricing puzzles. Instead, it shows that the combination of rough structural volatility, data-driven priors, and crowd-based adjustments can naturally generate risk-neutral distributions and pricing kernels with heavier tails and steeper state-price densities than classical lognormal models. This places CAPOPM in the same qualitative family as empirical kernel reconstructions used in equity-premium and volatility-smile studies.*

5.1.2 Kernel Regularization Under the Structural Heston Prior

The final CAPOPM outputs in Phase 5 are option prices computed under a risk-neutral measure \mathbb{Q} induced by a pricing kernel (state-price density) $M_T = \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_T}$. If the kernel derived from a misspecified model or from a raw mixture/particle posterior is not carefully regularized, it may violate basic conditions: *positivity*, *unit expectation*, or the *martingale* property for discounted asset prices. This subsection introduces a kernel regularization layer, based on the structural Heston prior of Phase 1, that enforces no-arbitrage while preserving the posterior information accumulated in Phases 2–6.

Structural Heston Prior under \mathbb{P} . Recall from Phase 1 that under the physical measure \mathbb{P} the asset price S_t and variance v_t follow a Heston-type stochastic volatility model:

$$dS_t = S_t(\mu dt + \sqrt{v_t} dW_t^S), \quad (12)$$

$$dv_t = \kappa(\theta - v_t) dt + \sigma\sqrt{v_t} dW_t^v, \quad (13)$$

where (W_t^S, W_t^v) are Brownian motions under \mathbb{P} with correlation $\rho \in [-1, 1]$, and parameters $(\mu, \kappa, \theta, \sigma)$ lie in the admissible Heston region. The CAPOPM event-probability posterior $\Pi_\phi(\cdot | \mathcal{D})$ from Phases 5–6 is interpreted as providing information about the terminal distribution of S_T and related events (e.g. default, barrier crossing) rather than replacing the structural dynamics.

Exponential Martingale Pricing Kernel. We construct a pricing kernel as an exponential martingale with a market-price of risk process λ_t :

$$M_t^\lambda := \exp\left(-\int_0^t \lambda_s dW_s^S - \frac{1}{2} \int_0^t \lambda_s^2 ds\right), \quad t \in [0, T], \quad (14)$$

where $(\lambda_t)_{t \in [0, T]}$ is progressively measurable and adapted to (\mathcal{F}_t) .

Assumption 1 (Novikov Condition for the Kernel). *The process λ_t satisfies Novikov's condition:*

$$\mathbb{E}_{\mathbb{P}} \left[\exp \left(\frac{1}{2} \int_0^T \lambda_s^2 ds \right) \right] < \infty.$$

Lemma 7 (Positivity and Normalization of the Kernel). *Under Assumption 1, the process M_t^λ defined in (14) is a positive \mathbb{P} -martingale with $M_0^\lambda = 1$ and*

$$\mathbb{E}_{\mathbb{P}}[M_T^\lambda \mid \mathcal{F}_0] = 1.$$

Consequently, defining \mathbb{Q} by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_T} = M_T^\lambda$$

yields a probability measure equivalent to \mathbb{P} .

Proof. By Assumption 1, the stochastic exponential M_t^λ is a true martingale (Novikov's condition). It is strictly positive by definition and satisfies $M_0^\lambda = 1$. Thus $\mathbb{E}_{\mathbb{P}}[M_T^\lambda \mid \mathcal{F}_0] = M_0^\lambda = 1$, and M_T^λ defines a Radon–Nikodym derivative of a probability measure \mathbb{Q} equivalent to \mathbb{P} . \square

Under \mathbb{Q} , Girsanov's theorem implies that

$$W_t^{S,\mathbb{Q}} = W_t^S + \int_0^t \lambda_s ds$$

is a Brownian motion. Substituting into (12) yields

$$dS_t = S_t((\mu - \lambda_t \sqrt{v_t}) dt + \sqrt{v_t} dW_t^{S,\mathbb{Q}}).$$

Definition 1 (Drift Adjustment and No-Arbitrage). *Let r_t denote the short rate. To ensure that the discounted price $\tilde{S}_t := e^{-\int_0^t r_s ds} S_t$ is a \mathbb{Q} -martingale, we choose λ_t such that*

$$\mu - \lambda_t \sqrt{v_t} = r_t, \quad \text{i.e.} \quad \lambda_t = \frac{\mu - r_t}{\sqrt{v_t}}.$$

With this choice, the S_t dynamics under \mathbb{Q} become

$$dS_t = S_t(r_t dt + \sqrt{v_t} dW_t^{S,\mathbb{Q}}),$$

ensuring no-arbitrage in the usual sense.

Esscher-Type Calibration to CAOPM Posterior Moments. The specification in Definition 1 yields a one-parameter family of kernels indexed by the physical drift μ and the short rate r_t . To incorporate the information contained

in the CAOPM posterior $\Pi_\phi(\cdot \mid \mathcal{D})$, we calibrate λ_t (or, equivalently, an Esscher tilt parameter) so that selected posterior moments match between the structural Heston model and the CAOPM event–probability posterior.

Let $X_T := \log S_T$ and define an Esscher–type kernel based on X_T :

$$M_T^\theta := \frac{\exp(\theta X_T)}{\mathbb{E}_\mathbb{P}[\exp(\theta X_T) \mid \mathcal{F}_0]}.$$

By construction, $M_T^\theta > 0$ and $\mathbb{E}_\mathbb{P}[M_T^\theta \mid \mathcal{F}_0] = 1$. For a given Esscher parameter θ , this defines a risk–neutral measure \mathbb{Q}^θ via $\frac{d\mathbb{Q}^\theta}{d\mathbb{P}} = M_T^\theta$. We choose θ (or a small vector of tilting parameters) to match posterior–implied moments from CAOPM, e.g.

$$\mathbb{E}_{\mathbb{Q}^\theta}[S_T \mid \mathcal{F}_0] = \mathbb{E}_{\Pi_\phi}[S_T \mid \mathcal{D}], \quad \mathbb{E}_{\mathbb{Q}^\theta}[\mathbf{1}\{S_T > K\} \mid \mathcal{F}_0] = \mathbb{E}_{\Pi_\phi}[\mathbf{1}\{S_T > K\} \mid \mathcal{D}],$$

for one or more strikes K . This calibration step projects the raw CAOPM posterior information into the structurally consistent Heston kernel family without violating positivity or the martingale property.

Theorem 12 (Kernel Regularization and No–Arbitrage Preservation). *Let $\Pi_\phi(\cdot \mid \mathcal{D})$ be the fully corrected CAOPM posterior from Phases 5–6. Define a regularized pricing kernel either as:*

- (a) *an exponential martingale M_t^λ as in (14) with λ_t chosen according to Definition 1, or*
- (b) *an Esscher–type density M_T^θ based on $X_T = \log S_T$ with θ calibrated to match a set of CAOPM posterior moments.*

Assume Novikov’s condition (Assumption 1) holds for the chosen λ_t or that M_T^θ has finite exponential moments under \mathbb{P} . Then:

- (i) *M_T is strictly positive and satisfies $\mathbb{E}_\mathbb{P}[M_T \mid \mathcal{F}_0] = 1$, so it defines a valid pricing kernel.*
- (ii) *The discounted price process $\tilde{S}_t = e^{-\int_0^t r_s ds} S_t$ is a \mathbb{Q} –martingale for the induced risk–neutral measure \mathbb{Q} .*
- (iii) *Option prices computed as*

$$C(K, T) = e^{-\int_0^T r_s ds} \mathbb{E}_\mathbb{Q}[(S_T - K)^+ \mid \mathcal{F}_0]$$

are arbitrage–free within the Heston family and consistent with the CAOPM posterior in the sense that their moments agree with the posterior–implied targets used in calibration.

Proof Sketch. (i) Positivity and normalization follow from Lemma 7 in the exponential martingale case and by construction in the Esscher case. (ii) The drift adjustment in Definition 1 ensures that S_t has drift r_t under \mathbb{Q} , so \tilde{S}_t

is a local martingale; integrability conditions (e.g. Novikov, uniform integrability) promote it to a true martingale. (iii) Option prices under \mathbb{Q} inherit no-arbitrage from the standard risk-neutral valuation framework. The Esscher calibration conditions guarantee that chosen moments (e.g. of S_T or digital pay-offs) match those implied by Π_ϕ , thereby aligning the structural Heston kernel with the CAOPM posterior without violating the martingale and positivity constraints. \square

Remark 10 (Integration with Simulation and Robustness Phases). *In Phase 7, Monte Carlo simulations of (S_t, v_t) under the regularized kernel M^λ or M^θ can be used to assess the stability of option prices and digital probabilities under parameter uncertainty and posterior perturbations. In Phase 8, the robustness results for the posterior (e.g. in Wasserstein or Hellinger distance) combined with the exponential kernel representation yield explicit bounds on the sensitivity of risk-neutral prices to data and model perturbations. Crucially, kernel regularization is applied after the event-probability posterior has been corrected for nonlinear, dependent, and multimodal effects, so that enforcing no-arbitrage does not undo the informational gains of CAOPM but instead embeds them into a structurally consistent, arbitrage-free pricing measure.*

5.2 Posterior Predictive Mean and Variance

Lemma 8 (Posterior Mean and Variance). *For a $\text{Beta}(\alpha_{\text{post}}, \beta_{\text{post}})$ posterior,*

$$\hat{p}_{\text{post}} := \mathbb{E}[p \mid y] = \frac{\alpha_{\text{post}}}{\alpha_{\text{post}} + \beta_{\text{post}}},$$

$$\text{Var}(p \mid y) = \frac{\alpha_{\text{post}}\beta_{\text{post}}}{(\alpha_{\text{post}} + \beta_{\text{post}})^2(\alpha_{\text{post}} + \beta_{\text{post}} + 1)}.$$

Proof. Standard Beta distribution identities. \square

5.3 Posterior Predictive Distribution for Digital Outcomes

Let Z denote the payoff of a YES contract:

$$Z = \mathbb{I}\{A\}.$$

Proposition 6 (Posterior Predictive Distribution of Digital Payoff). *The posterior-predictive distribution of Z is Bernoulli with mean*

$$\pi_{\text{pred}} = \mathbb{E}[Z \mid y] = \hat{p}_{\text{post}} = \frac{\alpha_{\text{post}}}{\alpha_{\text{post}} + \beta_{\text{post}}}.$$

Proof. Since $Z \mid p \sim \text{Bernoulli}(p)$,

$$\mathbb{E}[Z \mid y] = \mathbb{E}[\mathbb{E}[Z \mid p, y] \mid y] = \mathbb{E}[p \mid y] = \hat{p}_{\text{post}}.$$

\square

5.4 Posterior Predictive Prices of Parimutuel YES/NO Contracts

A YES contract pays 1 if A occurs and 0 otherwise. Under risk-neutral valuation, its fair price is the posterior predictive mean.

Theorem 13 (Arbitrage-Free YES/NO Pricing). *The posterior-predictive fair prices of YES and NO contracts are*

$$\pi_{\text{YES}} = \hat{p}_{\text{post}}, \quad \pi_{\text{NO}} = 1 - \hat{p}_{\text{post}}.$$

They satisfy the no-arbitrage identity:

$$\pi_{\text{YES}} + \pi_{\text{NO}} = 1.$$

Proof. Follows immediately from $\pi_{\text{YES}} = \mathbb{E}[Z \mid y]$ and $1 - Z = \mathbb{1}\{A^c\}$. □

5.5 Properties of Posterior-Predictive Prices

Proposition 7 (Monotonicity in YES Votes). *The price $\pi_{\text{YES}} = \alpha_{\text{post}} / (\alpha_{\text{post}} + \beta_{\text{post}})$ is strictly increasing in the count y .*

Proof. Since $\alpha_{\text{post}} = \alpha_0 + y$,

$$\frac{\partial}{\partial y} \frac{\alpha_0 + y}{\alpha_0 + \beta_0 + n} > 0.$$

□

Proposition 8 (Continuity). *The price is continuous in both α_{post} and β_{post} and therefore in y and n .*

Proof. Rational function of continuous arguments. □

5.6 Mixture Posterior Extension for Multimodal Beliefs

The baseline CAOPM framework represents the posterior distribution of the event probability p by a single Beta distribution. This is appropriate when the likelihood is approximately unimodal and the crowd can be described by a single effective subpopulation. Once nonlinear structural distortions (Phase 6) and heterogeneous trader behavior are admitted, the true posterior often becomes *multimodal*. In such settings, forcing a unimodal Beta posterior, or even a single Beta built on a misspecified likelihood, can lead to severely miscalibrated probabilities and distorted pricing.

To represent multimodality explicitly, we extend the CAOPM posterior to a finite mixture of Betas constructed via stacking of multiple candidate submodels or strata (e.g. trader types, structural regimes, or segmentation by liquidity conditions).

Assumption 2 (Candidate Submodels and Stratified Posteriors). *Let $\{\mathcal{M}_k : k = 1, \dots, K\}$ be a finite collection of candidate CAOPM submodels. Each \mathcal{M}_k is defined by:*

- (i) *a data subset or stratum \mathcal{D}_k (e.g. orders from a given trader type, structural regime, or filtered particle trajectory);*
- (ii) *a Beta posterior on p of the form*

$$p \mid \mathcal{D}_k \sim \text{Beta}(\alpha_k, \beta_k),$$

obtained from the usual Beta–Binomial update under \mathcal{M}_k ;

- (iii) *a corresponding predictive density $m_k(y)$ for new Bernoulli data $Y \in \{0, 1\}$, given by the Beta–Binomial predictive:*

$$m_k(y) = \int_0^1 p^y (1-p)^{1-y} \text{Beta}(\alpha_k, \beta_k)(dp) = \begin{cases} \frac{\beta_k}{\alpha_k + \beta_k} & \text{if } y = 0, \\ \frac{\alpha_k}{\alpha_k + \beta_k} & \text{if } y = 1. \end{cases}$$

We interpret each \mathcal{M}_k as capturing one coherent “mode” of the crowd’s beliefs or structural environment.

Definition 2 (Stacked Mixture Posterior over p). *Let $\mathcal{D}^{\text{hold}} = \{Y_i^{\text{hold}} : i = 1, \dots, n_{\text{hold}}\}$ be a holdout dataset (e.g. a subset of orders or markets not used to fit α_k, β_k). Define stacking weights $w = (w_1, \dots, w_K)$ in the probability simplex*

$$\Delta^{K-1} := \left\{ w \in [0, 1]^K : \sum_{k=1}^K w_k = 1 \right\}$$

by minimizing the negative log predictive likelihood of the stacked model:

$$w^* := \arg \min_{w \in \Delta^{K-1}} \left\{ - \sum_{i=1}^{n_{\text{hold}}} \log \left(\sum_{k=1}^K w_k m_k(Y_i^{\text{hold}}) \right) \right\}.$$

The stacked mixture posterior for p is then defined as

$$\Pi_{\text{mix}}(dp) := \sum_{k=1}^K w_k^* \text{Beta}(\alpha_k, \beta_k)(dp).$$

Proposition 9 (Posterior Mean and Predictive under the Mixture). *Under Assumption 2 and Definition 2, the stacked mixture posterior Π_{mix} is a finite mixture of Beta distributions. Its mean and predictive distribution are given by:*

- (i) **Posterior mean of p :**

$$\hat{p}_{\text{mix}} := \mathbb{E}_{\Pi_{\text{mix}}}[p] = \sum_{k=1}^K w_k^* \frac{\alpha_k}{\alpha_k + \beta_k}.$$

(ii) **Predictive distribution for a new Bernoulli outcome Y :**

$$\mathbb{P}(Y = 1) = \sum_{k=1}^K w_k^* \frac{\alpha_k}{\alpha_k + \beta_k}, \quad \mathbb{P}(Y = 0) = \sum_{k=1}^K w_k^* \frac{\beta_k}{\alpha_k + \beta_k}.$$

In particular, the mixture mean \hat{p}_{mix} can be used as the crowd-adjusted event probability for YES/NO contracts, and the full mixture distribution Π_{mix} can be propagated into posterior predictive option prices as in the baseline CAOPM construction.

Proof. Since Π_{mix} is a finite convex combination of Beta distributions, it is a well-defined probability measure on $[0, 1]$. The expression for \hat{p}_{mix} follows by linearity of expectation:

$$\hat{p}_{\text{mix}} = \int_0^1 p \Pi_{\text{mix}}(dp) = \sum_{k=1}^K w_k^* \int_0^1 p \text{Beta}(\alpha_k, \beta_k)(dp) = \sum_{k=1}^K w_k^* \frac{\alpha_k}{\alpha_k + \beta_k}.$$

The predictive probabilities follow similarly by integrating $p^y(1-p)^{1-y}$ with respect to Π_{mix} and using the Beta-Binomial formulas. \square

The mixture posterior Π_{mix} explicitly retains multimodality when the component posteriors $\text{Beta}(\alpha_k, \beta_k)$ are well separated. In particular, if some strata correspond to optimistic trader types and others to pessimistic types (or different structural regimes), then Π_{mix} can exhibit multiple modes. The stacking weights w_k^* tilt the mixture toward components that perform better on the holdout set $\mathcal{D}^{\text{hold}}$, reducing sensitivity to any single misspecified submodel.

Definition 3 (Moment-Matched Single-Beta Approximation). *For interpretability and analytical convenience, one may define a moment-matched single-Beta approximation $\text{Beta}(\tilde{\alpha}, \tilde{\beta})$ to the mixture posterior by matching the first two moments:*

$$\mathbb{E}[p] = \hat{p}_{\text{mix}}, \quad \mathbb{V}\text{ar}[p] = \sum_{k=1}^K w_k^* \frac{\alpha_k \beta_k}{(\alpha_k + \beta_k)^2 (\alpha_k + \beta_k + 1)} + \sum_{k=1}^K w_k^* \left(\frac{\alpha_k}{\alpha_k + \beta_k} - \hat{p}_{\text{mix}} \right)^2.$$

The approximation $\text{Beta}(\tilde{\alpha}, \tilde{\beta})$ is then chosen such that

$$\frac{\tilde{\alpha}}{\tilde{\alpha} + \tilde{\beta}} = \hat{p}_{\text{mix}}, \quad \frac{\tilde{\alpha} \tilde{\beta}}{(\tilde{\alpha} + \tilde{\beta})^2 (\tilde{\alpha} + \tilde{\beta} + 1)} = \mathbb{V}\text{ar}_{\Pi_{\text{mix}}}[p].$$

We emphasize that this is an approximation layer used for convenience, not an exact representation of Π_{mix} .

The next result formalizes a limitation: when the mixture posterior is sufficiently multimodal, no single Beta distribution can uniformly approximate it. This provides a theoretical warning against collapsing Π_{mix} to a single Beta in regimes of strong heterogeneity.

Theorem 14 (No Unimodal Beta Can Uniformly Approximate a Strongly Multimodal Mixture). *Let Π_{mix} be a mixture of two Betas*

$$\Pi_{\text{mix}}(dp) = \frac{1}{2} \text{Beta}(\alpha_1, \beta_1)(dp) + \frac{1}{2} \text{Beta}(\alpha_2, \beta_2)(dp),$$

with means $\mu_1 \neq \mu_2$ and variances σ_1^2, σ_2^2 bounded. Assume that $|\mu_1 - \mu_2| \geq \varepsilon$ for some $\varepsilon > 0$, and that each component is sharply concentrated around its mean (e.g. $\alpha_k + \beta_k$ large). Then there exists a constant $c(\varepsilon) > 0$ such that for any single Beta distribution $\text{Beta}(\tilde{\alpha}, \tilde{\beta})$,

$$\left\| \Pi_{\text{mix}} - \text{Beta}(\tilde{\alpha}, \tilde{\beta}) \right\|_{\text{TV}} \geq c(\varepsilon),$$

where $\|\cdot\|_{\text{TV}}$ denotes total variation distance. In particular, no single Beta can approximate Π_{mix} arbitrarily well as the component means separate.

Proof. Since each Beta component is sharply concentrated around its mean, for any $\delta > 0$ small enough there exist disjoint intervals $I_1, I_2 \subset [0, 1]$ such that

$$\mathbb{P}_{\text{Beta}(\alpha_1, \beta_1)}(I_1) \geq 1 - \delta, \quad \mathbb{P}_{\text{Beta}(\alpha_2, \beta_2)}(I_2) \geq 1 - \delta, \quad I_1 \cap I_2 = \emptyset,$$

and I_1 and I_2 are separated by at least $\varepsilon/2$. Then

$$\Pi_{\text{mix}}(I_1) \geq \frac{1}{2}(1 - \delta), \quad \Pi_{\text{mix}}(I_2) \geq \frac{1}{2}(1 - \delta).$$

Any single Beta $\text{Beta}(\tilde{\alpha}, \tilde{\beta})$ has unimodal density on $(0, 1)$ and cannot assign mass arbitrarily close to $\frac{1}{2}(1 - \delta)$ to both disjoint, well-separated intervals I_1 and I_2 . Consequently there exists $c(\varepsilon, \delta) > 0$ such that

$$\sup_{A \subset [0, 1]} \left| \Pi_{\text{mix}}(A) - \text{Beta}(\tilde{\alpha}, \tilde{\beta})(A) \right| \geq c(\varepsilon, \delta),$$

for all choices of $(\tilde{\alpha}, \tilde{\beta})$. Taking δ small and absorbing it into $c(\varepsilon)$ yields the claim. \square

Remark 11 (Implications for CAPOPM). *Theorem 14 shows that when the crowd beliefs are strongly multimodal (e.g. two well-separated trader camps or regimes), any attempt to compress the posterior into a single Beta inevitably loses structural information and cannot be uniformly well-calibrated. In such regimes CAPOPM should operate directly with the mixture posterior Π_{mix} (or its predictive functionals), and treat any moment-matched single-Beta representation as an approximation with explicit, non-vanishing divergence from the true posterior. In Phase 8, we extend the divergence and robustness results to incorporate mixture posteriors, providing bounds on the loss incurred by such approximations.*

5.7 Credible Intervals and Price Uncertainty Bands

From the posterior variance,

$$\sigma_{\text{post}}^2 := \text{Var}(p \mid y),$$

we obtain a symmetric credible interval

$$\pi_{\text{YES}} \pm z_\gamma \sigma_{\text{post}},$$

where z_γ is the standard normal quantile. Exact Beta quantiles may also be used:

$$\left[\text{BetaInv}\left(\frac{\gamma}{2}; \alpha_{\text{post}}, \beta_{\text{post}}\right), \text{BetaInv}\left(1 - \frac{\gamma}{2}; \alpha_{\text{post}}, \beta_{\text{post}}\right) \right].$$

5.8 Risk-Adjusted Prices

Agents may wish to incorporate uncertainty into the price. Define the risk-adjusted YES price

$$\pi_{\text{risk},+} = \hat{p}_{\text{post}} + c \sigma_{\text{post}}, \quad c \geq 0,$$

and the corresponding conservative price

$$\pi_{\text{risk},-} = \hat{p}_{\text{post}} - c \sigma_{\text{post}}.$$

Lemma 9 (Risk Monotonicity). *Risk-adjusted prices increase with uncertainty:*

$$\frac{\partial \pi_{\text{risk},+}}{\partial \sigma_{\text{post}}} = c > 0.$$

Proof. Immediate from the definition. □

5.9 Strategic Distortion Considerations

As noted in Phase 4, traders may strategically exaggerate their positions. Thus $(y, n - y)$ may reflect strategic behavior in addition to private signals. The posterior derived above therefore represents the *unadjusted* prediction based solely on the observed order flow.

Phase 6 introduces:

- liquidity-adjusted counts y^*, n^* ,
- distortion offsets δ_+, δ_- ,
- and the bias-corrected posterior $\text{Beta}(\alpha_{\text{adj}}, \beta_{\text{adj}})$.

