

Estimating Latent Volatility Regimes via Bayesian Markov Switching

A Probabilistic Approach to Regime Identification in Equity Markets

Pavlo Mysak

MA578: Bayesian Statistics
Professor AmirEmad Ghassami
Boston University
December 05, 2025

1 Introduction

Financial markets exhibit complex dynamics that cannot be adequately captured by models that assume stable variance or constant structural behavior over time. Periods of heightened volatility, rapid price dislocations, or persistent calm often cluster in ways that reflect underlying shifts in market conditions. These evolving environments (i.e. regimes) represent discrete-time latent states of the data-generating process in which the statistical properties of asset returns differ between states. Identifying and characterizing such regimes is of fundamental importance in financial econometrics, as risk is not constant but instead varies distinctly across these states.

In the context of equity markets, the most commonly studied regimes correspond to “low-volatility” and “high-volatility” periods, which typically align with broader macroeconomic conditions, liquidity constraints, or shifts in investor sentiment. Downstream financial decisions (e.g. portfolio allocation, risk management, option pricing, and market stress testing) are all sensitive to volatility, making accurate regime identification both practical and consequential.

Traditional approaches to volatility modeling often rely on continuous-time formulations, such as stochastic volatility models, which impose continuous evolution of latent variance. While powerful, these frameworks can struggle to represent abrupt structural changes or persistent shifts between distinct market conditions. By contrast, Markov regime-switching models explicitly encode discontinuous changes via a latent Markov chain that governs transitions between states. This discrete structure naturally captures the empirical observation that volatility tends to remain elevated or subdued for extended periods before transitioning rapidly.

A Bayesian formulation of regime-switching models offers several advantages. First, Bayesian inference provides a principled mechanism for quantifying uncertainty in both model parameters and the latent regime sequence. This is essential when regimes are unobserved, as point estimates alone can obscure ambiguity regarding regime boundaries or transition behavior. Second, hierarchical priors allow the modeler to incorporate structural knowledge. Third, Bayesian methods are well suited for situations in which regime membership is only weakly identified from the data, as posterior samples naturally reflect the degree of uncertainty about regime assignments, transition probabilities, and volatility levels.

In this project, I develop and estimate a Bayesian two-state Markov switching model for daily S&P 500 returns. The modeling framework centers on learning a latent categorical regime sequence, where each regime is associated with a distinct level of volatility, and transitions occur according to a first-order Markov process with state-dependent persistence. By placing hierarchical priors on the volatility parameters and probabilistic priors on the transition matrix, the model leverages both data-driven inference and structural assumptions rooted in financial theory. The resulting posterior distribution yields a full probabilistic characterization of volatility regimes and their dynamics over time.

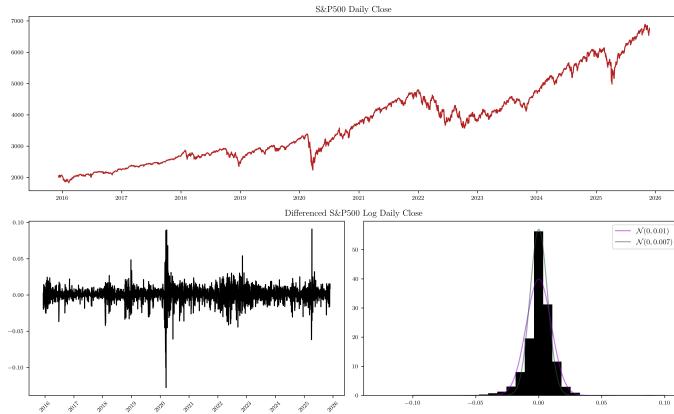


Figure 1: Daily S&P 500 closing prices, corresponding log-returns, and empirical return distribution with two candidate normal densities overlaid to illustrate central concentration and tail behavior.

primary input to the regime-switching model and is visualized in Figure 1 alongside the raw price series and the empirical distribution of returns.

