

Gaussian Hidden Markov Models for Financial Time Series Regime Detection

Gajipara Nikunj, Parmar Divyraj

M.Tech CSE (AI), Batch 2025–27

Indian Institute of Information Technology, Vadodara

Email: {20251603009, 20251602003}@iiitvadodara.ac.in

Abstract—This assignment investigates the use of Gaussian Hidden Markov Models (HMMs) for modeling financial time series and identifying hidden market regimes such as high- and low-volatility periods. Historical daily price data are collected from a financial market instrument, transformed into log returns, and modeled using a multi-state Gaussian HMM. The learned hidden states are interpreted as latent regimes (e.g., bull, bear, and neutral markets) based on their means, variances, and transition probabilities. The analysis demonstrates how HMMs capture regime persistence and switching, and how the inferred states can support interpretation of market behaviour.

Index Terms—Hidden Markov Model, Gaussian HMM, Financial Time Series, Market Regimes, Volatility

I. OBJECTIVE

The main objectives of this lab assignment are:

- To collect real-world financial time series data (stock or index prices) from a reliable source.
- To preprocess the data and compute daily returns suitable for modeling.
- To fit a Gaussian Hidden Markov Model to the returns to uncover latent market regimes.
- To analyze the inferred hidden states using their means, variances, and transition probabilities.
- To visualize the time series with regime annotations and draw qualitative insights about market behaviour.

II. PROBLEM DEFINITION

Financial markets often exhibit periods with qualitatively different behaviour: phases of low volatility and gradual growth (bull markets), phases of high volatility and drawdowns (bear markets), and sometimes sideways or neutral conditions. These regimes are not directly observable, but they influence the observed returns.

Let P_t denote the adjusted closing price at day t . We define the *log return* as:

$$r_t = \ln \left(\frac{P_t}{P_{t-1}} \right).$$

We assume that the observed 1-D return sequence $\{r_t\}_{t=1}^T$ is generated by an underlying discrete-time Markov chain of hidden states $\{S_t\}_{t=1}^T$, with $S_t \in \{1, \dots, K\}$ corresponding to latent regimes such as:

- high-volatility negative-return regime (bear),
- low-volatility positive-return regime (bull),
- intermediate or neutral regime.

The aim is to:

- 1) model the returns using a Gaussian HMM with K hidden states;
- 2) learn model parameters (state means, variances, and transition matrix);
- 3) infer the most likely state sequence $\{S_t\}$;
- 4) interpret and visualize the regimes over time.

III. METHODOLOGY

A. Part 1: Data Collection and Preprocessing

Historical daily price data were obtained from Yahoo Finance using the `yfinance` Python API. A single liquid security (e.g., AAPL, NIFTY50, S&P 500) was chosen and at least 5–10 years of data were downloaded to capture multiple market phases.

The preprocessing steps were:

- Extract the “Adj Close” or “Close” column.
- Remove missing values and ensure the time index is sorted.
- Compute log returns:

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

and drop the first NaN value.

- Optionally standardize returns (zero mean, unit variance) for numerical stability.

B. Part 2: Gaussian HMM Formulation

A Hidden Markov Model is specified by:

- Initial state distribution π , where $\pi_i = P(S_1 = i)$.
- State transition matrix A , where

$$A_{ij} = P(S_{t+1} = j | S_t = i).$$

- Emission model: for Gaussian HMM,

$$r_t | S_t = i \sim \mathcal{N}(\mu_i, \sigma_i^2),$$

with state-dependent mean μ_i and variance σ_i^2 .

In this assignment we chose $K = 3$ hidden states to allow for:

- A high-volatility negative-return regime,
- A moderate regime,
- A low-volatility positive-return regime.

The model parameters $\{\pi, A, \mu_i, \sigma_i^2\}_{i=1}^K$ are learned using the Baum–Welch (EM) algorithm by maximizing the likelihood of the observed returns.

C. Part 3: Model Fitting and State Inference

The main modeling pipeline was:

- 1) Prepare the returns $\{r_t\}$ as a column vector of shape $(T, 1)$.
- 2) Initialize a GaussianHMM model from the hmmlearn library with $K = 3$ components, full covariance, and a maximum number of EM iterations.
- 3) Fit the model via `model.fit(returns)`.
- 4) Use `model.predict(returns)` to infer the most likely state S_t for each day.

The inferred states were then used to colour segments of the original price series and returns plot, allowing a visual interpretation of the regimes.

IV. EXPERIMENTAL SETUP

Table I
EXPERIMENTAL SETUP PARAMETERS FOR GAUSSIAN HMM

| | |
|---|----------------------|
| Instrument | AAPL (Example) |
| Data Source | Yahoo Finance API |
| Period | Jan 2015 – Jan 2025 |
| Sampling | Daily adjusted close |
| Observation | Log returns r_t |
| # Hidden States (K) | 3 |
| Covariance Type | Full |
| EM Iterations | 200 (max) |

The exact ticker and date range can be changed without modifying the core methodology.