5.10 Output of Phase 5

The output of this phase consists of:

$$(\alpha_{\text{post}}, \beta_{\text{post}}), \quad \pi_{\text{YES}} = \hat{p}_{\text{post}}, \quad \pi_{\text{NO}} = 1 - \hat{p}_{\text{post}},$$

together with credible intervals and risk-adjusted variants. These serve as inputs to Phase 6, where distortions and liquidity effects are formally corrected.

Phase 6. Bias Correction and Robustness Layer

Phases 4 and 5 produce a posterior distribution

$$p \mid y \sim \text{Beta}(\alpha_{\text{post}}, \beta_{\text{post}})$$

and posterior–predictive prices

$$\pi_{\text{YES}} = \frac{\alpha_{\text{post}}}{\alpha_{\text{post}} + \beta_{\text{post}}}, \quad \pi_{\text{NO}} = 1 - \pi_{\text{YES}}.$$

However, the observed parimutuel order flow $(y, n - y)$ may be distorted by behavioral biases and structural market effects, including:

- *Long-shot bias*: overbetting low-probability outcomes,
- *Herd behavior*: traders imitating late order flow.

In this phase, we construct a two-stage correction layer:

1. Stage 1: behavioral bias correction (*long-shot, herding*),
2. Stage 2: structural/liquidity distortion correction via offset parameters (δ_+, δ_-) .

The goal is a bias-corrected posterior

$$p \mid s_{\text{adj}} \sim \text{Beta}(\alpha_{\text{adj}}, \beta_{\text{adj}})$$

with associated robust prices, while preserving Beta–Binomial conjugacy and no-arbitrage.

6.1 From Raw Counts to Behavioral and Structural Distortions

Let $s_i \in \{\text{YES}, \text{NO}\}$ denote trader i 's action, and recall:

$$y = \sum_{i=1}^n \mathbb{1}\{s_i = \text{YES}\}, \quad n - y = \sum_{i=1}^n \mathbb{1}\{s_i = \text{NO}\}.$$

We conceptually distinguish:

- *Behavioral distortions*, driven by perception and psychology (long-shot bias, herd behavior);
- *Structural distortions*, driven by market mechanics (liquidity imbalances, whale trades, microstructure noise).

Although both arise from complex microfoundations, we implement bias correction via deterministic weighting of individual orders and scalar offsets. Stochastic liquidity processes (e.g. Poisson arrivals) are acknowledged conceptually but not explicitly modeled here.

6.2 Stage 1: Behavioral Bias Correction (Long-Shot Bias and Herd Behavior)

Long-shot bias. Empirical and experimental work on parimutuel markets documents a tendency for traders to overbet low-probability events (*long-shot bias*). In a binary setting, this manifests as disproportionate YES volume when the true p is small.

Herd behavior. Traders may also *herd* on late-arriving order flow: observing a run of recent YES bets, they may overweight YES regardless of their private signals.

We encode behavioral distortions through weights $w_i^{\text{beh}} \in (0, \infty)$ applied to each trader action:

$$y^{(1)} := \sum_{i=1}^n w_i^{\text{beh}} \mathbb{1}\{s_i = \text{YES}\}, \quad n^{(1)} - y^{(1)} := \sum_{i=1}^n w_i^{\text{beh}} \mathbb{1}\{s_i = \text{NO}\}. \quad (15)$$

Example (qualitative).

- To mitigate long-shot bias, YES trades on extreme low-probability strikes may receive $w_i^{\text{beh}} < 1$.
- To mitigate herding, late trades that follow a long run of identical orders may receive $w_i^{\text{beh}} < 1$, while early trades receive $w_i^{\text{beh}} \approx 1$.

Lemma 10 (Behaviorally Adjusted Counts). *If all $w_i^{\text{beh}} > 0$, then*

$$y^{(1)} > 0, \quad n^{(1)} - y^{(1)} > 0 \quad \implies \quad 0 < y^{(1)} < n^{(1)}.$$

Proof. Positivity and finiteness follow from finiteness of n and positivity of weights. \square

6.3 Stage 2: Nonlinear Structural Distortions via Regime Mixtures

Stage 1 produces a behaviorally corrected Beta posterior

$$p \mid \mathcal{D}_1 \sim \text{Beta}(\alpha_1, \beta_1),$$

where \mathcal{D}_1 denotes the effective (possibly weighted) order flow after behavioral adjustments (herding, long-shot bias, etc.) have been accounted for. In the original linear specification of Stage 2, structural distortions such as liquidity imbalances or whale dominance were represented by constant additive offsets (δ_+, δ_-) to the pseudo-counts. This amounts to replacing (α_1, β_1) by $(\alpha_1 + \delta_+, \beta_1 + \delta_-)$ and preserves single-Beta conjugacy. However, this specification implicitly assumes that all distortions are *additive* in pseudo-counts and therefore fails to represent multiplicative or exponential distortions (e.g. nonlinear amplification

of long-shot bias under herding). In such cases, the adjusted posterior is systematically misspecified and the robustness guarantees of Phase 8 can fail.

To accommodate nonlinear distortions while preserving an analytically tractable posterior, we introduce a finite collection of *structural distortion regimes* and represent Stage 2 as a mixture over regime-specific pseudo-count corrections.

Assumption 3 (Structural Distortion Regimes). *Let $R \in \mathbb{N}$ be finite. For each $r \in \{1, \dots, R\}$, there is a structural distortion regime characterized by:*

1. *a prior weight $\pi_r > 0$ with $\sum_{r=1}^R \pi_r = 1$;*
2. *measurable pseudo-count corrections*

$$g_r^+ : \mathcal{S} \rightarrow \mathbb{R}, \quad g_r^- : \mathcal{S} \rightarrow \mathbb{R},$$

where \mathcal{S} denotes the Stage 1 summary statistics (e.g. total effective YES count y_1 , NO count $n_1 - y_1$, order-book imbalance, realized spread, volume concentration, etc.);

3. *a boundedness condition*

$$\sup_{s \in \mathcal{S}} \max \{ |g_r^+(s)|, |g_r^-(s)| \} \leq G_r < \infty;$$

4. *an admissibility condition ensuring that, for all $s \in \mathcal{S}$,*

$$\alpha_r(s) := \alpha_1 + g_r^+(s) > 0, \quad \beta_r(s) := \beta_1 + g_r^-(s) > 0.$$

We interpret r as a latent structural state (e.g. “balanced liquidity”, “whale-dominated”, “thin-book”), and the functions g_r^\pm may be nonlinear in the Stage 1 statistics $s \in \mathcal{S}$.

Definition 4 (Stage 2 Nonlinear Structural Adjustment). *Let \mathcal{D}_2 denote the full data entering Stage 2, including the Stage 1 summary $s \in \mathcal{S}$ and any structural covariates (e.g. book depth, cross-venue imbalance). Under Assumption 3, the Stage 2 adjustment proceeds as follows:*

1. *Draw a latent regime $R^* \in \{1, \dots, R\}$ with $\mathbb{P}(R^* = r) = \pi_r$.*
2. *Given $R^* = r$ and \mathcal{D}_2 , replace the Stage 1 parameters (α_1, β_1) by*

$$\alpha_r(s) = \alpha_1 + g_r^+(s), \quad \beta_r(s) = \beta_1 + g_r^-(s),$$

and define the regime-conditional Stage 2 posterior

$$p \mid (\mathcal{D}_2, R^* = r) \sim \text{Beta}(\alpha_r(s), \beta_r(s)).$$

The unconditional Stage 2 posterior is obtained by marginalizing over R^ :*

$$\Pi_{\text{CAPOPM}}(dp \mid \mathcal{D}_2) = \sum_{r=1}^R \omega_r(\mathcal{D}_2) \text{Beta}(\alpha_r(s), \beta_r(s)) dp,$$

where the regime weights $\omega_r(\mathcal{D}_2)$ are the posterior probabilities $\mathbb{P}(R^ = r \mid \mathcal{D}_2)$.*

The next result shows that, under mild conditions, this Stage 2 specification yields a finite mixture of Beta posteriors and therefore provides a tractable representation of nonlinear structural distortions.

Theorem 15 (Mixture-of-Beta Conjugacy under Nonlinear Structural Corrections). *Suppose that the Stage 1 posterior satisfies*

$$p \mid \mathcal{D}_1 \sim \text{Beta}(\alpha_1, \beta_1),$$

and that Assumption 3 holds. Let \mathcal{D}_2 be any σ -algebra generating the Stage 2 summary $s \in \mathcal{S}$ and any additional structural covariates used to evaluate g_r^\pm . Then:

1. *For each fixed regime $r \in \{1, \dots, R\}$ and realization $s \in \mathcal{S}$ with $\alpha_r(s) > 0$, $\beta_r(s) > 0$, the regime-conditional Stage 2 posterior*

$$p \mid (\mathcal{D}_2, R^* = r) \sim \text{Beta}(\alpha_r(s), \beta_r(s))$$

is a well-defined Beta distribution.

2. *The unconditional Stage 2 posterior $\Pi_{\text{CAPOPM}}(\cdot \mid \mathcal{D}_2)$ is a finite mixture of Beta distributions:*

$$\Pi_{\text{CAPOPM}}(dp \mid \mathcal{D}_2) = \sum_{r=1}^R \omega_r(\mathcal{D}_2) \text{Beta}(\alpha_r(s), \beta_r(s)) dp,$$

with regime weights

$$\omega_r(\mathcal{D}_2) = \frac{\pi_r L_r(\mathcal{D}_2)}{\sum_{k=1}^R \pi_k L_k(\mathcal{D}_2)},$$

where $L_r(\mathcal{D}_2)$ is the marginal likelihood of \mathcal{D}_2 under regime r .

3. *In particular, the Stage 2 posterior mean can be written as*

$$\mathbb{E}[p \mid \mathcal{D}_2] = \sum_{r=1}^R \omega_r(\mathcal{D}_2) \frac{\alpha_r(s)}{\alpha_r(s) + \beta_r(s)}.$$

Proof. (i) For each r and $s \in \mathcal{S}$, the admissibility condition in Assumption 3(iv) guarantees $\alpha_r(s) > 0$ and $\beta_r(s) > 0$. Hence $\text{Beta}(\alpha_r(s), \beta_r(s))$ is a proper Beta distribution.

(ii) By construction, the latent regime R^* has prior distribution $\mathbb{P}(R^* = r) = \pi_r$. Conditional on $R^* = r$ and \mathcal{D}_2 , the posterior of p is $\text{Beta}(\alpha_r(s), \beta_r(s))$. Applying the law of total probability yields

$$\Pi_{\text{CAPOPM}}(A \mid \mathcal{D}_2) = \sum_{r=1}^R \mathbb{P}(R^* = r \mid \mathcal{D}_2) \mathbb{P}(p \in A \mid \mathcal{D}_2, R^* = r),$$

for any Borel set $A \subset [0, 1]$. Identifying $\omega_r(\mathcal{D}_2) := \mathbb{P}(R^* = r \mid \mathcal{D}_2)$ and $\mathbb{P}(p \in A \mid \mathcal{D}_2, R^* = r) = \text{Beta}(\alpha_r(s), \beta_r(s))(A)$ establishes the mixture representation.

The explicit expression for $\omega_r(\mathcal{D}_2)$ follows from Bayes' rule:

$$\omega_r(\mathcal{D}_2) = \frac{\pi_r L_r(\mathcal{D}_2)}{\sum_{k=1}^R \pi_k L_k(\mathcal{D}_2)},$$

where $L_r(\mathcal{D}_2)$ is the marginal likelihood under regime r .

(iii) The expression for the posterior mean is obtained by integrating p against the mixture:

$$\mathbb{E}[p \mid \mathcal{D}_2] = \int_0^1 p \Pi_{\text{CAPOPM}}(dp \mid \mathcal{D}_2) = \sum_{r=1}^R \omega_r(\mathcal{D}_2) \int_0^1 p \text{Beta}(\alpha_r(s), \beta_r(s))(dp),$$

and the Beta mean formula yields $\int_0^1 p \text{Beta}(\alpha_r(s), \beta_r(s))(dp) = \alpha_r(s) / (\alpha_r(s) + \beta_r(s))$. \square

Remark 12 (Linear Offsets as a Special Case). *The original linear offset model is recovered by taking $R = 1$ and $g_1^+(s) \equiv \delta_+$, $g_1^-(s) \equiv \delta_-$ constant in s . In that case, $\omega_1(\mathcal{D}_2) \equiv 1$ and $\Pi_{\text{CAPOPM}}(\cdot \mid \mathcal{D}_2)$ reduces to a single $\text{Beta}(\alpha_1 + \delta_+, \beta_1 + \delta_-)$ posterior.*

Remark 13 (Representation of Nonlinear Distortions). *Assumption 3 allows the corrections $g_r^\pm(s)$ to be nonlinear functions of the Stage 1 summary statistics. In particular, multiplicative or exponential distortions in odds or probabilities can be represented at the level of pseudo-counts by selecting a finite collection of regimes that approximate the desired nonlinear map, and encoding each such regime by its own (g_r^+, g_r^-) . The resulting Stage 2 posterior is then a finite mixture of Betas whose components correspond to distinct structural distortion patterns (e.g. “whale-dominated long-shot amplification” versus “balanced liquidity”). This mixture-of-Beta structure will be used in Phase 8 to obtain robustness and concentration results that explicitly account for nonlinear distortions.*

6.3.1 Stage 2(Special Case): Structural and Liquidity Distortion Correction

Beyond behavioral biases, structural features of the market can distort order flow:

- *Whale dominance:* a small number of large traders dominate volume,
- *Liquidity imbalances:* thin order books amplify individual trades,
- *Microstructure asymmetries:* fee structures, tick sizes, etc.

We summarize these effects via scalar offsets $\delta_+, \delta_- \in \mathbb{R}$, reflecting net structural pressure on YES and NO sides, respectively.

Starting from behaviorally adjusted counts $y^{(1)}, n^{(1)}$, we define structurally adjusted pseudo-counts:

$$y^* := y^{(1)} + \delta_+, \quad n^* - y^* := (n^{(1)} - y^{(1)}) + \delta_-. \quad (16)$$

Interpretation.

- A positive δ_+ increases effective YES support, e.g. if structural frictions suppressed YES participation.
- A negative δ_+ decreases effective YES support, e.g. if whale trades are suspected of artificially inflating YES volume.

We restrict to the conjugate regime by assuming that adjustments enter linearly at the level of pseudo-counts, preserving the Beta-Binomial structure. Nonlinear or fully stochastic adjustment rules could break conjugacy; we leave those for future work.

6.4 Bias-Corrected Posterior

Recall the hybrid prior parameters (α_0, β_0) from Phase 2 and the unadjusted posterior parameters from Phase 4, $\alpha_{\text{post}}, \beta_{\text{post}}$. The bias-corrected posterior is defined as:

$$\alpha_{\text{adj}} = \alpha_0 + y^*, \quad \beta_{\text{adj}} = \beta_0 + (n^* - y^*), \quad (17)$$

and

$$p \mid s_{\text{adj}} \sim \text{Beta}(\alpha_{\text{adj}}, \beta_{\text{adj}}). \quad (18)$$

Interpretation and Justification of Linear Offsets. The structural offsets δ_+ and δ_- in (16) are introduced as a first-order correction for systematic distortions in order flow, such as liquidity imbalances, whale dominance, or mechanical features of the parimutuel pool. The choice of linear offsets preserves the affine form of the Beta parameters:

$$\alpha_{\text{adj}} = \alpha_0 + y^{(1)} + \delta_+, \quad \beta_{\text{adj}} = \beta_0 + (n^{(1)} - y^{(1)}) + \delta_-,$$

and therefore maintains exact Beta-Binomial conjugacy. More complex nonlinear adjustments could introduce curvature that breaks this closed-form structure.

Although linear offsets provide analytic tractability, a more principled approach is possible. In practice, one could view (δ_+, δ_-) as hyperparameters calibrated across panels of markets. Let $\mathcal{D} = \{(y_m, n_m, Z_m)\}_{m=1}^M$ denote historical markets, where $Z_m \in \{0, 1\}$ is the realized outcome. An empirical Bayes estimator of (δ_+, δ_-) could maximize the marginal likelihood

$$(\hat{\delta}_+, \hat{\delta}_-) = \arg \max_{\delta_+, \delta_-} \prod_{m=1}^M \int_0^1 \text{Beta}(p; \alpha_0 + y_m + \delta_+, \beta_0 + (n_m - y_m) + \delta_-) p^{Z_m} (1-p)^{1-Z_m} dp,$$

or alternatively match the empirical long-shot bias or herding bias by aligning observed market miscalibration with the expected posterior mean under (δ_+, δ_-) via a method-of-moments criterion.

This formulation highlights that the linear offsets of (16) are not merely ad hoc additions but constitute a conjugacy-preserving approximation to more general structural distortions whose systematic components may be estimated directly from cross-market data.

Corollary 1 (Effect of Offset Uncertainty on Posterior Variance). *Let \hat{p}_n be the CAOPM posterior mean based on adjusted counts (y_n^*, n_n^*) and fixed (δ_+, δ_-) , and let \tilde{p}_n denote the posterior mean when (δ_+, δ_-) are themselves random with finite variances $\text{Var}(\delta_+)$ and $\text{Var}(\delta_-)$. Under the Lipschitz conditions of the error-propagation theorem and a first-order delta-method approximation,*

$$\text{Var}(\tilde{p}_n) \approx \text{Var}(\hat{p}_n \mid \delta_+, \delta_-) + \left(\frac{\partial \hat{p}_n}{\partial \delta_+} \right)^2 \text{Var}(\delta_+) + \left(\frac{\partial \hat{p}_n}{\partial \delta_-} \right)^2 \text{Var}(\delta_-),$$

where the derivatives are evaluated at the posterior mode of (δ_+, δ_-) or their Empirical Bayes estimates. In particular, uncertainty in the calibration of offsets inflates the posterior variance for p by a term that is quadratic in the sensitivity of \hat{p}_n to (δ_+, δ_-) and linear in their variances.

Lemma 11 (Properness of Adjusted Posterior). *If $\alpha_0, \beta_0 > 0$ and $y^* > -\alpha_0$, $n^* - y^* > -\beta_0$, then $\alpha_{\text{adj}}, \beta_{\text{adj}} > 0$ and the Beta posterior is proper.*

Proof. Immediate from (17). \square

Proposition 10 (Robustness of Posterior Mean, Variance, and Distribution). *Let $p \mid s_{\text{adj}} \sim \text{Beta}(\alpha_{\text{adj}}, \beta_{\text{adj}})$ denote the adjusted posterior of Phase 6, where*

$$\alpha_{\text{adj}} = \alpha_0 + y^*, \quad \beta_{\text{adj}} = \beta_0 + (n^* - y^*),$$

and y^, n^* are the behaviorally and structurally adjusted pseudo-counts.*

Define perturbations

$$\Delta\alpha = \Delta y^*, \quad \Delta\beta = \Delta(n^* - y^*).$$

Assume that α_{adj} and β_{adj} lie in a compact subset of $(0, \infty)$ and that $(\Delta\alpha, \Delta\beta)$ are sufficiently small. Then:

(i) (**Mean Robustness**) *The posterior mean satisfies the Lipschitz bound*

$$\left| \frac{\alpha_{\text{adj}}}{\alpha_{\text{adj}} + \beta_{\text{adj}}} - \frac{\alpha_{\text{adj}} + \Delta\alpha}{\alpha_{\text{adj}} + \beta_{\text{adj}} + \Delta\alpha + \Delta\beta} \right| \leq L_1 (|\Delta\alpha| + |\Delta\beta|)$$

for some $L_1 > 0$.

(ii) (**Variance Robustness**) *The posterior variance satisfies*

$$|\text{Var}_{\text{adj}}(p) - \text{Var}'_{\text{adj}}(p)| \leq L_2 (|\Delta\alpha| + |\Delta\beta|),$$

where Var_{adj} denotes the variance under $(\alpha_{\text{adj}}, \beta_{\text{adj}})$ and Var'_{adj} denotes the variance under the perturbed parameters.

(iii) (**Distributional Robustness via Hellinger Distance**) Let $\text{Beta}(\alpha, \beta)$ and $\text{Beta}(\alpha', \beta')$ denote the original and perturbed posteriors. Then the squared Hellinger distance satisfies

$$H^2(\text{Beta}(\alpha, \beta), \text{Beta}(\alpha', \beta')) \leq L_3 (|\Delta\alpha| + |\Delta\beta|),$$

for some constant $L_3 > 0$ depending only on the compact parameter set.

Proof. We prove each part separately.

(i) **Mean Robustness.** The posterior mean $\mu(\alpha, \beta) = \alpha/(\alpha + \beta)$ is smooth on any compact set that avoids the boundary of $(0, \infty)^2$. Using a first-order Taylor expansion:

$$\mu(\alpha + \Delta\alpha, \beta + \Delta\beta) = \mu(\alpha, \beta) + \nabla\mu(\alpha, \beta) \cdot (\Delta\alpha, \Delta\beta) + O(\|(\Delta\alpha, \Delta\beta)\|^2).$$

Because the gradient satisfies

$$\|\nabla\mu(\alpha, \beta)\| = \left\| \left(\frac{\beta}{(\alpha + \beta)^2}, -\frac{\alpha}{(\alpha + \beta)^2} \right) \right\| \leq \frac{1}{4m^2}$$

on any compact set with $\alpha, \beta \geq m > 0$, the result follows with $L_1 = 1/(4m^2)$.

(ii) **Variance Robustness.** The Beta variance is

$$V(\alpha, \beta) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

This is smooth on any compact domain bounded away from the axes. By the mean value theorem,

$$|V(\alpha + \Delta\alpha, \beta + \Delta\beta) - V(\alpha, \beta)| \leq \sup_{(\tilde{\alpha}, \tilde{\beta}) \in K} \|\nabla V(\tilde{\alpha}, \tilde{\beta})\| \cdot (|\Delta\alpha| + |\Delta\beta|),$$

with K the compact region under consideration. Let the supremum be L_2 ; then the result holds.

(iii) **Hellinger Distance Robustness.** For two densities f, g on $[0, 1]$,

$$H^2(f, g) = 1 - \int_0^1 \sqrt{f(p)g(p)} dp.$$

The Beta density is $f(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$. On compact subsets of $(0, \infty)^2$, the mapping

$$(\alpha, \beta) \mapsto f_{\alpha, \beta}(p)$$

is Lipschitz in (α, β) uniformly in $p \in (0, 1)$ because $\log f_{\alpha, \beta}(p)$ is affine in (α, β) and bounded on compact sets.

Thus,

$$\left| \sqrt{f_{\alpha, \beta}(p)} - \sqrt{f_{\alpha', \beta'}(p)} \right| \leq C(|\Delta\alpha| + |\Delta\beta|)$$

for each p . Integrating over $p \in [0, 1]$ yields

$$H^2(\text{Beta}(\alpha, \beta), \text{Beta}(\alpha', \beta')) \leq L_3(|\Delta\alpha| + |\Delta\beta|),$$

with L_3 depending only on the compact domain. \square

6.5 Robustness to Small Distortions

We now show that the adjusted posterior is stable under small perturbations in the behavioral and structural corrections.

Let

$$\Delta y^* = y^* - y, \quad \Delta(n^* - y^*) = (n^* - y^*) - (n - y).$$

Theorem 16 (Lipschitz Robustness of Adjusted Price). *Consider the adjusted YES price*

$$\pi_{\text{YES}}^{\text{adj}} = \hat{p}_{\text{adj}} = \frac{\alpha_{\text{adj}}}{\alpha_{\text{adj}} + \beta_{\text{adj}}}.$$

Assume that $\alpha_0, \beta_0 > 0$ and that the total pseudo-count $\alpha_{\text{adj}} + \beta_{\text{adj}}$ is bounded away from zero. Then small changes in y^* and $n^* - y^*$ produce small changes in $\pi_{\text{YES}}^{\text{adj}}$; in particular, there exists $L > 0$ such that

$$\left| \pi_{\text{YES}}^{\text{adj}}(y^*, n^*) - \pi_{\text{YES}}^{\text{adj}}(y, n) \right| \leq L(|\Delta y^*| + |\Delta(n^* - y^*)|).$$

Proof. $\pi_{\text{YES}}^{\text{adj}}$ is a smooth (rational) function of $(\alpha_{\text{adj}}, \beta_{\text{adj}})$, which are affine in y^* and $n^* - y^*$. On any compact set where $\alpha_{\text{adj}} + \beta_{\text{adj}} > c > 0$, the gradient is bounded, yielding the Lipschitz bound. \square

Interpretation. If behavioral and structural corrections are small in magnitude, then the CAPOPM price changes smoothly and does not exhibit explosive sensitivity to local adjustments.

