Check for updates

Market state transitions and crash early warning in the Chinese stock market

Wu-Yue Pang[1,2] and Li Lin[1,3]*

[1]Department of Finance, Business School, East China University of Science and Technology, Shanghai, China, [2]School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China, [3]Risk-Center, ETH Zürich, Zürich, Switzerland

Understanding systemic risk in financial markets requires tools that capture both structural complexity and dynamic evolution. This study adopts a complex systems perspective to analyze the Chinese stock market through correlation-based state classification. Using multidimensional scaling and K-means clustering on rolling stock return correlations, we identify five distinct market states that reflect varying degrees of systemic co-movement and exhibit strong temporal persistence and local transition patterns. We find that transitions among these states encode meaningful information about market structural shifts and are closely linked to the emergence of crash conditions. Building on this insight, we construct an early warning model using decision trees trained on temporal features derived from market state transitions—including medium-term state distributions, directional change ratios, and short-term evolutionary paths. The model achieves high recall and precision across configurations, and supports real-time adaptability via projection-based state labeling. These results highlight the value of market state dynamics and their transitions as a basis for systemic risk monitoring and crash anticipation in complex financial systems.

KEYWORDS

market state transition, systemic risk, early warning, correlation structure, stock market

1 Introduction

Financial markets are widely regarded as complex adaptive systems, in which the interactions among heterogeneous agents give rise to emergent phenomena such as volatility clustering, contagion, and crashes [1–4]. Capturing these dynamics requires moving beyond isolated asset-level indicators to examine the structure and evolution of market-wide interactions. In this context, correlation-based and network-theoretic representations have emerged as powerful tools for modeling systemic behavior. Early studies constructed networks from return correlation matrices to quantify collective dynamics and market fragility [5, 6], while more recent work has introduced higher-order interaction structures, including Granger-causality networks [7], speculative influence networks [8], portfolio overlap networks [9], and Ricci-curvature-based geometric approaches [10, 11]. These approaches collectively highlight the importance of modeling financial markets as evolving structures whose topological and temporal features govern their stability—serving, in effect, as emergent order parameters that summarize the global configuration of the system [12, 13]. In parallel, recent advances in econometric modeling have enriched this structural perspective through dynamic correlation and network

connectivity analyses [14], contagion mechanism estimation [15], and cross-sectional uncertainty metrics [16], offering additional insights into the spatiotemporal propagation of systemic stress.

Building on these foundations, a growing body of research has demonstrated that financial markets can be decomposed into a small number of discrete market states, derived from clustering time-dependent correlation matrices [17–19]. These states are understood to reflect varying levels of market-wide coordination, with high-correlation regimes typically aligned with heightened systemic risk. Crucially, transitions between these states are not random but exhibit structured, often persistent patterns—indicative of path-dependent dynamics rooted in collective behavior and feedback effects [20, 21]. This suggests that financial instability is less a sudden shock than a gradual process, emerging from shifts in the system's internal organization and the evolution of its interaction topology. Such state transitions are also closely related to latent patterns of risk contagion and inter-market connectivity, as highlighted in recent econometric models of dynamic transmission mechanisms and time-zone-adjusted VAR frameworks [15]. Moreover, advances in modeling cross-sectional uncertainty have shown promise in enhancing correlation-based classification methods, offering insights into the predictive structure of latent regimes [16]. These perspectives collectively underscore the need to treat state transitions not merely as clustering artifacts, but as meaningful structural shifts reflecting endogenous risk propagation.

Despite these advances, several important aspects remain underexplored in the existing literature. First, most studies are retrospective in nature—focusing on identifying and interpreting market states ex post—without translating these structural insights into forward-looking monitoring tools. Second, while emerging research suggests that market state transitions encode valuable early signals of systemic stress [22, 23], relatively few studies have formally integrated such temporal patterns into operational early warning systems. This issue is especially relevant in the context of emerging markets such as China, where distinctive market microstructures from developed market and investor behaviors complicate risk assessment. Existing models often lack structural interpretability and real-time adaptability, underscoring the need for monitoring frameworks that reflect the endogenous evolution of systemic fragility [24–26].

To address the aforementioned limitations, this paper proposes a structure-aware early warning framework that explicitly incorporates the temporal dynamics of market state transitions. Using the CSI 300 Index as a representative system, we first identify five distinct market states through a combination of correlation-based similarity measures, multidimensional scaling (MDS), and K-means clustering. These states correspond to different levels of systemic coupling and exhibit empirically stable transition patterns: transitions occur predominantly between adjacent states, while high-correlation regimes are disproportionately associated with historical crash periods.

Leveraging these insights, we develop a decision tree-based early warning model constructed from temporally structured features of market state dynamics. These include medium-term state distributions, directional transition ratios, and recent evolutionary motifs. The model demonstrates high recall and precision under cross-validation and remains robust when extended to new data via an efficient projection-based labeling mechanism. Crucially,

unlike traditional indicators based on instantaneous correlation structure—such as the Absorption Ratio (AR) [6] and its extension ARR [27], eigenvalue-based early warning signals [20], or curvature-derived fragility measures [10, 11]—our framework emphasizes the predictive value of how structural regimes evolve over time. This transition-centric approach provides a dynamic lens on market fragility, enhancing the model's interpretability and its potential for real-time monitoring in complex financial environments.

To the best of our understanding, few studies have systematically leveraged market state transitions—interpreted as emergent order-parameter dynamics within complex systems—to inform supervised early warning models, particularly in the context of emerging markets. This study seeks to bridge complex systems modeling with practical risk monitoring by operationalizing structurally grounded state transitions as interpretable signals for crash forecasting.

The remainder of the paper is organized as follows. Section 2.1 introduces the methodology for identifying correlation-driven market states and modeling their temporal transitions. Section 2.2 details the decision tree-based early warning model, including the design of interpretable state-informed features and projection-based labeling of new data. Section 3.2 presents empirical analysis of 15 years of CSI 300 data, covering market state classification, transition dynamics, and their link to crash periods. Section 4 evaluates the early warning model, focusing on feature configuration tuning, performance assessment, and interpretation of decision paths related to pre-crash signals. Section 5 concludes with key findings and implications for systemic risk monitoring and future research.

2 Methods

2.1 Market state classification and labeling

We conceptualize the financial market as a complex system, where its internal structure is reflected in the correlation patterns among individual stocks. A market state, in this context, is defined by the collective behavior of these correlations. To distinguish different states over time, we assess how the correlation structure evolves across different time windows, mainly following the approach proposed by Pharasi et al. [17, 19]. Specifically, we measure the similarity of correlation matrices computed over these windows, treating similar correlation patterns as indicators of the same underlying market state. To construct the correlation matrices, we begin with daily post-rights-adjusted prices of all stocks and assemble them into a data matrix. The logarithmic return of each stock i is calculated as

$$r_i(t) = \ln P_i(t) - \ln P_i(t-1),$$

where t denotes the end time point of the corresponding time window. Using these returns, we derive the correlation matrix $\rho(\tau)$ for each time window, with matrix elements given by the Pearson correlation coefficients between stock return pairs.

$$\rho_{ij}(\tau) = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{\langle r_i^2 \rangle - \langle r_i \rangle^2} \sqrt{\langle r_j^2 \rangle - \langle r_j \rangle^2}}. \quad (1)$$

The average $\langle\ldots\rangle$ is evaluated within a fixed-length time window ending at time point τ, spanning M discrete time intervals. The subscripts $i, \quad j = 1,\ldots,N$ denote the stocks. To avoid ambiguity, we distinguish between two time-related indices: t refers to the actual trading day, while τ denotes the index of the rolling window, each of which spans a fixed number of trading days (e.g., $M = 22$). Thus, each window τ aggregates market information across the interval $[t - M + 1, t]$, enabling the construction of smoothed correlation matrices $C(\tau)$ for dynamic structural analysis. Since correlation matrices derived from short time series tend to be highly singular, we adopt the power mapping method to suppress noise in these matrices. This approach improves the stability and discriminative power of the correlation matrix spectra, as discussed in previous works like Guhr and Kälber [28]. The nonlinear exponentiation effectively reduces spurious correlations while preserving genuine structural relationships. In this approach, each Pearson correlation coefficient within a time window is transformed nonlinearly as

$$C_{ij}(\tau) = \text{sign}\left(\rho_{ij}(\tau)\right)\left|\rho_{ij}(\tau)\right|^{1+\epsilon},$$

where $\epsilon \in (0,1)$ is a parameter that controls the degree of noise suppression.

By applying this transformation, we obtain a sequence of noise-suppressed correlation matrices $C(\tau) = \left[C_{ij}(\tau)\right]_{N \times N}$ that capture the dynamic structure of stock return correlations over time. In our analysis, we use a rolling time window of $M = 22$ trading days (approximately 1 month), with a sliding step of $\Delta = 1$ day to ensure high temporal resolution.

