# Latent Structure in Equity Volatility

An Exploratory Study of Systemic Risk Using Linear-Algebraic Methods

2025-12-15

## Introduction

This project began as a self-directed learning exercise motivated by coursework in quantitative methods and network analysis. While studying topics such as linear algebra, dimensionality reduction, and network representations, I became interested in whether these tools could be meaningfully applied to real financial data rather than remaining abstract techniques encountered only in assignments and examples.

The goal was therefore not to produce a polished research contribution or a deployable model, but to deeply understand the mechanics, strengths, and limitations of these methods through implementation. Each component of the analysis from principal component analysis to clustering, network construction, and latent state modeling was chosen primarily because it offered an opportunity to engage with a different aspect of applied computational modeling.

Rather than optimizing for predictive accuracy or novelty, the project emphasizes interpretability and conceptual coherence. The guiding question throughout was simple. If relatively standard mathematical tools are applied carefully to market data, what kinds of structure become visible, and what do those structures suggest about the nature of systemic risk?

The work should therefore be read as exploratory. It documents a process of learning by building, using financial markets as a complex and meaningful domain in which to test ideas, methods, and representations. The value of the project lies less in any single result and more in the accumulation of structure across layers of analysis, and in the clarity gained about what these tools can and cannot reveal.

## Motivation

This project set out to explore what kinds of structure become visible in volatility when relatively simple mathematical tools are applied carefully. Rather than aiming for causal claims, prediction, or factor discovery, the focus was on understanding structure, representation, and the behavior of the system under stress.

Risk is commonly summarized through volatility, typically measured as the annualized standard deviation of returns. Yet volatility contains richer information than its average level alone. During periods of stress, assets do not simply become more volatile in isolation. Their volatility becomes more coordinated. Correlations rise, dispersion patterns shift, and markets begin to behave as if they had fewer effective degrees of freedom. This suggests that systemic risk is less about isolated shocks and more about how the structure of co-movement evolves over time.

Many standard approaches impose structure ex ante. Assets are grouped by sectors, factors are predefined, and regimes are assumed to be discrete. These assumptions are practical and often defensible. Technology stocks, for instance, tend to move together due to shared business models and macro sensitivities.

However, during periods of stress, co-movement frequently extends beyond such classifications. Assets from unrelated sectors may respond similarly to latent shocks, while firms within the same sector may decouple. Sector-based groupings capture economic similarity, but not necessarily the pathways through which stress propagates.

If systemic risk is at least partly endogenous, its structure should be allowed to emerge from the data rather than being imposed in advance.

This project explores an alternative perspective. Instead of asking which regime the market is in, it studies how the geometry of volatility co-movement evolves over time. Using linear-algebraic tools such as principal component analysis, clustering, and network representations, the focus shifts from forecasting or classification to structure. The goal is not to predict crises, but to understand how coordination, dimensionality, and coupling change as stress builds and disperses.

The methods used are intentionally simple and transparent. They are not presented as optimal or exhaustive, but as a vehicle for learning applied computational modeling. No attempt was made to optimize predictive performance, and the models are not validated out-of-sample. The emphasis is on how much structure can already be revealed when the modeling question is framed around geometry and interaction rather than labels and regimes.

## Data

Before introducing any structure, we look at volatility behavior in the S&P 500. The dataset consists of daily returns for S&P 500 constituents from 2007 to 2025. Volatility is proxied using 20-day realized log-volatility. Assets are required to have at least two years of observations to ensure reasonably stable estimates.

