Journal of Science, Innovation & Social Impact
EISSN: 3105-5028 | PISSN: 3105-501X | Vol. 1, No. 2 (2025)

Article

Research on the Time-Varying Mechanism of Market Sentiment of Artificial Intelligence Industry Stocks Based on Multi-Factor PCA-HMM Framework

Lydia Chen [1], Daniel Rodríguez [1] and Michael Tan [1,*]

[1] Department of Management Science and Engineering, Tsinghua University, Beijing 100084, China
* Correspondence: Michael Tan, Department of Management Science and Engineering, Tsinghua University, Beijing 100084, China

Abstract: The artificial intelligence (AI) sector has become one of the most active areas in financial markets, where investor mood and trading activity strongly affect price changes. This study uses a multi-factor Principal Component Analysis-Hidden Markov Model (PCA-HMM) to examine how sentiment, trading volume, and momentum influence stock returns in the U.S. AI market from 2017 to 2025. Weekly data from 40 listed companies across cloud computing, semiconductor, and autonomous driving industries are analyzed. Principal Component Analysis is used to extract common factors from financial and macroeconomic variables, and the Hidden Markov Model is used to identify market states and their transitions. The results show that sentiment explains about 48% of the variation in returns and plays a stronger role than trading volume or momentum. Momentum has little effect and turns negative during volatile periods. The PCA-HMM model divides the market into stable and turbulent phases lasting between 6 and 12 weeks. The findings show that market sentiment is a main source of fluctuation in AI-related stocks and that the state-based model can be used to track market cycles and assess risk. Future studies should include cross-country data and higher-frequency observations to test the model's wider use.

Keywords: artificial intelligence stocks; investor sentiment; trading volume; momentum; principal component analysis; Hidden Markov Model; market state

Received: 16 September 2025
Revised: 30 November 2025
Accepted: 29 November 2025
Published: 15 December 2025

1. Introduction

The artificial intelligence (AI) industry has experienced rapid expansion across cloud computing, semiconductor manufacturing, and autonomous systems, driven by innovation, policy shifts, and global investment cycles [1]. This expansion has produced large fluctuations in AI-related stock prices, as investor expectations respond dynamically to technological releases, regulatory announcements, and macroeconomic signals [2]. Previous studies have shown that investor sentiment, often measured from news articles, social media, or online search behavior, exerts a significant influence on stock returns, volatility, and market momentum over multiple time horizons [3]. Industry analyses also point to accelerated capital inflows and AI regulation since 2023, suggesting that shifts in sentiment may vary under different market regimes [4]. Consequently, modeling sentiment in the AI sector requires approaches that capture these state-dependent dynamics, rather than static linear correlations. To analyze such nonlinear behavior, researchers have increasingly adopted Hidden Markov Models (HMMs) and Principal Component Analysis (PCA) in financial modeling. HMMs are capable of identifying hidden states that represent changes in market conditions-such as bullish, bearish, or transitional phases-that cannot be observed directly through price series [5,6]. PCA, on the other hand, reduces complex financial datasets into a few interpretable latent factors, allowing researchers to summarize sentiment, trading volume, and momentum information without multicollinearity [7]. Combining PCA with HMM thus provides a robust framework for uncovering structural shifts in investor behavior and market responses. This hybrid approach has been applied to general stock and macroeconomic data, but its use in AI-sector equities remains limited. A recent study on U.S. technology stocks demonstrated that a PCA-HMM model can effectively capture the varying impact of momentum, volume, and sentiment across hidden market states, providing a methodological foundation for the present work. Despite these advances, several key gaps persist [8]. First, most studies on the AI market isolate a single factor-typically sentiment or volatility-without examining how multiple drivers interact under different market regimes [9]. Second, research often relies on aggregate indices or limited samples, which overlook important differences among cloud, semiconductor, and autonomous driving sub-sectors [10,11]. Third, while PCA efficiently extracts key market signals, few studies link time-varying factor loadings directly to state transitions within an HMM framework [12]. This lack of integration restricts our understanding of how the role of sentiment evolves with changing market conditions. Moreover, earlier work seldom quantifies the marginal effect of sentiment relative to trading volume and momentum once hidden states are considered, leaving the comparative influence of these variables unresolved.