We now define a similarity measure (or distance) between two correlation matrices, $C(\tau_1)$ and $C(\tau_2)$, evaluated at different ending time points τ_1 and τ_2, as:

$$d(\tau_1, \tau_2) \equiv \overline{|C(\tau_1) - C(\tau_2)|}, \tag{2}$$

where the overline denotes the average over all matrix elements. These pairwise distances are then used as input for Multidimensional Scaling (MDS), which projects the matrices into a low-dimensional space while preserving their mutual distances within an acceptable tolerance. MDS aims to map each correlation matrix $C(\tau)$ to a point in a lower-dimensional space, such that the pairwise distances between matrices are represented by a distance matrix $\zeta(\tau_1, \tau_2)$. This matrix is defined as:

$$\zeta(\tau_1, \tau_2) = \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,F_r} \\ \delta_{2,1} & \delta_{2,2} & \cdots & \delta_{2,F_r} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{F_r,1} & \delta_{F_r,2} & \cdots & \delta_{F_r,F_r} \end{pmatrix}, \tag{3}$$

where each entry $\delta_{i,j}$ represents the computed distance between matrices $C(\tau_i)$ and $C(\tau_j)$. The goal of MDS is to assign coordinates $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{F_r}$ in a low-dimensional Euclidean space such that the Euclidean distance $\|\mathbf{x}_i - \mathbf{x}_j\|$ approximates the original dissimilarities $\delta_{i,j}$, i.e.,

$$\|\mathbf{x}_i - \mathbf{x}_j\| \approx \delta_{i,j}, \quad \forall i,j \in \{1,2,\ldots,F_r\},$$

where $\|\cdot\|$ denotes the Euclidean norm. In doing so, MDS provides a spatial representation of the F_r correlation matrices in a reduced-dimensional space $\mathbb{R}^D$.

```
1: Input: Set of Fr data points from MDS, number
   of clusters K
2: Output: Cluster assignments and centroids
3: Randomly initialize K cluster centroids among
   the Fr data points
4: repeat
5: Compute the Euclidean distance from each data
   point to all centroids
6: Assign each point to the cluster with the
   nearest centroid
7: Update each centroid by averaging the positions
   of all points in the cluster
8: until cluster assignments and centroid
   positions converge
```

Algorithm 1. *K*-Means Clustering for Market State Classification.

In this study, we select $D = 2$ or $D = 3$ to enable effective visualization. The resulting coordinates $\mathbf{x}_i$ are then used to generate two- or three-dimensional mappings that capture the structural relationships among correlation matrices over time. Structurally similar matrices are projected to nearby points, whereas dissimilar ones appear farther apart. Thus, MDS serves as a powerful tool to visualize and explore the dynamic evolution of market states through the similarity structure encoded in $\zeta(\tau_1, \tau_2)$.

We then apply *K*-means clustering to partition the F_r data points obtained from the MDS mapping into K distinct clusters, each defined by a cluster centroid. The algorithm proceeds through the following iterative steps:

To determine the optimal number of clusters K and the appropriate noise suppression parameter ϵ, we evaluate the clustering quality using the intra-cluster distance d_{intra}. This metric is defined as the average of the mean Euclidean distances from each point to its respective cluster centroid, aggregated across all clusters. For each candidate value of K, we perform the clustering under 1,000 different random initializations and compute the standard deviation of the resulting intra-cluster distances, denoted as $\text{std}_{\text{intra}}$. This procedure enables robust assessment of clustering stability and helps identify a reliable choice for K.

To determine the optimal number of clusters K, we introduce a pseudo-AIC criterion that balances intra-cluster dispersion and model complexity:

$$\text{AIC}_{\text{pseudo}} = 2K + n \ln\left(\text{std}_{\text{intra}}\right),$$

where K is the number of clusters, n is the number of samples, and $\text{std}_{\text{intra}}$ is the average intra-cluster standard deviation.

While this formulation resembles the structure of the Akaike Information Criterion (AIC), it is not derived from a traditional likelihood framework. Rather, it draws on a probabilistic interpretation of K-means under isotropic statistical assumptions and inspired by the information-theoretic model selection principle of balancing model complexity and goodness-of-fit [29]. To avoid confusion, we refer to it as a *pseudo*-AIC, and provide a detailed justification of this formulation in Supplementary Appendix A. The value of *pseudo*-AIC balances model fit and complexity, penalizing

```
 1: Input: Daily closing prices P_t of a selected
    market index
 2: Parameters: Time window length N, drawdown
    threshold η
 3: Output: Labels y_t ∈ {0,1} for each trading day
 4: for each time window of length N (using sliding
    or fixed approach) do
 5:   for each trading day t in the window do
 6:     Compute drawdown: Drawdown(t) = P_t / max(P_window)
 7:   end for
 8:   if there exists t such that Drawdown(t) < η and t
      is not the first day then
 9:     Let MaxDate be the date of peak price
        in the window
10:     Mark the day before MaxDate as start of
        drawdown (Start_i), assign y = 1
11:     Find the first day t' after Start_i such that
        Drawdown(t') ≥ η
12:     Mark the day before t' as end of drawdown
        (End_i)
13:     Store drawdown window [Start_i, End_i]
14:   end if
15: end for
16: Merge all intervals [Start_i, End_i] to form
    disjoint market crash periods
17: Assign y_t = 0 for all days t outside of
    these periods
```

Algorithm 2. Labeling Market Crash Cases from Market Price Data.

models with a large number of clusters. We select the optimal number of clusters K and the noise suppression parameter ϵ by minimizing the AIC value. This ensures that the selected model achieves both compact and stable intra-cluster structure while avoiding overfitting caused by an excessively large K.

In our analysis, we primarily consider cases where $K \geq 4$, as configurations with fewer than four clusters ($K \leq 3$) are deemed insufficient to capture the diversity of market states in a realistic manner. Once the optimal values of K and ϵ are determined, we compute the average correlation matrix within each cluster. These clusters are then sorted in ascending order based on their average correlation levels and assigned sequential labels corresponding to market states: $S1, S2, S3, \ldots$

2.2 Early warning model for market crash using machine learning methods

The first step in constructing the machine learning-based early warning model is to define the binary classification labels: market crash cases ($y = 1$) and non-crash (normal) cases ($y = 0$). These labels can be generated based on the historical daily closing prices of a representative market index. The labeling algorithm proceeds as follows:

In this paper, we employ the decision trees to construct the market crash early warning model. This modeling choice is motivated by several considerations. Decision trees are inherently simple yet highly interpretable, offering a clear hierarchical structure of decision rules that makes them particularly well-suited for applications requiring transparent reasoning. Second, they are relatively robust to overfitting, especially when the model complexity (e.g., tree depth) is properly controlled. Third, and most importantly in our context, decision trees are well-suited to scenarios with a limited number of input features. Since our early warning model is constructed based on market state transitions—resulting in a compact, semantically meaningful feature set—the ability of decision trees to perform reliable classification without requiring high-dimensional input is a critical advantage. These properties allow the model to maintain strong generalization performance while ensuring interpretability for financial analysts and policy researchers.

To enhance predictive performance, there are three types of features to be implemented to extract patterns associated with market state transitions in pre-crash dynamics:

- Proportions of Specific States in the Past L_1 Windows: $p_k^{L_1}$. Market states are first categorized into K distinct types. Over the most recent L_1 time windows, the frequency of each state is recorded and converted into proportions to reflect their relative prevalence. Let $n_{i,k}$ denote the number of times for market state k appears in the i-th window. The proportion of state k across these windows is computed as:

$$p_k^{L_1} = \frac{\sum_{i=1}^{L_1} n_{i,k}}{L_1} \quad (k = 1, 2, \ldots, K) \qquad (4)$$

This process yields K features, represented by the vector $(p_1, p_2, \ldots, p_K)$, which captures the medium-term structural distribution of market states. These features are designed to reflect the prevailing level of market stability within the observed period.

- Ratio of Upward to Downward State Transitions in the Past L_2 Windows: r^{L_2}. An *upward transition* refers to a shift from a state with a lower index to one with a higher index (e.g., from S_3 to S_5), while a *downward transition* refers to the reverse (e.g., from S_4 to S_3). Let m_i and n_i denote the number of upward and downward transitions observed in the i-th window. The ratio over the most past L_2 windows is computed as:

$$r^{L_2} = \frac{\sum_{i=1}^{L_2} m_i}{\sum_{i=1}^{L_2} n_i}. \qquad (5)$$

Although the state indices themselves are agnostic to specific economic meanings, empirical observations—both from our subsequent empirical results and related literature—suggest that transitions toward higher-indexed states tend to coincide with increased structural complexity or market disorder. Hence, a high value of r^{L_2} may indicate an ongoing shift in market conditions that precedes a turbulent phase. Conversely, a lower ratio may imply a trend toward greater structural coherence or reduced systemic stress. By capturing the relative directional trend in state transitions, this feature provides the model with a coarse-grained but informative indicator of evolving market stability.

(3) Evolutionary Path of Market States in the Nearest L_3 Windows: π^{L_3}. The sequence of market states over the most recent L_3 time windows is encoded as a categorical transition pattern. Let S_i denote the identified market state in the i-th window. The observed pattern is then represented as an ordered sequence:

$$\pi^{L_3} = \left(S_1 \rightarrow S_2 \rightarrow \cdots \rightarrow S_{L_3} \right). \tag{6}$$

Each unique sequence of L_3 consecutive states forms a categorical feature that reflects the short-term evolutionary path of market structure. Some simple examples include $\pi^3 = (S_2 \rightarrow S_3 \rightarrow S_2)$, $\pi^3 = (S_3 \rightarrow S_4 \rightarrow S_5)$, or $\pi^4 = (S_1 \rightarrow S_3 \rightarrow S_4 \rightarrow S_4)$. The total number of distinct features generated depends on both L_3 and the total number of state categories K, with longer sequences enabling more granular pattern differentiation.

These pattern-based features are particularly useful for capturing local dynamics, including sudden structural shifts, oscillatory behavior, or persistent regime patterns within stable/unstable conditions. By encoding the short-run evolutionary trajectory of market states, this feature set complements other indicators focused on medium- and long-term structural characteristics.

Given the asymmetric costs associated with classification outcomes in this context—where missing a true crash event (false negative) may result in significantly more severe consequences than issuing a false alarm (false positive)—the model selection process is designed to prioritize recall. Among models achieving the highest recall, further selection is made by jointly considering overall classification accuracy and minimizing the false positive rate (FPR), thereby ensuring both sensitivity and precision in practical application.