The empirical analysis relies on daily closing prices of the S&P 500 index obtained from the Federal Reserve Bank of St. Louis (FRED). These data represent one of the most widely used benchmarks for U.S. equity market performance and capture broad market dynamics across large-capitalization firms. The dataset spans a multiyear period and is well suited to volatility-regime analysis. To model volatility, raw closing prices are first transformed into daily log-returns, defined as

$$r_t = \ln(P_t) - \ln(P_{t-1}),$$

a standard transformation that yields a stationary series more consistent with the assumptions of Gaussian observation models.

The differenced log-return series provides the

Several limitations of the dataset must be acknowledged. First, the S&P 500 represents only a single equity index and therefore provides a narrow view of market-wide volatility dynamics. The model does not incorporate information from other asset classes, macroeconomic indicators, or sector-specific series that may jointly influence market regimes. Second, daily closing prices may omit within-day variation that could affect volatility estimates. Finally, as with all market data sourced from public repositories such as FRED, any anomalies, revisions, or holiday effects inherent in the published series propagate into the analysis. These limitations do not invalidate the modeling approach but do restrict the scope of inference to the aggregate behavior of this specific index.

2 Methods and Analysis

2.1 Model Specification

Let Y_t denote the daily log-return of the S&P 500 on day t over the period 2016–2025. We model Y_t as arising from a two-state latent volatility process $Z_t \in \{1, 2\}$ where each state corresponds to a distinct level of market volatility. Conditional on the latent state, returns are assumed normally distributed with mean zero and state-dependent standard deviation:

$$Y_t|Z_t = k \sim \mathcal{N}(0, \sigma_k), \quad k \in \{1, 2\}$$

The latent state sequence $\{Z_t\}_{t=1}^T$ follows a first-order Markov process with transition matrix P :

$$Z_0 \sim \text{Bernoulli}\left(\frac{1}{2}\right), \quad Z_t|Z_{t-1} \sim P(Z_{t-1}), \quad t = 1, \dots, T$$

where

$$P = \begin{pmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{pmatrix}, \quad p_{00}, p_{11} \sim \text{Beta}(8, 1.2)$$

and p_{00}, p_{11} encode a broad preference for persistent regimes, reflecting the empirical observation that volatility clusters over multiple trading days.

We place a hierarchical Gamma prior on the state-dependent volatilities, σ_k , to capture two key features of financial returns. First, the high-volatility regime is expected to exhibit both a larger mean and greater dispersion than the low-volatility regime, reflecting intuitive financial reasoning and consistent with the empirical distribution of returns (see Figure 1). Second, the hierarchy partially pools information across regimes to improve computational stability while maintaining regime-specific flexibility:

$$\begin{aligned} \theta_1 &\sim \text{Gamma}(\mu = 0.003, \sigma = 0.001), & \theta_2 &\sim \text{Gamma}(\mu = 0.01, \sigma = 0.005) \\ \tau &\sim \text{Exponential}(25), & \sigma_k|\theta_k, \tau &\sim \text{Gamma}(\mu = \theta_k, \sigma = \tau), \quad k = 1, 2. \end{aligned}$$

To ensure identifiability of latent regimes and prevent label-switching, the priors on θ_1 and θ_2 are specified to induce an ordering on the regime-specific volatility states ($\theta_1 < \theta_2$), and are partly informed by the preliminary empirical observation in figure 1.

Posterior inference is conducted via a hybrid Markov chain Monte Carlo (MCMC) procedure. Continuous parameters $\tau, \theta_k, \sigma_k, p_{00}, p_{11}$ are sampled using the No-U-Turn Sampler (NUTS), which provides efficient exploration of correlated, high-dimensional posterior spaces. The latent discrete sequence $Z = \{Z_t\}$ is updated via Gibbs-Metropolis sampling conditional on the current values of the continuous parameters.

2.2 Model Checking

Model adequacy and robustness were assessed through a combination of convergence diagnostics, posterior predictive checks, and prior sensitivity analysis. Convergence of the Markov chain Monte Carlo sampler was evaluated using standard diagnostics, including \hat{R} statistics and visual inspection of trace and autocorrelation function plots for all continuous model parameters. Effective sample sizes were monitored to ensure sufficient exploration of the posterior distribution, and chains were run long enough to ensure proper mixing.