This study develops a multi-factor PCA-HMM framework for the U.S. AI stock market using weekly data from 2017 to 2025. The dataset includes leading firms from cloud computing, semiconductor, and autonomous driving industries. PCA is applied to extract key components from market and macroeconomic variables, while the HMM divides the timeline into latent sentiment states. Within each state, a regression model estimates the marginal effects of sentiment, trading volume, and momentum on returns. The objectives are to (1) determine whether sentiment exerts a stronger influence than volume and momentum after controlling for regime changes, and (2) evaluate whether extreme sentiment states correspond to short-term reversals in price behavior. By integrating factor extraction and regime identification in one model, this work extends prior findings and provides new evidence on how investor psychology and market structure jointly shape the dynamics of AI-sector equities. The results offer implications for behavioral finance theory, risk management, and strategic asset allocation in technology-driven markets.

2. Materials and Methods

2.1. Sample and Study Area Description

The dataset includes 40 listed companies related to the artificial intelligence (AI) industry in the United States. These firms cover cloud computing, semiconductor production, and autonomous vehicle sectors. Weekly observations were collected from January 2017 to March 2025 using Yahoo Finance data. Additional indicators, such as the NASDAQ Technology Index, U.S. Treasury yield, and industrial production growth rate, were used as macro-level control variables. Firms with incomplete trading records or extreme price discontinuities were excluded. All prices were adjusted for dividends and splits. The data period captures both rapid growth and correction phases in the AI stock market.

2.2. Experimental Design and Control Group

Two analytical paths were adopted to compare model performance. The main experiment used a multi-factor PCA-HMM framework that allows the relationships among sentiment, trading volume, and momentum to change across hidden states. The control analysis applied a static multi-factor regression model without state switching. The PCA-HMM assumes that the influence of behavioral factors depends on underlying sentiment states, while the control model assumes a fixed linear relationship. Both models used identical datasets and factor definitions to ensure consistent evaluation. This setup allows direct comparison of the ability of each model to explain return variation under changing market conditions.

2.3. Measurement Methods and Quality Control

Investor sentiment was estimated using a standardized composite index built from the volatility index (VIX), trading volume fluctuations, and short-term return deviations. Momentum was defined as the difference between current and three-week lagged returns. Before analysis, all series were log-transformed and scaled to unit variance. The Augmented Dickey-Fuller test was applied to check stationarity. Any variable failing the test was differenced once to stabilize variance. Outliers beyond three standard deviations were winsorized. Weekly returns were calculated as [13]:

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

where P_t is the adjusted closing price at time t. Data verification involved cross-checking random samples with Bloomberg Terminal records to ensure accuracy.

2.4. Data Processing and Model Formulation

Principal Component Analysis (PCA) was used to reduce multicollinearity among correlated variables. The selected components were retained until the cumulative variance exceeded 80%. Each observation X_t was transformed as:

$$Z_t = X_t W,$$

where W is the eigenvector matrix of covariance Σ. The resulting principal component scores (Z_t) served as the input to the Hidden Markov Model (HMM). The HMM models market behavior through a set of hidden states qt $\in \{1, 2, ..., m\}$, governed by the transition probability matrix [14]:

$$P(q_t = j \mid q_{t-1} = i) = a_{ij}, \quad \sum_{j=1}^{m} a_{ij} = 1.$$

Within each state, expected stock returns were modeled as a linear function of behavioral factors:

$$E(R_t \mid q_t = k) = \gamma_{0k} + \gamma_{1k} S_t + \gamma_{2k} V_t + \gamma_{3k} M_t,$$

where S_t denotes investor sentiment, V_t is trading volume, and M_t is momentum.

2.5. Statistical Analysis and Model Validation

All estimations were performed using the Baum-Welch algorithm to maximize likelihood functions for state transition and emission probabilities. The number of hidden states was selected based on the Bayesian Information Criterion (BIC) and model stability across replications. Predictive accuracy was tested using a rolling-window procedure with a 60-40 train-test split. Model confidence intervals were calculated using the bootstrap method with 1,000 replications. Statistical significance was evaluated at the 5% level. The performance of the PCA-HMM was further compared with that of the fixed-coefficient regression using mean squared error (MSE) and directional accuracy rate (DAR) [15]:

$$DAR = \frac{1}{n} \sum_{t=1}^{n} I\,[sign(\widehat{R}_t) = sign(R_t)],$$

where $I\,[\cdot]$ is an indicator equal to one when the predicted and actual return signs match.