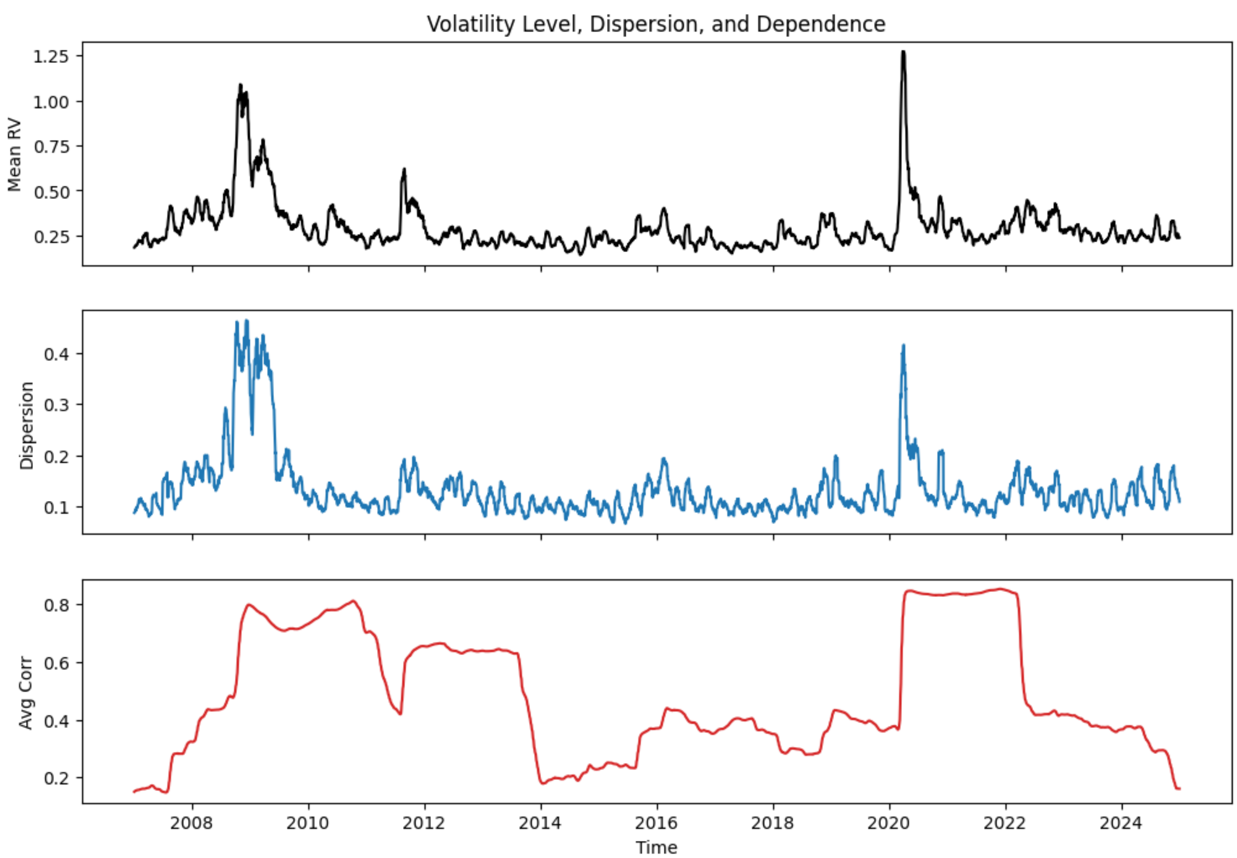


Figure 1: Volatility statistics

Figure 1 above tracks the historic movement of mean realized volatility, cross-sectional dispersion, and average pairwise volatility correlation over time. Unsurprisingly, major stress episodes stand out clearly. Infamous economic shocks such as the global financial crisis (GFC) and the COVID shock in 2020 show sharp spikes in average volatility, increases in dispersion and, more notably, strong rises in cross-asset correlation.

What is already visible at this level is that volatility does not simply rise uniformly. During stress, assets become volatile together. Correlations increase persistently, often remaining elevated even after volatility levels begin to normalize. This suggests that systemic risk is as much about coordination as it is about magnitude.