To address the issue of model adaptability over time, we further consider how to incorporate newly arriving data without compromising model consistency. A key challenge lies in assigning market state labels to new incoming time points. Re-clustering the entire dataset whenever new data becomes available is computationally expensive and introduces instability, particularly near cluster boundaries, where minor perturbations may lead to significant reassignments and undermine the validity of previously extracted features. Moreover, substantial updates in data may necessitate a recalibration of clustering hyper-parameters, including the optimal number of clusters and the noise suppression coefficient. To mitigate these issues, we propose the following procedure: compute the similarity (or distance) between the new data and the historical correlation matrices defined in Section 2.1, apply MDS algorithms to project the new observation into the same low-dimensional space, and subsequently assign its cluster label using a computationally efficient classifier (e.g., SVM) trained on the original MDS-embedded data.

3 The state transition behavior of the Chinese Stock Market

3.1 Data

This study utilizes the daily post-rights-adjusted prices of constituent stocks from the CSI 300 Index. The CSI 300, comprising 300 large-cap and highly liquid stocks from both the Shanghai and Shenzhen stock exchanges, represents approximately 60% of the total market capitalization of China's A-share market. As such, it serves as a key indicator for tracking the overall performance of the Chinese stock market. Its constituents are typically industry-leading enterprises that are highly responsive to sectoral and macroeconomic developments. Fluctuations in their stock prices tend to promptly reflect market sentiment and investor expectations, making them particularly suitable for investigating market state transitions.

Focusing on CSI 300 constituents allows the study to capture systemic market dynamics using a representative sample set, balancing analytical validity with computational efficiency. For the empirical analysis, there are 285 stocks selected from the CSI 300 Index, covering a 15-year period from January 2010 to December 2024, corresponding to $T = 3643$ trading days. To ensure the temporal consistency and comparability of correlation matrices $C_{ij}(\tau)$, we included only those stocks that were part of the CSI 300 on 4 January 2010, and remained continuously listed throughout the entire study period. The data is sourced from the RESSET database. All stock prices are post-rights-adjusted, a standard adjustment method that incorporates the effects of corporate actions including stock splits, cash and stock dividends, and rights offerings. This adjustment ensures that historical price series reflect true economic returns and are free from artificial jumps, thereby allowing for consistent return calculation across the full 15-year sample period.

3.2 Identification of market states

We compute the logarithmic returns of the daily post-adjusted closing prices for each stock and generate a sequence of 3,621 correlation matrices of dimension 285 using a sliding window approach, with the window length set to $M = 22$—approximately one trading month—and the sliding step $\Delta = 1$ to ensure sufficient granularity in state identification. To determine the optimal number of market states K, we employ the K-means clustering procedure as introduced in Algorithm 1 of Section 2.1, and apply the Akaike Information Criterion (AIC) to penalize excessive model complexity. Specifically, we examine how the number of clusters K and the noise-suppression parameter ϵ interact under varying initialization conditions, while tracking the stability of the clustering outcomes. The convergence condition is defined as:

- Convergence threshold: relative change in centroid position $< 10^{-5}$;
- Maximum iterations: 20 (to avoid oscillation in unstable cases).

Figure 1 presents the landscape of pseudo AIC-adjusted standard deviations across a grid of cluster counts ($K = 1$ to 10) and noise-suppression levels ($\epsilon = 0.0$ to 0.95). Color intensity reflects the magnitude of the standard deviation, with darker regions corresponding to more stable clustering outcomes. Each parameter configuration is evaluated over 1,000 independent K-means initializations, and the minimum observed pseudo AIC-adjusted standard deviation is interpreted as identifying the most robust and reliable clustering structure.
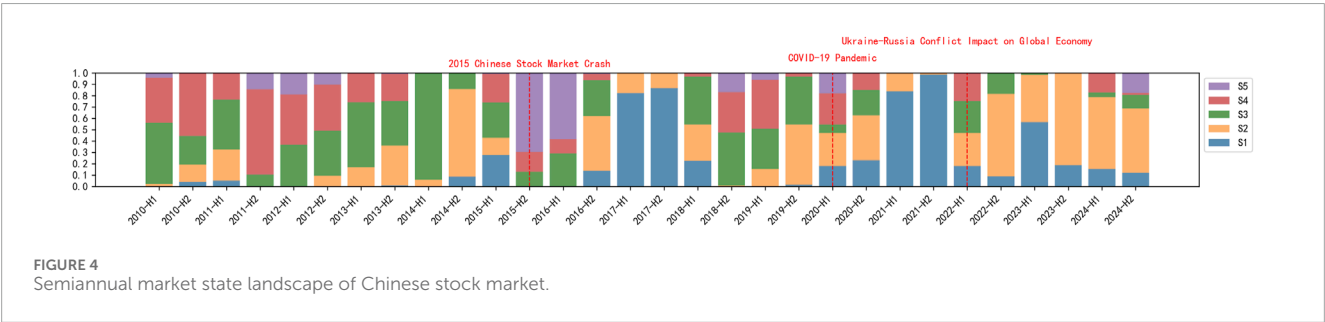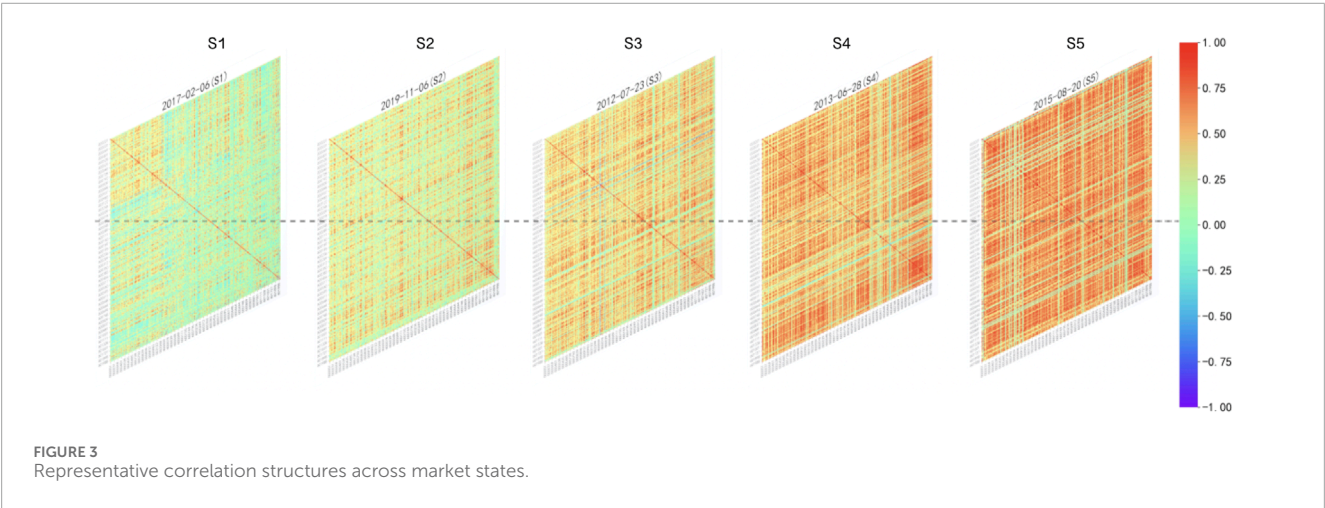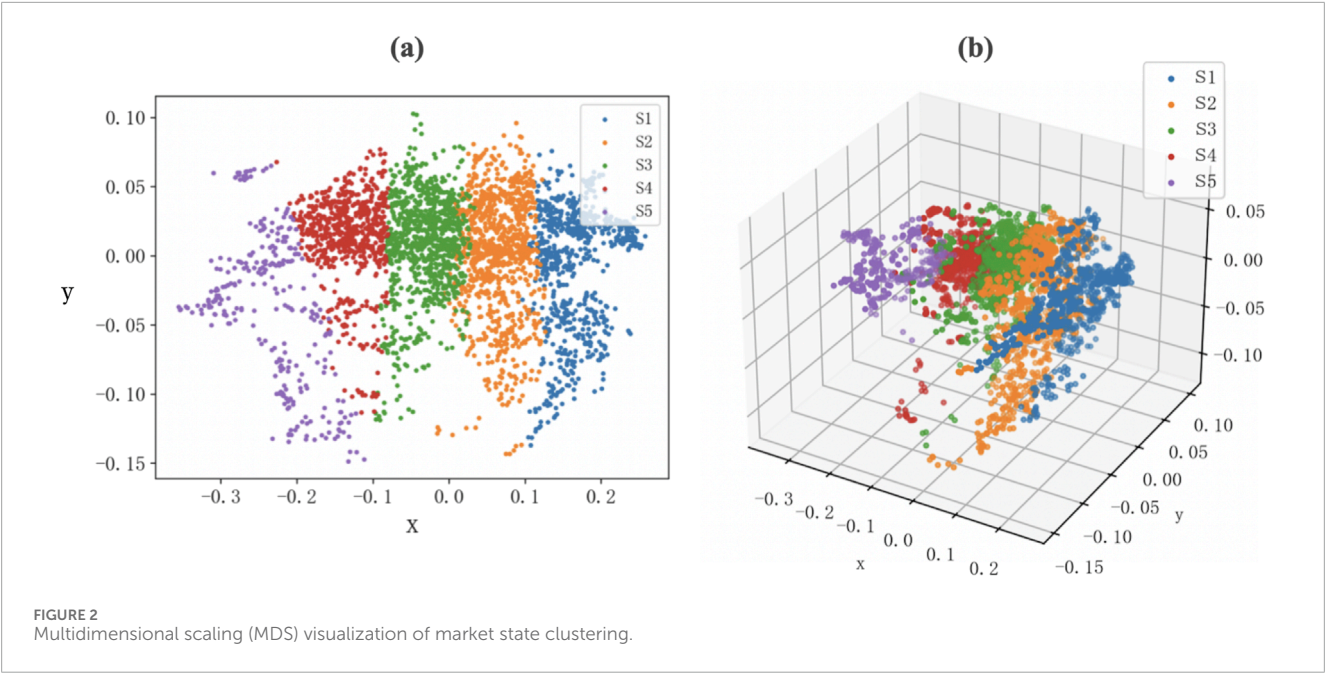
FIGURE 1
Stability landscape for clustering: Pseudo AIC-Adjusted standard deviation across K and ϵ.
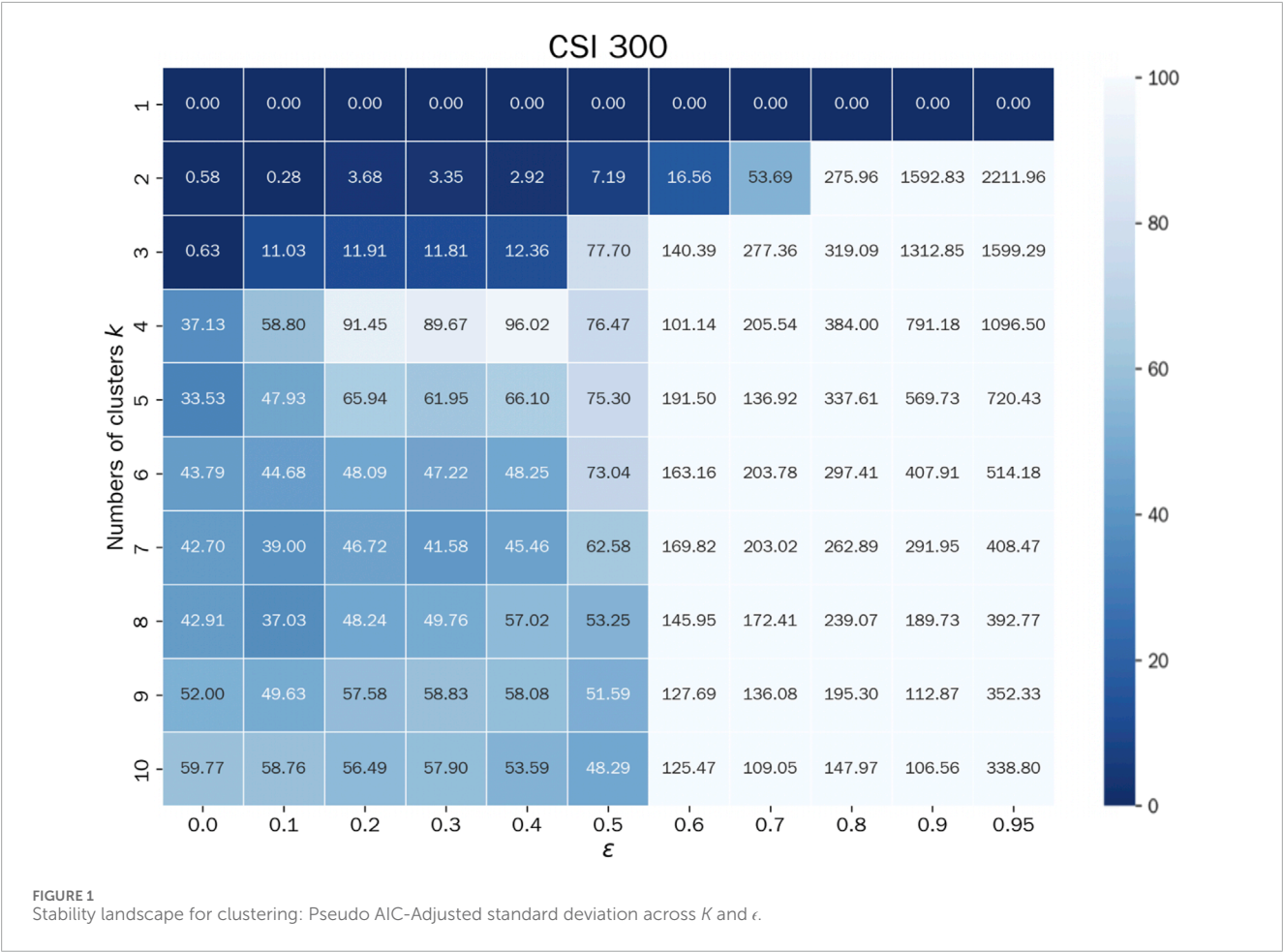
As discussed in Section 2.1, although configurations with $K <$ 4 sometimes yield lower intra-cluster standard deviations, we restrict our analysis to $K \geq 4$ to ensure sufficient interpretability and structural resolution. This choice is supported by prior studies in complex systems and financial market dynamics, which suggest that markets typically evolve through multiple qualitatively distinct regimes which are characterized by different levels of systemic co-movement and risk [11, 30]. Clustering configurations with fewer than four states often fail to capture this heterogeneity, resulting in overly coarse classifications that obscure meaningful structural transitions. For the CSI 300 Index, the configuration yielding the most stable and interpretable clustering outcome is ultimately determined to be $K = 5$ with $\epsilon = 0.0$.