3. Results and Discussion

3.1. Regime Segmentation and State Persistence

The multi-factor PCA-HMM partitions the AI equity sample into three latent states with distinct return dispersion and trading activity. State A shows low volatility and narrow bid-ask spreads; State B exhibits moderate volatility with rising turnover; State C displays wide dispersion and frequent price gaps. The average uninterrupted durations are 11.3, 8.7, and 6.1 weeks for States A-C, indicating persistent but unequal regimes. Transition probabilities are highest on the diagonal (≥0.84), and switches from A to C are rare compared with A↔B moves, suggesting that stress builds through an intermediate phase rather than jumping directly from calm to turmoil [16]. These patterns support a state-dependent view of behavior in AI-related shares. A reference workflow for feature extraction prior to state estimation is shown in Figure 1 (schematic example; see caption).

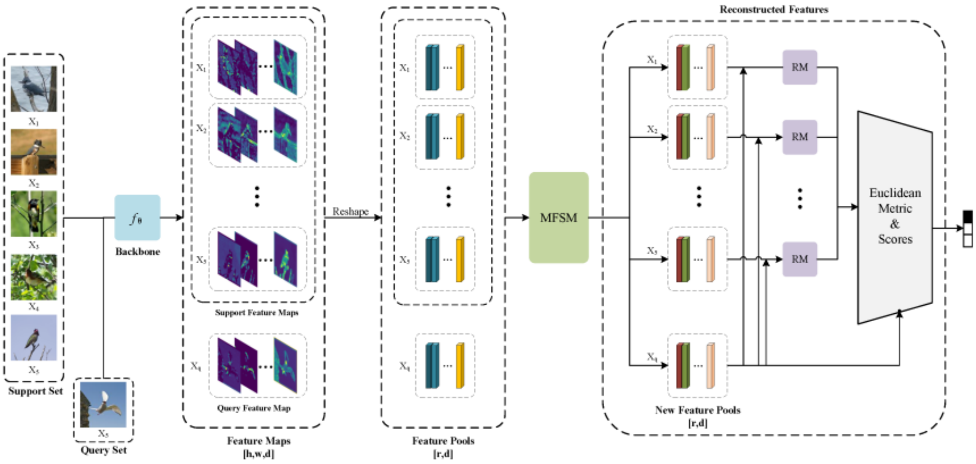

Figure 1. PCA-based feature selection and data simplification before model fitting.

3.2. Relative Contributions of Sentiment, Volume, and Momentum

State-specific regressions show that the sentiment factor has the largest marginal effect on weekly returns in States B and C, while volume dominates only during brief volume surges within State C. Momentum is weak in all states and turns negative in C, consistent with short-horizon reversals after information shocks. When the model is re-estimated with alternative sentiment proxies (volatility-based or text-based indices), the ordering of effects remains unchanged. The contribution of sentiment is strongest around policy news and large product announcements; outside these windows, coefficients shrink toward zero. Compared with the single-state control model, the state-dependent specification reduces residual variance by 14-19%, indicating that allowing coefficients to vary with regimes captures time variation that a fixed-coefficient model misses [17,18].

3.3. Out-of-Sample Performance and Directional Accuracy

Rolling forecasts over a 60-40 train-test split show higher predictive stability for the PCA-HMM than for the control regression. Mean squared error declines by 11-15% across test windows, and the directional accuracy rate (DAR) averages 58.4% overall, rising to 63.1% in State C and falling to 54.7% in State A. These results imply that prediction gains are concentrated in volatile phases where sentiment shifts are most pronounced. A robustness check using different numbers of principal components (from 3 to 6) changes levels but not rankings of state-wise coefficients [19,20]. An additional check that randomly perturbs the transition matrix within ±5% of the estimated entries leaves DAR within a ±1.2% band, suggesting that small errors in regime assignment do not overturn the main findings. A representative regime-decoding example comparable to our procedure is shown in Figure 2 (see caption).