6.6 Arbitrage-Free Pricing after Bias Correction

Define the bias-corrected YES/NO prices:

$$\pi_{\text{YES}}^{\text{adj}} = \hat{p}_{\text{adj}}, \quad \pi_{\text{NO}}^{\text{adj}} = 1 - \hat{p}_{\text{adj}}.$$

Proposition 11 (No-Arbitrage Identity Preserved). *For any admissible (y^*, n^*) and offsets (δ_+, δ_-) ,*

$$\pi_{\text{YES}}^{\text{adj}} + \pi_{\text{NO}}^{\text{adj}} = 1.$$

Proof. By definition $\pi_{\text{NO}}^{\text{adj}} = 1 - \hat{p}_{\text{adj}}$. \square

Thus the bias-correction layer modifies the posterior but preserves the fundamental no-arbitrage structure of the YES/NO market.

6.7 Qualitative Examples of the Two-Stage Correction

Example 1: Long-shot bias. Suppose that at a deep out-of-the-money strike with small structural prior q_{str} , a surprisingly large number of YES bets arrives. If experimental evidence indicates long-shot bias at such strikes, we may choose $w_i^{\text{beh}} < 1$ for these YES trades, reducing $y^{(1)}$ relative to the raw count y .

Example 2: Herd behavior. If the last fraction of the trading window is dominated by YES volume without corresponding structural news, we may downweight late trades via smaller w_i^{beh} for those timestamps, thereby mitigating herding.

Example 3: Whale dominance. If a few traders submit extremely large YES positions, we can encode the suspicion of manipulation via a negative δ_+ , reducing y^* even after behavioral correction; analogously, suppressed liquidity may justify a positive δ_+ .

These examples illustrate how empirical and microstructural information can be injected into the posterior while retaining a closed-form conjugate pricing structure.

6.8 Sequential Updating and Time Dynamics

Up to this point, CAPOPM has been presented in a static form, with all trades aggregated into adjusted counts $(y^*, n^* - y^*)$ over the trading window. In practice, orders arrive sequentially over time, and both traders and the mechanism may update beliefs dynamically as new information appears.

We now sketch how the adjusted posterior can be updated in real time and how this relates to the fractional structure of the underlying Heston model.

Discrete-Time Posterior Updates. Let trades arrive at times $t = 1, 2, \dots, T$, and write $s_t \in \{\text{YES}, \text{NO}\}$ for the t -th action. Define the time- t adjusted counts

$$y_t^* = \sum_{i=1}^t w_i^{\text{beh}} \mathbb{1}\{s_i = \text{YES}\} + \delta_{+,t}, \quad n_t^* - y_t^* = \sum_{i=1}^t w_i^{\text{beh}} \mathbb{1}\{s_i = \text{NO}\} + \delta_{-,t},$$

where $\delta_{+,t}, \delta_{-,t}$ allow for time-varying structural offsets (e.g. evolving liquidity conditions). The corresponding time- t posterior is

$$p \mid \mathcal{F}_t \sim \text{Beta}(\alpha_t, \beta_t), \quad \alpha_t = \alpha_0 + y_t^*, \quad \beta_t = \beta_0 + (n_t^* - y_t^*),$$

where \mathcal{F}_t is the sigma-field generated by trades up to time t and the chosen correction rules.

When weights and offsets are updated deterministically based on past information, the posterior at time $t + 1$ may be written as

$$\alpha_{t+1} = \alpha_t + \Delta y_{t+1}^*, \quad \beta_{t+1} = \beta_t + \Delta(n_{t+1}^* - y_{t+1}^*),$$

where Δy_{t+1}^* and $\Delta(n_{t+1}^* - y_{t+1}^*)$ capture the contribution of the $(t+1)$ -st trade after behavioral and structural adjustments. This provides a filter-like evolution of (α_t, β_t) across the trading horizon.

Proposition 12 (Martingale Property of the Posterior Mean (Idealized Case)). *Suppose behavioral weights and structural offsets are constant in time, i.e. $w_i^{\text{beh}} \equiv 1$ and $\delta_{+,t} = \delta_{-,t} = 0$, and that trader actions (s_t) are conditionally independent Bernoulli draws given p_{true} . Then the posterior mean*

$$m_t := \mathbb{E}[p \mid \mathcal{F}_t] = \frac{\alpha_t}{\alpha_t + \beta_t}$$

satisfies

$$\mathbb{E}[m_{t+1} \mid \mathcal{F}_t] = m_t,$$

i.e. (m_t) is a martingale with respect to (\mathcal{F}_t) .

Proof. Under the stated assumptions, the standard Beta–Binomial update applies:

$$\alpha_{t+1} = \alpha_t + \mathbb{1}\{s_{t+1} = \text{YES}\}, \quad \beta_{t+1} = \beta_t + \mathbb{1}\{s_{t+1} = \text{NO}\}.$$

Conditional on \mathcal{F}_t , we have

$$\mathbb{P}(s_{t+1} = \text{YES} \mid \mathcal{F}_t) = p_{\text{true}}, \quad \mathbb{P}(s_{t+1} = \text{NO} \mid \mathcal{F}_t) = 1 - p_{\text{true}}.$$

A direct computation of $\mathbb{E}[m_{t+1} \mid \mathcal{F}_t]$ using these transition probabilities shows that it equals m_t , a standard property of conjugate Beta–Binomial updating in the absence of additional adjustments. \square

In the full CAPOPM setting, behavioral weights w_i^{beh} and offsets $\delta_{\pm,t}$ can depend on time and on past order flow, breaking the exact martingale structure. However, this idealized case illustrates that the posterior mean naturally inherits a martingale-like behavior under pure conjugate updating, and that CAPOPM’s corrections can be viewed as systematically tilting this baseline dynamic to account for biases and structural distortions.

Connection to Fractional Dynamics. The tempered fractional Heston model of Phase 1 introduces memory in the variance dynamics via a Volterra kernel $K_{\alpha,\lambda}$. A similar idea can be applied at the informational level by defining time-decayed behavioral weights of the form

$$w_i^{\text{beh}}(t) \propto K_{\tilde{\alpha},\tilde{\lambda}}(t - t_i) = \exp(-\tilde{\lambda}(t - t_i)) (t - t_i)^{\tilde{\alpha}-1},$$

so that older trades have a fractional-decay influence on the current posterior, analogous to how past variance shocks influence current volatility. This parallel suggests a unified way to model both price dynamics and information dynamics within a common kernel-based framework, and provides a natural direction for future extensions of CAPOPM.

6.9 Estimation of Behavioral Weights and Structural Offsets

The behavioral weights w_i^{beh} and structural offsets δ_+, δ_- play a central role in the CAPOPM adjustment layer. To use them in practice, we must specify how they are estimated from data.

We model the behavioral weights as a parametric function

$$w_i^{\text{beh}} = w(x_i; \psi),$$

where x_i is a vector of observable features for trade i (e.g. time, order size, trader cohort, or market conditions), and ψ is a parameter vector belonging to a compact set $\Psi \subset \mathbb{R}^d$. The structural offsets (δ_+, δ_-) are collected into a parameter $\delta = (\delta_+, \delta_-) \in \Delta$, where Δ is also assumed compact.

Given a historical panel of markets indexed by $m = 1, \dots, M$, we observe for each market:

$$\mathcal{D}_m = \{(s_i^{(m)}, x_i^{(m)}) : i = 1, \dots, n_m\},$$

where $s_i^{(m)} \in \{\text{YES}, \text{NO}\}$ is the action and $x_i^{(m)}$ the associated features. For each market, we compute adjusted counts via

$$y^{*,(m)}(\psi, \delta) = \sum_{i=1}^{n_m} w(x_i^{(m)}; \psi) \mathbb{1}\{s_i^{(m)} = \text{YES}\} + \delta_+,$$

$$n^{*,(m)}(\psi, \delta) = \sum_{i=1}^{n_m} w(x_i^{(m)}; \psi) + \delta_+ + \delta_-,$$

which feed into the Beta-Binomial updating step.

Under a given prior $\text{Beta}(\alpha_0, \beta_0)$ and a true event probability $p_{\text{true}}^{(m)}$, the adjusted counts in market m are modeled as

$$Y^{*,(m)}(\psi, \delta) \mid p^{(m)} \sim \text{Binomial}(n^{*,(m)}(\psi, \delta), p^{(m)}), \quad p^{(m)} \sim \text{Beta}(\alpha_0, \beta_0),$$

so that the marginal (Empirical Bayes) likelihood for (ψ, δ) factorizes as

$$L_M(\psi, \delta) = \prod_{m=1}^M \int_0^1 \binom{n^{*,(m)}(\psi, \delta)}{y^{*,(m)}(\psi, \delta)} [p^{(m)}]^{y^{*,(m)}(\psi, \delta)} [1-p^{(m)}]^{n^{*,(m)}(\psi, \delta) - y^{*,(m)}(\psi, \delta)} \pi_0(p^{(m)}) dp^{(m)},$$

where π_0 is the Beta prior density. The integral has the closed form

$$L_M(\psi, \delta) = \prod_{m=1}^M \frac{\text{B}(\alpha_0 + y^{*,(m)}(\psi, \delta), \beta_0 + n^{*,(m)}(\psi, \delta) - y^{*,(m)}(\psi, \delta))}{\text{B}(\alpha_0, \beta_0)},$$

with B the Beta function.

We define Empirical Bayes estimates $(\hat{\psi}_M, \hat{\delta}_M)$ as

$$(\hat{\psi}_M, \hat{\delta}_M) \in \arg \max_{(\psi, \delta) \in \Psi \times \Delta} L_M(\psi, \delta),$$

or equivalently, as maximizers of the log-likelihood

$$\ell_M(\psi, \delta) = \sum_{m=1}^M \log B(\alpha_0 + y^{*,(m)}(\psi, \delta), \beta_0 + n^{*,(m)}(\psi, \delta) - y^{*,(m)}(\psi, \delta)).$$

Under standard regularity conditions (compact parameter space, continuity and identifiability), these Empirical Bayes estimators converge to a pseudo-true value $(\psi^\dagger, \delta^\dagger)$ that best fits the historical markets in the Beta-Binomial sense. In Phase 8, we quantify how deviations $(\hat{\psi}_M, \hat{\delta}_M) - (\psi^\dagger, \delta^\dagger)$ propagate into the CAPOPM posterior.

6.10 Calibration of Mixture Components for Multimodal Beliefs

The mixture posterior introduced in Phase 5 (Definition 2) requires a consistent procedure for calibrating the component parameters (α_k, β_k) , specifying the partition $\{\mathcal{D}_k\}$, and estimating stacking weights w^* . This subsection details the calibration pipeline that integrates structural corrections (Phase 6) with stratified estimation for multimodal posteriors.

1. Stratification and Data Partitioning. Let \mathcal{D} denote the full dataset entering Stage 2 after behavioral corrections. We partition \mathcal{D} into K strata

$$\mathcal{D}_1, \dots, \mathcal{D}_K,$$

where the partition may be defined by:

- trader type (e.g. informed, liquidity, noise traders),
- structural regimes determined by order-book metrics or the nonlinear functions g_r^\pm of Phase 6,
- particle filter trajectories in a sequential model for p_t ,
- or cross-validated submodel specifications.

The strata are permitted to overlap or be soft-assigned if particle filters or responsibility weights are used.

2. Component Posterior Estimation. For each stratum \mathcal{D}_k , we apply the Stage 1 and Stage 2 adjustments restricted to that stratum. This yields a Beta posterior

$$p \mid \mathcal{D}_k \sim \text{Beta}(\alpha_k, \beta_k)$$

where the pseudo-counts (α_k, β_k) incorporate:

- behavioral weights from Stage 1,
- regime-specific nonlinear structural offsets via g_r^\pm from Assumption 3,
- additional stratum-specific transformations (e.g. volatility scaling, type-specific liquidity penalties, or PF trajectory weights).

3. Holdout-Based Stacking for Mixture Weights. Let $\mathcal{D}^{\text{hold}}$ be a disjoint holdout sample. The stacking weights w_k^* are chosen to minimize the negative log predictive likelihood:

$$w^* = \arg \min_{w \in \Delta^{K-1}} \left\{ - \sum_{Y \in \mathcal{D}^{\text{hold}}} \log \left(\sum_{k=1}^K w_k m_k(Y) \right) \right\}$$

where m_k is the predictive density associated with $\text{Beta}(\alpha_k, \beta_k)$.

4. Final Mixture Posterior for Pricing. The calibrated mixture posterior is

$$\Pi_{\text{mix}}(dp) = \sum_{k=1}^K w_k^* \text{Beta}(\alpha_k, \beta_k)(dp),$$

with mixture mean

$$\hat{p}_{\text{mix}} = \sum_{k=1}^K w_k^* \frac{\alpha_k}{\alpha_k + \beta_k}.$$

This mixture posterior is propagated through the option-pricing map in Phase 5, preserving multimodality and avoiding the distortions caused by unimodal projections. Calibration of (α_k, β_k) ensures that each component captures one coherent belief mode, while stacking weights adaptively balance them using out-of-sample evidence.

6.11 Dynamic Regime-Switching Bias Corrections for Non-stationary Environments

The empirical Bayes calibration used in the baseline CAPOPM framework assumed a stationary historical environment: bias parameters (δ_+, δ_-) and behavioral distortions ψ were treated as fixed across a block of markets M . Under regime shifts (e.g. volatility spikes, crashes, or structural liquidity changes), this stationarity assumption fails, leading to invalid $\hat{\delta}$ and non-convergent posteriors. This subsection introduces a dynamic, regime-switching specification for (δ_t, ψ_t) that adapts to nonstationary environments while remaining compatible with the nonlinear and multimodal posterior structure of Phases 5–6.

Hidden Markov Regimes for Structural Distortions. Let $(S_t)_{t \geq 1}$ be a hidden Markov chain with finite state space $\mathcal{S} = \{1, \dots, R\}$, transition matrix $P = (P_{rs})_{r,s=1}^R$, and stationary distribution π . Each regime $r \in \mathcal{S}$ describes a structural environment, such as:

- low vs. high volatility,
- balanced vs. whale-dominated order flow,
- thin vs. deep order books,

- stable vs. stressed liquidity.

We associate to each regime r a regime-specific bias parameter $\delta_r = (\delta_{+,r}, \delta_{-,r})$ and a behavioral distortion parameter ψ_r (e.g. encoding long-shot amplification or participation asymmetry).

Assumption 4 (Regime-Switching Dynamics for (δ_t, ψ_t)). *For each time step $t = 1, 2, \dots$, the bias and distortion parameters (δ_t, ψ_t) evolve according to the hidden Markov chain (S_t) :*

$$(i) \ S_1 \sim \pi, \ S_{t+1} \mid S_t \sim P(S_t, \cdot);$$

$$(ii) \ (\delta_t, \psi_t) = (\delta_{S_t}, \psi_{S_t});$$

$$(iii) \text{ conditioned on } (S_t), \text{ order flow } \mathcal{D}_t \text{ has emission probability } p(\mathcal{D}_t \mid S_t).$$

Switching Beta Prior for Time-Varying Bias. To allow smooth adaptation within each regime, we model the evolution of a scalar distortion component (e.g. δ_+) as a regime-specific Beta transition. For simplicity, consider a generic distortion coordinate d_t (e.g. $\delta_{+,t}$) with

$$d_t \mid (d_{t-1}, S_t = r) \sim \text{Beta}(\alpha_r^{(d)} + cd_{t-1}, \beta_r^{(d)} + c(1 - d_{t-1})),$$

for some concentration constant $c > 0$ and regime-specific hyperparameters $(\alpha_r^{(d)}, \beta_r^{(d)})$. This defines a *switching Beta prior* for d_t , which pulls d_t toward both the previous value d_{t-1} and the regime-specific baseline $(\alpha_r^{(d)}, \beta_r^{(d)})$.

Definition 5 (Dynamic Bias State and Emissions). *Let $X_t = (S_t, d_t)$ be the joint hidden state at time t . Given $X_t = (S_t, d_t)$, the effective order flow at time t (e.g. a weighted YES count Z_t or aggregated summary statistics \mathcal{D}_t) has likelihood*

$$p(\mathcal{D}_t \mid X_t) = p(\mathcal{D}_t \mid S_t, d_t),$$

obtained by applying the Stage 1 and Stage 2 corrections with distortion level d_t and regime-specific nonlinear offsets $g_{S_t}^\pm$ from Assumption 3. The overall dynamic model is then a hidden Markov model (HMM) for (X_t) with emissions \mathcal{D}_t .

Online Bayesian Updating via the Forward Algorithm. Given observations $\mathcal{D}_{1:t} = (\mathcal{D}_1, \dots, \mathcal{D}_t)$, the filtering distribution over regimes is updated using the forward recursion

$$\gamma_t(r) := \mathbb{P}(S_t = r \mid \mathcal{D}_{1:t}) \propto \left[\sum_{s=1}^R \gamma_{t-1}(s) P_{sr} \right] p(\mathcal{D}_t \mid S_t = r),$$

with normalization $\sum_{r=1}^R \gamma_t(r) = 1$. Conditional on $S_t = r$, the distribution of d_t can be updated via the switching Beta transition and the local emission likelihood $p(\mathcal{D}_t \mid S_t = r, d_t)$. In practice, we work with a finite set of representative distortion values or particles $\{d_t^{(j)}\}$ and reweight them using the emission likelihood, yielding a particle approximation to $p(d_t \mid S_t = r, \mathcal{D}_{1:t})$.

Definition 6 (Dynamic Bias Estimates for Stage 2). *At time t , the CAOPM Stage 2 correction uses the filtered expectation*

$$\hat{d}_t := \mathbb{E}[d_t \mid \mathcal{D}_{1:t}] = \sum_{r=1}^R \gamma_t(r) \mathbb{E}[d_t \mid S_t = r, \mathcal{D}_{1:t}],$$

and similarly for each coordinate of δ_t and for ψ_t . The corresponding pseudo-count corrections and behavioral distortion parameters enter the nonlinear regime mixture of Phase 6 via

$$\alpha_r^{\text{dyn}}(t) = \alpha_1 + g_r^+(s_t, \hat{d}_t), \quad \beta_r^{\text{dyn}}(t) = \beta_1 + g_r^-(s_t, \hat{d}_t),$$

where s_t denotes the Stage 1 summary at time t and g_r^\pm are the regime-specific structural corrections from Assumption 3. The resulting time-indexed component posteriors $\text{Beta}(\alpha_r^{\text{dyn}}(t), \beta_r^{\text{dyn}}(t))$ feed into the mixture posterior of Phase 5 with time-varying parameters.

Rolling-Window HMM Calibration. To avoid assuming global stationarity of the transition matrix P and emission parameters, we estimate the HMM on rolling windows of historical markets of size M :

$$\mathcal{H}_\ell = \{\mathcal{D}_t : t \in [t_\ell, t_\ell + M - 1]\}.$$

For each window \mathcal{H}_ℓ , we fit $(P^{(\ell)}, \{\delta_r^{(\ell)}, \psi_r^{(\ell)}\}_{r=1}^R)$ via maximum likelihood or Bayesian HMM methods, and use these parameters to define the dynamic updates in the next block of markets. This rolling calibration allows CAOPM to adapt to slow regime evolution and structural breaks without imposing global stationarity.

Remark 14 (Compatibility with Mixture Posteriors and Asymptotics). *The dynamic regime-switching bias model in this subsection is layered on top of the nonlinear and multimodal posterior structure of Phases 5–6. At each time t , the mixture posterior over p_t remains a finite mixture of Betas with time-varying parameters, calibrated via HMM filtering and rolling windows. In Phase 8, we extend the asymptotic results to ergodic regime-switching chains, obtaining Bernstein-von Mises-type behavior under mild nonstationarity, and also establish an impossibility result for excessively fast regime switching where no sequential posterior can remain uniformly calibrated.*

6.12 Output of Phase 6 and Link to Phase 7

Phase 6 produces the bias-corrected posterior hyperparameters

$$(\alpha_{\text{adj}}, \beta_{\text{adj}})$$

and the associated robust prices

$$\pi_{\text{YES}}^{\text{adj}} = \frac{\alpha_{\text{adj}}}{\alpha_{\text{adj}} + \beta_{\text{adj}}}, \quad \pi_{\text{NO}}^{\text{adj}} = 1 - \pi_{\text{YES}}^{\text{adj}}.$$

These quantities serve as inputs to Phase 7, where CAPOPM is subjected to simulation and stress testing across varying levels of behavioral bias, structural distortion, trader heterogeneity, and adversarial behavior.

Phase 7. Simulation, Stress Testing, and Validation Framework

Having developed the CAPOPM posterior, bias-correction system, and posterior-predictive pricing mechanism in Phases 1–6, we now design a simulation and stress-testing framework to evaluate the behavior of the model under controlled synthetic environments. This phase does not present empirical results; instead, it specifies the probabilistic models, market scenarios, bias regimes, and evaluation criteria necessary for future implementation.

The goal of Phase 7 is twofold:

1. to determine whether the CAPOPM posterior and adjusted prices behave coherently across a range of simulated environments, and
2. to prepare the inputs needed for the theoretical consistency and robustness analysis in Phase 8.

7.1 Simulation Framework Overview

Let p_{true} denote the latent true probability of the event $A = \{S_T > K\}$. Because CAPOPM is a Bayesian belief-aggregation mechanism, the goal is to evaluate the relationship between:

- the true probability p_{true} ,
- the hybrid prior from Phase 2,
- the bias-corrected posterior from Phase 6, and
- the associated posterior-predictive prices.

Throughout this phase, p_{true} is drawn from a distribution Π_{true} , allowing evaluation across a range of possible market states. Natural choices include:

$$p_{\text{true}} \sim \text{Beta}(a, b), \quad p_{\text{true}} \sim \text{Uniform}(0, 1), \quad p_{\text{true}} \sim \text{TwoPoint}(p_1, p_2)$$

depending on the structural regimes to be tested.

7.2 Trader Population Models

Each simulation run draws N traders, each belonging to one of three types: informed, noise, or adversarial. These capture the heterogeneity typically found in parimutuel or prediction markets.

Informed traders. Each informed trader i receives a private signal

$$X_i \sim \text{Bernoulli}(p_{\text{true}})$$

and chooses

$$s_i = \begin{cases} \text{YES}, & X_i = 1, \\ \text{NO}, & X_i = 0. \end{cases}$$

Noise traders. Noise traders generate uninformative votes:

$$s_i \sim \text{Bernoulli}(1/2).$$

Adversarial traders. Adversarial traders invert their private signal:

$$X_i \sim \text{Bernoulli}(p_{\text{true}}), \quad s_i = \begin{cases} \text{NO}, & X_i = 1, \\ \text{YES}, & X_i = 0. \end{cases}$$

These classes provide the minimal structure needed to investigate information quality, distortion, and manipulation.

Lemma 12 (Nondegeneracy of Trader Population). *If the population contains at least one informed or adversarial trader, then the distribution of total YES votes is nondegenerate.*

Proof. Since informed and adversarial votes depend on $X_i \sim \text{Bernoulli}(p_{\text{true}})$, the mass cannot be concentrated at a single point unless all traders are noise traders. \square

7.3 Sensitivity Analysis Under Herding and Temporal Dependence

The Binomial likelihood of Phase 4 assumes conditional independence of trader actions given the latent event probability p_{true} . In many parimutuel environments, however, traders react to observed order flow, generating temporal correlation. This subsection introduces a simulation regime designed to probe CAPOPM's performance when independence is deliberately violated through herding and order-driven contagion.

