

Hidden Markov Models for Volatility Regime Detection

Jim Denne

January 21, 2026

1 Introduction

Financial time series rarely display constant volatility. Equity indices exhibit prolonged tranquil episodes intertwined with short-lived but intense turbulent periods, giving rise to the well-documented phenomenon of volatility clustering. Capturing this time variation is central to risk management, derivatives pricing, and portfolio construction.

Using daily log-returns of SPY from 2000 to 2025, we fit a Gaussian Hidden Markov Model (HMM) with state-dependent variances and compare its performance against standard benchmarks. Specifically, we aim to:

1. validate the presence of persistent latent volatility regimes in SPY returns through likelihood-based diagnostics and regime stability,
2. assess whether incorporating discrete volatility regimes improves both in-sample fit and out-of-sample density forecasting relative to single-regime models.

The benchmark models include a baseline i.i.d. Gaussian specification; an i.i.d. Student- t model, ensuring that any improvement is not merely driven by static heavy tails; and the canonical GARCH(1, 1) model with Gaussian innovations, to evaluate whether discrete regimes offer comparable predictive value to continuous volatility dynamics.

2 Methodology

All data is sourced from Yahoo Finance. To evaluate the out-of-sample robustness of volatility regimes across structural market breaks, we partition the sample into a pre-COVID training period (1 Jan 2000 to 31 Dec 2019, 5029 observations) and a post-COVID test period (1 Jan 2020 to 31 Dec 2025, 1506 observations). This split isolates the unprecedented volatility shock of the COVID-19 pandemic, providing a strict out-of-sample evaluation of model robustness and regime persistence absent from historical training data. Following standard practice in financial econometrics, we work with daily log-returns to achieve approximate stationarity (Cont, 2001), further scaled by a factor of 100 to ensure numerical stability. As documented in the stylized facts of asset returns, raw returns exhibit little serial dependence, whereas absolute returns display persistent autocorrelation associated with volatility clustering, motivating the use of a regime-switching framework.

2.1 Hidden Markov Models

In this section, we describe the Markov regime switching setup. Let $\{Z_t\}_{t=1}^T$ be a discrete-time Markov chain defined on a state space $\{1, 2, \dots, N\}$, where N denotes the number of regimes. We assume a first-order Markov property

$$p(Z_t | Z_{1:t-1}) = p(Z_t | Z_{t-1}),$$

reflecting the assumption that regime persistence is adequately captured by the current state. Define the one-step transition probabilities as $a_{ij} = p(Z_t = j | Z_{t-1} = i)$, with $A = (a_{ij})$ denoting the $N \times N$ transition matrix. The initial distribution is given by the row vector $\pi = (\pi_1, \pi_2, \dots, \pi_N)$, where $\pi_i = p(Z_1 = i)$ for $i = 1, 2, \dots, N$.

Let $\{X_t\}_{t=1}^T$ be the log-return time series generated by this hidden process. For a Gaussian HMM, conditional on the hidden state $Z_t = i$ the observation X_t is Normally distributed with a unique mean μ_i and variance σ_i^2

$$X_t | (Z_t = i) \sim \mathcal{N}(\mu_i, \sigma_i^2).$$

We denote all HMM parameters with $\lambda = (A, \pi, \theta)$, where $\theta = \{(\mu_i, \sigma_i^2)\}_{i=1}^N$ represents the collection of emission distribution parameters. We further assume conditional independence of observations given the current hidden state

$$p(X_t | Z_{1:t}, X_{1:t-1}) = p(X_t | Z_t),$$

so that all temporal dependence in the observations is mediated through the latent regime process.

We implement the canonical algorithms outlined in Rabiner (1989) to deploy the model. Namely, likelihoods are computed using the forward pass, hidden states are inferred from Viterbi decoding, and parameters are estimated via the Baum-Welch (EM) algorithm (see Appendix A for each algorithm). Since globally optimal convergence isn't guaranteed by Baum-Welch, we fit all HMMs by initialising from 50 random seeds and retain the solution with the highest log-likelihood. Using 1000 EM algorithm iterations for each seed, convergence is declared when the log-likelihood increase is less than 10^{-4} .