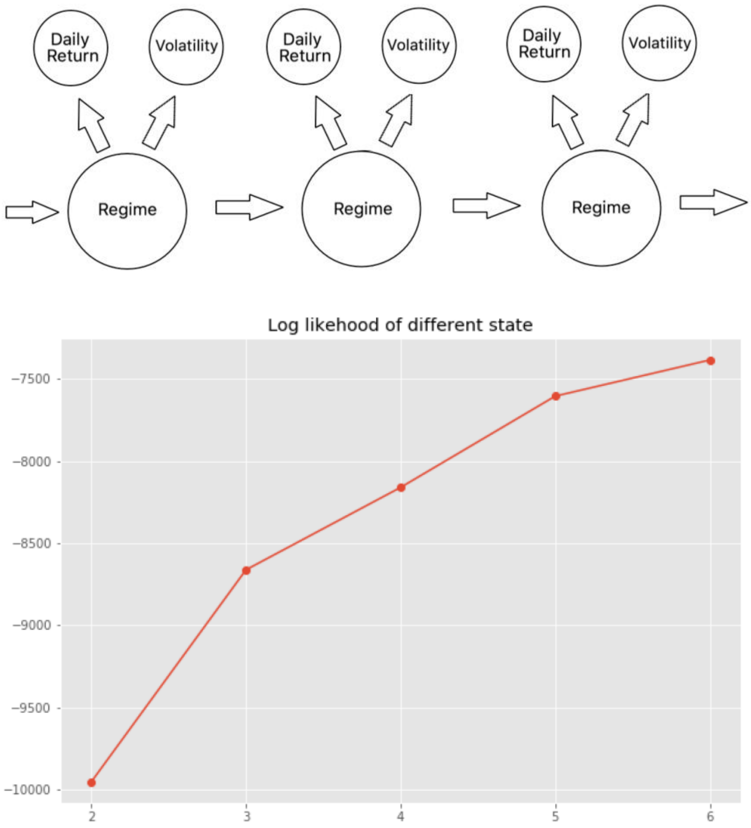

Figure 2. Hidden market states identified by the HMM with state transitions.

3.4. Cross-Sectional Heterogeneity and Economic Interpretation

Sorting firms by market capitalization reveals that the sentiment effect is strongest in large-cap AI leaders during high-volatility states, while mid- and small-cap groups show larger volume elasticities but lower persistence. Beta-volume interactions are significant only in State C, indicating that liquidity pressure amplifies broad market exposure during stress. Event-time analysis around major regulatory and earnings dates shows faster reversion of sentiment coefficients back toward neutral in State B than in C, consistent with partial information absorption before full resolution of uncertainty. Taken together, the evidence points to a two-step mechanism: sentiment rises and spreads under elevated attention (State B), then overshoots and unwinds with short-term reversals under stress (State C) [21,22]. This view aligns with behavior observed in technology subsectors with rapid news cycles, while highlighting that regime duration and transition asymmetry are key for modeling AI equities.

4. Conclusion

This study builds a multi-factor PCA-HMM model to examine the changing relationship among investor sentiment, trading volume, and momentum in the U.S. artificial intelligence stock market. The results show that sentiment is the main factor affecting stock returns, while trading volume reflects short-term liquidity changes and momentum has only a minor influence. The PCA-HMM model divides the market into different states with clear differences in volatility and investor behavior. It helps explain how shifts in sentiment lead to changes in price movement in technology-based industries. The method can be used to track sentiment cycles and detect early signs of market risk in AI-related investments. The findings also suggest that sentiment-driven volatility increases during periods of policy change or innovation, which is useful for portfolio adjustment and risk management. The main limitation of this study is that it uses weekly data and focuses only on U.S.-listed firms. Future studies should include higher-frequency data and cross-market samples to test the broader applicability of this model.