To assess robustness with respect to prior specification, a diffuse-prior sensitivity analysis was conducted. The baseline informative priors on regime-specific volatility parameters and transition probabilities were replaced with weakly informative alternatives that remove prior assumptions regarding volatility scale, regime separation, and state persistence. In particular, the hierarchical scale parameter governing volatility dispersion was assigned a HalfNormal prior, regime-level volatility hyperparameters were given diffuse HalfNormal priors, and transition probabilities were assigned uniform Beta distributions. Posterior summaries under the alternative specification were compared to those obtained under the baseline priors to evaluate the stability of regime identification, volatility ordering, and transition behavior.

Model fit was further evaluated using posterior predictive checks. Replicated return series were generated from the posterior predictive distribution and compared to the observed data. Given the model's focus on volatility dynamics, particular attention was paid to the tail behavior of the return distribution. Tail-oriented test statistics were constructed from both observed and replicated data, and posterior predictive p-values were computed to assess the model's ability to reproduce extreme returns. This approach directly evaluates whether the regime-switching structure adequately captures the empirical heaviness of return tails without overstating dispersion in the center of the distribution.

3 Results

3.1 Posterior Inference on Volatility, Regime Dynamics and State Persistence

Figure 2 summarizes posterior inference for the regime-specific volatility parameters. The posterior distribution of the low-volatility regime, σ_1 , is tightly concentrated, yielding a narrow credible interval with posterior mean $\bar{\sigma}_1 = 0.0062$ [0.0059, 0.0066]. In contrast, the posterior for the high-volatility regime, σ_2 , exhibits substantially greater dispersion, with posterior mean $\bar{\sigma}_2 = 0.0187$ [0.0175, 0.0200] and a noticeably wider credible interval.

This asymmetry in posterior uncertainty is economically and statistically intuitive. Periods of elevated market volatility are characterized by heterogeneous and episodic shocks, leading to greater uncertainty in volatility estimation relative to more tranquil market conditions. The posterior distributions therefore reflect not only level differences across regimes but also fundamentally distinct uncertainty environments.

The clear separation between the posterior mass of σ_1 and σ_2 indicates that the model successfully identifies two distinct volatility regimes rather than attributing variation to sampling noise. This separation supports the interpretation of the latent state variable as a meaningful characterization of persistent volatility dynamics rather than as a transient clustering artifact.

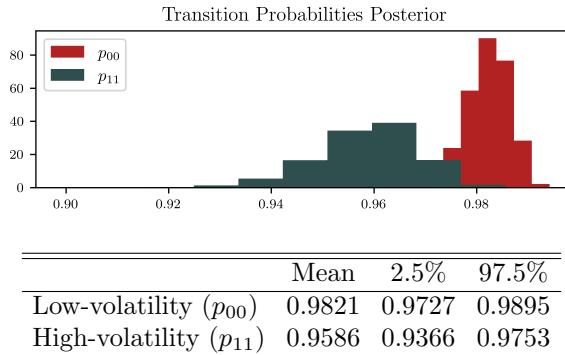


Figure 3: Posterior summary of transition probabilities.

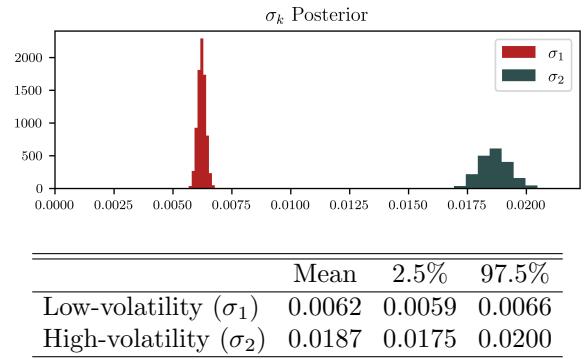


Figure 2: Posterior summary of regime-specific volatilities.

Posterior inference on the transition probabilities reinforces that the probability of remaining in the low-volatility regime, p_{00} , exhibits a posterior mean of 0.9821, implying highly persistent stable states. In contrast, the probability of remaining in the high-volatility regime, p_{11} , has a lower posterior mean of 0.9586, indicating that turbulent market conditions are less persistent and tend to revert more quickly. Consistent with posterior inference on the volatility parameters, uncertainty surrounding p_{11} is notably larger than that of p_{00} , reflecting greater variability in the duration and dynamics of high-volatility episodes.