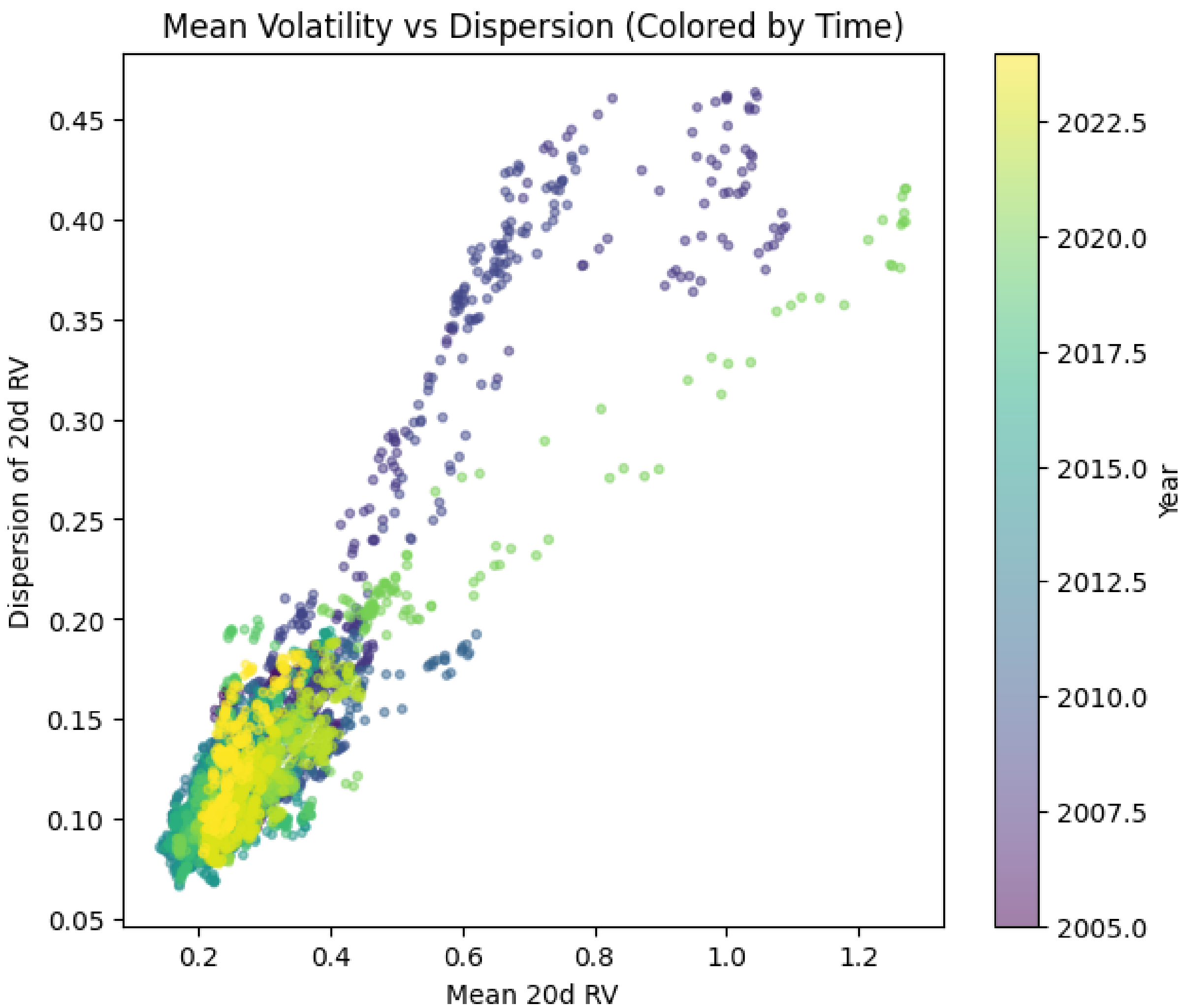


Figure 2: Mean volatility versus cross-sectional dispersion

The scatter plot (figure 2) of mean volatility versus cross-sectional dispersion adds a complementary view. While higher average volatility is generally associated with greater dispersion, the relationship is not one-to-one. Different periods occupy distinct regions of this space, hinting that similar volatility levels can correspond to very different internal structures. In other words, the market can be equally volatile but organized in meaningfully different ways. Taken together, these plots motivate the central question of this project. Rather than asking when volatility is high or low, it may be more informative to ask how volatility co-movement is structured, how that structure changes over time, and whether periods of stress are associated with a collapse in effective dimensionality.