To determine the optimal number of states, we conduct model selection between HMMs with 2, 3, 4 and 5 states. For a stronger over-parameterisation penalty, we use the Bayesian information criterion,

$$\text{BIC} = -2\hat{\ell} + n \log T,$$

where $\hat{\ell}$ is the maximised likelihood and n is the number of parameters. To ensure economically meaningful regimes, we compute expected regime durations

$$\mathbb{E}[T_i] = \frac{1}{1 - a_{ii}}$$

for each state i . HMMs with regimes persisting fewer than 5 trading days are classified as degen-

erate and excluded, as shorter regimes are weakly identified in noisy return data and inconsistent with the empirically observed multi-day persistence of financial risk regimes.

2.2 Benchmarking

i.i.d. Gaussian. The baseline assumes log-returns are independent and identically distributed drawings from a single Gaussian with constant mean and variance. Rejecting this establishes that volatility is time-varying rather than static.

i.i.d. Student-t. Retaining independence but replacing the Gaussian with a heavy-tailed Student-*t* tests whether any improvement from the HMM is merely due to unconditional kurtosis rather than discrete regimes.

GARCH(1,1). The work-horse GARCH specification lets conditional variance evolve continuously over time as a function of past innovations and its own lagged values,

$$\begin{aligned} X_t &= \sigma_t \epsilon_t, & \epsilon_t &\sim \mathcal{N}(0, 1) \\ \sigma_t^2 &= \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$. Comparing the HMM to GARCH answers whether discrete regime shifts offer explanatory power beyond the standard continuous-volatility framework.

Models are evaluated using in-sample information criteria and out-of-sample density forecasting. In-sample fit is assessed using the Akaike and Bayesian information criteria,

$$\text{AIC} = -2\hat{\ell} + 2n, \quad \text{BIC} = -2\hat{\ell} + n \log T,$$

where $\hat{\ell}$ is the maximized log-likelihood, n the number of parameters, and T the sample size. Out-of-sample evaluation from $t = T+1, T+2, \dots, T+S$ is determined using the average one-step-ahead predictive log-score,

$$\frac{1}{S} \sum_{t=T+1}^{T+S} \log p(x_t | \mathcal{F}_{t-1}),$$

a proper scoring rule that evaluates full predictive distributions. All model parameters are estimated once on the pre-COVID training period (2000–2019) and held fixed throughout the post-COVID test period (2020–2025), yielding a strict fixed-origin forecasting scheme. While model parameters are held fixed after 2019, state variables – including filtered regime probabilities in the HMM and conditional variance in GARCH – are updated recursively using post-COVID observations. Precise forecasting algorithms are found in Appendix B.

This design eliminates parameter leakage, isolates the role of regime persistence in volatility dynamics, and ensures a fair comparison across models with differing sources of time variation, including recursive volatility updating in GARCH and latent-state filtering in HMMs. While periodic re-estimation could improve operational forecasting performance, the fixed-origin framework provides a conservative and structurally informative benchmark aligned with the study’s objectives.

3 Results

3.1 Model Selection

Model selection results based on BIC and regime persistence are reported in Table 1. BIC strongly favors the 4-state HMM ($BIC = 13,598.67$) over both the 2-state (14,044.43), 3-state (13,639.09) and 5-state (13,614.23) specifications. While the 5-state model achieves a somewhat comparable BIC, it contains degenerate regimes with expected durations of only 1.07 trading days, which fails our 5-day minimum threshold for economic interpretability. The 4-state HMM, by contrast, exhibits economically meaningful regime persistence with a minimum duration of 16.18 trading days, consistent with the notion of distinct volatility regimes rather than transient artifacts. This combination of statistical parsimony and economic plausibility validates our choice of a 4-state specification for subsequent forecasting analysis.