Figure 4 displays the posterior mean state probability $\mathbb{E}[\mathbb{I}\{Z_t = 2\}|Y]$ over time, overlaid on the observed return series. Darker blue regions correspond to higher posterior probability of belonging to the high-volatility regime, while darker red regions indicate higher posterior probability of the low-volatility regime. The resulting pattern reveals clear temporal clustering, with extended periods of high confidence in the low-volatility state punctuated by shorter, episodic transitions into the high-volatility state.

3.2 Model Checking and Robustness

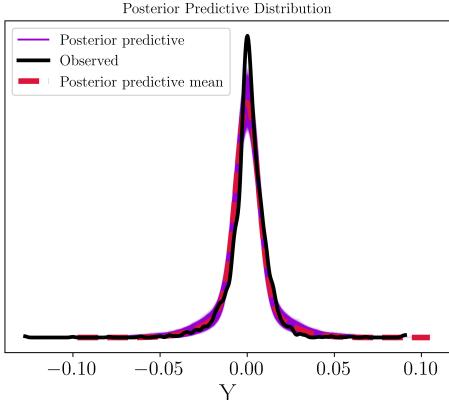


Figure 5: Posterior predictive distributions.

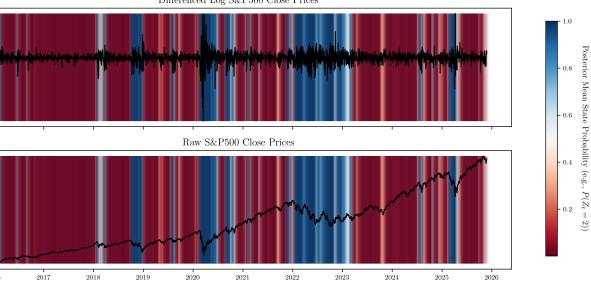


Figure 4: Timeseries plot with overlayed posterior mean of regime probabilities.

Figure 5 compares the empirical distribution of observed returns with posterior predictive draws generated under the fitted model. Overall, the predictive distribution aligns closely with the observed data, particularly in the tails, indicating that the regime-switching specification adequately captures extreme return behavior. Minor discrepancies are observed near the center of the distribution, where the empirical density exhibits slightly greater mass around zero relative to the posterior predictive mean. This pattern suggests a modest tendency of the model to attribute a small fraction of low-volatility observations to the high-volatility regime, consistent with minor overestimation of high-volatility states.

To formally assess tail fit, a posterior predictive p-value is computed using the test statistic

$$T(Y) = P(|Y_t| > q_{0.9}(|Y|))$$

where $q_{0.9}(|Y|)$ denotes the empirical 90th percentile of the absolute observed returns. The resulting posterior predictive p-value of 0.8663 indicates that replicated datasets tend to exhibit slightly more extreme behavior than observed, providing quantitative support for the visual assessment while on the edge for acceptable bounds for model adequacy.

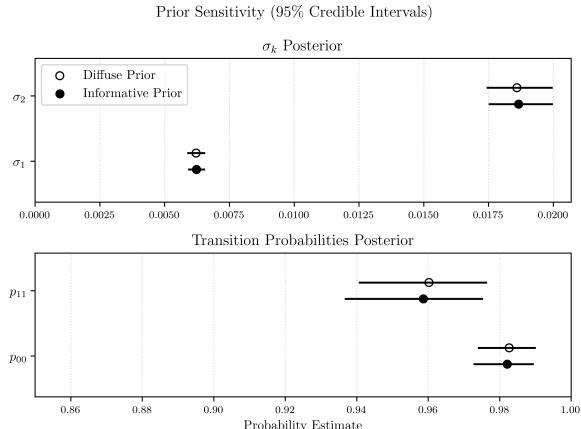


Figure 6: Posterior distributions under informative vs diffuse priors.