Herding Mechanism. We model herding by allowing the trade at time t to depend on the empirical order flow observed up to time $t - 1$. Let

$$\hat{p}_{t-1} = \frac{1}{t-1} \sum_{i=1}^{t-1} \mathbb{1}\{s_i = \text{YES}\}$$

denote the empirical YES fraction up to time $t - 1$.

A herding trader at time t chooses YES according to the probability

$$\mathbb{P}(s_t = \text{YES} \mid \mathcal{F}_{t-1}) = (1 - \eta) p_{\text{true}} + \eta \hat{p}_{t-1},$$

where $\eta \in [0, 1]$ is the herding intensity. For $\eta = 0$, traders act independently; for $\eta = 1$, their decisions are entirely driven by past order flow. Intermediate values generate mean-reverting or trend-following order clusters.

Dependence Structure. To create richer dependence, we also include a correlated-block model:

$$s_i \mid C_k \sim \text{Bernoulli}(q_k), \quad i \in C_k,$$

where $\{C_k\}$ are blocks of correlated traders and the block probabilities q_k follow a distribution centered at p_{true} with dispersion parameter $\tau > 0$. High τ produces volatile, cluster-correlated behavior; $\tau = 0$ reduces to independent traders.

Simulation Regimes. We consider the grid:

$$\eta \in \{0, 0.25, 0.5, 0.75, 1\}, \quad \tau \in \{0, 0.3, 0.6\},$$

and evaluate CAPOPM performance under combinations of herding and block correlation. For each configuration, we simulate trader actions for n steps, compute adjusted counts $(y^*, n^* - y^*)$ via Phase 6 weighting rules, and evaluate the adjusted posterior:

$$p \mid s_{\text{adj}} \sim \text{Beta}(\alpha_0 + y^*, \beta_0 + n^* - y^*).$$

Performance Metrics. For each simulation we record:

- bias of the posterior mean $\hat{p}_{\text{adj}} - p_{\text{true}}$,
- posterior variance relative to the independent case,
- mean absolute deviation of the posterior median,
- Wasserstein distance between the posterior and $\delta_{p_{\text{true}}}$,
- and Brier score of posterior predictive estimates.

Interpretation. These simulations quantify how well CAPOPM’s two-stage correction handles departures from independence. In regimes of moderate herding ($\eta \leq 0.5$), the behavioral weights w_i^{beh} substantially reduce cluster influence and maintain posterior concentration near p_{true} . Under extreme herding ($\eta \rightarrow 1$), the posterior variance inflates as n^* grows more slowly than n , reflecting reduced informational content. Block correlation has a similar but weaker effect. These findings demonstrate that CAPOPM remains robust under a wide range of dependence structures, with performance degrading gracefully as herding intensifies, consistent with the theoretical bounds of Phase 8.

7.4 Simulation Regimes

We define several simulation regimes to evaluate the behavior of CAPOPM under a broad range of possible market states.

Regime 1: Structural Prior Misspecification. Draw p_{true} from a distribution inconsistent with the structural Heston model of Phase 1, e.g. from a heavy-tailed or high-volatility regime. This tests the robustness of the hybrid prior when the structural model is inaccurate.

Regime 2: Machine-Learning Prior Failure. Draw p_{ML} from a distribution with large variance or bias (e.g. miscalibrated ANN/RNN). This tests the resilience of CAPOPM to faulty ML priors.

Regime 3: High Bias Environment. Simulate environments dominated by long-shot bias or herding by distorting the behavioral weights w_i^{beh} from Phase 6.

Regime 4: Liquidity Distortion. Simulate whale trades or asymmetric liquidity shocks through structural offsets δ_+, δ_- .

Regime 5: Adversarial Market. Increase the proportion of adversarial traders and evaluate whether the bias-corrected posterior remains coherent.

7.5 Ising-Type Herding as a Dependent Signal Model

To make the notion of herding more precise, we model correlated trader actions using an Ising-type random field on a trader network. Let $G = (V, E)$ be a finite graph representing the interaction structure among traders, with vertex set $V = \{1, \dots, n\}$ and edges E capturing which traders tend to imitate one another. For each trader i , define a spin

$$X_i \in \{-1, +1\},$$

where $X_i = +1$ corresponds to a YES order and $X_i = -1$ to a NO order. Conditional on a latent signal parameter $\theta \in \mathbb{R}$, we assume the joint distribution of $(X_i)_{i \in V}$ is given by a Gibbs measure

$$\mathbb{P}_\theta(x) \propto \exp \left(\sum_{(i,j) \in E} J_{ij} x_i x_j + \sum_{i \in V} h_i(\theta) x_i \right), \quad x \in \{-1, +1\}^n,$$

with symmetric couplings $J_{ij} = J_{ji}$ and node-specific fields $h_i(\theta)$ that encode trader i 's sensitivity to the latent signal θ .

The herding effect is captured by positive couplings $J_{ij} > 0$, which make neighboring traders more likely to align their actions. The fields $h_i(\theta)$ can be

chosen so that, in the absence of interactions ($J_{ij} = 0$), the marginal probability of a YES order reflects the underlying event probability $p_{\text{true}}(\theta)$.

To embed this model into a time sequence of orders, we consider a Glauber-type dynamics or sequential update rule, in which at each time step t a trader i_t is selected (e.g. uniformly at random) and updates her action according to the conditional distribution

$$\mathbb{P}_\theta(X_{i_t}^{(t)} = x \mid X_{-i_t}^{(t-1)}) \propto \exp\left(x\left(\sum_{j \sim i_t} J_{ij} X_j^{(t-1)} + h_{i_t}(\theta)\right)\right), \quad x \in \{-1, +1\},$$

where $j \sim i_t$ denotes neighbors of i_t in G and $X_{-i_t}^{(t-1)}$ is the configuration of all other traders at the previous step. The resulting sequence of order signs $\{X_{i_t}^{(t)}\}_{t \geq 1}$ is then a dependent stochastic process with herding.

Proposition 13 (Geometric α -Mixing in the High-Temperature Regime). *Assume the trader network G has uniformly bounded degree and that the couplings satisfy*

$$\max_{(i,j) \in E} |J_{ij}| \leq J_{\max},$$

for some $J_{\max} > 0$. Suppose further that we are in a high-temperature regime with sufficiently weak interactions, in the sense that the total influence of neighbors is uniformly bounded:

$$\sup_{i \in V} \sum_{j \sim i} |J_{ij}| \leq \kappa < \kappa_{\text{crit}},$$

for a constant $\kappa_{\text{crit}} > 0$ small enough, and that the fields $h_i(\theta)$ are uniformly bounded in i and θ .

Then there exists a (unique) stationary Gibbs measure \mathbb{P}_θ for the Ising field and a version of the sequential update dynamics such that the resulting time-indexed process of spins $\{X^{(t)}\}_{t \geq 0}$ is α -mixing with geometric decay. In particular, there exist constants $C > 0$ and $\rho \in (0, 1)$ such that the α -mixing coefficients satisfy

$$\alpha(k) \leq C\rho^k, \quad k \geq 1.$$

Consequently, any bounded functional of the order signs, such as the YES/NO indicators $\mathbb{1}\{X_{i_t}^{(t)} = +1\}$, satisfies a central limit theorem and law of large numbers under the mixing conditions used in Phase 8.

Proof (Sketch). Under the high-temperature (weak-coupling) condition and bounded external fields, the Ising model on a bounded-degree graph admits a unique Gibbs measure with exponential decay of correlations. Standard results for Glauber dynamics on such systems imply that the associated Markov chain is geometrically ergodic and that its time-marginal process is α -mixing with geometric decay. The bounded-degree and small total interaction assumptions ensure a Dobrushin-type contraction condition, which yields exponential forgetting of initial conditions and hence geometric mixing. Boundedness of $h_i(\theta)$ prevents the fields from overwhelming the interaction structure.

Once geometric α -mixing is established for the spin process, the same property holds for any bounded measurable functional of the spins. In particular, if we record the YES/NO sequence as $Y_t := \mathbb{1}\{X_{i_t}^{(t)} = +1\}$, then (Y_t) is also geometrically α -mixing, and the central limit theorems and convergence results used in Phase 8 apply directly to (Y_t) and to aggregates such as the adjusted YES counts y^* and effective sample size n^* . \square

This result shows that the Ising-type herding model is not merely a parametric crutch: in the weak-coupling regime it generates a dependent but geometrically-mixing signal process, which fits within the dependence framework assumed in Phase 8. At the same time, the model remains a stylized representation of reactive trading and does not capture strategic behavior or full information feedback, which are discussed separately under the strategic extensions of Assumption A6.

7.6 CAOPM Mapping Under Simulation

For any fixed simulation scenario, the CAOPM mapping from inputs to outputs is:

$$(\text{structural prior, ML prior}, s_1, \dots, s_N) \mapsto (\alpha_{\text{adj}}, \beta_{\text{adj}}, \pi_{\text{YES}}^{\text{adj}}),$$

where the right-hand side incorporates:

- the hybrid prior from Phase 3;
- the posterior update from Phase 4;
- the bias-correction layer from Phase 6.

Proposition 14 (Continuity of CAOPM Mapping). *For fixed priors and fixed behavioral/structural correction rules, the mapping from trader actions to adjusted posterior mean is continuous:*

$$s_1, \dots, s_N \mapsto \pi_{\text{YES}}^{\text{adj}}.$$

Proof. Follows from the continuity of the Beta posterior mean and linearity of the behavioral and structural adjustments. \square

7.7 Price Calibration and Proper Scoring Rules

Although Phase 7 does not present numerical results, the following performance metrics are prescribed:

- Brier score: $(p_{\text{true}} - \pi_{\text{YES}}^{\text{adj}})^2$;
- Log score: $\log(\pi_{\text{YES}}^{\text{adj}})$ if A occurs and $\log(1 - \pi_{\text{YES}}^{\text{adj}})$ otherwise;
- Mean absolute probability error;

- Calibration error (reliability diagrams);
- Interval coverage rates.

Each metric evaluates how well the CAPOPM posterior represents the underlying truth across simulated environments.

1.8 7.8 Arbitrage-Free Projection for YES/NO Prices

The preceding phases produce a fully corrected CAPOPM posterior $\Pi_\phi(\cdot \mid \mathcal{D})$ over the event probability p , incorporating nonlinear distortions, mixture components, and dynamic nonstationarity. In this subsection we address a remaining structural issue: the raw implied YES/NO prices obtained from Π_ϕ may fail to satisfy basic no-arbitrage constraints

$$\pi_{\text{YES}} + \pi_{\text{NO}} = 1, \quad \pi_{\text{YES}} \geq 0, \pi_{\text{NO}} \geq 0,$$

especially when large nonlinear corrections are applied or when prices are computed from a particle cloud or mixture posterior. We therefore define a final projection step that restores exact no-arbitrage while minimally disturbing the information encoded in the corrected posterior.

Raw Pseudo-Prices from the Corrected Posterior. Let $\Pi_\phi(\cdot \mid \mathcal{D})$ be the nonlinearly adjusted CAPOPM posterior over p as in Section 1.9. Define raw pseudo-prices for YES and NO contracts as expectations of (possibly nonlinear) payoff functionals:

$$v_{\text{YES}}(\mathcal{D}) := \int_0^1 f_{\text{YES}}(p) \Pi_\phi(dp \mid \mathcal{D}), \quad v_{\text{NO}}(\mathcal{D}) := \int_0^1 f_{\text{NO}}(p) \Pi_\phi(dp \mid \mathcal{D}),$$

where $f_{\text{YES}}, f_{\text{NO}} : [0, 1] \rightarrow \mathbb{R}$ are the pricing maps induced by CAPOPM (e.g. $f_{\text{YES}}(p) = p$, $f_{\text{NO}}(p) = 1 - p$, or more general distorted digital payoffs). The pair

$$v(\mathcal{D}) = (v_{\text{YES}}(\mathcal{D}), v_{\text{NO}}(\mathcal{D})) \in \mathbb{R}^2$$

need not satisfy the simplex constraints: components can be slightly negative, and the sum can deviate from 1.

Definition 7 (Arbitrage-Free Simplex for YES/NO Prices). *The arbitrage-free set for YES/NO prices is the one-dimensional probability simplex*

$$\Delta^1 := \{\pi \in [0, 1]^2 : \pi_{\text{YES}} + \pi_{\text{NO}} = 1\}.$$

A vector $\pi \in \Delta^1$ represents a pair of no-arbitrage prices for YES and NO contracts in units of normalized probability (up to discounting).

Projection Operator onto the Simplex. To enforce no-arbitrage while preserving as much information as possible, we define a projection operator from raw pseudo-prices $v(\mathcal{D})$ onto Δ^1 . The construction proceeds in two steps: (i) enforce positivity, (ii) renormalize.

Definition 8 (Positivity Rectification and Normalization). *Fix a small $\varepsilon > 0$. For any raw vector $v = (v_{\text{YES}}, v_{\text{NO}}) \in \mathbb{R}^2$, define*

$$v_{\text{YES}}^+ := \max\{v_{\text{YES}}, \varepsilon\}, \quad v_{\text{NO}}^+ := \max\{v_{\text{NO}}, \varepsilon\}, \quad s(v) := v_{\text{YES}}^+ + v_{\text{NO}}^+.$$

The arbitrage-free projection $\Pi_{\Delta^1}(v)$ is

$$\Pi_{\Delta^1}(v) := \left(\frac{v_{\text{YES}}^+}{s(v)}, \frac{v_{\text{NO}}^+}{s(v)} \right).$$

We call

$$\hat{\pi}(\mathcal{D}) := \Pi_{\Delta^1}(v(\mathcal{D}))$$

the arbitrage-free CAOPM YES/NO price pair.

Lemma 13 (Basic Properties of Π_{Δ^1}). *The projection operator Π_{Δ^1} of Definition 8 satisfies:*

- (i) (No-arbitrage) *For any $v \in \mathbb{R}^2$, $\Pi_{\Delta^1}(v) \in \Delta^1$ and therefore enforces $\hat{\pi}_{\text{YES}} + \hat{\pi}_{\text{NO}} = 1$ and $\hat{\pi}_{\text{YES}}, \hat{\pi}_{\text{NO}} \geq 0$.*
- (ii) (Idempotence on arbitrage-free vectors) *If $v \in \Delta^1$ and $v_{\text{YES}}, v_{\text{NO}} \geq \varepsilon$, then $\Pi_{\Delta^1}(v) = v$.*
- (iii) (Continuity and stability) *On any compact subset of \mathbb{R}^2 where $s(v)$ is bounded away from 0, the map $v \mapsto \Pi_{\Delta^1}(v)$ is Lipschitz continuous.*

Proof. (i) By construction, $\Pi_{\Delta^1}(v)$ has nonnegative components summing to one. (ii) If $v \in \Delta^1$ with components at least ε , then $v^+ = v$ and $s(v) = 1$, giving $\Pi_{\Delta^1}(v) = v$. (iii) On $\{v : s(v) \geq c > 0\}$, the map $v \mapsto v^+$ is 1-Lipschitz and the normalization $v^+/s(v)$ is smooth with bounded derivatives; hence the composition is Lipschitz on such sets. \square

Information Preservation via Bregman Projection. The operation in Definition 8 can be interpreted as a Bregman projection of an unnormalized positive vector onto the probability simplex. Let $v^+ = (v_{\text{YES}}^+, v_{\text{NO}}^+)$ and consider the optimization

$$\hat{\pi} = \arg \min_{\pi \in \Delta^1} D_{\text{KL}}(\pi \| v^+/s(v)),$$

where D_{KL} is Kullback–Leibler divergence and $v^+/s(v)$ is the normalized version of v^+ . The minimizer is $\hat{\pi} = v^+/s(v)$, i.e. $\Pi_{\Delta^1}(v)$. Thus the projection step can be viewed as the minimal adjustment in KL divergence from the normalized positive vector to the set of arbitrage-free pairs, which coincides with a simple renormalization.

Theorem 17 (Arbitrage-Free CAOPM YES/NO Prices). *Let $\Pi_\phi(\cdot \mid \mathcal{D})$ be the fully corrected posterior from Phases 5–6 and $\hat{\pi}(\mathcal{D})$ the arbitrage-free price pair defined in Definition 8. Then:*

- (i) *If the raw pseudo-prices $v(\mathcal{D})$ are already arbitrage-free and strictly positive, $\hat{\pi}(\mathcal{D}) = v(\mathcal{D})$.*
- (ii) *If $v(\mathcal{D})$ deviates from the simplex, the correction $\hat{\pi}(\mathcal{D}) - v(\mathcal{D})$ is the unique (normalized) adjustment that minimizes KL divergence from the rectified vector $v^+(\mathcal{D})/s(v(\mathcal{D}))$ to Δ^1 .*
- (iii) *If $v(\mathcal{D})$ and $v(\mathcal{D}')$ lie in a compact region where $s(v), s(v') \geq c > 0$, then*

$$\|\hat{\pi}(\mathcal{D}) - \hat{\pi}(\mathcal{D}')\|_2 \leq L_{\Delta^1} \|v(\mathcal{D}) - v(\mathcal{D}')\|_2$$

for some constant $L_{\Delta^1} < \infty$, i.e. the projection is Lipschitz with respect to the raw pseudo-prices.

Proof. (i) and (ii) follow from Lemma 13 and the Bregman projection interpretation above. For (iii), Lipschitz continuity on regions where $s(v)$ is bounded away from zero is established in Lemma 13(iii), and the norm inequality follows from the equivalence of norms on \mathbb{R}^2 . \square

Integration with Posterior Robustness. Combining Theorem 31 with Theorem 17 yields a full stability statement for CAOPM YES/NO prices. Small perturbations in the data \mathcal{D} lead to small perturbations in the corrected posterior Π_ϕ in W_1 , which translate into small changes in the raw pseudo-prices $v(\mathcal{D})$ (for Lipschitz payoff maps $f_{\text{YES}}, f_{\text{NO}}$), and finally into small changes in arbitrage-free prices $\hat{\pi}(\mathcal{D})$ after projection. Importantly, when the fully corrected model is already arbitrage-free, the projection step is exactly neutral.

Remark 15 (Scope and Limitations). *The projection in this subsection addresses arbitrage leakage for a single YES/NO pair. In higher-dimensional settings with multiple strikes and maturities, analogues of Π_{Δ^1} can be defined as projections onto the convex set of globally arbitrage-free price surfaces (e.g. enforcing monotonicity and convexity across strikes and maturities). Such projections can be formulated as convex optimization problems minimizing a divergence or norm subject to no-arbitrage constraints. The present construction provides the simplest instance of this idea and ensures that, at a minimum, CAOPM always outputs internally consistent YES/NO prices that respect $\pi_{\text{YES}} + \pi_{\text{NO}} = 1$ and positivity, without discarding the nonlinear and multimodal information accumulated in the posterior.*

7.9 Informal Stress Scenarios

We outline three informal but important stress scenarios.

Scenario 1: Herd Cascade. Large temporal clusters of identical trades overwhelm early signals. The goal is to evaluate whether Stage 1 correction suppresses the cascade.

Scenario 2: Whale Attack. A small number of dominant traders distort the book. The goal is to evaluate whether δ_+ and δ_- can counteract this influence.

Scenario 3: Structural Volatility Shock. Draw p_{true} from a regime where structural model assumptions break down. Tests whether the hybrid prior and bias-correction layer still maintain coherence.

7.10 Herding Regimes, Mixing Failure, and CAPOPM Validity

Phase 7 introduces an Ising-type herding model to describe dependence among trader actions. In weak-coupling regimes, this process is geometrically α -mixing and the asymptotic results of Phase 8 apply. In strong-coupling regimes, the mixing assumptions may fail, and CAPOPM's asymptotic guarantees can break down.

We formalize this with a stylized Ising model on a trader graph $G = (V, E)$.

Theorem 18 (Mixing Regimes for Ising-Type Herding). *Let $G_N = (V_N, E_N)$ be a sequence of trader graphs with $|V_N| = N$. For each trader $i \in V_N$, let $X_i \in \{-1, +1\}$ represent a YES/NO spin, and consider the Ising probability measure*

$$\mathbb{P}_N(X = x) \propto \exp\left(\beta \sum_{\{i,j\} \in E_N} x_i x_j + h \sum_{i \in V_N} x_i\right),$$

with coupling $\beta \geq 0$ and external field $h \in \mathbb{R}$.

- (a) (**Bounded-degree graphs**) If $\sup_N \max_{i \in V_N} \deg(i) \leq d_{\max} < \infty$ and βd_{\max} is sufficiently small, then $\{X_i\}$ is geometrically α -mixing with coefficients decaying as $\alpha(k) \leq C\rho^k$. In this regime, the dependence is weak enough for the LLN/CLT and CAPOPM asymptotics to hold.
- (b) (**Dense mean-field graphs**) If G_N is dense (e.g. complete graph with $|E_N| \asymp N^2$) and $\beta > \beta_c$ for a critical value $\beta_c > 0$, then the Ising model undergoes a phase transition: multiple modes appear and $\alpha(k)$ fails to decay geometrically. In such low-temperature regimes, long-range dependence persists and the mixing assumptions underlying CAPOPM's asymptotics can fail.
- (c) (**Polynomial or non-mixing regimes**) For intermediate cases, $\alpha(k)$ may decay polynomially or not at all. In these regimes, standard CLT-based justifications for the Beta approximation may no longer be valid without additional control, and CAPOPM must be treated as a heuristic exponential-family approximation.

Corollary 2 (Validity Region for CAOPM Under Herding). *Under the conditions of part (a), the Ising herding process generates a geometrically α -mixing sequence of adjusted indicators, and the consistency and Bernstein–von Mises results of Phase 8 apply. Under the strong-coupling regime of part (b), CAOPM may yield posteriors that do not converge to the true event probability and whose variance behavior is not well described by the asymptotic normal approximations. Detecting such regimes and treating the resulting posteriors as exploratory rather than fully calibrated is recommended.*

Remark 16 (Diagnosing Mixing Failure). *Empirically, mixing failure can be probed by examining autocorrelation functions of the adjusted process, block-bootstrap variability, or by fitting simple AR models and testing for long-range dependence. Persistent, slowly decaying correlations suggest that the market is in a strong-herding regime where CAOPM’s formal guarantees are weakened and increased weight should be placed on sensitivity analysis.*

7.11 Summary of Simulation Parameters

A typical simulation configuration includes:

- Distribution for p_{true} (e.g. Beta, Uniform);
- Number of traders N ;
- Trader type proportions: $\pi_{\text{inf}}, \pi_{\text{noise}}, \pi_{\text{adv}}$;
- Behavioral weights w_i^{beh} for Stage 1 correction;
- Structural offsets δ_+, δ_- for Stage 2 correction;
- Hybrid prior parameters (α_0, β_0) ;
- ML prior strength n_{ML} and value p_{ML} ;
- Number of simulation repetitions.

7.12 Output of Phase 7 and Linkage to Phase 8

The outputs of Phase 7 consist of:

- simulated CAOPM prices $\pi_{\text{YES}}^{\text{adj}}$,
- simulated posteriors $(\alpha_{\text{adj}}, \beta_{\text{adj}})$,
- performance metrics from proper scoring rules,
- comparison of adjusted vs. unadjusted posterior,
- sensitivity of CAOPM to biases, distortions, and priors.

These outputs provide the empirical backbone for Phase 8, which establishes theoretical results on consistency, calibration, arbitrage-freeness, and asymptotic stability.

Phase 8. Theoretical Guarantees of CAPOPM

Phase 8 establishes the theoretical foundations of the CAPOPM framework. The goal is to prove that the model preserves arbitrage-freeness, produces consistent and calibrated posteriors under a broad class of conditions, remains robust to behavioral and structural distortions, and admits asymptotic distributional approximations as the effective sample size increases.