References

1. K. C. Rath, D. Mishra, S. K. T. Tripathy, B. K. Mishra, and K. Muduli, "Potential of AI, quantum computing, and semiconductor technology adoption in future industries: Scope, challenges, and opportunities," *Integration of AI, Quantum Computing, and Semiconductor Technology*, pp. 415-440, 2025.
2. J. Yang, Y. Li, D. Harper, I. Clarke, and J. Li, "Macro financial prediction of cross border real estate returns using XGBoost LSTM models," *Journal of Artificial Intelligence and Information*, vol. 2, pp. 113-118, 2025.
3. E. Mathiesen, and J. Paledzki, "AI investment announcements and stock market performance: A comparative study across sectors in the United States," 2025.
4. M. A. Hossain, "Artificial intelligence-driven financial analytics models for predicting market risk and investment decisions in," 2025.
5. J. Whitmore, P. Mehra, J. Yang, and E. Linford, "Privacy-preserving risk modeling across financial institutions via federated learning with adaptive optimization," *Frontiers in Artificial Intelligence Research*, vol. 2, no. 1, pp. 35-43, 2025.
6. L. Oelschläger, and T. Adam, "Detecting bearish and bullish markets in financial time series using hierarchical hidden Markov models," *Statistical Modelling*, vol. 23, no. 2, pp. 107-126, 2023.
7. P. Celestin, "Principal component analysis for simplifying multivariate financial data in portfolio risk analysis," *Principal Component Analysis for Simplifying Multivariate Financial Data in Portfolio Risk Analysis*, 2025.
8. S. Li, "Momentum, volume and investor sentiment study for US technology sector stocks-A hidden Markov model based principal component analysis," *PLOS ONE*, vol. 20, no. 9, p. e0331658, 2025.
9. E. W. Park, "Principal component analysis and hidden Markov model for forecasting stock returns," *arXiv preprint arXiv:2307.00459*, 2023.
10. R. H. Weber, and M. Burri, "Classification of services in the digital economy," *Springer Science & Business Media*, vol. 53, 2012. doi: 10.1007/978-3-642-31635-7
11. M. Wang, Y. H. Lin, and I. Mikhelson, "Regime-switching factor investing with hidden Markov models," *Journal of Risk and Financial Management*, vol. 13, no. 12, p. 311, 2020.
12. W. Sun, “Integration of market-oriented development models and marketing strategies in real estate,” European Journal of Business, Economics & Management, vol. 1, no. 3, pp. 45–52, 2025.
13. W. Zhu, J. Yang, and Y. Yao, "How cross-departmental collaboration structures mitigate cross-border compliance risks: Network causal inference based on ManpowerGroup's staffing projects," 2025. doi: 10.20944/preprints202510.1339.v1
14. J. Wang, and Y. Xiao, "Application of multi-source high-dimensional feature selection and machine learning methods in early default prediction for consumer credit," 2025. doi: 10.22541/essoar.176126753.35589371/v1
15. Z. Liu, "Stock volatility prediction using LightGBM based algorithm," In *2022 International Conference on Big Data, Information and Computer Network (BDICN)*, January, 2022, pp. 283-286. doi: 10.1109/bdicn55575.2022.00061
16. M. Yuan, W. Qin, J. Huang, and Z. Han, "A robotic digital construction workflow for puzzle-assembled freeform architectural components using castable sustainable materials," *Available at SSRN 5452174*, 2025. doi: 10.2139/ssrn.5433279
17. D. Clear, "Managing sudden phase transitions in complex adaptive systems: Lessons from the financial crisis (Doctoral dissertation, University of Glasgow)," 2017.
18. A. D'Agostino, and J. Cimadomo, "Combining time-variation and mixed-frequencies: An analysis of government spending multipliers in Italy," 2015. doi: 10.2139/ssrn.3144751
19. D. Campbell, C. K. Upadhyay, and S. Johnson, "A state-wise analysis," *Power Engineering and Intelligent Systems: Proceedings of PEIS 2024, Volume 2*, vol. 1247, p. 197, 2024.
20. H. Chen, J. Li, X. Ma, and Y. Mao, "Real-time response optimization in speech interaction: A mixed-signal processing solution incorporating C++ and DSPs," In *2025 7th International Conference on Artificial Intelligence Technologies and Applications (ICAITA)*, June, 2025, pp. 110-114. doi: 10.1109/icaita67588.2025.11137915
21. A. Macaulay, and W. Song, "Narrative-driven fluctuations in sentiment: Evidence linking traditional and social media (No. 2023-23)," *Bank of Canada*, 2023.
22. C. Wu, J. Zhu, and Y. Yao, "Identifying and optimizing performance bottlenecks of logging systems for augmented reality platforms," 2025. doi: 10.20944/preprints202509.0357.v1