Figure 6 summarizes posterior sensitivity of the structural parameters to alternative prior specifications. Posterior inference for the regime-specific volatility parameters, σ_1 and σ_2 , is effectively unchanged under diffuse priors, indicating that volatility estimates are strongly data-driven and robust to prior perturbation.

Posterior inference for the transition probabilities is similarly stable across prior choices. Under both the informative Beta(8, 1.2) specification and the diffuse Beta(1, 1) prior, posterior means and credible intervals for p_{00} and p_{11} are nearly identical, suggesting that persistence estimates are primarily identified by the data rather than imposed by prior assumptions.

3.3 Summary

This project demonstrates that a Bayesian Markov switching model identifies two distinct and highly persistent volatility regimes in equity returns, with clearly separated posterior distributions for regime-specific volatility levels. Low-volatility states dominate the sample and exhibit longer expected durations, while high-volatility states arise episodically and revert more quickly, albeit with greater uncertainty.

From a methodological perspective, achieving reliable MCMC convergence in the presence of latent state processes is difficult and requires careful prior specification and substantial computational effort. These considerations are central to the successful application of Bayesian methods in dynamic economic settings.

4 Appendix

Trace plots and autocorrelation diagnostics indicate reasonable mixing across all chains for structural parameters, although the minimum effective sample size across the latent state sequence $\{Z_t\}$ is relatively low. Summaries are reported in Table 1.

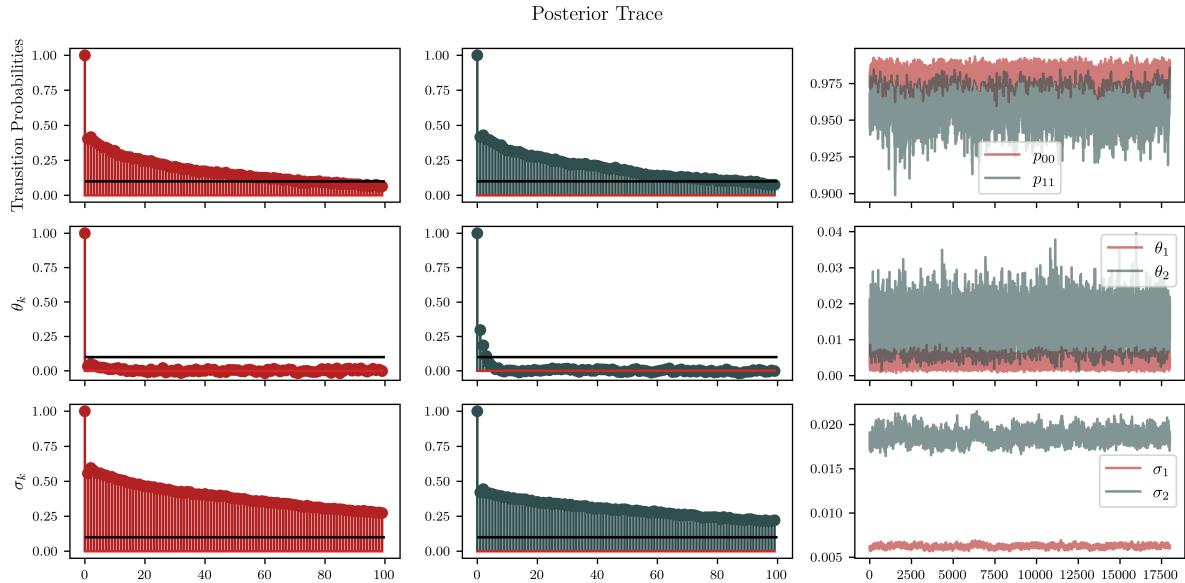


Figure 7: Trace plot and posterior distribution for $p_{00}, p_{11}, \theta_k, \sigma_k$.

Table 1: MCMC Convergence Diagnostics (Effective Sample Size and \hat{R})

Parameter	ESS _{bulk}	ESS _{tail}	\hat{R}
τ	5123	3572	1.00
θ_1	11907	9708	1.00
θ_2	8168	7387	1.00
σ_1	122	455	1.02
σ_2	159	442	1.01
p_{00}	420	1407	1.01
p_{11}	273	1117	1.01
Z_t (min/max)	68	68	1.05