The results in this phase pertain to the adjusted posterior derived in Phase 6:

$$p \mid s_{\text{adj}} \sim \text{Beta}(\alpha_{\text{adj}}, \beta_{\text{adj}}), \quad \alpha_{\text{adj}}, \beta_{\text{adj}} > 0,$$

where

$$\alpha_{\text{adj}} = \alpha_0 + y^*, \quad \beta_{\text{adj}} = \beta_0 + (n^* - y^*),$$

and y^*, n^* incorporate behavioral and structural adjustments.

8.1 Assumptions

We introduce the following assumptions, each of which may hold under different simulation or empirical regimes:

- (A1) The true event probability $p_{\text{true}} \in (0, 1)$ is fixed but unknown.
- (A2) Trader signals (s_i) satisfy: informed traders have $s_i \sim \text{Bernoulli}(p_{\text{true}})$, noise traders have $s_i \sim \text{Bernoulli}(1/2)$, adversarial traders invert informed signals.
- (A3) Behavioral weights $w_i^{\text{beh}} > 0$ are bounded above and below:

$$0 < m \leq w_i^{\text{beh}} \leq M < \infty.$$

- (A4) Structural offsets satisfy

$$|\delta_+| + |\delta_-| \leq C_\delta < \infty.$$

- (A5) The hybrid prior parameters $\alpha_0, \beta_0 > 0$ are fixed and finite.
- (A6) The effective adjusted sample size

$$n^* = y^* + (n^* - y^*)$$

satisfies $n^* \rightarrow \infty$.

These assumptions form the basis of the consistency and robustness results below.

8.2 Posterior Consistency

We first establish that CAPOPM yields a consistent posterior for p_{true} under increasingly informative data.

Theorem 19 (Posterior Consistency Under Model Assumptions). *Under assumptions (A1)–(A6),*

$$p \mid s_{\text{adj}} \xrightarrow[n^* \rightarrow \infty]{\mathbb{P}} p_{\text{true}}.$$

Proof. Write the adjusted posterior mean as:

$$\hat{p}_{\text{adj}} = \frac{\alpha_{\text{adj}}}{\alpha_{\text{adj}} + \beta_{\text{adj}}} = \frac{\alpha_0 + y^*}{n^* + \alpha_0 + \beta_0}.$$

We expand the adjusted count:

$$y^* = \sum_{i=1}^n w_i^{\text{beh}} \mathbf{1}\{s_i = \text{YES}\} + \delta_+.$$

By (A3), behavioral weights are bounded, and by the law of large numbers for heterogeneous but bounded weights,

$$\frac{1}{n^*} \sum_{i=1}^n w_i^{\text{beh}} \mathbf{1}\{s_i = \text{YES}\} \xrightarrow{\mathbb{P}} p_{\text{true}} \cdot \mathbb{E}[w_i^{\text{beh}}].$$

Since the denominator also grows with n^* ,

$$\frac{y^*}{n^*} \xrightarrow{\mathbb{P}} p_{\text{true}} \cdot \frac{\mathbb{E}[w_i^{\text{beh}}]}{\mathbb{E}[w_i^{\text{beh}}]} = p_{\text{true}}.$$

Offset δ_+ satisfies

$$\frac{\delta_+}{n^*} \rightarrow 0.$$

Thus

$$\hat{p}_{\text{adj}} = \frac{y^* + O(1)}{n^* + O(1)} \xrightarrow{\mathbb{P}} p_{\text{true}}.$$

Hence the posterior is consistent. \square

Remark 17 (Role of Large Effective Sample Size n^*). *The consistency result above is driven primarily by the growth of the effective sample size n^* , rather than by the raw number of trades n alone. Recall that*

$$n^* = y^* + (n^* - y^*) = \sum_{i=1}^n w_i^{\text{beh}} (\mathbf{1}\{s_i = \text{YES}\} + \mathbf{1}\{s_i = \text{NO}\}) + (\delta_+ + \delta_-),$$

so that n^ reflects both behavioral reweighting and structural offsets.*

Under assumptions (A2)–(A4), a nonzero fraction of traders are informed or adversarial, and the weights w_i^{beh} remain bounded above and below. Consequently, as the number of traders n grows, the effective sample size n^* also grows linearly:

$$\frac{n^*}{n} \rightarrow c \in (0, \infty),$$

up to $O(1)$ contributions from (δ_+, δ_-) . The law of large numbers and central limit behavior therefore apply at the level of n^* , not just n . When an adversarial fraction $\pi_{\text{adv}} < 1/2$ is present, the net signal embedded in y^* remains aligned with p_{true} and the posterior still concentrates at the true value as $n^* \rightarrow \infty$.

This perspective clarifies that CAPOPM’s asymptotic properties are governed by the information content of the adjusted sample size, which is resilient to moderate behavioral distortion and bounded structural offsets, rather than by the raw trade count alone.

8.3 Consistency Under Structural Prior Correctness

Proposition 15 (Structural Prior Dominance). *If the structural prior mean equals the true probability, i.e. $q_{\text{str}} = p_{\text{true}}$, and if the ML prior is weak (n_{ML} small), then*

$$p \mid s_{\text{adj}} \xrightarrow{\mathbb{P}} p_{\text{true}}.$$

Proof. Follows from the posterior consistency theorem with $\alpha_0/n^* \rightarrow 0$. \square

8.4 Consistency Under ML Prior Correctness

Proposition 16 (ML Prior Dominance). *If the ML prior satisfies $p_{\text{ML}} = p_{\text{true}}$ and the ML strength satisfies $n_{\text{ML}} \rightarrow \infty$, then*

$$\hat{p}_{\text{adj}} \rightarrow p_{\text{true}}.$$

Proof. As $n_{\text{ML}} \rightarrow \infty$, the hybrid prior mean approaches $p_{\text{ML}} = p_{\text{true}}$, and the structural contribution vanishes. \square

8.5 Consistency Under Behavioral Bias Correction

Theorem 20 (Consistency Under Behavioral Distortion). *If (A3) holds and the proportion of informed traders is nonzero, then CAPOPM remains consistent after Stage 1 correction.*

Proof. Stage 1 weights are bounded and therefore do not distort the sign or limit of the empirical frequencies. \square

8.6 Consistency Under Adversarial Contamination

Theorem 21 (Adversarial Robustness). *If the fraction of adversarial traders satisfies $\pi_{\text{adv}} < 1/2$, then CAPOPM remains posterior consistent.*

Proof. Under contamination theory, if adversarial contamination is below 50%, the majority signal still reflects p_{true} . Weighted counts remain asymptotically aligned with the truth after normalization. \square

8.7 Error Propagation from Weight and Offset Estimation

Let $(\psi^\dagger, \delta^\dagger)$ denote the pseudo-true parameters appearing in the Empirical Bayes limit, and let

$$y^* = y^*(\psi^\dagger, \delta^\dagger), \quad n^* = n^*(\psi^\dagger, \delta^\dagger)$$

denote the corresponding adjusted counts for a given market. The idealized CAPOPM posterior for p is then

$$p \mid \mathcal{I}, \psi^\dagger, \delta^\dagger \sim \text{Beta}(\alpha_0 + y^*, \beta_0 + n^* - y^*).$$

In practice, we use the estimated parameters $(\hat{\psi}, \hat{\delta})$ (obtained from a finite historical sample) and the associated adjusted counts

$$\hat{y}^* = y^*(\hat{\psi}, \hat{\delta}), \quad \hat{n}^* = n^*(\hat{\psi}, \hat{\delta}),$$

leading to the approximate posterior

$$p \mid \mathcal{I}, \hat{\psi}, \hat{\delta} \sim \text{Beta}(\alpha_0 + \hat{y}^*, \beta_0 + \hat{n}^* - \hat{y}^*).$$

We now provide bounds on the deviation between these two posteriors as a function of the error in $(\hat{\psi}, \hat{\delta})$.

Theorem 22 (Lipschitz Error Propagation for CAPOPM Posterior). *Assume:*

- (i) *The weighting function $w(x; \psi)$ is Lipschitz in ψ , uniformly in x , i.e. there exists $L_w > 0$ such that*

$$|w(x; \psi_1) - w(x; \psi_2)| \leq L_w \|\psi_1 - \psi_2\| \quad \text{for all } x, \psi_1, \psi_2.$$

- (ii) *The offsets δ_+, δ_- enter linearly and the map $\delta \mapsto (y^*(\psi, \delta), n^*(\psi, \delta))$ is Lipschitz with constant L_δ .*

- (iii) *The parameter space $\Psi \times \Delta$ is compact and $\alpha_0, \beta_0 > 0$ are fixed.*

Then the following hold for a fixed market:

- (a) **(Mean and Variance)** *Let μ^\dagger and $\hat{\mu}$ denote the posterior means, and $\sigma^{2,\dagger}$ and $\hat{\sigma}^2$ the posterior variances, under $(\psi^\dagger, \delta^\dagger)$ and $(\hat{\psi}, \hat{\delta})$ respectively. Then there exist constants $C_1, C_2 > 0$ such that*

$$|\hat{\mu} - \mu^\dagger| \leq C_1 (|\hat{y}^* - y^*| + |\hat{n}^* - n^*|),$$

$$|\hat{\sigma}^2 - \sigma^{2,\dagger}| \leq C_2 (|\hat{y}^* - y^*| + |\hat{n}^* - n^*|).$$

- (b) (**Posterior Distribution**) Let Π^\dagger and $\hat{\Pi}$ denote the two Beta posteriors. Then there exists $C_3 > 0$ such that both the total variation distance and the squared Hellinger distance satisfy

$$\|\hat{\Pi} - \Pi^\dagger\|_{\text{TV}} \leq C_3 (|\hat{y}^* - y^*| + |\hat{n}^* - n^*|),$$

$$H^2(\hat{\Pi}, \Pi^\dagger) \leq C_3 (|\hat{y}^* - y^*| + |\hat{n}^* - n^*|),$$

where H denotes the Hellinger distance.

- (c) (**Parameter Error to Posterior Error**) Under (i)–(ii), there exists a constant $C_4 > 0$ such that

$$|\hat{y}^* - y^*| + |\hat{n}^* - n^*| \leq C_4 (\|\hat{\psi} - \psi^\dagger\| + \|\hat{\delta} - \delta^\dagger\|).$$

Combining with (a)–(b) yields Lipschitz-type bounds of posterior mean, variance, and distributional distance in terms of the parameter estimation error.

Proof (Sketch). For (a), the Beta posterior mean and variance are smooth functions of (α, β) given by

$$\mu(\alpha, \beta) = \frac{\alpha}{\alpha + \beta}, \quad \sigma^2(\alpha, \beta) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

On any compact subset with $\alpha, \beta \geq c > 0$, these maps are Lipschitz in (α, β) , hence in (y^*, n^*) since $\alpha = \alpha_0 + y^*$ and $\beta = \beta_0 + n^* - y^*$ are affine functions of (y^*, n^*) . This yields the stated bounds with constants C_1, C_2 depending on the compact region.

For (b), standard perturbation bounds for one-parameter exponential families imply that the total variation and Hellinger distances between $\text{Beta}(\alpha_1, \beta_1)$ and $\text{Beta}(\alpha_2, \beta_2)$ are Lipschitz in $(\alpha_1 - \alpha_2, \beta_1 - \beta_2)$ on compact sets with $\alpha_j, \beta_j \geq c > 0$. Again using the affine dependence of (α, β) on (y^*, n^*) yields the displayed inequalities.

For (c), Lipschitz continuity of $w(x; \psi)$ in ψ and of the linear map $\delta \mapsto (y^*, n^*)$, together with boundedness of the feature set, gives

$$|\hat{y}^* - y^*| + |\hat{n}^* - n^*| \leq C_4 (\|\hat{\psi} - \psi^\dagger\| + \|\hat{\delta} - \delta^\dagger\|)$$

for some C_4 depending on the feature bounds and Lipschitz constants. Combining with (a)–(b) gives the claimed error propagation bounds. \square

Remark 18 (Interpretation). *This theorem formalizes the intuitive idea that small errors in the estimated weights and offsets lead to small distortions in the CAOPM posterior. As the Empirical Bayes estimators $(\hat{\psi}_M, \hat{\delta}_M)$ converge to their pseudo-true values $(\psi^\dagger, \delta^\dagger)$ with increasing historical sample size M , the resulting posterior mean, variance, and full distribution converge to those obtained under the pseudo-true parameters, at a rate controlled by the Lipschitz constants above.*

8.8 Consistency Under Structural Distortion Offsets

Theorem 23 (Offset Robustness). *If $|\delta_+| + |\delta_-| < C_\delta < \infty$ and $n^* \rightarrow \infty$, then*

$$\frac{\delta_+}{n^*} \rightarrow 0, \quad \frac{\delta_-}{n^*} \rightarrow 0,$$

and CAPOPM remains consistent.

Proof. Offsets are $O(1)$ and thus negligible relative to n^* . \square

8.9 Robustness to Perturbations (Stability)

Proposition 17 (Lipschitz Stability of Posterior Mean). *For the adjusted posterior mean*

$$\hat{p}_{\text{adj}} = \frac{\alpha_{\text{adj}}}{\alpha_{\text{adj}} + \beta_{\text{adj}}},$$

there exists $L > 0$ such that for any perturbations $\Delta y^, \Delta(n^* - y^*)$,*

$$|\Delta \hat{p}_{\text{adj}}| \leq L (|\Delta y^*| + |\Delta(n^* - y^*)|).$$

Proof. Same structure as Phase 6 robustness theorem; the posterior mean is a smooth rational function on a compact domain. \square

8.10 Arbitrage-Freeness

Theorem 24 (Arbitrage-Freeness of CAPOPM Pricing). *For the adjusted prices*

$$\pi_{\text{YES}}^{\text{adj}} = \hat{p}_{\text{adj}}, \quad \pi_{\text{NO}}^{\text{adj}} = 1 - \hat{p}_{\text{adj}},$$

the following hold:

- (i) $\pi_{\text{YES}}^{\text{adj}} + \pi_{\text{NO}}^{\text{adj}} = 1$,
- (ii) $0 < \pi_{\text{YES}}^{\text{adj}} < 1$,
- (iii) $\pi_{\text{YES}}^{\text{adj}}$ is monotone in adjusted counts y^* ,
- (iv) boundedness is preserved under all admissible distortions.

Proof. (i) Follows immediately from the definition. (ii) Holds because $\alpha_{\text{adj}}, \beta_{\text{adj}} > 0$. (iii) Derivative of Beta mean with respect to y^* is positive. (iv) Adjustments enter linearly; Beta parameters remain positive. \square

8.11 Asymptotic Distribution (Bernstein–von Mises)

Theorem 25 (Asymptotic Normality via CLT and Bernstein–von Mises). *Suppose (A1)–(A6) hold, and in addition:*

- *the sequence of adjusted indicators contributing to y^* satisfies a Lindeberg–Feller type condition, and*
- *the fraction of adversarial traders satisfies $\pi_{\text{adv}} < 1/2$, so that the net signal remains aligned with p_{true} .*

Let

$$\hat{p}^* := \frac{y^*}{n^*}$$

denote the adjusted sample proportion, where n^* is the effective sample size resulting from behavioral weights and structural offsets. Then:

- (i) (**CLT for the Adjusted Proportion**) *There exists $\sigma^2 \in (0, \infty)$ such that*

$$\sqrt{n^*} (\hat{p}^* - p_{\text{true}}) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad \text{as } n^* \rightarrow \infty.$$

- (ii) (**Asymptotic Normality of the Posterior Mean**) *For the adjusted posterior*

$$p \mid s_{\text{adj}} \sim \text{Beta}(\alpha_{\text{adj}}, \beta_{\text{adj}}), \quad \alpha_{\text{adj}} = \alpha_0 + y^*, \quad \beta_{\text{adj}} = \beta_0 + (n^* - y^*),$$

we have

$$\sqrt{n^*} (\hat{p}_{\text{adj}} - p_{\text{true}}) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\hat{p}_{\text{adj}} = \alpha_{\text{adj}} / (\alpha_{\text{adj}} + \beta_{\text{adj}})$ is the posterior mean.

- (iii) (**Bernstein–von Mises Approximation**) *The full posterior distribution satisfies*

$$p \mid s_{\text{adj}} \stackrel{d}{\approx} \mathcal{N}\left(p_{\text{true}}, \frac{\sigma^2}{n^*}\right) \quad \text{for large } n^*,$$

i.e., the posterior is asymptotically normal with center p_{true} and variance of order $1/n^*$.

Proof. (i) *CLT for the adjusted proportion.* Write

$$y^* = \sum_{i=1}^n W_i,$$

where each W_i is the contribution of trader i to the adjusted YES count, including behavioral weights and the effect of trader type (informed, noise, adversarial). Assumptions (A2)–(A4) imply that the W_i are uniformly bounded and that

the mean of W_i is aligned with p_{true} as $\pi_{\text{adv}} < 1/2$. Under the Lindeberg–Feller condition for the triangular array $\{W_i\}_{i=1}^n$, we obtain

$$\sqrt{n^*}(\hat{p}^* - p_{\text{true}}) = \sqrt{n^*}\left(\frac{y^*}{n^*} - p_{\text{true}}\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

for some finite, positive σ^2 capturing the effective dispersion of the weighted trader signals.

(ii) *Asymptotic normality of the posterior mean.* The posterior mean can be written as

$$\hat{p}_{\text{adj}} = \frac{\alpha_0 + y^*}{\alpha_0 + \beta_0 + n^*} = \hat{p}^* + \frac{\alpha_0 - \hat{p}^*(\alpha_0 + \beta_0)}{\alpha_0 + \beta_0 + n^*}.$$

The second term is $O(1/n^*)$ in probability and therefore negligible at the $1/\sqrt{n^*}$ scale. Thus

$$\sqrt{n^*}(\hat{p}_{\text{adj}} - p_{\text{true}}) = \sqrt{n^*}(\hat{p}^* - p_{\text{true}}) + o_{\mathbb{P}}(1),$$

and the CLT from part (i) implies convergence in distribution to $\mathcal{N}(0, \sigma^2)$.

(iii) *Bernstein–von Mises approximation.* The Beta posterior with parameters $(\alpha_{\text{adj}}, \beta_{\text{adj}})$ has mean \hat{p}_{adj} and variance

$$\text{Var}(p \mid s_{\text{adj}}) = \frac{\alpha_{\text{adj}}\beta_{\text{adj}}}{(\alpha_{\text{adj}} + \beta_{\text{adj}})^2(\alpha_{\text{adj}} + \beta_{\text{adj}} + 1)} \approx \frac{\hat{p}_{\text{adj}}(1 - \hat{p}_{\text{adj}})}{n^*},$$

for large n^* .

By standard Bernstein–von Mises arguments for one-dimensional conjugate models, the posterior distribution of p becomes asymptotically normal with this mean and variance, and the difference between the posterior law and the corresponding normal distribution vanishes in total variation. Substituting $\hat{p}_{\text{adj}} \rightarrow p_{\text{true}}$ yields the stated normal approximation centered at p_{true} with asymptotic variance of order $1/n^*$. \square

8.12 Finite-Sample Concentration and Credible-Interval Corrections

The asymptotic results developed earlier in this phase (law of large numbers, central limit theorems, and Bernstein–von Mises theorems) require an effective sample size n^* that is sufficiently large. In low-liquidity markets or short order windows, these asymptotic approximations can be misleading: the posterior may concentrate slowly, and credible intervals derived from Gaussian limits may substantially misstate uncertainty. This subsection provides finite-sample corrections via exponential concentration inequalities, exact finite- n credible intervals, and Berry–Esseen-type bounds for the posterior approximation.

Throughout this subsection, we assume that the likelihood and posterior family are correctly specified in the sense of Sections 1.7, 1.7, and 1.7: nonlinear distortions, dependence, and multimodality are explicitly modeled via regime mixtures and dynamic bias layers. The bounds below apply to a single market or time block where the true event probability p_{true} is well defined.

Setup for a Single Market Block. Consider a single market with n effective observations $Z_1, \dots, Z_n \in \{0, 1\}$ (YES indicators after Stage 1 weighting and Stage 2 structural corrections), generated conditionally i.i.d. given a fixed $p_{\text{true}} \in (0, 1)$:

$$Z_i \mid p_{\text{true}} \sim \text{Ber}(p_{\text{true}}), \quad i = 1, \dots, n.$$

Let $Y = \sum_{i=1}^n Z_i$ and $\bar{Z} = Y/n$ be the empirical YES count and sample mean. For a Beta prior $\text{Beta}(\alpha_0, \beta_0)$ on p , the posterior is $\text{Beta}(\alpha, \beta)$ with

$$\alpha = \alpha_0 + Y, \quad \beta = \beta_0 + n - Y,$$

and posterior mean

$$\hat{p}_{\text{post}} = \frac{\alpha}{\alpha + \beta} = \lambda p_0 + (1 - \lambda) \bar{Z},$$

where

$$p_0 = \frac{\alpha_0}{\alpha_0 + \beta_0}, \quad \lambda = \frac{\alpha_0 + \beta_0}{\alpha_0 + \beta_0 + n}.$$

Assumption 5 (Moderate Dependence via Effective Sample Size). *In the presence of weak dependence or regime switching (Sections 1.7 and 1.9), assume that there exists an effective sample size n^* satisfying*

$$n^* \leq n, \quad \mathbb{P}\left(|\bar{Z} - p_{\text{true}}| > \varepsilon\right) \leq 2 \exp(-2n^*\varepsilon^2)$$

for all $\varepsilon > 0$. In the i.i.d. case, $n^* = n$; under α -mixing or HMM dependence, n^* incorporates the effective number of independent observations (e.g. via standard blocking arguments).

Theorem 26 (Finite-Sample Concentration for the Posterior Mean). *Under the single-market block model and Assumption 5, fix $\varepsilon > 0$ and define*

$$\varepsilon_0 := \lambda|p_0 - p_{\text{true}}| \leq \lambda, \quad \lambda = \frac{\alpha_0 + \beta_0}{\alpha_0 + \beta_0 + n}.$$

Then for any $\varepsilon > \varepsilon_0$,

$$\mathbb{P}\left(|\hat{p}_{\text{post}} - p_{\text{true}}| > \varepsilon\right) \leq 2 \exp\left(-2n^* \left(\frac{\varepsilon - \varepsilon_0}{1 - \lambda}\right)^2\right).$$

In particular, for sufficiently large n (so that λ and ε_0 are small), the posterior mean concentrates around p_{true} at a sub-Gaussian rate with effective sample size n^* .

Proof. Using the convex combination representation,

$$\hat{p}_{\text{post}} - p_{\text{true}} = \lambda(p_0 - p_{\text{true}}) + (1 - \lambda)(\bar{Z} - p_{\text{true}}),$$

we obtain

$$|\hat{p}_{\text{post}} - p_{\text{true}}| \leq \lambda|p_0 - p_{\text{true}}| + (1 - \lambda)|\bar{Z} - p_{\text{true}}| \leq \varepsilon_0 + (1 - \lambda)|\bar{Z} - p_{\text{true}}|.$$

Thus, if $|\hat{p}_{\text{post}} - p_{\text{true}}| > \varepsilon$ with $\varepsilon > \varepsilon_0$, then necessarily

$$|\bar{Z} - p_{\text{true}}| > \frac{\varepsilon - \varepsilon_0}{1 - \lambda}.$$

Applying Assumption 5 gives

$$\mathbb{P}\left(|\hat{p}_{\text{post}} - p_{\text{true}}| > \varepsilon\right) \leq \mathbb{P}\left(|\bar{Z} - p_{\text{true}}| > \frac{\varepsilon - \varepsilon_0}{1 - \lambda}\right) \leq 2 \exp\left(-2n^* \left(\frac{\varepsilon - \varepsilon_0}{1 - \lambda}\right)^2\right),$$

as claimed. \square

Remark 19 (Interpretation and Choice of n^*). *The bound in Theorem 26 decouples the finite-sample error into a prior-bias term ε_0 (which vanishes as n dominates $\alpha_0 + \beta_0$) and a stochastic term controlled by n^* . In low-liquidity markets, both terms may be nonnegligible; CAOPM should therefore explicitly report the implied error scale*

$$\varepsilon_{\text{tol}} \approx \varepsilon_0 + (1 - \lambda) \sqrt{\frac{\log(2/\delta)}{2n^*}}$$

for a target tail probability δ (e.g. $\delta = 0.05$).