## Methodology

This section walks through the modeling choices used to study the evolving structure of volatility co-movement. The emphasis is not on deriving new algorithms, but on understanding how a small set of linear-algebra based tools can be combined to reveal meaningful structure in high-dimensional financial data.

First, principal component analysis is used to characterize dominant modes of volatility dependence and track changes in system dimensionality over time. The leading component is interpreted as a market-wide volatility mode, while residual structure captures non-market interactions. These residual representations are then used to construct endogenous volatility communities, which serve as the nodes of a reduced system. Interactions between these nodes are summarized using simple network statistics computed from rolling correlations. Finally, a latent state-space model is introduced to compress the evolving network structure into a single continuous measure of systemic stress.

### PCA on volatility correlations

Principal component analysis (PCA) is a linear transformation that re-expresses a set of correlated variables in terms of orthogonal directions that successively maximize variance. Given a correlation matrix $C$ of size $n \times n$ where $n$ is the number of assets, PCA solves the eigenvalue problem:

$$Cv_i = \lambda_i v_i$$

where $v_i$ is the i-th eigenvector and $\lambda_i$ is the corresponding eigenvalue ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

PCA provides a natural notion of effective dimensionality. If a small number of eigenvalues dominate, the system behaves as if it has fewer independent degrees of freedom.

When applied to correlation matrices, PCA does not identify causal factors. Instead, it describes dominant modes of co-movement. A large leading eigenvalue $\lambda_1$ indicates strong coordination across variables, while a flatter spectrum suggests more heterogeneous, weakly coupled behavior.

The PCA is applied to rolling correlation matrices of log-scaled 20-day realized volatility across S&P 500 equities. For each time window $t$, we compute the correlation matrix $C_t$ and extract its eigenvalues $\lambda_i^{(t)}$ and eigenvectors $v_i^{(t)}$. These summerize the structure of volatility co-movement.

The first eigenvector $v_1^{(t)}$ is interpreted as a market wide volatility mode similarly to the returns beta in the capital asset pricing model. The loadings of each asset on $v_1$ describe their sensitivity to this common volatility factor.

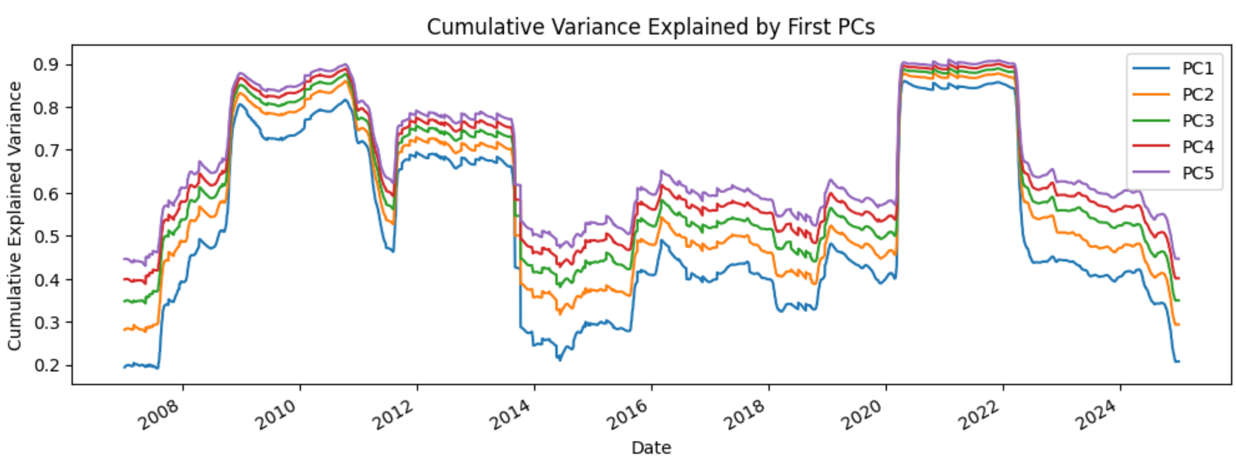


Figure 3: First Principal Components

The explained-variance profile as demonstrated in figure 3 shows how quickly variance concentrates in the leading components, while cumulative variance curves highlight the fraction of total dependence captured by the first few PCs. Most notably, the time series of the largest eigenvalue exhibits pronounced spikes during known stress episodes.