Figure 2a presents a two-dimensional projection of the clustering result onto the xz-plane, derived from the optimal three-dimensional embedding obtained via MDS. This projection facilitates intuitive visualization of the cluster separation in reduced dimensions. The full three-dimensional representation of the MDS result, as shown in Figure 2b, provides a more comprehensive view of the geometric structure underlying the correlation matrices. Together, these visualizations demonstrate that the five identified market states are well-separated in the embedded space, suggesting the presence of latent structural patterns in market behavior over time.

Figure 3 presents example heatmaps of stock return correlation matrices for 285 constituent stocks, each selected from a representative trading day corresponding to one of the five identified market states under the configuration $\epsilon = 0.0$. A notable progression is observed across the states: as the market transitions upward (i.e., from lower- to higher-indexed states), the red intensity in the heatmaps increases, reflecting a rising level of average pairwise correlation among stocks. This pattern suggests a shift in market structure—from a relatively fragmented condition with weak inter-stock dependencies to a more cohesive regime marked by stronger co-movement. Such transitions often coincide with phases of elevated systemic dynamics, potentially signaling the emergence of market-wide trends or collective reactions to macro shocks.

To capture the temporal dynamics of market conditions, we examine the evolution of state occupancy probabilities over time. The color blocks in Figure 4 represent the relative frequency of each market state within semiannual windows. Specifically, the probability of a state in each window reflects the proportion of trading days classified under that state. Blue, orange, and green blocks correspond to lower-indexed market states, while red and purple denote higher-indexed states. Several noteworthy patterns emerge. Periods such as 2014, the second half of 2016 through 2017, and the second half of 2020–2021 are dominated by lower-indexed states, suggesting relatively stable market conditions. In

FIGURE 2
Multidimensional scaling (MDS) visualization of market state clustering.

FIGURE 3
Representative correlation structures across market states.

FIGURE 4
Semiannual market state landscape of Chinese stock market.

contrast, the window spanning 2015 to early 2016 is characterized by a pronounced transition to high-indexed states, aligning with the 2015 A-share market crash—one of the most severe episodes of turbulence in recent years. This temporal clustering of market states provides intuitive validation for the proposed state classification and its relevance to real-world market events.

FIGURE 5
Empirical transition matrix and probabilities among market states.
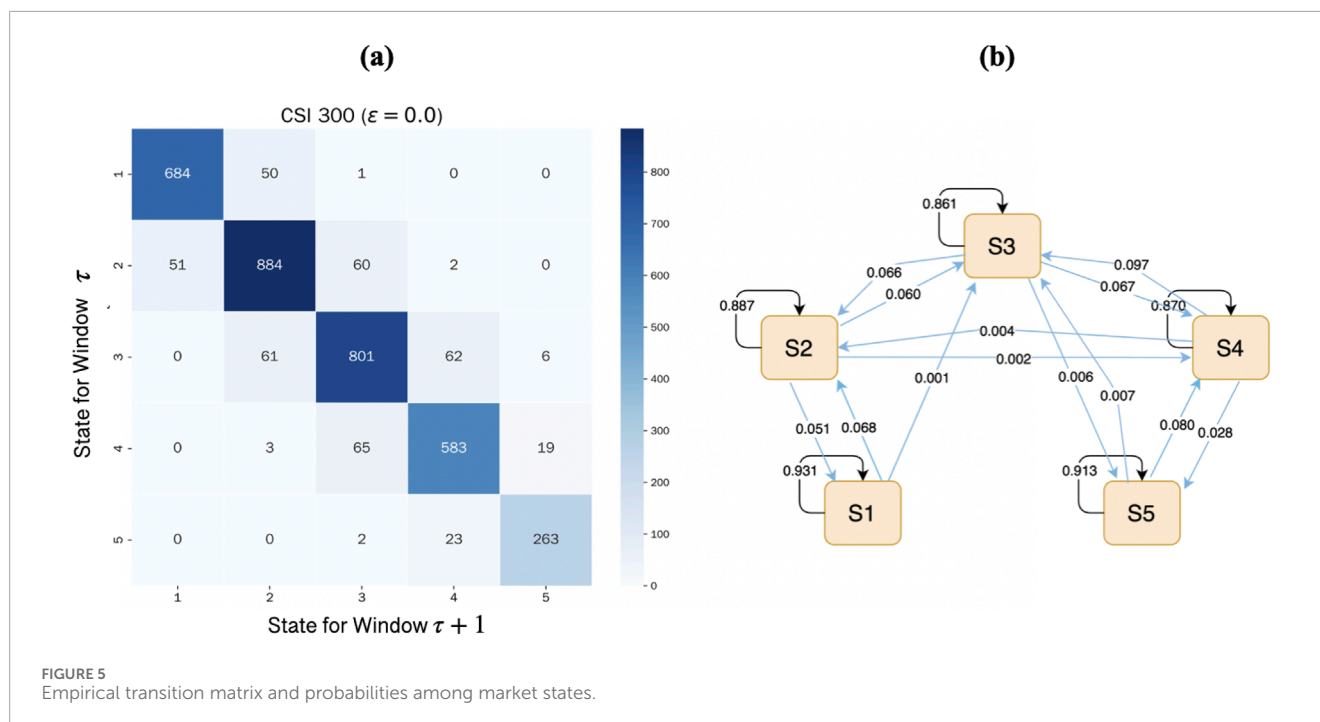
3.3 Market states transition