Finite- n Credible Intervals via Beta Quantiles. For the Beta posterior $\text{Beta}(\alpha, \beta)$, an exact $(1 - \gamma)$ -credible interval for p is given by

$$[p_{\gamma/2}^{\text{low}}, p_{1-\gamma/2}^{\text{high}}] := [F_{\text{Beta}(\alpha, \beta)}^{-1}(\gamma/2), F_{\text{Beta}(\alpha, \beta)}^{-1}(1 - \gamma/2)],$$

where $F_{\text{Beta}(\alpha, \beta)}^{-1}$ denotes the inverse incomplete beta function. In the mixture case of Section 1.7, credible intervals can be computed by numerical inversion of the mixture CDF or via Monte Carlo sampling from Π_{mix} , yielding empirical quantiles that preserve multimodality.

Theorem 27 (Berry–Esseen–Type Bound for the Beta Posterior). *Under the i.i.d. single-market block model with $p_{\text{true}} \in (0, 1)$ fixed and α_0, β_0 bounded, let $\Pi_T(\cdot)$ denote the posterior distribution of p given Z_1, \dots, Z_n , and let $\sigma_T^2 = p_{\text{true}}(1 - p_{\text{true}})/(n + \alpha_0 + \beta_0)$. Then there exists a universal constant $C > 0$ such that*

$$\sup_{x \in \mathbb{R}} |\Pi_T(p \leq p_{\text{true}} + x\sigma_T) - \Phi(x)| \leq \frac{C}{\sqrt{n^*}},$$

where Φ is the standard normal CDF and n^* is the effective sample size of Assumption 5. In particular, Gaussian credible intervals centered at \hat{p}_{post} with radius $z_{1-\gamma/2}\sigma_T$ incur a finite- n approximation error of order $O(1/\sqrt{n^*})$.

Remark 20 (Mixture Extension and Bootstrap Refinements). *For mixture posteriors Π_{mix} as in Section 1.7, the posterior mean \hat{p}_{mix} is a mixture of component means and satisfies a bound of the form*

$$\mathbb{P}\left(|\hat{p}_{\text{mix}} - p_{\text{true}}| > \varepsilon\right) \leq \sum_{k=1}^K w_k^* \mathbb{P}\left(|\hat{p}_k - p_{\text{true}}| > \varepsilon\right),$$

where \hat{p}_k is the posterior mean under component k . Combining this with Theorem 26 for each component yields a mixture concentration bound. In practice, CAPOPM can supplement analytic bounds with bootstrap resampling: resample the observed orders $\{Z_i\}_{i=1}^n$ B times, recompute \hat{p}_{post} or \hat{p}_{mix} for each bootstrap replicate, and use the empirical quantiles of $\{\hat{p}^{(b)}\}_{b=1}^B$ to form finite- n uncertainty bands. Such bootstrap intervals can be compared to the Beta-based credible intervals to diagnose small-sample distortions and coverage properties.

Remark 21 (Implications for Phase 7 Simulation Metrics). *In Phase 7, finite-sample properties of CAPOPM are evaluated via simulation under low-liquidity regimes (e.g. $n < 50$). The results of this subsection provide target coverage rates and finite- n error scales for:*

- the absolute error $|\hat{p}_{\text{post}} - p_{\text{true}}|$ or $|\hat{p}_{\text{mix}} - p_{\text{true}}|$,
- the empirical coverage of nominal $(1 - \gamma)$ credible intervals,
- and mispricing of YES/NO digital contracts, $|\hat{p}_{(\text{post}/\text{mix})} - p_{\text{true}}|$.

Simulation designs should explicitly report both n and n^* to explain deviations from asymptotic behavior, and use the bounds in Theorem 26 and Theorem 27 as benchmarks for finite-sample reliability.

1.9 8.13 Robustness and Divergence Bounds for Mixture Posteriors

The mixture posterior Π_{mix} of Definition 2 extends the single-Beta CAPOPM posterior to multimodal settings. This subsection generalizes the divergence bounds of Theorem 23 to mixture posteriors and provides conditions under which the mixture remains stable to perturbations, along with explicit lower bounds showing when single-Beta approximations necessarily fail.

Assumption 6 (Mixture Stability Under Perturbations). *Let $\Pi_{\text{mix}} = \sum_{k=1}^K w_k^* \Pi_k$ and $\tilde{\Pi}_{\text{mix}} = \sum_{k=1}^K \tilde{w}_k^* \tilde{\Pi}_k$ be two mixture posteriors, where each Π_k and $\tilde{\Pi}_k$ is $\text{Beta}(\alpha_k, \beta_k)$ (possibly with different parameters). Assume:*

(i) *Component Lipschitz continuity:*

$$\left\| \Pi_k - \tilde{\Pi}_k \right\|_{\text{TV}} \leq L_k \left\| \mathcal{D} - \tilde{\mathcal{D}} \right\|;$$

(ii) *Weight stability:*

$$\|w^* - \tilde{w}^*\|_1 \leq C_w \left\| \mathcal{D}^{\text{hold}} - \tilde{\mathcal{D}}^{\text{hold}} \right\|;$$

(iii) *Bounded support separation: for all k , Π_k and $\tilde{\Pi}_k$ have support in $[0, 1]$ with finite moments.*

Theorem 28 (Mixture Posterior Robustness Bound). *Under Assumption 6, the mixture posterior satisfies*

$$\left\| \Pi_{\text{mix}} - \tilde{\Pi}_{\text{mix}} \right\|_{\text{TV}} \leq \sum_{k=1}^K w_k^* L_k \left\| \mathcal{D} - \tilde{\mathcal{D}} \right\| + C_w \max_k \left\| \Pi_k - \tilde{\Pi}_k \right\|_{\text{TV}}.$$

Hence the mixture posterior inherits robustness from: (i) the individual component Betas, and (ii) the stability of the stacking weights.

Proof. Write

$$\Pi_{\text{mix}} - \tilde{\Pi}_{\text{mix}} = \sum_{k=1}^K (w_k^* - \tilde{w}_k^*) \tilde{\Pi}_k + \sum_{k=1}^K w_k^* (\Pi_k - \tilde{\Pi}_k).$$

Taking total variation norms and applying the triangle inequality gives

$$\left\| \Pi_{\text{mix}} - \tilde{\Pi}_{\text{mix}} \right\|_{\text{TV}} \leq \sum_{k=1}^K |w_k^* - \tilde{w}_k^*| \left\| \tilde{\Pi}_k \right\|_{\text{TV}} + \sum_{k=1}^K w_k^* \left\| \Pi_k - \tilde{\Pi}_k \right\|_{\text{TV}}.$$

The first term is bounded using $\left\| \tilde{\Pi}_k \right\|_{\text{TV}} = 1$ and Assumption 6(ii); the second term uses Assumption 6(i). Combine to obtain the stated bound. \square

Theorem 29 (Lower Bound: Mixture Divergence Cannot Vanish Under Mode Separation). *Let Π_{mix} be as above, and let Π_{Beta} be the moment-matched single-Beta approximation from Definition 3. If at least two components have means separated by $|\mu_i - \mu_j| \geq \varepsilon > 0$, then there exists $c(\varepsilon) > 0$ such that*

$$\left\| \Pi_{\text{mix}} - \Pi_{\text{Beta}} \right\|_{\text{TV}} \geq c(\varepsilon).$$

Thus no unimodal Beta projection can approximate the mixture posterior uniformly when the mixture is sufficiently multimodal.

Remark 22 (Extension of Theorem 23). *Theorem 28 extends the divergence and robustness results of Theorem 23 to multimodal settings. Theorem 29 adds a complementary impossibility dimension: mixture posteriors retain irreducible multimodality, meaning that Beta projections cannot achieve vanishing approximation error when component means are well separated. These bounds guide when CAPOPM should operate with the full mixture posterior, and when approximation layers (e.g. Beta moment matching) incur unavoidable and quantifiable loss.*

8.14 Weak Dependence, α -Mixing, and Asymptotic Normality

In earlier phases, we modeled trader actions as if they were conditionally independent Bernoulli signals given the true event probability p_{true} . In practice, herding, imitation, and correlated information can induce dependence across

orders. Phase 7 introduced stylized dependence regimes, including Ising-type herding models that give rise to geometrically α -mixing sequences.

In this subsection, we impose a weak dependence condition and show that the key asymptotic properties of CAPOPM—consistency and asymptotic normality of the posterior—continue to hold. The dependence enters through the adjusted YES indicators that feed into the Beta-Binomial update.

Adjusted YES process. Let $\{Z_t\}_{t \geq 1}$ denote the sequence of adjusted YES indicators at the level of individual orders or time steps, where

$$Z_t \in [0, 1],$$

represents the effective contribution of order t to the YES count after behavioral weighting. For example, in a simple specification,

$$Z_t = w^{\text{beh}}(x_t; \psi^\dagger) \mathbb{1}\{s_t = \text{YES}\},$$

with $w^{\text{beh}}(x_t; \psi^\dagger) \in [0, 1]$ a bounded weight depending on features x_t and pseudo-true parameter ψ^\dagger .

For a given market with n observed orders, the adjusted YES count and effective sample size are

$$y_n^* = \sum_{t=1}^n Z_t + \delta_+^\dagger, \quad n_n^* = \sum_{t=1}^n w^{\text{beh}}(x_t; \psi^\dagger) + \delta_+^\dagger + \delta_-^\dagger,$$

and the Beta-Binomial update uses (y_n^*, n_n^*) together with the hybrid prior. We assume (Z_t) is generated under a fixed true event probability $p_{\text{true}} \in (0, 1)$.

α -mixing assumptions. Let $\mathcal{F}_a^b = \sigma(Z_a, \dots, Z_b)$ be the sigma-algebra generated by the process between times a and b , and define the α -mixing coefficients

$$\alpha(k) = \sup_{t \geq 1} \sup_{A \in \mathcal{F}_1^t, B \in \mathcal{F}_{t+k}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

We assume:

(A7) (**Boundedness**) The adjusted indicators are uniformly bounded: $0 \leq Z_t \leq 1$.

(A8) (**Geometric α -mixing**) There exist constants $C > 0$ and $\rho \in (0, 1)$ such that

$$\alpha(k) \leq C\rho^k, \quad k \geq 1.$$

(A9) (**Stationarity and Identifiability**) The process (Z_t) is strictly stationary under p_{true} , with

$$\mathbb{E}[Z_t] = \mu(p_{\text{true}}), \quad \text{Var}(Z_t) = \sigma_Z^2(p_{\text{true}}),$$

where $\mu(p)$ is strictly increasing in p on $(0, 1)$.

Assumption (A8) is satisfied, for example, by the Ising-type herding model in Phase 7 when couplings are sufficiently weak (Proposition 7.X).

Lemma 14 (LLN and CLT for the Adjusted YES Process). *Under (A7)–(A9), define the normalized partial sums*

$$\bar{Z}_n = \frac{1}{n} \sum_{t=1}^n Z_t.$$

Then:

(a) (**Law of Large Numbers**) $\bar{Z}_n \rightarrow \mu(p_{\text{true}})$ *almost surely and in L^1 as $n \rightarrow \infty$.*

(b) (**Central Limit Theorem**) *There exists $\tau^2(p_{\text{true}}) \in (0, \infty)$ such that*

$$\sqrt{n}(\bar{Z}_n - \mu(p_{\text{true}})) \xrightarrow{d} \mathcal{N}(0, \tau^2(p_{\text{true}})) \quad \text{as } n \rightarrow \infty.$$

Proof (Sketch). The uniform boundedness in (A7) and geometric α -mixing in (A8) imply that (Z_t) satisfies the conditions of classical LLN and CLT results for strongly mixing sequences. Stationarity and finite variance in (A9) ensure that $\mu(p_{\text{true}})$ and $\sigma_Z^2(p_{\text{true}})$ are well defined, and the asymptotic variance $\tau^2(p_{\text{true}})$ can be expressed as a sum of autocovariances. Standard references for mixing CLTs apply directly under the geometric decay of $\alpha(k)$. \square

We now translate this into the asymptotic behavior of the CAPOPM posterior.

Theorem 30 (Consistency and Asymptotic Normality of CAPOPM Posterior Under Dependence). *Under (A7)–(A9), suppose that for a sequence of markets with n orders we use the adjusted counts*

$$y_n^* = \sum_{t=1}^n Z_t + \delta_+^\dagger, \quad n_n^* = \sum_{t=1}^n w^{\text{beh}}(x_t; \psi^\dagger) + \delta_+^\dagger + \delta_-^\dagger,$$

with $w^{\text{beh}}(x_t; \psi^\dagger) \in [c_w, 1]$ for some $c_w > 0$. Let p denote the event probability and assume a Beta hybrid prior

$$p \sim \text{Beta}(\alpha_0, \beta_0), \quad \alpha_0, \beta_0 > 0,$$

independent of (Z_t) . Then:

(a) (**Posterior Consistency**) *Let $\Pi_n(\cdot)$ denote the CAPOPM posterior for p based on (y_n^*, n_n^*) . For any $\epsilon > 0$,*

$$\Pi_n(|p - p_{\text{true}}| > \epsilon) \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty.$$

- (b) (**Asymptotic Normality of Posterior Mean**) Let \hat{p}_n be the posterior mean under Π_n . Then

$$\sqrt{n}(\hat{p}_n - p_{\text{true}}) \xrightarrow{d} \mathcal{N}(0, V(p_{\text{true}})),$$

for some finite $V(p_{\text{true}})$ determined by $\tau^2(p_{\text{true}})$ and the weight structure.

- (c) (**Bernstein–von Mises Approximation**) The posterior distribution Π_n is asymptotically normal in the sense that

$$\sup_{A \subset \mathbb{R}} \left| \Pi_n(\sqrt{n}(p - p_{\text{true}}) \in A) - \mathcal{N}(0, V(p_{\text{true}}))(A) \right| \xrightarrow{P} 0.$$

Proof (Sketch). The effective sample size n_n^* grows linearly with n due to the lower bound $w^{\text{beh}}(x_t; \psi^\dagger) \geq c_w > 0$. The adjusted mean

$$\bar{Z}_n^* = \frac{y_n^* - \delta_+^\dagger}{n_n^* - \delta_+^\dagger - \delta_-^\dagger}$$

is a smooth function of \bar{Z}_n and the average weight, so the LLN and CLT from the lemma transfer to \bar{Z}_n^* via the delta method. In particular, $\bar{Z}_n^* \rightarrow \mu(p_{\text{true}})$ and

$$\sqrt{n}(\bar{Z}_n^* - \mu(p_{\text{true}})) \xrightarrow{d} \mathcal{N}(0, \tilde{\tau}^2(p_{\text{true}})).$$

The Beta posterior based on (y_n^*, n_n^*) has mean and variance that can be written as smooth functions of (\bar{Z}_n^*, n_n^*) . As $n_n^* \sim cn$ for some $c > 0$, standard Bayesian asymptotics for one-dimensional parameters under weak dependence (together with the mixing LLN/CLT) yield posterior consistency and a Bernstein–von Mises type result. The asymptotic variance $V(p_{\text{true}})$ incorporates both the intrinsic variance of Z_t and the effect of dependence through the long-run variance $\tilde{\tau}^2(p_{\text{true}})$. \square

Remark 23 (Small Samples and Long-Shot Regimes). *The results above are asymptotic in nature and require the effective sample size n_n^* to grow without bound. In low-liquidity markets (small n) or extreme-probability regimes (long-shot events with p_{true} near 0 or 1), the Beta posterior can be highly skewed and heavy-tailed, and the normal approximations may be poor. In such regimes, CAOPM should be used with caution, relying on full posterior credible intervals rather than Gaussian approximations, and sensitivity analysis with respect to the prior and adjustment parameters is particularly important.*

Proposition 18 (Finite-Sample Concentration under Mixing). *Let (Z_t) be the adjusted YES process satisfying the boundedness and geometric α -mixing conditions of Phase 8. Let*

$$\bar{Z}_n = \frac{1}{n} \sum_{t=1}^n Z_t, \quad \mu = \mathbb{E}[Z_t].$$

Then, for all $x > 0$ and $n \geq 1$, there exist constants $C_1, C_2 > 0$ (depending on the mixing coefficients and bounds on Z_t) such that

$$\mathbb{P}(|\bar{Z}_n - \mu| > x) \leq C_1 \exp(-C_2 n x^2),$$

i.e. a Bernstein-type exponential inequality holds for the empirical average. In particular, for any confidence level $\delta \in (0, 1)$, with probability at least $1 - \delta$,

$$|\bar{Z}_n - \mu| \lesssim \sqrt{\frac{\log(1/\delta)}{n}},$$

up to constants depending on the mixing structure.

Remark 24 (Practical Implications in Low- n and Long-Shot Regimes). *The concentration bound above is asymptotic in spirit: it guarantees that, for moderate n and weak dependence, \bar{Z}_n will concentrate around μ at a rate comparable to the i.i.d. case. However, in low-liquidity markets (small n), or when p is very close to 0 or 1, three issues arise:*

- *The constants C_1, C_2 may be unfavorable, leading to loose finite-sample bounds.*
- *The Beta posterior can be highly skewed, so Gaussian approximations to credible intervals may be misleading.*
- *Long-shot events make the empirical process more volatile relative to the natural scale of p , further weakening normal approximations.*

In these situations, CAOPM posteriors should be interpreted via full credible intervals and sensitivity checks, rather than relying solely on asymptotic normality or point estimates.

8.15 Metric-Based Robustness for Nonlinear Posterior Updates

The original Lipschitz robustness results in this phase implicitly assumed a linear or affine adjustment of pseudo-counts, so that the posterior mapping from data to distribution was essentially linear. Once nonlinear distortions, mixture posteriors, and dynamic bias layers are introduced (Sections 1.7, 1.7, 1.7), the adjustment of the posterior can no longer be modeled as a simple additive transformation. This subsection replaces the linear Lipschitz arguments with metric-based robustness results formulated in Wasserstein and Hellinger distances for general nonlinear updates.

Setup. Let $\Theta = [0, 1]$ denote the parameter space for the event probability p . For a given dataset \mathcal{D} , let $\Pi_{\text{base}}(\cdot \mid \mathcal{D})$ be the *base* CAOPM posterior over p , constructed as in Phases 5–6 (e.g. a mixture of Beta components calibrated via

stacking and dynamic bias layers). We model nonlinear bias correction at the posterior level via a measurable map

$$\phi : \Theta \rightarrow \Theta,$$

which acts on p to produce an adjusted parameter $\phi(p)$.²

Definition 9 (Nonlinear Posterior Update via Pushforward). *For a given dataset \mathcal{D} and base posterior $\Pi_{\text{base}}(\cdot \mid \mathcal{D})$, the nonlinearly adjusted posterior is defined as the pushforward measure*

$$\Pi_{\phi}(\cdot \mid \mathcal{D}) := \phi_{\#} \Pi_{\text{base}}(\cdot \mid \mathcal{D}),$$

i.e. for any Borel set $A \subseteq \Theta$,

$$\Pi_{\phi}(A \mid \mathcal{D}) = \Pi_{\text{base}}(\phi^{-1}(A) \mid \mathcal{D}).$$

We are interested in how sensitive $\Pi_{\phi}(\cdot \mid \mathcal{D})$ is to small changes in \mathcal{D} , measured in appropriate probability metrics (e.g. Wasserstein W_1 or Hellinger distance d_H).

Assumption 7 (Base Posterior Robustness in Wasserstein Distance). *There exists a data metric $d_{\mathcal{D}}$ on the space of datasets such that for any two datasets $\mathcal{D}, \mathcal{D}'$,*

$$W_1(\Pi_{\text{base}}(\cdot \mid \mathcal{D}), \Pi_{\text{base}}(\cdot \mid \mathcal{D}')) \leq C_{\text{base}} d_{\mathcal{D}}(\mathcal{D}, \mathcal{D}'),$$

for some constant $C_{\text{base}} < \infty$. Here W_1 is the 1-Wasserstein distance on Θ with respect to the Euclidean metric.

Assumption 7 is a metric formulation of the robustness results established earlier (e.g. those analogous to Theorem 13 and Theorem 28), expressed at the level of the base posterior Π_{base} .

Assumption 8 (Lipschitz Nonlinear Adjustment). *The nonlinear adjustment map $\phi : \Theta \rightarrow \Theta$ is globally Lipschitz with constant $L_{\phi} < \infty$:*

$$|\phi(p_1) - \phi(p_2)| \leq L_{\phi} |p_1 - p_2| \quad \text{for all } p_1, p_2 \in \Theta.$$

Theorem 31 (Wasserstein Robustness of Nonlinear Posterior Updates). *Under Assumptions 7 and 8, the adjusted posterior mapping $\mathcal{D} \mapsto \Pi_{\phi}(\cdot \mid \mathcal{D})$ is Lipschitz in W_1 :*

$$W_1(\Pi_{\phi}(\cdot \mid \mathcal{D}), \Pi_{\phi}(\cdot \mid \mathcal{D}')) \leq L_{\phi} C_{\text{base}} d_{\mathcal{D}}(\mathcal{D}, \mathcal{D}') \quad \text{for all } \mathcal{D}, \mathcal{D}'.$$

Proof. By Definition 9, $\Pi_{\phi}(\cdot \mid \mathcal{D}) = \phi_{\#} \Pi_{\text{base}}(\cdot \mid \mathcal{D})$. The 1-Wasserstein distance is contracting under Lipschitz maps: for any two measures μ, ν on Θ and any L -Lipschitz map ϕ ,

$$W_1(\phi_{\#} \mu, \phi_{\#} \nu) \leq L W_1(\mu, \nu).$$

²In applications, ϕ may depend on additional covariates or summaries of \mathcal{D} ; here we treat these as fixed when conditioning on \mathcal{D} .

Applying this with $\mu = \Pi_{\text{base}}(\cdot \mid \mathcal{D})$, $\nu = \Pi_{\text{base}}(\cdot \mid \mathcal{D}')$, and $L = L_\phi$, we get

$$W_1(\Pi_\phi(\cdot \mid \mathcal{D}), \Pi_\phi(\cdot \mid \mathcal{D}')) \leq L_\phi W_1(\Pi_{\text{base}}(\cdot \mid \mathcal{D}), \Pi_{\text{base}}(\cdot \mid \mathcal{D}')).$$

Combining this with Assumption 7 yields the claimed bound. \square

Remark 25 (Lipschitz Regularization in Calibration). *Theorem 31 shows that the robustness constant for the adjusted posterior is the product $L_\phi C_{\text{base}}$. Thus, in calibrating ϕ (e.g. via empirical risk minimization on historical markets), it is natural to penalize large Lipschitz constants. A practical approach is to include a regularization term of the form*

$$\lambda \text{Lip}(\phi) \quad \text{or} \quad \lambda \|\nabla \phi\|_{L^2},$$

in the calibration objective, trading off fit against robustness. This yields an explicit statistical justification for Lipschitz regularization of nonlinear posterior updates.

Extension to Hellinger Distance. In some arguments, it is convenient to work with Hellinger distance d_H between posterior densities. For one-dimensional models with sufficiently smooth densities and strictly monotone ϕ , we can relate Hellinger distances before and after the nonlinear transformation.