As shown in figure 4 below. Across the sample, periods of elevated market stress are associated with a sharp increase in the leading eigenvalue and a corresponding collapse in effective dimensionality. In calmer periods, variance is distributed more evenly across components, indicating weaker global coordination and richer internal structure.

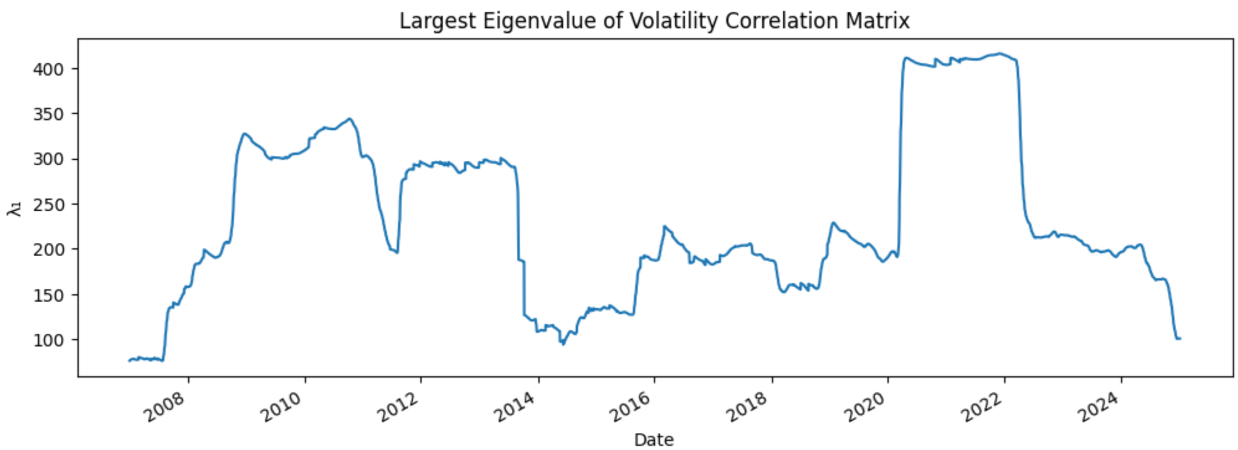


Figure 4: Largest Eigenvalue of Correlation Matrix

At this stage, PCA serves a descriptive role. It establishes that volatility co-movement is not stable over time and that stress episodes are characterized by increasing dominance of a single global mode. This motivates two subsequent steps: first, removing the market-wide component to study residual structure, and second, asking how that residual structure organizes into interacting subunits.

### Residualization and Non-Market Volatility Structure

Once a dominant common component has been identified, a natural next step is to look at the structure after it is removed. This corresponds to projecting observations onto the subspace orthogonal to the leading principal component.

Let $x_a$ represent the standardized volatility vector for asset $a$ over the estimation window. Its projection onto the first principal component $v_1$ is:

$$x_a^{\parallel} = (x_a \cdot v_1)v_1$$

Then the residual, which is the part that is orthogonal to the market mode is:

$$r_a = x_a - x_a^{\parallel}$$

Equivalently if we arrange the eigenvectors $v_1, v_2, \ldots, v_n$ into a matrix $V$, the residual volatility in the full n-dimensional space can be expressed by zeroing out the contribution of $v_1$ in the eigenbasis representation.

In practice, we work directly with the residual correlation matrix, which is derived from the residuals $r_a$ across all assets. This removes the rank-one contribution associated with $v_1$ isolating co-movement that is orthogonal to the dominant market factor.

However, this operation does not eliminate dependence altogether. It isolates variation that cannot be explained by a single global factor, which makes "fine-grained" structure more visible.

These residual loadings describe how each asset responds to latent volatility shocks once the dominant system-wide effect has been accounted for. Assets with similar residual loading vectors occupy nearby locations in this reduced factor space, indicating similar patterns of volatility behavior beyond the market mode.