	2 States	3 States	4 States	5 States
BIC	14044.43	13639.09	13598.67	13614.23
Min. Duration	36.41	28.73	16.18	1.07

Table 1: Model Selection Results

3.2 Model Validation

The parameters of the 4-state HMM are reported in Table 2. The selected HMM exhibits strong internal coherence and economically meaningful regime structure. Estimated transition probabilities imply high regime persistence, with diagonal elements exceeding 0.900 across all states, resulting in expected durations of approximately 20, 16, 21, and 39 trading days respectively. The decoded states capture four economically distinct regimes characterized by their volatility levels: two low-volatility tranquil states ($\hat{\sigma}_0^2 = 0.214\%$, $\hat{\sigma}_1^2 = 0.835\%$), an extreme-volatility turbulent state ($\hat{\sigma}_3^2 = 12.460\%$), and an intermediate volatile state ($\hat{\sigma}_2^2 = 2.342\%$). Notably, all regime mean returns are economically negligible, ranging from -0.263% to 0.124% for daily log-returns. This reflects the near-zero daily drift of equity index returns and confirms that regime identification is dominated by variance dynamics. Empirical regime statistics closely match the estimated parameters, indicating stable estimation and minimal overfitting.

State i	0	1	2	3	State i	$\hat{\mu}_i$	$\hat{\sigma}_i^2$	\bar{x}_i	s_i^2
0	0.951	0.047	0.000	0.002	0	0.124	0.214	0.110	0.218
1	0.051	0.938	0.000	0.011	1	0.020	0.835	0.032	0.828
2	0.000	0.000	0.953	0.047	2	-0.263	12.460	-0.271	14.210
3	0.000	0.020	0.006	0.974	3	-0.075	2.342	-0.068	2.285

(a) Transition matrix; entry (i, j) is \hat{a}_{ij} .

(b) Conditional parameters & sample statistics (%).

Table 2: 4-State HMM parameters.

Moreover, we overlay the Viterbi-decoded states with empirical returns, and plot the smoothed state probabilities in Figure 1. Viterbi-decoded state sequences align closely with known periods of market stress, including the Dot-Com Bubble and Global Financial Crisis, while state probabilities illustrate high confidence with clustering around 0 and 1, suggesting well-identified regimes rather than frequent state ambiguity. Together, these diagnostics support the regime-switching interpretation and suggests that the model captures persistent volatility states consistent with observed market dynamics.