To better understand the dynamic evolution of market conditions, we analyze the empirical transition behavior among the identified market states. Figure 5a presents the empirical state transition matrix of the CSI 300 Index, obtained under the optimal clustering setting $k = 5, \epsilon = 0.0$. The transition matrix exhibits a near-tridiagonal structure, suggesting that state changes tend to occur gradually—either remaining in the same state or shifting to adjacent states. This reflects the market's inertia and path-dependent nature in its structural evolution. Transitions that leap multiple levels—from lower-indexed states directly to a high-indexed state (e.g., from S_1 or S_2 to S_5)—are virtually nonexistent, highlighting the rarity of abrupt structural shifts without intermediate transitions. Instead, gradual escalations are more common. Notably, we observe a limited number of transitions from S_2 to S_4 (2 times), from S_3 to S_4 (62 times), from S_3 to S_5 (6 times), and from S_4 to S_5 (19 times), which may be interpreted as structural precursors to periods of elevated market stress.

Figure 5b visualizes the corresponding transition probabilities using a directed graph, where edge annotations indicate the estimated transition probabilities. Most of the mass is concentrated along self-loops and adjacent-state arrows, reinforcing the local stability of market states. Furthermore, the long-run stationary distribution derived from the estimated Markov chain closely aligns with the empirical frequency distribution:

$$P(S1, S2, S3, S4, S5) = (0.20, 0.28, 0.26, 0.18, 0.08),$$

which closely mirrors the empirical frequency of observed states. This consistency reinforces the robustness of the five-state market classification and suggests that the estimated state transitions capture meaningful and persistent structural dynamics in the market.

3.4 Relationship between market states and market crashes

To investigate whether market crashes tend to occur under specific structural states of the market, we examine the relationship between identified market states and empirically observed crash periods. The five market states of the CSI 300 Index are ordered by ascending average pairwise correlation, with the mean correlations given by $[\overline{C^{S1}(\tau)}, \overline{C^{S2}(\tau)}, \overline{C^{S3}(\tau)}, \overline{C^{S4}(\tau)}, \overline{C^{S5}(\tau)}] = [0.19, 0.30, 0.41, 0.52, 0.61]$, displaying a clear linear progression from weakly to strongly correlated market conditions.

Table 1 summarizes the start and end dates of crash episodes, the number of crash cases detected in each episode, the average correlation level during each crash period, and the corresponding market state assigned. These results are based on the proposed crash cases labeling procedure, i. e., Algorithm 2, with time window length $N = 10$ and a drawdown threshold of $\eta = 0.08$. To map each crash episode to a market state, we calculated the average correlation within the episode and classified it using the midpoint between adjacent state correlation means, i.e., $\frac{1}{2}(\overline{C^{S_i}(\tau)} + \overline{C^{S_{i+1}}(\tau)})$.

Based on Table 1, different crash episodes are mapped to different market states, forming an initial distribution of state assignments as $(S_1, S_2, S_3, S_4, S_5) = (0, 3, 4, 6, 9)$. However, some of these episodes are notably brief—lasting fewer than three time windows—and are likely driven by exogenous shocks rather than the result of endogenous market dynamics such as herding behavior or feedback amplification. To focus on more persistent and structurally significant crashes, we exclude these short-lived cases by removing periods with fewer than four cases. The adjusted distribution then becomes $(S_1, S_2, S_3, S_4, S_5) = (0, 1, 0, 4, 8)$, clearly indicating that most sustained crashes occur under higher-indexed states, particularly States 4 and 5. This result reinforces the association between elevated systemic co-movement and the

TABLE 1 Crash period characteristics and corresponding market states.

No. of crash periods	Start	End	No. of crash cases	Average correlation	State	Max drawdown
1	2010-04-14	2010-05-18	10	0.498	4	20.2%
2	2010-06-22	2010-07-05	3	0.467	4	9.7%
3	2010-11-08	2010-11-24	7	0.617	5	12.5%
4	2011-12-02	2011-12-15	1	0.329	2	8.5%
5	2013-05-28	2013-07-01	11	0.542	4	18.3%
6	2015-01-26	2015-02-06	1	0.383	3	8.2%
7	2015-06-08	2015-07-01	15	0.612	5	31.6%
8	2015-07-23	2015-08-06	7	0.664	5	10.3%
9	2015-08-10	2015-09-07	10	0.657	5	25.9%
10	2015-12-22	2016-02-01	18	0.714	5	26.4%
11	2018-01-26	2018-02-12	3	0.429	3	12.3%
12	2018-06-13	2018-06-28	2	0.422	3	9.6%
13	2018-07-24	2018-08-06	1	0.387	3	8.6%
14	2018-09-28	2018-10-18	5	0.624	5	11.5%
15	2019-04-19	2019-05-09	4	0.548	4	12.6%
16	2020-01-13	2020-02-05	3	0.714	5	12.3%
17	2020-03-05	2020-03-24	7	0.554	4	16.1%
18	2021-02-10	2021-03-10	5	0.288	2	14.4%
19	2022-03-01	2022-03-16	4	0.678	5	13.8%
20	2022-04-14	2022-04-26	2	0.542	4	9.7%
21	2022-10-18	2022-10-31	1	0.291	2	8.6%
22	2024-10-08	2024-10-17	4	0.693	5	11.0%

onset of significant market downturns. Figure 6 illustrates the time series of daily closing prices for the CSI 300 Index, with shaded areas indicating crash periods. Each color corresponds to the market state assigned to the crash episode: orange for S_2, green for S_3, red for S_4, and purple for S_5. As shown, these periods align well with major modification of the index, further validating the classification.

Overall, the results suggest a strong association between elevated-correlation market states and the occurrence of crashes. In particular, States 4 and 5 appear to signal heightened structural instability. This implies the Markov transition diagram of states in Figure 5b can be served to synchronously monitor the market's stability status. However, it is important to emphasize that such state classifications should be interpreted as contemporaneous reflections of market conditions, rather than forward-looking

predictors. Whether they provide sufficient information for lead time warning is addressed in the following section.

4 Early warning of stock market crashes based on state transition information

This section explores the predictive value of market state transitions in generating early warnings for stock market crashes. Building on the classification framework established in previous sections, we aim to determine whether temporal patterns in market state dynamics can be effectively utilized to identify imminent periods of elevated market risk.

Based on the results in Table 1, we identify a total of 124 crash cases. For each of these, we label the trading day immediately
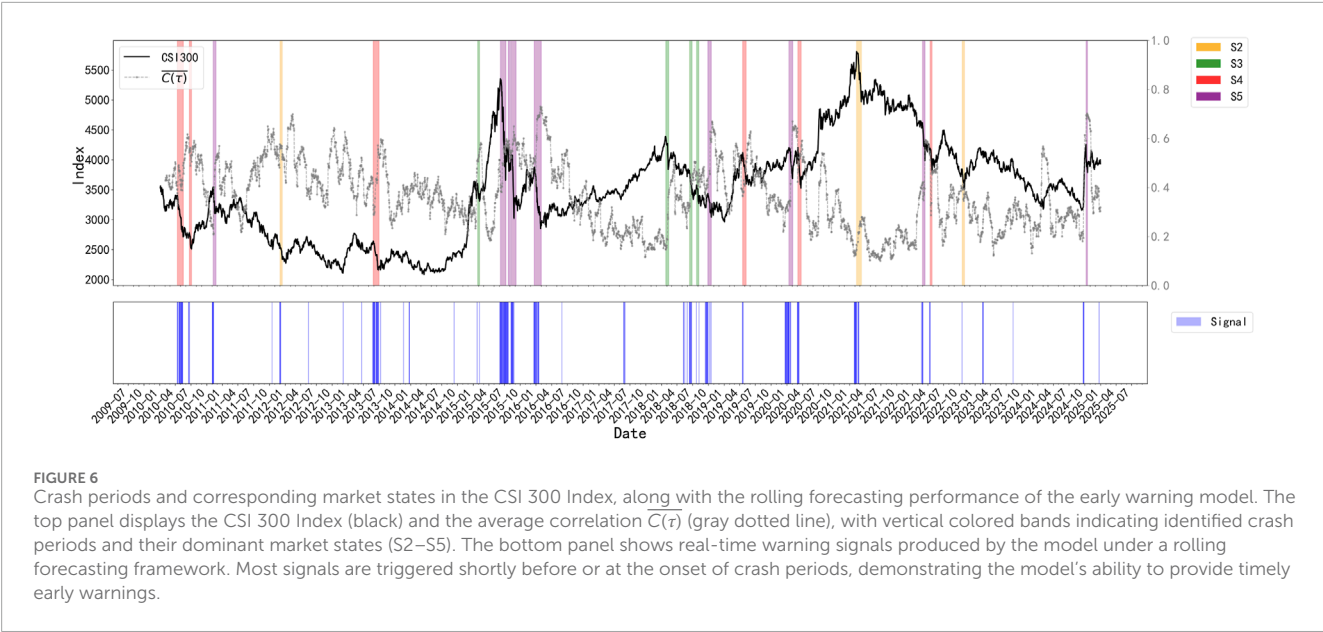
FIGURE 6
Crash periods and corresponding market states in the CSI 300 Index, along with the rolling forecasting performance of the early warning model. The top panel displays the CSI 300 Index (black) and the average correlation $\overline{C(\tau)}$ (gray dotted line), with vertical colored bands indicating identified crash periods and their dominant market states (S2−S5). The bottom panel shows real-time warning signals produced by the model under a rolling forecasting framework. Most signals are triggered shortly before or at the onset of crash periods, demonstrating the model's ability to provide timely early warnings.