After removing the market component, substantial structure remains in the residual space. While overall correlations are reduced, assets do not become independent. Instead, clusters of assets emerge that share similar residual loading profiles.This suggests that volatility dependence is not exhausted by a single common factor. Even after accounting for market-wide coordination, there exist coherent patterns of co-movement that reflect internal organization within the system.

The residualization serves two purposes in this analysis. It separates global stress effects from more localized interactions and provides a natural feature space in which assets can be compared and grouped without relying on predefined classifications.

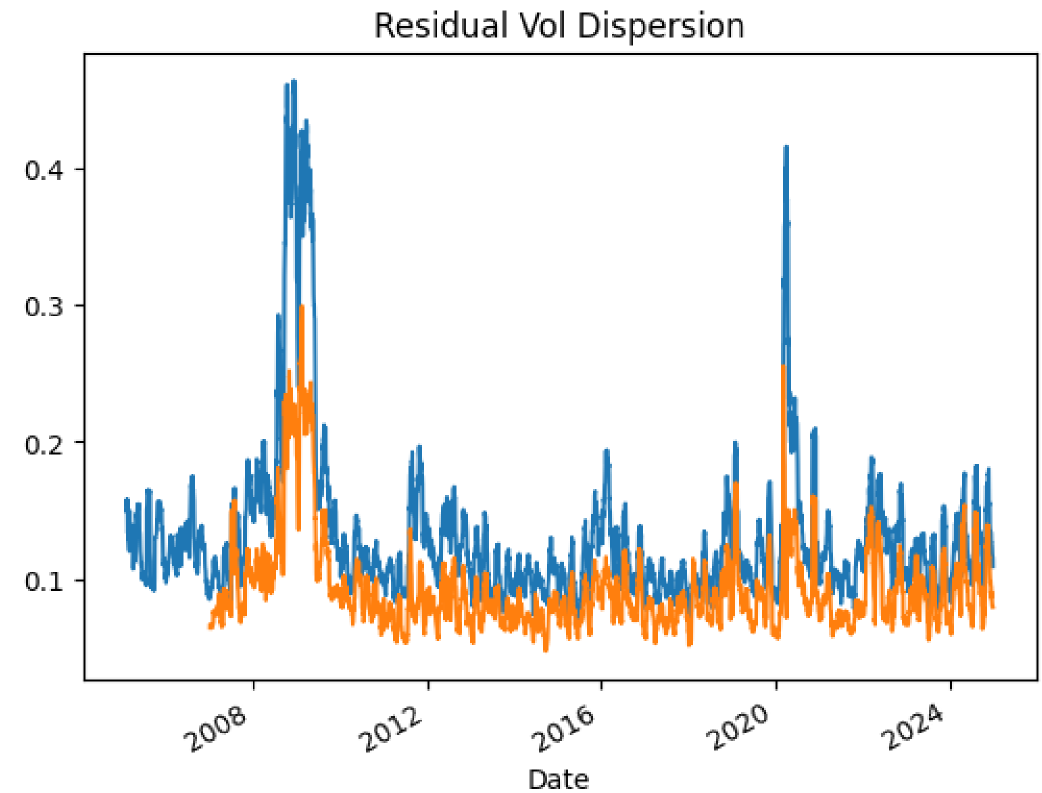


Figure 5: Residual Volatility Dispersion

**Endogenous Clustering**

The next goal is to reduce dimensionality for network analysis. Rather than tracking hundreds of pairwise relationships, we group assets into a smaller number of endogenous clusters based on their residual loading similarity. Each cluster will become a node in the subsequent network model.

Using Euclidean distance between residual loading vectors and hierarchical clustering with Ward linkage, we partition the S&P 500 into $k$ groups. The choice of $k$ (6) is pragmatic. Large enough to capture meaningful substructure, small enough to yield an interpretable network. Exploratory checks across $k \in [4, 10]$ revealed that k = 6 provided a stable clustering structure without excessive fragmentation, balancing robustness and interpretability.