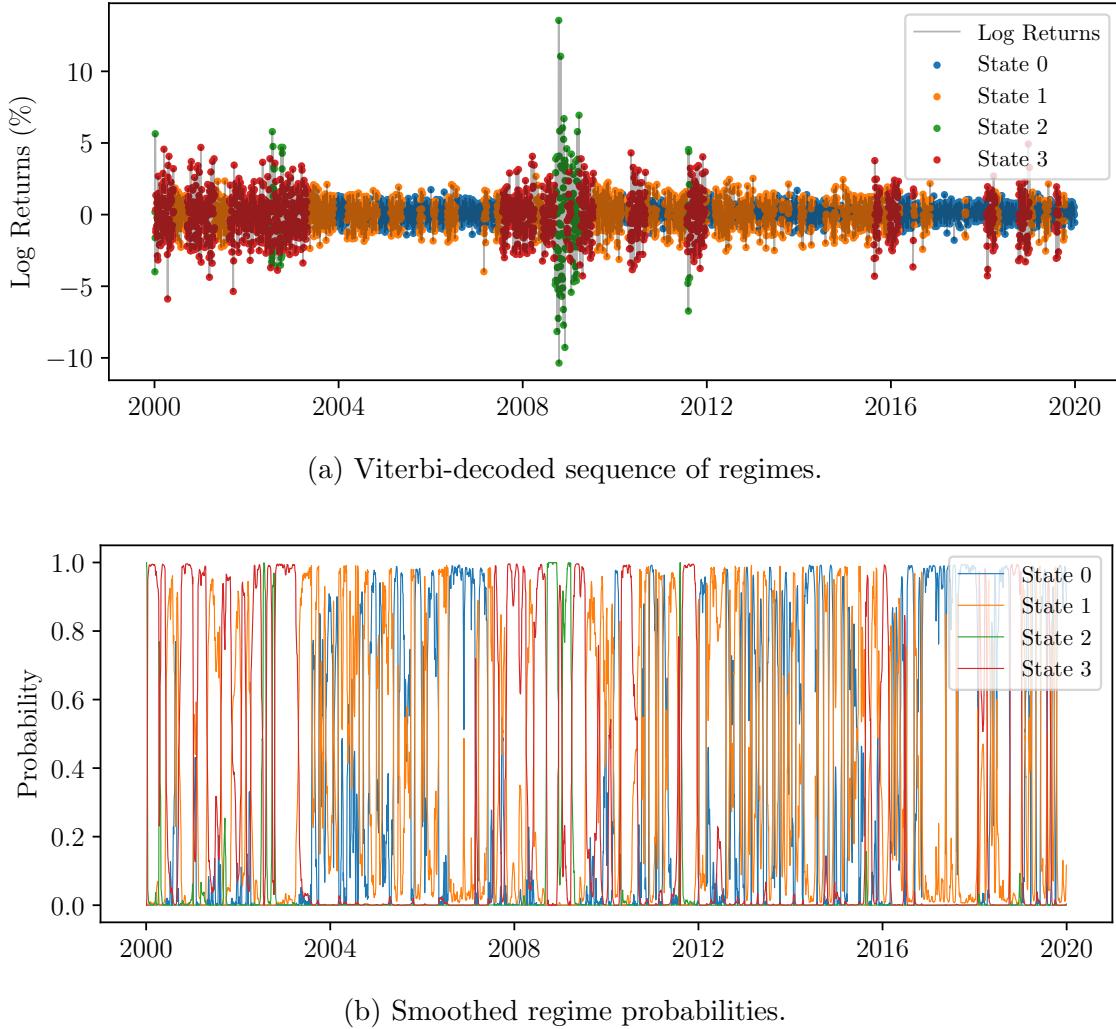


Figure 1: 4-State HMM diagnostic plots.

3.3 Benchmark Comparisons

We refer to Table 3 for empirical results. Both the 4-state HMM and GARCH(1,1) substantially outperform the i.i.d. Gaussian and Student- t benchmarks in terms of in-sample information criteria and out-of-sample predictive log-scores, strongly supporting the presence of persistent time-varying volatility dynamics. The HMM achieves a markedly superior in-sample fit, reflecting its ability to capture discrete latent volatility regimes, while the GARCH model delivers the

strongest out-of-sample density forecasts, consistent with efficient short-horizon volatility updating. Taken together, these results support the hypothesis that SPY returns are governed by latent volatility regimes, and that explicitly modelling regime persistence improves statistical fit beyond static heavy-tailed alternatives. However, for short-horizon density forecasting, continuous autoregressive volatility dynamics remain highly competitive, suggesting that regime-switching and GARCH models capture complementary aspects of volatility clustering rather than serving as strict substitutes.

Model	Parameters	AIC	BIC	Log-Score
4-State HMM	See Section 3.2	13448.654	13598.682	-1.412
i.i.d. Gaussian	$\hat{\mu} = 0.023, \hat{\sigma}^2 = 1.427$	16063.563	16076.609	-1.698
i.i.d. Student- t	$\hat{\nu} = 2.572, \hat{\mu} = 0.072, \hat{\sigma}^2 = 0.470$	14663.975	14683.544	-1.523
GARCH(1, 1)	$\hat{\omega} = 0.021, \hat{\alpha} = 0.112, \hat{\beta} = 0.871$	13673.570	13699.662	-1.326

Table 3: Benchmark parameter estimates and results.

4 Discussion and Limitations

This study evaluated whether volatility clustering in equity returns is better explained by discrete latent regimes or by continuous autoregressive volatility dynamics. The results decisively reject static return distributions: both regime-switching and GARCH models deliver substantial improvements in fit and predictive accuracy. The HMM achieves the strongest in-sample performance, supporting the existence of persistent latent volatility regimes that align with major market stress episodes, while GARCH provides superior short-horizon out-of-sample density forecasts through efficient recursive volatility updating.