TABLE 2 Top 10 best feature configurations.

Rank no.	L_1	L_2	L_3	Precision	Recall	FPR
1	35	40	6	0.9636	0.9677	0.0359
2	35	40	5	0.9633	0.9677	0.0362
3	40	40	6	0.9630	0.9677	0.0365
4	35	40	4	0.9621	0.9677	0.0374
5	35	40	3	0.9615	0.9677	0.0380
6	40	40	5	0.9603	0.9677	0.0392
7	40	35	6	0.9587	0.9677	0.0410
8	40	40	4	0.9582	0.9677	0.0413
9	40	40	3	0.9574	0.9677	0.0423
10	40	30	4	0.9533	0.9677	0.0465

preceding the start of the case as $y = 1$, representing a positive (crash-imminent) sample. All other trading days outside the crash periods are labeled as $y = 0$. This procedure yields 124 positive samples and 3,289 negative samples. To mitigate the severe class imbalance in the dataset, the $y = 1$ samples are replicated 26 times, thereby constructing a balanced classification dataset suitable for supervised learning.

4.1 Model tuning and implication

To systematically examine how the temporal structure of state-based inputs affects the performance of crash early warning, we adopt the methodology outlined in Section 2.2, varying the window lengths used to construct the three categories of state-derived variables. Specifically, the candidate window lengths L_1 for the proportion of each market state $p_k^{L_1}$ and window lengths L_2 for the ratio of upward to downward state transitions r^{L_2} are both set to $(10, 15, 20, 25, 30, 35, 40)$. For the third feature type—capturing the recent evolutionary paths of market states—the window size L_3 is selected from $(3, 4, 5, 6)$. The smaller values of L_3 are chosen deliberately to reduce the combinatorial complexity of the resulting categorical pattern space, thereby enhancing model generalization and mitigating the risk of overfitting. These three temporal window features are designed to capture complementary structural and temporal dimensions of market dynamics. Specifically, L_1 encodes the medium-term proportion of time the market resides in each state, which reflects persistent structural tendencies; L_2 captures the directional momentum of market transitions—quantifying how frequently the market moves into more fragile or stable states—thus acting as a proxy for regime-level drift; and L_3 records short-term transition motifs, designed to detect recent local fluctuations that may precede regime shifts. The design draws inspiration from complex systems thinking, where multiscale memory effects and structural inertia often play a critical role in the lead-up to systemic events.

In total, 196 different (L_1, L_2, L_3) combinations are evaluated using a decision tree classifier with 5-fold cross-validation. Table 2 lists the top 10 configurations first ranked by recall and then ranked by precision. The best-performing configuration is $(L_1, L_2, L_3) = (35, 40, 6)$, which achieves a highest recall of 96.77%, along with the precision of 96.36%, and a low FPR of 3.59%. This setting is selected as the optimal feature configures for the model, balancing sensitivity to crash signals with robustness to false alarms. All performance metrics reported in this section are obtained via 5-fold cross-validation, where the dataset is randomly partitioned into five subsets. In each round, four subsets are used for training and one for validation, ensuring that every observation is tested exactly once. This repeated re-training procedure provides an estimate of model
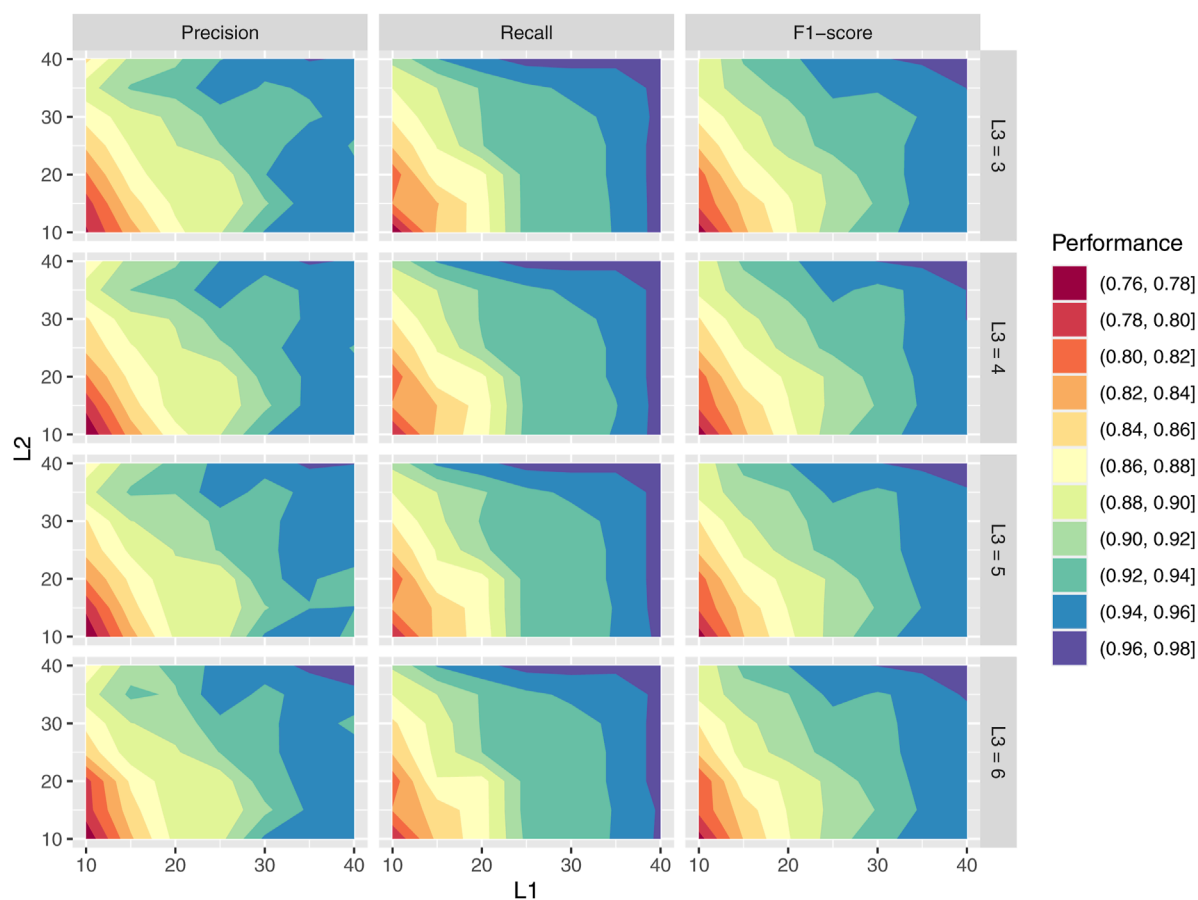
FIGURE 7
Heatmaps of precision, recall, and F1-Score with hyper-parameters combinations L_1, L_2, L_3.
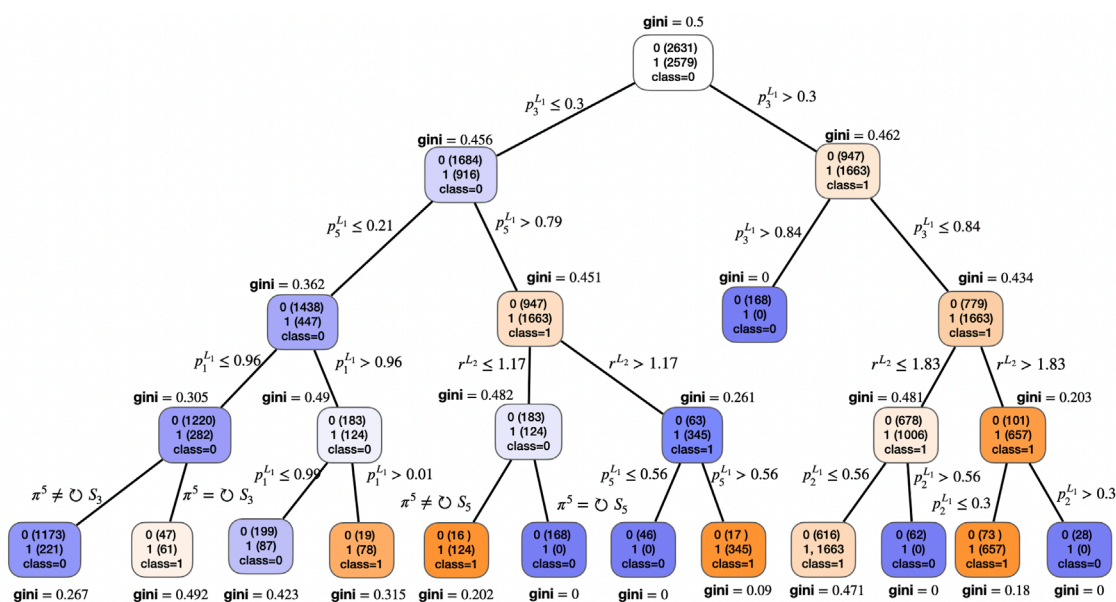

FIGURE 8
Decision tree structure under the optimal window configuration $(L_1, L_2, L_3) = (35, 40, 6)$. Each node shows the predicted class, Gini coefficient, the number of samples labeled as crash ($y = 1$) and non-crash ($y = 0$), and the total number of samples in parentheses. Warm-colored nodes indicate classification into the crash category ($y = 1$), while cool-colored nodes correspond to non-crash classification ($y = 0$). The notation $\circlearrowleft S_i$ denotes persistent stay in state S_i, represented as a sequence such as $(S_i \rightarrow S_i \rightarrow S_i \rightarrow S_i \rightarrow S_i)$.