Assumption 9 (Smooth Monotone Nonlinear Adjustment). *Assume $\phi : (0, 1) \rightarrow (0, 1)$ is a C^1 diffeomorphism with derivative bounded away from 0 and ∞ :*

$$0 < m \leq \phi'(p) \leq M < \infty \quad \text{for all } p \in (0, 1).$$

Let $\pi_{\text{base}}(p \mid \mathcal{D})$ and $\pi_{\text{base}}(p \mid \mathcal{D}')$ denote posterior densities, and $\pi_\phi(q \mid \mathcal{D})$ the density of $\Pi_\phi(\cdot \mid \mathcal{D})$ under the change of variables $q = \phi(p)$.

Theorem 32 (Hellinger Stability Under Smooth Monotone Transformations). *Under Assumption 9, there exists a constant $C_H = C_H(m, M)$ such that for any two datasets $\mathcal{D}, \mathcal{D}'$,*

$$d_H(\Pi_\phi(\cdot \mid \mathcal{D}), \Pi_\phi(\cdot \mid \mathcal{D}')) \leq C_H d_H(\Pi_{\text{base}}(\cdot \mid \mathcal{D}), \Pi_{\text{base}}(\cdot \mid \mathcal{D}')).$$

In particular, if the base posterior mapping is Hellinger–Lipschitz in \mathcal{D} , then so is the adjusted posterior mapping.

Proof Sketch. Under the change of variables $q = \phi(p)$, densities transform via

$$\pi_\phi(q \mid \mathcal{D}) = \pi_{\text{base}}(\phi^{-1}(q) \mid \mathcal{D}) |\phi^{-1}'(q)|.$$

The Hellinger distance between the transformed densities is controlled by the Hellinger distance between the original densities and the Jacobian factors, with constants depending only on bounds for ϕ' and $(\phi^{-1})'$. Using $m \leq \phi'(p) \leq M$ and standard change-of-variable bounds, one obtains the stated inequality with C_H depending only on (m, M) . \square

Implications for Pricing Functionals. Let $f : \Theta \rightarrow \mathbb{R}$ be a 1-Lipschitz payoff functional of p (e.g. $f(p) = \mathbf{1}\{p > K\}$ smoothed, or a bounded Lipschitz proxy for digital pricing). Then for any two datasets $\mathcal{D}, \mathcal{D}'$,

$$\left| \int f(p) \Pi_\phi(dp \mid \mathcal{D}) - \int f(p) \Pi_\phi(dp \mid \mathcal{D}') \right| \leq W_1(\Pi_\phi(\cdot \mid \mathcal{D}), \Pi_\phi(\cdot \mid \mathcal{D}')).$$

Combining with Theorem 31 yields a bound on the sensitivity of CAPOPM prices to data perturbations under nonlinear adjustments:

$$\left| \hat{C}_\phi(\mathcal{D}) - \hat{C}_\phi(\mathcal{D}') \right| \leq L_\phi C_{\text{base}} d_{\mathcal{D}}(\mathcal{D}, \mathcal{D}'),$$

for any Lipschitz payoff functional f used in price computation.

Remark 26 (Failure Modes and Relation to F2–F3). *The robustness results in this subsection rely critically on two properties: (i) the base posterior must be stable in Wasserstein or Hellinger distance, and (ii) the nonlinear update ϕ must be Lipschitz (or smooth monotone with bounded derivative). In the strong herding and fast regime-switching regimes (Sections 1.9 and 1.9), condition (i) fails: small changes in data can produce large changes in the base posterior. In such regimes, no choice of ϕ (Lipschitz or otherwise) can restore uniform robustness. Thus the metric-based Lipschitz results here apply to the “regular” region of the model space where dependence and nonstationarity remain within the bounds of the CAPOPM assumptions.*

8.16 Failure of Robustness Under Strong Herding: A Threshold Auto-Regressive Counterexample

This section establishes that the robustness guarantees of CAPOPM fail under *strong herding*, understood as threshold-based majority-following behavior in which order flow becomes nearly deterministic once the fraction of YES orders crosses a critical threshold. In this regime, the dependence in the adjusted order process Z_t violates the mixing assumptions (A7)–(A9), the likelihood becomes multimodal, and neither a single-Beta posterior nor any finite mixture of Betas can recover the true event probability uniformly over the strong-herding parameter region.

Definition 10 (Threshold Auto-Regressive Herding Model). *Let $(Z_t)_{t \geq 1}$ denote the effective YES-indicator process entering Stage 1 after behavioral weighting. For parameters (θ, ρ) with $\theta \in (0, 1)$ and $\rho \in [0, 1]$, define*

$$Z_{t+1} = \begin{cases} 1 & \text{with probability } \rho \quad \text{if } \frac{1}{t} \sum_{i=1}^t Z_i \geq \theta, \\ 0 & \text{with probability } \rho \quad \text{if } \frac{1}{t} \sum_{i=1}^t Z_i < \theta, \\ \text{Ber}(p_{\text{true}}) & \text{with probability } 1 - \rho. \end{cases}$$

When ρ is close to 1, the process follows a majority rule as soon as the empirical fraction of YES exceeds the threshold θ , and otherwise follows a pure “NO-cascade”. We call this the strong-herding regime.

Lemma 15 (Loss of Mixing Under Strong Herding). *Fix $\theta \in (0, 1)$ and let $\rho \rightarrow 1$. Under the TAR model of Definition 10, the process (Z_t) fails to satisfy α -mixing with any summable mixing rate. In particular,*

$$\alpha_Z(k) \not\rightarrow 0 \quad \text{as } k \rightarrow \infty$$

whenever the event $\{\frac{1}{t} \sum_{i=1}^t Z_i \geq \theta\}$ occurs with positive probability.

Proof. When $\rho \rightarrow 1$, the transition becomes

$$Z_{t+1} = \mathbf{1} \left\{ \frac{1}{t} \sum_{i=1}^t Z_i \geq \theta \right\} \quad \text{a.s.,}$$

so once the empirical mean crosses θ , the process becomes identically 1 thereafter. Similarly, if the empirical mean remains below θ , the process becomes identically 0. Hence (Z_t) becomes asymptotically constant and perfectly predictable from events arbitrarily far in the past, implying $\alpha_Z(k) = 1$ for all k . Summability fails, completing the proof. \square

Lemma 16 (Bimodal Likelihood Under Strong Herding). *Let $L_n(p)$ denote the likelihood of p_{true} based on Z_1, \dots, Z_n under the TAR model. If ρ is sufficiently close to 1, then with positive probability*

$$L_n(p) \text{ is asymptotically bimodal on } [0, 1],$$

with modes concentrated near 0 and 1 corresponding to NO-cascades and YES-cascades respectively.

Proof. From Lemma 15, the process enters a deterministic regime once the empirical mean crosses θ . If it crosses upward, $Z_t = 1$ for all large t , which for the Bernoulli likelihood behaves as if $p_{\text{true}} = 1$. If it crosses downward, $Z_t = 0$ eventually, behaving as if $p_{\text{true}} = 0$. Since both cascades occur with positive probability whenever p_{true} is not exactly equal to θ , the likelihood assigns mass to neighborhoods of both 0 and 1. Bimodality follows. \square

Theorem 33 (Impossibility of Uniform Consistency Under Strong Herding). *Let $\Pi_{\text{CAPOPM}}(\cdot \mid \mathcal{D}_1, \mathcal{D}_2)$ denote the Stage 2 posterior of CAPOPM, modeled as a finite mixture of Beta distributions. Under the TAR herding model (Definition 10), the following statements hold.*

- (i) (No uniform consistency) *For any finite mixture of Beta distributions with $R < \infty$ components,*

$$\sup_{p_{\text{true}} \in (0, 1)} \mathbb{E}_{p_{\text{true}}} [|\mathbb{E}[p \mid Z_1, \dots, Z_n] - p_{\text{true}}|] \not\rightarrow 0 \quad \text{as } \rho \rightarrow 1.$$

- (ii) (Failure of finite-mixture representation) *The bimodality of the likelihood (Lemma 16) implies that any finite mixture of Betas fails to approximate the true posterior uniformly over the strong-herding region $\{\rho > \rho^*\}$ for any fixed $\rho^* < 1$.*
- (iii) (Breakdown of robustness theorems) *The Lipschitz-type continuity results of Phase 8 cannot hold under strong herding: no constant $C < \infty$ can satisfy*

$$\|\Pi_{\text{CAPOPM}}(\cdot \mid \mathcal{D}) - \Pi_{\text{CAPOPM}}(\cdot \mid \mathcal{D}')\|_{\text{TV}} \leq C \|\mathcal{D} - \mathcal{D}'\|$$

for all ρ sufficiently close to 1. The deterministic cascades in strong herding force the total variation distance to jump by 1 when \mathcal{D} crosses the threshold θ .

Proof. (i) From Lemma 16, the likelihood assigns mass to neighborhoods of 0 and 1 with positive probability even when $p_{\text{true}} \in (\theta - \varepsilon, \theta + \varepsilon)$. A finite mixture of Betas cannot track both cascades simultaneously: its posterior mean necessarily lies in a compact subinterval of $(0, 1)$ independent of the data configuration producing the cascades. Thus, the posterior mean cannot converge uniformly to p_{true} as $\rho \rightarrow 1$, proving (i).

(ii) Any finite mixture of Betas has unimodal or mildly multimodal densities, but cannot represent a sequence of likelihoods whose mass splits between 0 and 1 in a way that depends discontinuously on the data path. Therefore no finite mixture can uniformly approximate the posterior, yielding (ii).

(iii) Let \mathcal{D} and \mathcal{D}' differ only by whether the empirical mean crosses θ at time t_0 . Under strong herding, the posteriors collapse to neighborhoods of 0 and 1 respectively, giving total variation distance equal to 1. Since $\|\mathcal{D} - \mathcal{D}'\|$ can be arbitrarily small (e.g. one flip of a single Bernoulli), no Lipschitz constant can satisfy the inequality uniformly. This proves (iii). \square

Remark 27 (Interpretation). *The strong-herding regime invalidates the data-generating assumptions required for CAPOPM's robustness theory. Deterministic cascades destroy mixing, induce bimodal likelihoods, and generate posterior discontinuities that no finite-mixture Beta representation can smooth uniformly. Thus the impossibility theorem above provides a fundamental limitation: CAPOPM can be consistent and Lipschitz-robust only on subregions of the parameter space where dependence is sufficiently weak and mixing conditions (A7)–(A9) hold.*

8.17 Boundary Behavior, Long-Shot Events, and Stabilization

The CAPOPM posterior for the event probability p is Beta with parameters

$$p \mid \mathcal{I} \sim \text{Beta}(\alpha_n, \beta_n), \quad \alpha_n = \alpha_0 + y_n^*, \quad \beta_n = \beta_0 + n_n^* - y_n^*,$$

where (y_n^*, n_n^*) are the adjusted counts and (α_0, β_0) are the hybrid prior parameters. In long-shot regimes (p_{true} near 0 or 1) or in very small samples, it is possible for the posterior mass to concentrate near 0 or 1, leading to heavy tails and numerical instability for functions of p (e.g. log-odds or certain risk measures).

We formalize this behavior and describe a simple stabilization based on either truncation or a logit transform.

Lemma 17 (Tail Behavior of Beta Posteriors Near the Boundaries). *Let Π_n denote the $\text{Beta}(\alpha_n, \beta_n)$ posterior for p .*

(a) *If $\alpha_n \leq 1$, then the density of Π_n behaves like*

$$\pi_n(p) \propto p^{\alpha_n-1}(1-p)^{\beta_n-1} \sim p^{\alpha_n-1} \quad \text{as } p \downarrow 0,$$

so that the left tail near 0 is heavy whenever $\alpha_n < 1$.

(b) *If $\beta_n \leq 1$, then*

$$\pi_n(p) \sim (1-p)^{\beta_n-1} \quad \text{as } p \uparrow 1,$$

and the right tail near 1 is heavy whenever $\beta_n < 1$.

(c) *If $\alpha_n, \beta_n \geq c > 1$ uniformly in n , then there exists $\varepsilon > 0$ such that*

$$\Pi_n([0, \varepsilon] \cup (1 - \varepsilon, 1]) \leq C\varepsilon^c,$$

for some constant $C > 0$ independent of n . In particular, the posterior places vanishing mass near 0 and 1 as $\varepsilon \downarrow 0$.

Proof (Sketch). Parts (a) and (b) follow from the Beta density

$$\pi_n(p) = \frac{1}{B(\alpha_n, \beta_n)} p^{\alpha_n-1}(1-p)^{\beta_n-1},$$

and standard asymptotics as $p \rightarrow 0$ and $p \rightarrow 1$. When $\alpha_n < 1$, the factor p^{α_n-1} diverges as $p \downarrow 0$, indicating a heavy left tail; an analogous statement holds for $\beta_n < 1$ near 1.

For (c), if $\alpha_n, \beta_n \geq c > 1$ and $p \in (0, \varepsilon)$, then

$$\pi_n(p) \leq \frac{1}{B(\alpha_n, \beta_n)} p^{c-1},$$

and integrating on $(0, \varepsilon)$ gives

$$\Pi_n([0, \varepsilon]) \leq C_1 \varepsilon^c$$

for some $C_1 > 0$ that can be chosen uniformly in n due to the compactness of the parameter region. A symmetric argument holds near 1. \square

The lemma shows that heavy tails near 0 and 1 arise precisely when the posterior parameters α_n and β_n are small, which can occur under three circumstances: (i) very small effective sample size n_n^* , (ii) extreme long-shot outcomes (e.g. no YES orders in a rare-event market), or (iii) extremely concentrated or misaligned priors.

To mitigate numerical instability and avoid overconfident long-shot posteriors in these regimes, we consider a simple stabilized transform.

Proposition 19 (Stabilized Posterior via Truncation or Logit Transform). *Fix a truncation parameter $\varepsilon \in (0, 1/2)$ and define the truncated interval*

$$I_\varepsilon = [\varepsilon, 1 - \varepsilon].$$

Let Π_n be the $\text{Beta}(\alpha_n, \beta_n)$ posterior and define the truncated posterior Π_n^ε by

$$\Pi_n^\varepsilon(A) = \frac{\Pi_n(A \cap I_\varepsilon)}{\Pi_n(I_\varepsilon)}, \quad A \subseteq [0, 1] \text{ measurable},$$

whenever $\Pi_n(I_\varepsilon) > 0$.

Then:

(a) *If $\alpha_n, \beta_n \geq c > 1$ uniformly in n , then*

$$\|\Pi_n - \Pi_n^\varepsilon\|_{\text{TV}} = \Pi_n([0, \varepsilon] \cup (1 - \varepsilon, 1]) \leq C\varepsilon^c,$$

for some $C > 0$ independent of n . Thus, for small ε , the truncated posterior is close in total variation to the original posterior.

(b) *Define the logit transform*

$$\theta = \log \frac{p}{1 - p},$$

and let Λ_n be the induced posterior distribution for θ under Π_n^ε . Then moments of all orders exist for Λ_n , and Λ_n is supported on a compact interval

$$\Theta_\varepsilon = \left[\log \frac{\varepsilon}{1 - \varepsilon}, \log \frac{1 - \varepsilon}{\varepsilon} \right].$$

Consequently, functionals of p that are Lipschitz in θ are uniformly bounded and well-behaved under Λ_n .

(c) *If $\varepsilon = \varepsilon_n \downarrow 0$ is chosen such that $\varepsilon_n^c \rightarrow 0$ and $n_n^* \rightarrow \infty$, then the truncated posterior $\Pi_n^{\varepsilon_n}$ remains asymptotically equivalent to Π_n for inference about p_{true} while preventing extreme numerical instability near 0 and 1 at finite n .*

Proof (Sketch). For (a), observe that truncation only removes mass near 0 and 1, so the total variation distance equals the probability of the removed regions:

$$\|\Pi_n - \Pi_n^\varepsilon\|_{\text{TV}} = \Pi_n([0, \varepsilon] \cup (1 - \varepsilon, 1]).$$

The bound then follows directly from part (c) of the lemma, with the same exponent c and an adjusted constant C .

For (b), the truncation ensures that $p \in I_\varepsilon$ almost surely under Π_n^ε . The logit map $p \mapsto \theta = \log(p/(1-p))$ sends I_ε to Θ_ε , a compact interval in \mathbb{R} , and is smooth on $(0, 1)$. As a result, all moments of θ under Λ_n exist and are bounded uniformly in n . Any functional of p that can be expressed as a Lipschitz function of θ thus inherits uniform boundedness and stability.

For (c), if $\varepsilon_n^c \rightarrow 0$, then the total variation distance between Π_n and $\Pi_n^{\varepsilon_n}$ tends to zero by (a). At the same time, the growing effective sample size n_n^* drives the posterior mass towards p_{true} , and the truncation can be chosen small enough that it does not distort the asymptotic concentration in typical cases (where $p_{\text{true}} \in (0, 1)$). Hence the truncated posterior is asymptotically equivalent to the original one for inference while improving finite-sample stability. \square

Remark 28 (Practical Guidance for Long-Shot Markets). *The analysis above suggests a simple stabilization strategy for CAOPM in long-shot or low-liquidity markets:*

- Choose a small truncation level ε (e.g. 10^{-4} or 10^{-3}) and work with the truncated posterior Π_n^ε .
- When transforming probabilities (e.g. to log-odds), perform the transform on $\theta = \log(p/(1-p))$ under Λ_n rather than directly on p near the boundaries.
- Report both the truncated posterior summaries and the original Beta summaries, especially in cases where α_n or β_n are close to 1 or below.

This keeps the asymptotic properties intact while explicitly addressing the finite-sample instabilities associated with heavy Beta tails near 0 and 1.

8.18 CAOPM as a KL Projection: A Formal Information-Theoretic Interpretation

The adjusted CAOPM posterior for the event probability p is a Beta distribution of the form

$$p \mid \mathcal{I} \sim \text{Beta}(\alpha_n, \beta_n), \quad \alpha_n = \alpha_0 + y_n^*, \quad \beta_n = \beta_0 + n_n^* - y_n^*,$$

constructed from the hybrid prior and adjusted parimutuel counts. Up to this point the Beta form has been motivated by conjugacy and interpretability. Here we show that it also admits a fundamental *information-theoretic characterization*: it is the KL-projection of a general posterior onto the Beta family.

Let Π^* denote the (hypothetical) posterior distribution for p that would arise under a fully specified, potentially nonparametric data-generating model with dependence, behavioral distortions, or heterogeneous signals. In a general market this Π^* may not be Beta, and may not be computationally tractable.

CAPOPM provides a tractable Beta posterior. The following theorem shows that the CAPOPM posterior coincides with the *I-projection* (information projection) of Π^* onto the Beta family $\mathcal{B} = \{\text{Beta}(a, b) : a, b > 0\}$.

Theorem 34 (KL Projection Theorem for CAPOPM). *Let Π^* be any posterior distribution on $p \in (0, 1)$ with finite mean and finite log-moment, and let \mathcal{B} be the family of Beta distributions. Consider the KL divergence*

$$D_{\text{KL}}(\Pi^* \parallel \text{Beta}(a, b)) = \int_0^1 \log \frac{d\Pi^*}{d\text{Beta}(a, b)}(p) d\Pi^*(p),$$

defined whenever Π^ is absolutely continuous with respect to $\text{Beta}(a, b)$.*

Define the KL-projection of Π^ onto \mathcal{B} as*

$$(a^\dagger, b^\dagger) \in \arg \min_{a, b > 0} D_{\text{KL}}(\Pi^* \parallel \text{Beta}(a, b)).$$

Assume that Π^ has mean m^* and inverse second moment $M^* = \mathbb{E}_{\Pi^*}[p^{-1} + (1-p)^{-1}] < \infty$. Then:*

- (a) *The minimizer (a^\dagger, b^\dagger) exists and is unique.*
- (b) *The KL-projection satisfies*

$$\frac{a^\dagger}{a^\dagger + b^\dagger} = m^*, \quad \frac{a^\dagger b^\dagger}{(a^\dagger + b^\dagger)^2(a^\dagger + b^\dagger + 1)} = \text{Var}_{\Pi^*}(p),$$

i.e. the minimizing Beta distribution matches the mean and variance of Π^ .*

- (c) *If Π^* is generated by a hybrid prior and adjusted counts y_n^*, n_n^* (with mixing or dependence), then the CAPOPM posterior $\text{Beta}(\alpha_n, \beta_n)$ coincides with the KL-projection:*

$$(\alpha_n, \beta_n) = (a^\dagger, b^\dagger).$$

- (d) *In particular, for large effective sample size n_n^* , CAPOPM selects, among all Beta distributions, the one closest in KL sense to the ideal but intractable Π^* .*

Proof (Sketch). Part (a) follows from strict convexity of the KL divergence in (a, b) on the Beta family. For (b), writing the KL divergence explicitly and differentiating under the integral yields two first-order conditions:

$$\frac{\partial}{\partial a} D_{\text{KL}} = 0 \quad \Rightarrow \quad \mathbb{E}_{\Pi^*}[\log p] = \psi(a) - \psi(a + b),$$

$$\frac{\partial}{\partial b} D_{\text{KL}} = 0 \quad \Rightarrow \quad \mathbb{E}_{\Pi^*}[\log(1 - p)] = \psi(b) - \psi(a + b),$$

where ψ is the digamma function. Using identities for Beta means and variances, these conditions imply the matching of mean and variance stated in (b).

For (c), under a Beta–Binomial model (even with adjusted counts), the exact posterior for p is $\text{Beta}(\alpha_n, \beta_n)$. Thus if Π^* arises from such updating under a fully specified likelihood, Π^* is already in \mathcal{B} and the unique KL minimizer is precisely $\text{Beta}(\alpha_n, \beta_n)$.

When Π^* is more general (due to dependence, behavioral structures, or nonparametric components), the CAPOPM posterior can be interpreted as the projection of Π^* onto the Beta family using the adjusted empirical mean and variance (m^* , Var_{Π^*}) induced by the CAPOPM adjustment rules. This identifies (α_n, β_n) with the unique minimizer (a^\dagger, b^\dagger) . \square

Remark 29 (Interpretation and Novelty). *This theorem provides an information-theoretic justification for the CAPOPM posterior beyond conjugacy. Even when the true posterior Π^* is nonparametric or analytically intractable, CAPOPM delivers the closest Beta distribution in KL divergence.*

Thus the Beta form is not merely a convenient algebraic choice, but the information-projection that preserves the two most important moments of the ideal posterior under the CAPOPM-adjusted signal process.

This interpretation also helps explain why the hybrid prior and adjustment mechanism remain stable even under dependence or behavioral distortions: CAPOPM selects the “least distorted” Beta posterior compatible with the adjusted empirical information.

Remark 30 (Limitations of the KL Projection View). *Interpreting CAPOPM as a KL projection onto the Beta family is useful but also restrictive. It guarantees that, among all Beta distributions, the chosen posterior preserves certain moments of a more complex underlying posterior. However, it does not claim that the Beta family is rich enough to capture all features of the true posterior under strong dependence, multimodal priors, or adversarial behavior. In settings where such features are important, the KL-projection perspective should be viewed as an approximation tool rather than a full description of market beliefs, and more flexible models (e.g. MCMC-based or variational) may be warranted.*

8.19 Alternatives to Beta Conjugacy under Dependence

While CAPOPM deliberately uses the Beta–Binomial structure for tractability and interpretability, other approaches can, in principle, accommodate richer dependence at the cost of computational complexity:

- **MCMC with dependent likelihoods.** One can specify an explicit dependent model for the order sequence, such as an Ising or Markov random field for YES/NO decisions, and sample from the posterior for p via Gibbs or Metropolis–Hastings. This yields a more flexible posterior but requires careful tuning and may be slow in large markets.
- **Variational approximations.** Variational Bayes can approximate complex posteriors with factored or low-rank distributions, trading off accuracy for speed. In this setting, one could approximate the joint posterior

over $(p, \text{latent states})$ with a product of a Beta distribution for p and a tractable family for the latent dependence structure.

- **Composite likelihoods.** Composite or pseudo-likelihood methods replace the full joint likelihood with products of low-dimensional marginals or conditionals, offering a compromise between full MCMC and the single-sufficient-statistic approach of CAPOPM.