Crucially, these clusters are not sectors or style factors. They emerge purely from volatility co-movement patterns. Members of the same cluster exhibit similar residual volatility dynamics, regardless of their industry or market cap.

The result is a compressed representation of the system of 500 initial individual assets into 6 volatility communities. This mapping allows us to study interactions at a structural level, moving from asset-level noise to system-level signal.

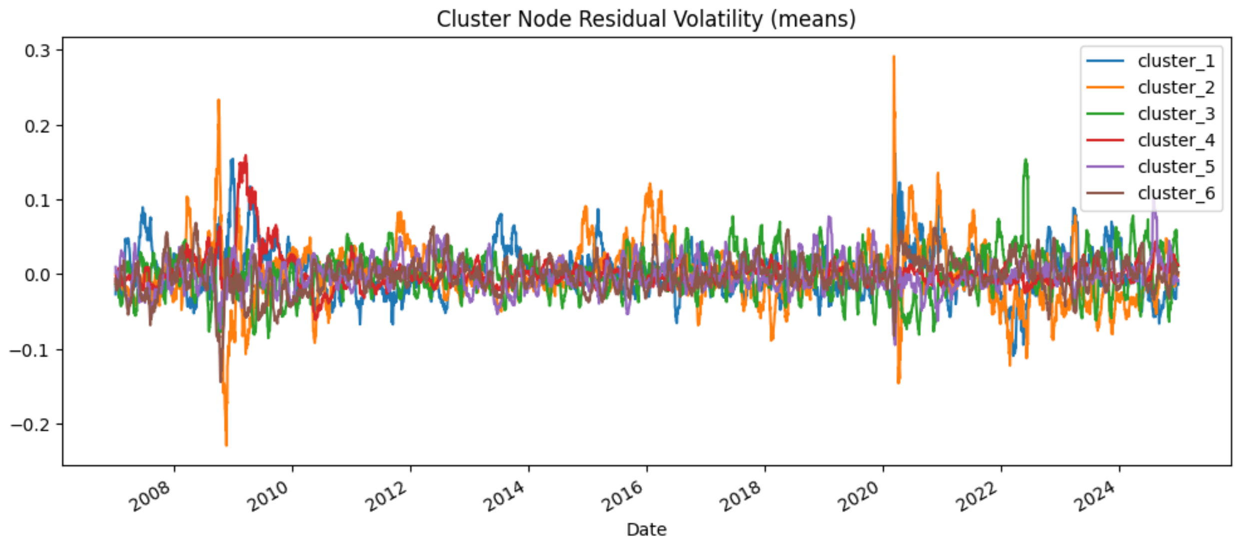


Figure 6: Clustered Volatility Means

**Network Construction**

Once endogenous communities have been defined, the system can be represented at a higher level of abstraction. Let $\bar{v_c}(t)$ denote the average residual volatility of cluster $c$ at time $t$, computed across all assets in that cluster.

Each community is treated as a node, and the relationship between nodes $c$ and $d$ is defined by their rolling correlation over a window of length $w$. A rolling window of 70 trading days was chosen as a compromise between temporal resolution and statistical stability. Shorter windows produced highly volatile correlation estimates, while longer windows smoothed over transitional dynamics and delayed the detection of structural shifts. The chosen window captures medium-term structural evolution without excessive estimation noise.

$$\rho_{cd}(t) = Corr(\bar{v_c}(t-w:t), \bar{v_d}(t-w:t))$$

In practice, we require at least 90% non-missing observations per node within the rolling window, and we discard windows where fewer than three clusters meet this criterion. The density threshold is set to $\mid \rho \mid \geq 0.2$. This intentionally low threshold captures modest co-movement between volatility clusters. After removing the dominant market-wide mode residual correlations are naturally subdued as most of the common signal has been removed in the residualization process.

This produces a time-varying network in which nodes correspond to internal volatility groups and edges reflect the strength of interaction between them. Rather than tracking hundreds of pairwise asset relationships, the analysis now focuses on how a small number of structural components interact.