TABLE 3 Comparison of early warning models using 5-fold cross-validation.

Model	Precision	Recall	F1-score	FPR
MST model	**0.964**	**0.968**	**0.966**	**0.037**
VIX-like model	0.863	0.755	0.803	0.120
Spectrum model	0.677	0.904	0.773	0.347
Absorption ratio	0.783	0.828	0.804	0.229

The maximum value in this column is indicated in bold.

generalization under moderate sample variability. To further assess the robustness of the selected feature configuration, we conducted an additional evaluation using repeated 5-fold cross-validation. Specifically, we repeated the 5-fold CV procedure 20 times with different random splits to estimate the variability of model performance. Based on these repetitions, we calculated confidence intervals for key performance metrics. The 95% confidence interval for precision is $[96.2\%, 96.5\%]$, for recall is $[96.7\%, 96.8\%]$, for F1-score is $[96.5\%, 96.7\%]$, and for the false positive rate (FPR) is $[3.4\%, 3.7\%]$. These narrow intervals confirm the model's stable and reliable predictive capacity under different sample partitions.

We did not construct a separate test set for final model evaluation. Given the rarity and temporal clustering of crash events, a fixed test split may lead to high variance and unreliable evaluation. Instead, our focus is on examining whether temporal features of market state transitions offer informative signals for early crash detection. In future work, more rigorous backtesting approaches—such as Combinatorial Purged Cross-Validation (CPCV) [31]—could be explored to validate model performance in a production-like environment.

To provide a more detailed and intuitive assessment of how different window length combinations contribute to model performance, we visualize the precision, recall, and F1-score across all 196 (L_1, L_2, L_3) configurations in Figure 7. Each panel corresponds to a fixed value of $L_3 \in 3, 4, 5, 6$, while the horizontal and vertical axes represent values of L_1 and L_2, respectively. The color shading in each cell reflects the performance level for a specific metric, with darker shades indicating better results.

The heatmaps reveal several consistent patterns. First, larger values of L_1 and L_2 generally yield stronger performance across all metrics, especially when $L_3 = 6$, confirming the advantage of incorporating longer historical context for state proportions and transition ratios. Second, the F1-score—balancing precision and recall—tends to peak in the upper-right regions of the grids, where both L_1 and L_2 lie in the range of 30–40. This aligns well with the previously identified optimal configuration $(L_1, L_2, L_3) = (35, 40, 6)$, providing visual validation for the selected model parameters. Importantly, the combination $(L_1, L_2, L_3) = (35, 40, 6)$ integrates features over both short- and medium-term windows, capturing persistent structural shifts as well as abrupt transitions.

Figure 8 illustrates the structure of a representative decision tree trained under the optimal window configuration $(L\_1, L\_2, L\_3) = (35, 40, 6)$, which displays the first four layers. Interpretation of the decision rules reveals that the most informative paths leading to a crash warning (i.e., terminal nodes classified as $y = 1$) tend

to involve long-term patterns in state proportions and transition ratios—particularly $p_5^{L_1}$ and r^{L_2}—rather than short-term state transition sequences.

Notably, one of the clearest decision paths leading to a crash classification occurs when $0.3 < p_3^{L_1} \le 0.84$, and $r^{L_2} > 1.83$, suggesting that elevated presence of mid-to-high index states combined with a strong net upward transition tendency serves as a robust early warning signal. In contrast, the specific forms of short-term state transition patterns—such as $\pi^5 = (S_i \to \cdots \to S_j)$—do not appear to provide significant predictive power on their own. However, when such patterns indicate that the market has not persistently remained in state S_5 (i.e., $\pi^5 \ne \cup S_5$), and are combined with medium- to long-term indicators such as $p_k^{L_1}$ and r^{L_2}, the model still achieves strong early warning performance. Specifically, long-term state occupancy proportions $p_5^{L_1}$ act as proxies for structural stress accumulation—indicating prolonged market exposure to vulnerable regimes. The transition ratio r^{L_2} captures directional asymmetries in regime shifts, with high values signaling drift toward higher fragility states. The short-term motif π^5 provides further granularity by detecting whether the market remains stable in a specific state (e.g., $\cup S_5$) or experiences disruptive shifts. These layered indicators together help uncover latent warning signals preceding systemic crashes, supporting the interpretability and theoretical consistency of the model.

To further examine the practical utility of the proposed early warning model, we apply the optimized classifier in a rolling forecasting mode. At each time step, the model uses only past information to predict whether a warning signal should be triggered for the next period. This simulates a realistic deployment setting where market operators make real-time decisions based on evolving structural dynamics. The lower panel of Figure 6 visualizes the resulting forecasted warning signals over the full sample period. Most warnings align closely with the start of crash periods identified by our drawdown-based labeling scheme, often appearing several days in advance. These results confirm that the model generalizes well to unseen data and is capable of generating timely early warnings in real-world settings.

To benchmark the performance of our proposed market state transition (MST) framework, we compare it with three representative baseline models of early warning signals that are widely used in the literature: (1) volatility-based indicators, (2) spectrum-based indicators and (3) absorption ratio.

For the volatility model, we construct a VIX-like indicator based on the implied volatility of the China ETF 50 options, often referred to as the "China fear index." This index serves as a functional proxy for market-perceived risk in the Chinese context. For the spectrum-based category, we follow the methodology of Chakraborti et al. [20] to compute the maximum eigenvalue of the correlation matrix. For the absorption ratio, we use the implementation proposed by Kritzman et al. [6], which quantifies the proportion of variance absorbed by a fixed number of principal components.

To enable a fair comparison, we construct decision-tree-based classifiers for each of these indicators by tuning thresholds to generate binary crash warnings. All models are evaluated using the same labeled crash dataset and assessed under 5-fold cross-validation. Table 3 presents the results. Our MST-based framework consistently outperforms the benchmark models across all metrics,
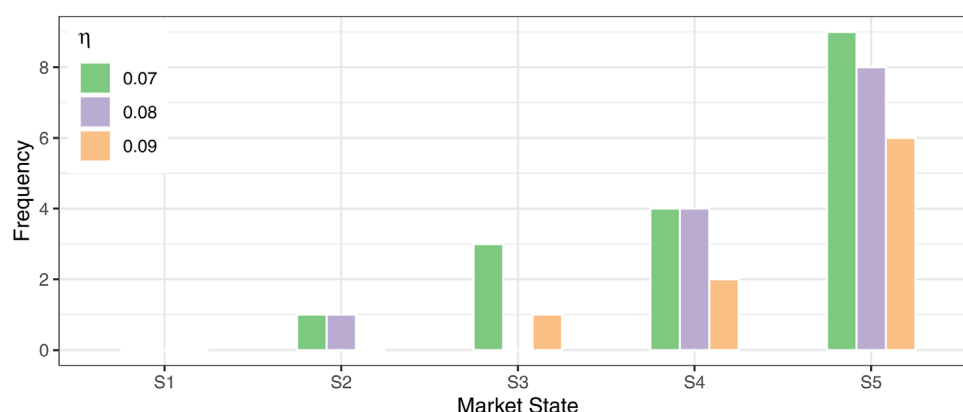
FIGURE 9
Sensitivity of crash-state distribution to drawdown threshold η. Frequencies of each market state observed during identified crash periods under three thresholds $\eta = 7\%$, 8%, 9% are compared. Results confirm the consistency of crash-state associations.
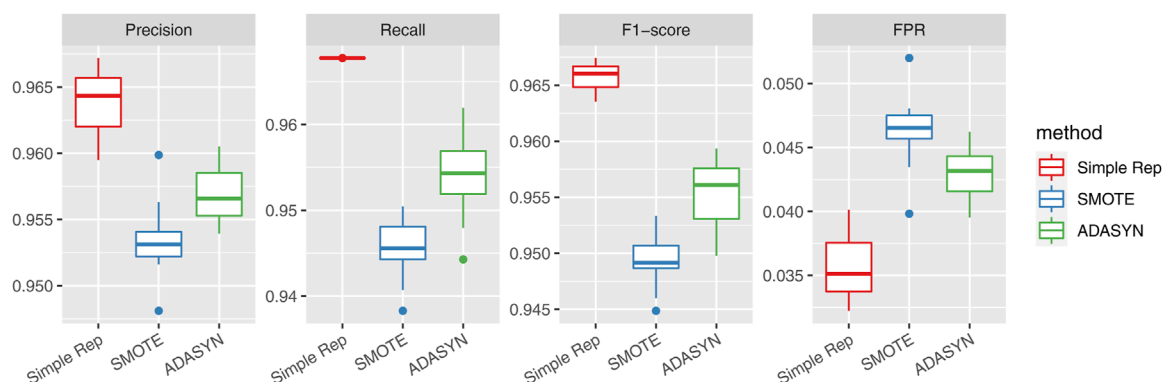

FIGURE 10
Model performance under different resampling methods. Bars represent precision, recall, F1-score, and false positive rate (FPR) for simple replication, SMOTE, and ADASYN. The results demonstrate that the model's performance remains stable across various class-balancing strategies.

achieving the highest precision (0.964), recall (0.968), and F1-score (0.966), along with the lowest false positive rate (3.7%). These results highlight the advantage of structurally grounded, transition-driven features in anticipating systemic stress.