CAPOPM chooses Beta conjugacy as a deliberate design decision: it provides closed-form updates, interpretable pseudo-counts, and an information-projection interpretation, while recognizing that more flexible likelihood-based methods are possible when computational resources and data volume permit.

8.20 Asymptotics Under Ergodic Regime Switching and an Impossibility Result

The dynamic bias layer of Section 1.7 models (δ_t, ψ_t) as a hidden Markov chain (S_t) with a finite state space $\mathcal{S} = \{1, \dots, R\}$ and regime-specific parameters (δ_r, ψ_r) . This section establishes (i) a Bernstein–von Mises–type result under ergodic regime switching (mild nonstationarity), and (ii) an impossibility result when regimes switch so quickly that no regime accumulates enough information for consistent learning.

Assumption 10 (Ergodic Regime Switching and Regularity). *Let $(S_t)_{t \geq 1}$ be an irreducible, aperiodic Markov chain on $\mathcal{S} = \{1, \dots, R\}$ with transition matrix P and unique stationary distribution $\pi = (\pi_1, \dots, \pi_R)$. Assume:*

- (i) (Ergodicity) *For each $r, s \in \mathcal{S}$ there exists $k \geq 1$ such that $(P^k)_{rs} > 0$, and the chain is aperiodic. Consequently, for any initial distribution,*

$$\mathbb{P}(S_t = r) \rightarrow \pi_r \quad \text{as } t \rightarrow \infty.$$

- (ii) (True parameter vector) *Each regime r has a true parameter $\theta_r^* = (\delta_r^*, \psi_r^*)$ in a compact subset $\Theta_r \subset \mathbb{R}^{d_r}$, and the emission model $p(\mathcal{D}_t \mid S_t = r, \theta_r^*)$ coincides with the Stage 1/Stage 2 correction structure of CAPOPM.*

- (iii) (Identifiability and smoothness) *The mapping $\theta_r \mapsto p(\mathcal{D}_t \mid S_t = r, \theta_r)$ is identifiable and C^2 in a neighborhood of θ_r^* , with Fisher information matrix $I_r(\theta_r^*)$ positive definite.*

- (iv) (Finite moments) *The emission log-likelihoods have finite second moments under the true model.*

We are interested in a smooth functional of the regime parameters, such as the long-run average bias

$$\bar{\delta}^* := \sum_{r=1}^R \pi_r \delta_r^*,$$

or more generally a differentiable functional $\varphi(\theta_1, \dots, \theta_R)$ with $\varphi : \Theta_1 \times \dots \times \Theta_R \rightarrow \mathbb{R}$.

Theorem 35 (Bernstein–von Mises Theorem under Ergodic Regime Switching). *Under Assumption 10, suppose the CAOPM posterior over $(\theta_1, \dots, \theta_R)$ is proper and assigns positive prior density in a neighborhood of the true parameter vector $(\theta_1^*, \dots, \theta_R^*)$. Let*

$$\hat{\theta}_r(T) \quad (r = 1, \dots, R)$$

denote the (quasi-)maximum likelihood or posterior mean estimator of θ_r based on data $\mathcal{D}_{1:T} = (\mathcal{D}_1, \dots, \mathcal{D}_T)$, and define $\hat{\varphi}(T) := \varphi(\hat{\theta}_1(T), \dots, \hat{\theta}_R(T))$. Then, as $T \rightarrow \infty$,

(i) (Consistency)

$$\hat{\varphi}(T) \xrightarrow{\mathbb{P}} \varphi(\theta_1^*, \dots, \theta_R^*).$$

(ii) (Asymptotic normality)

$$\sqrt{T} \left(\hat{\varphi}(T) - \varphi(\theta_1^*, \dots, \theta_R^*) \right) \xrightarrow{d} \mathcal{N}(0, V_\varphi),$$

where V_φ is the asymptotic variance obtained by the delta method applied to the joint asymptotic distribution of $(\hat{\theta}_1(T), \dots, \hat{\theta}_R(T))$.

(iii) (Bernstein–von Mises) *The posterior distribution of $\varphi(\theta_1, \dots, \theta_R)$ given $\mathcal{D}_{1:T}$ converges in total variation to the normal law $\mathcal{N}(\hat{\varphi}(T), V_\varphi/T)$:*

$$\left\| \Pi(\varphi(\theta_1, \dots, \theta_R) \in \cdot \mid \mathcal{D}_{1:T}) - \mathcal{N}(\hat{\varphi}(T), V_\varphi/T) \right\|_{\text{TV}} \xrightarrow{T \rightarrow \infty} 0.$$

Proof Sketch. Under Assumption 10, the hidden Markov model (S_t, \mathcal{D}_t) is an ergodic HMM with finite state space and regular parametric emission densities. Standard results for HMMs imply that the (quasi-)maximum likelihood estimator of each θ_r is consistent and asymptotically normal, with an information matrix determined by long-run frequencies π_r and the per-regime Fisher information $I_r(\theta_r^*)$. The joint asymptotic normality of $(\hat{\theta}_1(T), \dots, \hat{\theta}_R(T))$ then yields asymptotic normality of $\hat{\varphi}(T)$ via the delta method. The Bayesian Bernstein–von Mises conclusion follows from general BvM theorems for HMMs with finite state space and regular parametric families, together with the prior positivity condition. Full details follow the usual template for BvM in dependent data models; we omit these for brevity. \square

Remark 31 (Application to Dynamic CAOPM). *In the dynamic CAOPM setting of Section 1.7, the regime-specific parameters θ_r^* encode bias corrections (δ_r^*, ψ_r^*) and possibly additional structural quantities. The functional φ may be taken as the long-run average distortion $\bar{\delta}^* = \sum_r \pi_r \delta_r^*$, or as a mapping from the collection (θ_r^*) to a long-run effective event probability p_{eff}^* under dynamic corrections. Theorem 35 then ensures that the CAOPM posterior for $\bar{\delta}$ (or p_{eff}) concentrates and becomes asymptotically normal despite mild nonstationarity induced by regime switching.*

We now show that this positive result has a natural limit: if regime switching is too fast for any regime to accumulate information, no sequential estimator or posterior can remain uniformly well calibrated.

Assumption 11 (Fast Regime Switching with Vanishing Dwell Time). *Let (S_t) be a Markov chain on \mathcal{S} with transition matrix P depending on T such that:*

(i) *The minimum expected dwell time in each state is uniformly bounded:*

$$\sup_T \max_{r \in \mathcal{S}} \mathbb{E}_T[\text{time spent in state } r \text{ up to } T] < C < \infty.$$

(ii) *The regime-specific parameters $\theta_r^*(T)$ are allowed to vary with T , remaining in a compact set, and emissions are generated from $p(\mathcal{D}_t \mid S_t, \theta_{S_t}^*(T))$.*

In words, regimes switch so frequently that the average number of visits to any given state does not grow with T .

Theorem 36 (Impossibility of Uniform Consistency Under Fast Regime Switching). *Under Assumption 11, consider any sequential estimator sequence $\hat{\theta}(T)$ or any Bayesian posterior sequence $\Pi_T(\cdot \mid \mathcal{D}_{1:T})$ for a regime-dependent parameter functional $\varphi(\theta_1^*(T), \dots, \theta_R^*(T))$ (e.g. a regime-specific bias $\delta_r^*(T)$ or a regime-specific event probability $p_r^*(T)$). Then there exists a choice of parameter sequences $\{\theta_r^*(T)\}$ such that*

$$\limsup_{T \rightarrow \infty} \sup_{\{\theta_r^*(T)\}} \mathbb{E} \left[\left| \hat{\varphi}(T) - \varphi(\theta_1^*(T), \dots, \theta_R^*(T)) \right| \right] > 0,$$

and similarly, no sequence of posteriors Π_T can concentrate around the true functional uniformly over $\{\theta_r^(T)\}$. In particular, there is no uniformly consistent dynamic CAOPM estimator or posterior in this fast switching regime.*

Proof Sketch. Under Assumption 11, the expected number of observations generated in any fixed regime r up to time T is uniformly bounded by C . Therefore, for any estimator or posterior targeting a regime-specific parameter (or a functional that depends nontrivially on the per-regime values), the effective sample size per regime does not grow with T . By standard parametric lower bounds, no estimator can achieve vanishing risk uniformly over the parameter sequences $\{\theta_r^*(T)\}$ when each regime is observed only $O(1)$ times. One can construct pairs of parameter sequences that are indistinguishable from the data but induce separated values of the functional φ , forcing a nonzero lower bound on the estimation error. The same argument applies to Bayesian posteriors: with bounded regime information, the posterior cannot concentrate uniformly on the true functional. \square

Remark 32 (Interpretation for Dynamic CAOPM). *Theorem 35 and Theorem 36 identify a feasible and an infeasible nonstationary regime for CAOPM. Under ergodic regime switching with growing effective sample size per regime, dynamic CAOPM remains consistent and asymptotically normal for smooth*

functionals of the regime parameters. When regime switching becomes so fast that no regime accumulates information, no sequential estimator or posterior can be uniformly well calibrated. In practice, CAPOPM should therefore treat fast, high-frequency structural breaks as a regime where the bias layer and posterior must be explicitly flagged as fragile, and any pricing output should carry a warning about nonstationary uncertainty that cannot be statistically resolved.

8.21 Summary

Phase 8 demonstrates that the CAPOPM framework yields:

- posterior consistency across structural, ML, behavioral, and adversarial regimes;
- asymptotic normality of the posterior distribution;
- robustness to behavioral biases and structural distortions;
- preservation of arbitrage-freeness; and
- stability of the posterior under small perturbations.

These results complete the theoretical foundation of CAPOPM.

Ethical and Computational Considerations

Although CAPOPM is primarily a statistical and market-microstructure framework, its practical deployment raises two considerations that fall outside the purely mathematical scope: (i) ethical issues related to information interpretation and trader impact, and (ii) computational scalability in high-frequency or large-market environments.

Ethical considerations. The CAPOPM adjustment mechanism is designed to extract latent information from market behavior, not to infer personal traits or to profile individual traders. All behavioral weights and offsets operate on aggregated orders or coarse feature categories (such as order size, time of submission, or broad trader cohorts). In typical parimutuel settings, these inputs are public, event-level signals rather than personally identifying data.

Two ethical concerns still warrant attention:

- **Misinterpretation of crowd behavior.** Behavioral adjustments such as herding corrections, long-shot bias offsets, or systematic weight modulation must be disclosed transparently. CAPOPM does not judge the rationality of individual traders; it merely models aggregate effects. Misrepresenting these adjustments as normative statements about trader competence should be avoided.
- **Incentive effects.** If CAPOPM were used in a live market to set prices or guide payouts, the knowledge of the adjustment mechanism might influence trader strategies. The framework is therefore intended as a research and belief-elicitation tool rather than a prescriptive mechanism for market design or individual decision-making.

These considerations do not modify the mathematics of the model but help clarify its appropriate scope.

Computational scalability. In its simplest form, CAPOPM requires maintaining weighted sums

$$y_n^* = \sum_{t=1}^n Z_t + \delta_+, \quad n_n^* = \sum_{t=1}^n w(x_t; \psi) + \delta_+ + \delta_-,$$

which can be updated in $O(1)$ time per order. However, certain extensions—notably those involving dynamic behavioral weights or dependence structures that require recalculating pairwise interactions—may scale as $O(n)$ per update. Over a long sequence of orders, this can become burdensome.

A few practical strategies mitigate this issue:

- **Streaming updates.** Most CAPOPM quantities reduce to running sums or averages. When $w(x; \psi)$ is fixed within a market, updates remain $O(1)$.

- **Batch weighting.** If behavioral weights depend on slowly varying features (e.g. time of day or volatility regime), weights can be precomputed for batches, reducing per-order cost.
- **Sparse interaction structures.** In dependence models such as the Ising-type herding formulation, restricting the trader graph to a bounded-degree or sparse structure ensures that the effective cost of updating conditional probabilities remains manageable.
- **Approximate inference.** When n is very large, the limiting results of Phase 8 (LLN/CLT) support the use of approximate moment-based updates rather than exact counts, allowing amortized $O(1)$ updates.

In summary, the core CAPOPM update remains computationally inexpensive, while more elaborate behavioral or dependence components require thoughtful implementation to avoid unnecessary overhead. Nothing in the framework requires storing all historical orders or performing pairwise computations across traders.

Remark 33 (Scope). *The ethical and computational considerations presented here serve as practical guidance rather than limitations of the theory. They outline appropriate use cases and clarify that CAPOPM is best viewed as a research tool for structured belief elicitation, not a real-time trading protocol.*

Conclusion and Future Directions

This paper has developed CAOPM, a crowd-adjusted parimutuel option pricing mechanism designed to extract beliefs from a combination of structural financial modeling, machine learning signals, and observed parimutuel order flow. The framework is organized into eight phases that separate structural assumptions, prior construction, behavioral and structural correction, Bayesian updating, and theoretical evaluation. The intention is not to assert that any single component of the model is definitive, but rather to provide a structured setting in which these components can be examined individually and in combination.

A key feature of the approach is the construction of a hybrid prior that integrates a tempered fractional Heston model with neural-network-based probability estimates. This combination leverages the interpretability and market-consistent properties of the structural approach while allowing data-driven models to influence the prior when their predictive quality is high. The parimutuel order flow then acts as a sequence of informational signals, updating the hybrid prior through a Beta-Binomial mechanism. Behavioral and structural corrections are incorporated through a two-stage adjustment that preserves conjugacy and provides analytic clarity when analyzing robustness, sensitivity, and asymptotic behavior.

The theoretical analysis in Phase 8 shows that, under reasonable conditions on trader populations and bounded behavioral distortions, the adjusted posterior is both robust and consistent. Extensions of the robustness results to variance and Hellinger distance ensure that the entire posterior distribution remains stable under small perturbations in the adjusted counts. A CLT-based argument establishes that, as the effective sample size grows, the posterior becomes asymptotically normal and concentrates near the true event probability. These results are not presented as definitive guarantees for all market settings, but as formal properties that hold within the controlled environment defined in the model.

There are several natural directions for further development. First, the structural offsets used in the bias-correction layer could be estimated via an empirical Bayes procedure, using historical panels of markets to infer the systematic component of crowd miscalibration. This would provide a more principled basis for the correction terms while maintaining the closed-form posterior structure. Second, the behavioral weights applied to order flow could be generalized to incorporate time-varying or kernel-based decay, possibly using the same tempered fractional kernels that appear in the Heston variance dynamics. Such an extension would more closely align the informational and structural memory components of the model and create a unified kernel-based approach to volatility and information flow.

Third, the likelihood specification could be extended to admit correlated or exchangeable sequences of trades, replacing the independent Bernoulli model with an overdispersed or block-correlated alternative. This would allow the model to more directly represent herding, clustered trading, and contagion ef-

fects. In conjunction with the empirical Bayes approach described above, such a specification could be evaluated using hierarchical or mixed-effects models that incorporate trader heterogeneity at a more granular level.

Fourth, the machine learning component could be expanded beyond point estimates of p_{ML} to incorporate uncertainty quantification, such as Bayesian neural networks or ensemble variance. This would allow the hybrid prior to more flexibly represent model uncertainty rather than treating p_{ML} as purely deterministic.

Finally, an applied implementation would necessarily involve empirical validation. While real-world data analysis lies outside the scope of this paper, the eight-phase design is intended to make such studies modular. Researchers can select specific phases to refine, test, or replace, while preserving the overall Bayesian updating structure. For example, empirical studies could examine how different trader populations affect posterior concentration, how severe herding impacts inference, or how the hybrid prior adapts to markets with rapidly changing volatility.

Taken together, these observations highlight that CAPOPM should be viewed as a framework rather than a fixed model. Its value lies in the organization of structural, behavioral, and machine learning components into a coherent Bayesian mechanism for belief extraction in a parimutuel environment. The results presented here establish theoretical foundations upon which empirical studies and more detailed structural extensions can be developed.

Glossary of Symbols and Notation

This glossary summarizes the notation used throughout the CAPOPM framework, including the structural model, hybrid prior, behavioral adjustments, and asymptotic results. Symbols are defined once and used consistently in all phases.

A. Probability, Events, and Measures

- $\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}$ — underlying filtered probability space.
- Q — risk-neutral measure.
- P — physical (real-world) measure when needed.
- $A = \{S_T > K\}$ — CAPOPM event (“YES event”).
- $\theta \in \{0, 1\}$ — binary state of nature (Koessler–Noussair convention).
- $p = Q(S_T > K)$ — digital event probability.
- p_{true} — true event probability under the data-generating process.
- p_0 — CAPOPM hybrid prior mean.

B. Asset Price and Structural Model Parameters

- S_t — asset price process.
- V_t — variance process.
- (W_t, B_t) — correlated Brownian motions with $\rho \in [-1, 1]$.
- γ — mean-reversion intensity in the variance equation.
- θ (Heston) — long-run variance level (not to be confused with event state).
- σ — volatility-of-volatility.
- $\alpha \in (1/2, 1)$ — fractional exponent in the tempered fractional kernel.
- $\lambda \geq 0$ — tempering parameter.
- $K_{\alpha, \lambda}(t - s) = e^{-\lambda(t-s)}(t - s)^{\alpha-1}$ — Volterra kernel.
- $h(t), g(t)$ — Riccati–Volterra solution pair (Shi).
- $\Phi(u; T)$ — characteristic function of $\ln S_T$.
- $f_{\Theta, \alpha}(s; T)$ — structural risk-neutral density of S_T .
- $q_{\text{Shi}}(K, T; \Theta, \alpha)$ — structural digital tail probability.

C. Machine-Learning Prior and Hybrid Prior

- x — feature vector (ANN) or input sequence (RNN).
- $g_{\text{NN}}(x)$ — ML model output estimating p .
- $p_{\text{ML}} = g_{\text{NN}}(x)$ — ML-implied event probability.
- n_{ML} — pseudo-sample size for ML prior (performance-weighted).
- η_{str} — structural prior strength.
- $\eta = \eta_{\text{str}} + n_{\text{ML}}$ — effective hybrid prior mass.
- $\alpha_0 = \eta p_0, \beta_0 = \eta(1 - p_0)$ — hybrid Beta prior parameters.

D. Parimutuel Market Quantities

- $s_i \in \{\text{YES}, \text{NO}\}$ — trader i 's action.
- N — total number of traders (market size).
- Π_N — empirical YES fraction in market of size N .
- π_N — parimutuel odds derived from Π_N .
- $\Phi(\cdot)$ — odds mapping (aggregate-to-price function).
- π^* — fixed-point odds in the large-market limit.

E. Behavioral Adjustments (Phase 5–6)

- $w_i^{\text{beh}} = w(x_i; \psi)$ — behavioral weight for order i based on features x_i and parameters ψ .
- ψ — parameter vector governing the behavioral weighting function.
- δ_+, δ_- — structural offsets correcting systematic directional bias.
- $\delta = (\delta_+, \delta_-)$ — combined structural adjustment vector.
- $\hat{\psi}_M, \hat{\delta}_M$ — Empirical Bayes estimates from M historical markets.
- Z_t — adjusted YES contribution at time/order t .
- y_n^* — adjusted YES pseudo-count, including offsets.
- n_n^* — adjusted total pseudo-count.

F. Posterior Quantities

- Π_n — posterior distribution for p given n adjusted observations.
- $\alpha_n = \alpha_0 + y_n^*$ — posterior YES parameter.
- $\beta_n = \beta_0 + n_n^* - y_n^*$ — posterior NO parameter.
- \hat{p}_n — posterior mean.
- σ_n^2 — posterior variance.
- Π_n^ε — truncated posterior (boundary-stabilized).
- $\theta = \log(p/(1-p))$ — logit-transformed probability.

G. Dependence, Mixing, and Asymptotics

- Z_t — adjusted YES process (bounded).
- $\alpha(k)$ — α -mixing coefficient of the dependent sequence (Z_t) .
- C, ρ — constants for geometric mixing: $\alpha(k) \leq C\rho^k$.
- \bar{Z}_n — empirical average of adjusted indicators.
- $\mu(p_{\text{true}})$ — mean of Z_t under true probability p_{true} .
- $\tau^2(p_{\text{true}})$ — long-run variance for CLT under dependence.
- $V(p_{\text{true}})$ — asymptotic variance for posterior mean (Bernstein–von Mises).

H. Option Pricing and Pricing Kernel

- $D_{\text{CAP}}(K, T) = e^{-rT} \hat{p}(K, T)$ — CAPOPM-implied digital price.
- $F_{\text{CAP}}^Q(K, T)$ — CAPOPM-implied risk-neutral CDF.
- $f_{\text{CAP}}^Q(s, T)$ — CAPOPM-implied risk-neutral density.
- $C_{\text{CAP}}(K, T)$ — CAPOPM-implied call price.
- $\phi_{\text{CAP}}(s, T) = e^{-rT} f_{\text{CAP}}^Q(s, T)$ — state-price density.
- $m_{\text{CAP}}(s, T) \propto f_{\text{CAP}}^Q(s, T)/f^P(s)$ — pricing kernel when f^P is known.

References

- [1] Boris S. Axelrod, Ben J. Kulick, Charles R. Plott, and Kevin A. Roust. The design of improved parimutuel-type information aggregation mechanisms: Inaccuracies and the long-shot bias as disequilibrium phenomena. *Journal of Economic Behavior & Organization*, 69:170–181, 2009.
- [2] Mikkel Bennedsen, Asger Lunde, and Mikko S. Pakkanen. Decoupling the short- and long-term behavior of stochastic volatility. *Journal of Financial Econometrics*, 15(2):233–259, 2017.
- [3] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 2 edition, 1985.
- [4] Olivier Cappé, Eric Moulines, and Tobias Rydén. *Inference in Hidden Markov Models*. Springer, New York, 2005.
- [5] Kai Diethelm. *The Analysis of Fractional Differential Equations*, volume 2004 of *Lecture Notes in Mathematics*. Springer, Berlin, 2004.
- [6] Paul Doukhan. *Mixing: Properties and Examples*, volume 85 of *Lecture Notes in Statistics*. Springer, New York, 1994.
- [7] Angela Maria D’Uggento, Marta Biancardi, and Domenico Ciriello. Predicting option prices: From the black-scholes model to machine learning methods. *Big Data Research*, 40:100518, 2025.
- [8] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1994.
- [9] Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough heston models. *Mathematical Finance*, 29(1):3–38, 2019.
- [10] Ken-Ichi Funahashi and Yuichi Nakamura. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, 6(6):801–806, 1993.
- [11] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pages 1050–1059. PMLR, 2016.
- [12] Hans U. Gerber and Elias S. Shiu. Option pricing by esscher transforms. *Transactions of the Society of Actuaries*, 46:99–191, 1994.
- [13] Subhashis Ghosal and Aad W. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, Cambridge, 2017.
- [14] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

- [15] Emad Jaber, Michael Larsson, and Mathieu Rosenbaum. Affine volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200, 2019.
- [16] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, 1991.
- [17] Frédéric Koessler, Charles Noussair, and Anthony Ziegelmeyer. Information aggregation and belief elicitation in experimental parimutuel betting markets. *Journal of Economic Behavior & Organization*, 83:195–208, 2012.
- [18] Alan L. Lewis. *Option Valuation under Stochastic Volatility*. Finance Press, Newport Beach, CA, 2000.
- [19] David Pollard. *A User’s Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge, 2001.
- [20] Herbert Robbins. An empirical bayes approach to statistics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1:157–163, 1956.
- [21] Zhengguang Shi. Option pricing for heston model with tempered fractional brownian motion. *Results in Applied Mathematics*, 27:100623, 2025.
- [22] Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.
- [23] Cristiano Varin, Nancy Reid, and David Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42, 2011.
- [24] Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman. Using stacking to average bayesian predictive distributions. *Bayesian Analysis*, 13(3):917–1003, 2018.
- [25] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin, 6 edition, 2003.