These findings indicate that discrete regime models and continuous-volatility models capture complementary dimensions of volatility clustering. Regime-switching models offer structural interpretability and regime identification benefits beyond static heavy-tailed alternatives, whereas GARCH remains highly effective for short-term forecasting. Together, the evidence supports modelling volatility as time-varying and persistent, with the choice between discrete and continuous frameworks depending on whether interpretability or short-horizon predictive accuracy is the primary objective. Indeed, hybrid specifications that combine discrete regimes with continuous within-regime dynamics, such as Markov-switching GARCH, have been shown to improve risk-adjusted returns and drawdown control (Duriez, 2025).

Limitations: Several limitations should be acknowledged. First, while the selected HMM exhibits stable parameter estimates and persistent regimes, regime identification in finite samples is inherently fragile. Discrete volatility states are not uniquely defined and may depend on the assumed number of regimes, initialization, and distributional assumptions. As a result, regime labels should be interpreted descriptively rather than as structural economic states.

Second, the use of Gaussian emissions within each regime may partially conflate discrete regime

shifts with unconditional tail risk. Although comparisons against i.i.d. Student- t benchmarks mitigate this concern, the possibility remains that extreme volatility states absorb heavy-tailed behavior that could alternatively be captured through fat-tailed emission distributions. Extensions to Student- t or skewed emissions would provide a sharper separation between regime dynamics and tail risk.

Third, the fixed-origin forecasting design assumes stability of transition dynamics across structural market breaks. While this enables a stringent test of regime persistence, it implicitly assumes that post-COVID volatility dynamics are governed by the same latent process as the pre-COVID period – an assumption that cannot be formally validated within this framework.

Finally, the analysis is restricted to one-step-ahead density forecasts for a single equity index. Regime-switching models may exhibit greater advantages at longer horizons or under tail-focused loss functions relevant to risk management. Extending the analysis across multiple asset classes and forecasting horizons would be necessary to assess the generality of the results.

A Hidden Markov Model Implementation

All algorithms are implemented using the `hmmlearn` Python package wrapped in custom classes.

A.1 Forward Algorithm

Given an observation sequence $x_{1:T}$ and a model $\lambda = (A, \pi, \theta)$, direct likelihood computation by summing over all possible hidden state sequences is computationally infeasible. The forward algorithm provides a dynamic programming solution with polynomial complexity. For $t = 1, 2, \dots, T$ and $i = 1, 2, \dots, N$ define the forward variable

$$\alpha_t(i) = p(X_{1:t} = x_{1:t}, Z_t = i \mid \lambda).$$

This represents the joint likelihood of the partial observation sequence up to time t , where the underlying Markov process is in state i at time t . We can recursively compute $\alpha_t(i)$ as follows:

1. *Initialisation.* Let

$$\alpha_1(i) = \pi_i f_i(x_1), \quad i = 1, 2, \dots, N,$$

where $f_i(x) = \mathcal{N}(x; \mu_i, \sigma_i^2)$ denotes the emission density corresponding to state i .

2. *Recursion.* For $t = 2, 3, \dots, T$ compute

$$\alpha_t(i) = f_i(x_t) \sum_{j=1}^N \alpha_{t-1}(j) a_{ji}, \quad i = 1, 2, \dots, N.$$

3. *Termination.* The likelihood of the full observation sequence is given by marginalisation:

$$p(x_{1:T} \mid \lambda) = \sum_{i=1}^N \alpha_T(i).$$

The forward algorithm requires $\mathcal{O}(N^2T)$ operations, a dramatic reduction relative to naïve enumeration.

A.2 Viterbi Algorithm

Given an observation sequence $x_{1:T}$ and a model $\lambda = (A, \pi, \theta)$, the Viterbi algorithm finds the most likely hidden state sequence

$$z_{1:T}^* = \operatorname{argmax}_{z_{1:T}} p(Z_{1:T} = z_{1:T} \mid X_{1:T} = x_{1:T}, \lambda).$$

For $t = 1, 2, \dots, T$ and $i = 1, 2, \dots, N$ define the Viterbi score

$$\delta_t(i) = \max_{z_{1:t-1}} p(Z_{1:t-1} = z_{1:t-1}, Z_t = i, X_{1:t} = x_{1:t} \mid \lambda),$$

which represents the maximum joint likelihood of the observations up to time t among all state sequences that end in state i . To enable backtracking, define the corresponding backpointer

$$\psi_t(i) = \operatorname{argmax}_{j \in \{1, 2, \dots, N\}} \delta_{t-1}(j) a_{ji}, \quad t = 2, 3, \dots, T.$$