4.2 Adaptability to new data

To evaluate the model's adaptability in real-time environments, we assess its ability to incorporate newly arriving trading data using the procedure described at the end of Section 2.2. Specifically, new correlation matrices are projected into the original low-dimensional MDS space, and their market state labels are assigned using a Support Vector Machine (SVM) trained on the initial clustering results.

Results from 5-fold cross-validation indicate that the SVM with a radial basis function (RBF) kernel achieves a classification accuracy of 98.20%, demonstrating that new trading days can be accurately labeled with minimal computational cost. This confirms the model's suitability for real-time deployment and continuous market monitoring under dynamic conditions.

4.3 Sensitivity analysis

To evaluate the robustness of our results with respect to key design choices, we conduct two sets of sensitivity analyses focusing on crash labeling and resampling methods.

First, we examine the impact of changing the drawdown threshold η used in crash event identification. Figure 9 illustrates the distribution of market states associated with crash periods under three threshold settings: $\eta = 7\%$, 8%, and 9%. The patterns remain highly consistent, confirming that the core findings reported in Section 3.4—especially the concentration of crash periods in states S_4 and S_5—are not sensitive to the exact choice of η.

Second, we assess the influence of different resampling techniques used to address class imbalance in model training. In addition to the simple replication method (Simple Rep) used in the main analysis, we consider two widely used alternatives: SMOTE and ADASYN. As shown in Figure 10, the decision tree model maintains stable performance across all three approaches. While SMOTE and ADASYN yield slightly lower precision, recall and F1-score, the overall differences are modest, supporting the validity of

our original choice. These results demonstrate that the predictive effectiveness of our framework is not materially dependent on the specific sampling method adopted.

5 Conclusion

Grounded in the complex systems perspective, this paper applies correlation-based structural analysis to identify and characterize market states in the Chinese stock market, with the CSI 300 Index as a representative system. Using multidimensional scaling and K-means clustering on rolling correlation matrices, we identify five distinct market states that reflect varying levels of systemic co-movement. Our analysis confirms that these states are not merely statistical artifacts but correspond to meaningful structural configurations of the market.

Empirical results reveal that market state transitions are predominantly local, occurring between adjacent states, while abrupt multi-level jumps are rare. States also exhibit high temporal persistence, underscoring the gradual and inertial nature of structural evolution in financial systems. Notably, high-indexed states—those with elevated average correlations—are strongly associated with historically observed crashes, particularly when the market resides in State 4 or State 5, where systemic co-movement is most pronounced. While these states offer valuable insight into ongoing market fragility, they function primarily as contemporaneous indicators rather than forward-looking predictors. Together, these findings support the use of market state classification as a structurally grounded, real-time measure of systemic risk.

Based on the identified state structure, an early warning model is constructed using decision trees trained on state-derived temporal features. These features capture medium-term structural distributions, directional trends in state transitions, and recent evolutionary patterns. The model demonstrates high recall and precision across multiple validation settings, indicating strong robustness in identifying pre-crash conditions. Furthermore, real-time adaptability is ensured through a projection-based labeling procedure, allowing new data to be efficiently embedded and classified, thus enabling continuous and responsive market surveillance.

Several avenues for future research emerge from this work. One direction is to extend the proposed framework to cross-market or multi-asset systems, such as incorporating global indices, fixed income instruments, or cryptocurrency markets, to examine whether state transitions exhibit comparable structural patterns across different financial domains. Another promising line lies in enriching the modeling of state transitions themselves—e.g., using recurrent neural networks to capture nonlinear memory effects. Furthermore, combining structural state dynamics with behavioral signals, news flows may enhance the predictive power and interpretability of crash warnings. Finally, linking market state trajectories with macro-financial policy variables could offer new insights into the interplay between systemic risk and regulatory interventions.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

W-YP: Data curation, Formal Analysis, Software, Visualization, Writing – original draft. LL: Conceptualization, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2025.1647667/full#supplementary-material

References

1. Sornette D. Why stock markets crash: critical events. In: *Complex financial systems*. 2nd edn. Princeton, NJ: Princeton University Press (2017).

2. Mantegna RN, Stanley HE. *Introduction to econophysics: correlations and complexity in finance*. Cambridge University Press (2000).

3. Bouchaud J-P. The (unfortunate) complexity of the economy. *Phys World* (2009) 22:28–32. doi:10.1088/2058-7058/22/04/39

4. Hommes C. Complex evolutionary systems in behavioral finance. In: *Handbook of financial markets: dynamics and evolution*. Elsevier (2009). p. 217–76.

5. Münnix MC, Shimada T, Schäfer R, Leyvraz F, Seligman TH, Guhr T, et al. Identifying states of a financial market. *Scientific Rep* (2012) 2:644–6. doi:10.1038/srep00644

6. Kritzman M, Li Y, Page S, Rigobon R. Principal components as a measure of systemic risk. *The J Portfolio Management* (2011) 37:112–26. doi:10.3905/jpm.2011.37.4.112

7. Billio M, Getmansky M, Lo AW, Pelizzon L. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J Financial Econ* (2012) 104:535–59. doi:10.1016/j.jfineco.2011.12.010

8. Lin L, Sornette D. "speculative influence network" during financial bubbles: application to Chinese stock markets. *J Econ Interaction Coord* (2018) 13:385–431. doi:10.1007/s11403-016-0187-7

9. Lin L, Guo X. Identifying fragility for the stock market: perspective from the portfolio overlaps network. *J Int Financial Markets, Institutions Money* (2019) 62:132–51. doi:10.1016/j.intfin.2019.07.001

10. Sandhu R, Georgiou T, Tannenbaum A. Ricci curvature: an economic indicator for market fragility and systemic risk. *Sci Adv* (2016) 2:e1501495. doi:10.1126/sciadv.1501495

11. Samal A, Pharasi HK, Ramaiah SJ, Kannan H, Saucan E, Jost J, et al. Network geometry and market instability. *R Soc Open Sci* (2021) 8:rsos.201734. doi:10.1098/rsos.201734

12. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, et al. Early-warning signals for critical transitions. *Nature* (2009) 461:53–9. doi:10.1038/nature08227

13. Hoel EP. When the map is better than the territory. *Entropy* (2017) 19:188. doi:10.3390/e19050188

14. Chen M, Li N, Zheng L, Huang D, Wu B. Dynamic correlation of market connectivity, risk spillover and abnormal volatility in stock price. *Physica A: Stat Mech Its Appl* (2022) 587:126506. doi:10.1016/j.physa.2021.126506

15. Wu B, Huang D, Chen M. Estimating contagion mechanism in global equity market with time-zone effect. *Financial Management* (2023) 52:543–72. doi:10.1111/fima.12430

16. Yu D, Huang D. Cross-sectional uncertainty and expected stock returns. *J Empirical Finance* (2023) 72:321–40. doi:10.1016/j.jempfin.2023.04.001

17. Pharasi HK, Sharma K, Chatterjee R, Chakraborti A, Leyvraz F, Seligman TH Identifying long-term precursors of financial market crashes using correlation patterns. *Physica A: Stat Mech its Appl* (2018) 516:603–16.

18. Heckens A, Guhr T. A new attempt to identify long-term precursors for financial crisis in the market correlation structures. *arXiv preprint* (2021).

19. Pharasi HK, Sharma K, Chakraborti A. Dynamics of market states and risk assessment. *Physica A: Stat Mech its Appl* (2024) 631:129191.

20. Chakraborti A, Sharma K, Pharasi HK. Phase separation and scaling in correlation structures of financial markets. *J Phys Complexity* (2020) 2:015002. doi:10.1088/2632-072x/abbed1

21. Chetalova D, Schäfer R, Guhr T. Zooming into market states. *J Stat Mech Theor Exp* (2015) 2015:P01029. doi:10.1088/1742-5468/2015/01/p01029

22. Spelta A, Flori A, Pecora N, Buldyrev S, Pammolli F. A behavioral approach to instability pathways in financial markets. *Nat Commun* (2020) 11:1707–9. doi:10.1038/s41467-020-15356-z

23. Begušić S, Kostanjčar Z, Kovač D, Stanley H, Podobnik B. Information feedback in temporal networks as a predictor of market crashes. *Complexity* (2018) 2018:1–13. doi:10.1155/2018/2834680

24. Wang X, Zhao L, Zhang N, Feng L, Lin H. Stability of China's stock market: measure and forecast by ricci curvature on network. *Complexity* (2023) 2023:1–12. doi:10.1155/2023/2361405

25. Xing K, Yang X. How to detect crashes before they burst: evidence from Chinese stock market. *Physica A: Stat Mech Its Appl* (2019) 528:121392. doi:10.1016/j.physa.2019.121392

26. Cerqueti R, Gatfaoui H, Rotundo G. Resilience for financial networks under a multivariate garch model of stock index returns with multiple regimes. *Ann Operations Res* (2024). doi:10.1007/s10479-023-05756-x

27. Lim B, Zohren S, Roberts S. Detecting changes in asset co-movement using the autoencoder reconstruction ratio. *arXiv preprint arXiv:2002.02008* (2020).

28. Guhr T, Kälber B. A new method to estimate the noise in financial correlation matrices. *J Phys A: Math Gen* (2003) 36:3009–32. doi:10.1088/0305-4470/36/12/310

29. Burnham KP, Anderson DR. *Model selection and multimodel inference: a practical information-theoretic approach*. 2nd edn. Springer (2004).

30. Chakraborti A, Sharma K, Pharasi HK. Characterization of catastrophic instabilities: market crashes as paradigm. *arXiv preprint* (2018).

31. López de Prado M. *Advances in financial machine learning*. Wiley (2018).