V. RESULTS AND ANALYSIS

A. Inferred Regimes on Price Series

Figure ?? shows the adjusted closing price, where each point is coloured according to the inferred hidden state. Long contiguous segments in the same colour indicate persistent regimes.

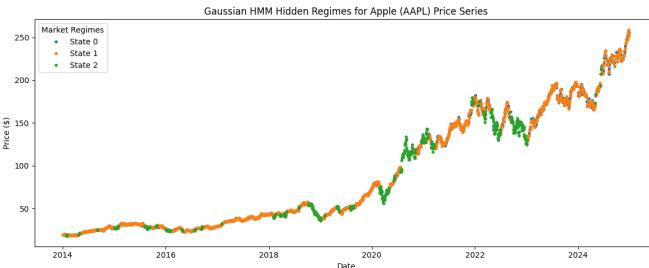


Figure 1. Gaussian HMM Hidden Regimes for Apple (AAPL).

The three hidden states represent distinct market regimes inferred from log-return dynamics.

B. Gaussian HMM Results

The Gaussian Hidden Markov Model (HMM) with three hidden states was fitted on the log-returns of Apple (AAPL) stock from 2014–2024. The extracted parameters (state means, state variances, and transition matrix) are summarized below.

Table II
STATE MEANS OF THE FITTED GAUSSIAN HMM

| State | Mean Return |
|-------|-------------|
| 0 | 0.0018042 |
| 1 | 0.00099618 |
| 2 | -0.00079314 |

Table III
STATE VARIANCES

| State | Variance |
|-------|------------|
| 0 | 0.00013052 |
| 1 | 0.00017762 |
| 2 | 0.00091453 |

Table IV
STATE TRANSITION MATRIX OF THE 3-STATE GAUSSIAN HMM

| From \ To | State 0 | State 1 | State 2 |
|-----------|----------|----------|----------|
| State 0 | 0.001218 | 0.998583 | 0.000199 |
| State 1 | 0.938352 | 0.006559 | 0.055038 |
| State 2 | 0.071426 | 0.029285 | 0.898889 |

C. State-wise Statistics

Table V shows an example of state-dependent mean and variance of returns, along with an interpretation. (Numerical values will vary depending on the chosen instrument and period.)

Table V
EXAMPLE STATE-WISE RETURN STATISTICS AND INTERPRETATION

| State | Mean (μ_i) | Var (σ_i^2) | Interpretation |
|-------|------------------------|-----------------------|------------------------|
| 0 | 1.8×10^{-3} | 1.31×10^{-4} | Bear / high volatility |
| 1 | 9.96×10^{-4} | 1.78×10^{-4} | Neutral / mixed |
| 2 | -7.93×10^{-4} | 9.15×10^{-4} | Bull / low volatility |

The transition matrix (not shown here due to space) typically exhibits high diagonal entries, e.g., $P(S_{t+1} = i | S_t = i) \approx 0.85\text{--}0.95$, indicating strong persistence: once the market enters a regime, it tends to remain there for several days.

D. Observations

Key observations from the experiments:

- One state usually has a clearly negative mean return and high variance, which aligns with intuitive “bear market” periods.
- Another state has small positive mean and low variance, corresponding to “calm” or “bull” regimes.
- The neutral state captures noisy days around zero with intermediate variance.
- The inferred regimes often align with known historical events such as crashes or rallies, suggesting that the HMM captures meaningful structure in the data.

VI. DISCUSSION

The Gaussian HMM framework treats returns as generated from a mixture of Gaussian distributions whose mixing proportions evolve according to a Markov chain. This allows:

- Explicit modeling of regime persistence through the transition matrix.
- Separate characterization of each regime via its mean and variance.
- A principled way to “decode” the most likely regime at each time.

However, there are also limitations:

- The Gaussian assumption may be too simplistic for heavy-tailed financial returns.
- The choice of number of hidden states K affects interpretability and fit; too many states overfit noise.
- The model assumes time-homogeneous transitions, which may not hold over very long horizons.

Despite these caveats, the experiment shows that even a simple 1-D Gaussian HMM can provide non-trivial insight into an asset’s volatility regimes.

VII. CONCLUSION

This assignment demonstrated the application of Gaussian Hidden Markov Models to financial time series. Starting from raw daily prices, we computed log returns, fit a 3-state Gaussian HMM, and interpreted the hidden states as different volatility regimes. The inferred regimes exhibited strong temporal persistence and were associated with distinct mean and variance structures.

Such models can be used as building blocks for risk management (e.g., adjusting exposure in high-volatility regimes) or for exploratory analysis of market behaviour. Extensions could include multivariate HMMs (for portfolios), heavy-tailed emission distributions, or regime-based trading rules.

REFERENCES

- [1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson, 2021.
- [2] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed., draft, 2020.
- [3] L. R. Rabiner, “A tutorial on Hidden Markov Models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, 1989.
- [4] CS367/659 Artificial Intelligence Lab Manual, IIIT Vadodara, Autumn 2025–26.