The Viterbi variables are computed recursively as follows:

1. *Initialisation.* Let

$$\delta_1(i) = \pi_i f_i(x_1), \quad \psi_1(i) = 0, \quad i = 1, 2, \dots, N,$$

where $f_i(x) = \mathcal{N}(x; \mu_i, \sigma_i^2)$ denotes the emission density corresponding to state i .

2. *Recursion.* For $t = 2, 3, \dots, T$ compute

$$\delta_t(i) = f_i(x_t) \max_{j \in \{1, 2, \dots, N\}} \left(\delta_{t-1}(j) a_{ji} \right), \quad i = 1, 2, \dots, N.$$

3. *Termination.* Take

$$z_T^* = \operatorname{argmax}_{i \in \{1, 2, \dots, N\}} \delta_T(i).$$

4. *Backtracking.* The most likely state sequence $z_{1:T}^*$ is recovered by backward recursion:

$$z_t^* = \psi_{t+1}(z_{t+1}^*) \quad t = T-1, T-2, \dots, 1.$$

The Viterbi algorithm returns a single globally most likely state path under the model. The computational complexity of the Viterbi algorithm is $\mathcal{O}(N^2T)$, matching that of the forward pass.

A.3 Baum-Welch Algorithm

Given an observation sequence $x_{1:T}$, the Baum-Welch (EM) algorithm provides an efficient iterative procedure to find the maximum likelihood estimate of HMM parameters $\lambda = (A, \pi, \theta)$. Based on the Expectation-Maximisation (EM) framework, it iteratively maximises a lower bound on the log-likelihood. Each iteration is guaranteed to produce a non-decreasing likelihood value and converges to a local maximiser.

The EM algorithm requires posterior expectations with respect to the latent states. These quantities are efficiently computed using the forward and backward algorithms. The forward variables $\alpha_t(i)$ are defined as in Section A.1. To complement these, for $t = 1, 2, \dots, T$ and $i = 1, 2, \dots, N$ define the backward variables

$$\beta_t(i) = p(X_{t+1:T} = x_{t+1:T} | Z_t = i, \lambda).$$

These represent the likelihood of future observations conditional on the current hidden state. The $\beta_t(i)$ can be computed recursively as follows:

1. *Initialisation.* Let

$$\beta_T(i) = 1, \quad i = 1, 2, \dots, N,$$

since there are no observations after time T .

2. *Recursion.* For $t = T - 1, T - 2, \dots, 1$ compute

$$\beta_t(i) = \sum_{j=1}^N a_{ij} f_j(x_{t+1}) \beta_{t+1}(j), \quad i = 1, 2, \dots, N.$$

Using the forward and backward variables, we obtain the posterior marginal distribution of the hidden state at each time point. For $t = 1, 2, \dots, T$ and $i = 1, 2, \dots, N$ define the smoothed state probabilities

$$\gamma_t(i) = p(Z_t = i | X_{1:T} = x_{1:T}, \lambda).$$

This quantity can be written in terms of the forward and backward variables as

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{p(x_{1:T} | \lambda)},$$

where the normalising constant is given by

$$p(x_{1:T} | \lambda) = \sum_{i=1}^N \alpha_T(i).$$

The Baum-Welch algorithm alternates between the following two steps.

