Gaussian HMM for Regime Detection in Financial Time Series: A Case Study on the Nikkei 225 Index (1965–2025)

Smit Patel, Mann Mehta, Nihar Patel

Department of Computer Science
Indian Institute of Information Technology Vadodara
Gandhinagar, Gujarat, India

Email: {20251602018, 20251602012, 20251603011}@iiitvadodara.ac.in

***Abstract*—Financial markets are known to exhibit regime-switching behavior, alternating between periods of growth, stagnation, drawdowns, and crashes. In this paper, we apply a Gaussian Hidden Markov Model (HMM) to the Nikkei 225 index using daily data from 1965–2025. The model identifies four hidden regimes with distinct return and volatility profiles, analyzes transition probabilities, and quantifies state persistence. We further demonstrate how regime inference can be used for risk management and portfolio decision-making. Our results show that the HMM successfully captures economically meaningful regimes and provides a data-driven framework for financial forecasting and asset allocation.**

***Index Terms*—Gaussian Hidden Markov Model, Financial Time Series, Volatility Regimes, Risk Management, Nikkei 225, Market Crashes.**

I. Introduction

Financial markets do not move in a single statistical regime. Periods of steady growth are often interrupted by short bursts of high volatility, crashes, and structural breaks. Traditional models such as ARIMA or constant-volatility models fail to capture these changing dynamics. Regime-switching models provide an alternative that allows *hidden states* to govern market behavior.

In this work, we apply a Gaussian Hidden Markov Model (HMM) to nearly six decades of historical data from the Nikkei 225 index (Japan) to:

- detect hidden market regimes,
- quantify return and volatility per regime,
- compute transition probabilities between regimes,
- infer investment strategies based on regime behavior.

II. Data and Preprocessing

The dataset consists of daily OHLC and volume data for the Nikkei 225 index from **1965-01-06 to 2025-11-21**.

A. Cleaning

The following steps were applied:

- Converted the `Date` column to `datetime` and set it as index.
- Handled missing data via forward/backward filling.
- Converted all numeric columns using `pd.to_numeric(...)`.

B. Returns

Since prices are non-stationary, log returns were computed:

$$\text{Log\_Return}_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

Log return was used as the primary feature for the HMM.

III. Gaussian HMM Model

We used the `GaussianHMM` from `hmmlearn`. Observations:

$$X = \texttt{df[['Log\_Return']].values}$$

A. Selecting the Number of States

Models were tested with $n \in \{2, 3, 4, 5\}$. The **Bayesian Information Criterion (BIC)** was used:

$$\text{BIC} = -2\log L + k\log(T)$$

The **4-state model** achieved the lowest BIC and was selected.

B. Model Fitting

```
model = GaussianHMM(n_components=4, n_iter=1000,
                    covariance_type='full', random_
model.fit(X)
hidden_states = model.predict(X)
```

Each day was assigned one state $\{0, 1, 2, 3\}$.

IV. Regime Analysis

We computed statistical properties per regime:

- mean return,
- variance,
- volatility,
- risk score (based on volatility ranking).

This research was conducted for academic purposes and uses publicly available financial data.

State	Mean	Var	Std	Count	Risk
2	0.000882	0.000035	0.005889	6193	1
1	0.000281	0.000160	0.012656	7540	2
0	-0.002095	0.000301	0.017360	319	3
3	-0.003550	0.000998	0.031593	917	4

TABLE I
REGIME-WISE STATISTICS FROM THE HMM

A. *Regime Statistics*

B. *Interpretation*

- **State 2** – Bull Market (Low Risk)
- **State 1** – Neutral / Sideways
- **State 0** – Mild Bear
- **State 3** – Crash / Panic (High Risk)

V. TRANSITION MATRIX

The transition probabilities were estimated as:

$$P = \begin{bmatrix} 0.7285 & 0.0000 & 0.2538 & 0.0177 \\ 0.0000 & 0.9815 & 0.0063 & 0.0122 \\ 0.0462 & 0.0035 & 0.9503 & 0.0000 \\ 0.0023 & 0.0833 & 0.0000 & 0.9144 \end{bmatrix}$$

High diagonal values indicate regime persistence. Crash states tend to shift into neutral states rather than bull states directly.

VI. FINANCIAL INTERPRETATION

Each regime supports a practical portfolio strategy:

- **Bull** $\rightarrow$ Increase equity exposure.
- **Neutral** $\rightarrow$ Market-neutral strategies.
- **Mild Bear** $\rightarrow$ Partial hedging.
- **Crash** $\rightarrow$ Move to bonds/cash.

VII. CONCLUSION

The Gaussian HMM captures distinct economic regimes in financial markets and can support risk-aware investment decisions. It identifies four interpretable states, quantifies regime durations, and enables short-term forecasting based on transition probabilities. Future extensions may include multivariate HMMs or heavy-tailed emission models such as Student-t HMMs.

APPENDIX

Full implementation is available at:
https://colab.research.google.com/drive/1drbIqcaWopVDfvA_yvGLQ8NwqaGFbXLm?usp=sharing

REFERENCES

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed., Prentice Hall, USA, 2009.
[2] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed., A Bradford Book, Cambridge, MA, USA, 2018.
[3] D. Khemani, *A First Course in Artificial Intelligence*, McGraw Hill Education, New Delhi, 2013.