E-step. In addition to $\gamma_t(i)$, the EM algorithm requires the joint posterior probability of successive hidden states. For $t = 1, 2, \dots, T - 1$ and $i, j \in \{1, 2, \dots, N\}$, define

$$\xi_t(i, j) = p(Z_t = i, Z_{t+1} = j | X_{1:T} = x_{1:T}, \lambda),$$

the posterior probability of occupying state i and transitioning to state j . Using the forward and backward variables, this quantity can be written as

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} f_j(x_{t+1}) \beta_{t+1}(j)}{p(x_{1:T} | \lambda)}.$$

The smoothed marginal probabilities satisfy

$$\gamma_t(i) = p(Z_t = i | X_{1:T} = x_{1:T}, \lambda) = \sum_{j=1}^N \xi_t(i, j), \quad t = 1, 2, \dots, T - 1.$$

M-step. Given $\{\gamma_t(i)\}$ and $\{\xi_t(i, j)\}$, updated parameter estimates $\lambda^* = (A^*, \pi^*, \theta^*)$ are obtained by maximising the expected complete-data log-likelihood:

1. *Initial distribution.*

$$\pi_i^* = \gamma_1(i), \quad i = 1, 2, \dots, N.$$

2. *Transition probabilities.*

$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad i, j = 1, 2, \dots, N.$$

3. *Emission parameters (Gaussian case).*

$$\mu_i^* = \frac{\sum_{t=1}^T \gamma_t(i) x_t}{\sum_{t=1}^T \gamma_t(i)}, \quad (\sigma_i^2)^* = \frac{\sum_{t=1}^T \gamma_t(i) (x_t - \mu_i^*)^2}{\sum_{t=1}^T \gamma_t(i)}, \quad i = 1, 2, \dots, N.$$

Iterating the E-step and M-step yields a sequence of parameter estimates with monotonically non-decreasing likelihood values.

B Fixed-Origin Density Forecasting Algorithms

B.1 Hidden Markov Model

Let $\hat{\lambda} = (\hat{A}, \hat{\pi}, \hat{\theta})$ be the parameters of a HMM trained on $x_{1:T}$. Initialise by computing the filtered state probabilities at time T

$$\gamma_T(i) = p(Z_T = i \mid X_{1:T} = x_{1:T}, \hat{\lambda}),$$

using the forward–backward algorithm. For each observation at $t = T + 1, T + 2, \dots, T + S$ in the test sample:

1. Propagate state probabilities forward:

$$\hat{\gamma}_t(j) = p(Z_t = j \mid X_{1:t-1} = x_{1:t-1}, \hat{\lambda}) = \sum_{i=1}^N \gamma_{t-1}(i) a_{ij}.$$

2. Compute the one-step-ahead predictive density:

$$p(x_t \mid \mathcal{F}_{t-1}) = \sum_{j=1}^N \hat{\gamma}_t(j) \mathcal{N}(x_t \mid \mu_j, \sigma_j^2),$$

and record the log-score: $\log p(x_t \mid \mathcal{F}_{t-1})$.

3. Update the filtered probabilities:

$$\gamma_t(j) = \frac{\hat{\gamma}_t(j) \mathcal{N}(x_t \mid \mu_j, \sigma_j^2)}{\sum_{k=1}^N \hat{\gamma}_t(k) \mathcal{N}(x_t \mid \mu_k, \sigma_k^2)}.$$

The log predictive score for the test set is $\frac{1}{S} \sum_{t=T+1}^{T+S} \log p(x_t | \mathcal{F}_{t-1})$.

B.2 GARCH(1,1)

Let $\hat{\omega}, \hat{\alpha}, \hat{\beta}$ be the estimated parameters and σ_T^2 the last in-sample conditional variance. For each observation x_t in the test set $x_{T+1}, x_{T+2}, \dots, x_{T+S}$:

1. Update variance:

$$\sigma_t^2 = \hat{\omega} + \hat{\alpha}x_{t-1}^2 + \hat{\beta}\sigma_{t-1}^2.$$

2. Compute Predictive density:

$$p(x_t | \mathcal{F}_{t-1}) = \mathcal{N}(x_t | 0, \sigma_t^2).$$

B.3 i.i.d. Models

For the Gaussian and Student- t models, the predictive density is simply the unconditional density with parameters estimated on the training data: $p(x_t | \mathcal{F}_{t-1}) = p(x_t | \hat{\theta})$.

References

- R. Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. doi: 10.1080/713665670.
- J. Duriez. Modeling financial volatility with hidden Markov and stochastic volatility models overlays. *SSRN Electronic Journal*, April 2025. doi: 10.2139/ssrn.5580230. URL <https://ssrn.com/abstract=5580230>.
- L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.