Several simple network statistics are used to summarize this evolving structure:

**Average absolute correlation across nodes**

$$\bar{\rho} = \frac{2}{k(k-1)} \sum_{c<d} \mid \rho_{cd}(t) \mid$$

where k is the number of clusters.

**Connection density above a threshold $\theta$**

$$D(t) = \frac{|\{(c,d) : |\rho_{cd}(t)| \geq \theta\}|}{k(k-1)/2}$$

where absolute correlations are used to capture both positive and negative strong linkages.

**The largest eigenvalue** $\lambda_1(t)$ in the $k \times k$ correlation matrix $R(t)$ was used to capture the dominance of the primary mode of co-movement between clusters.

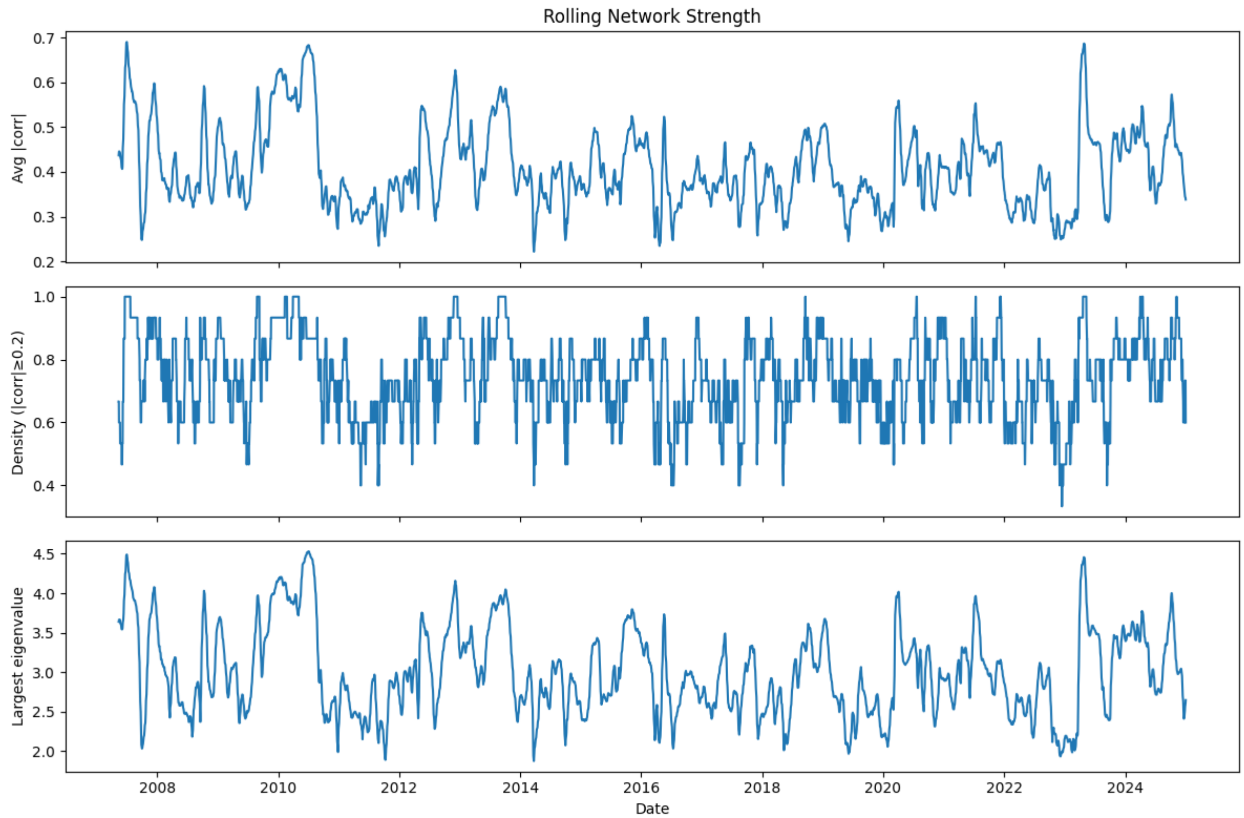


Figure 7: Rolling Network Metrics

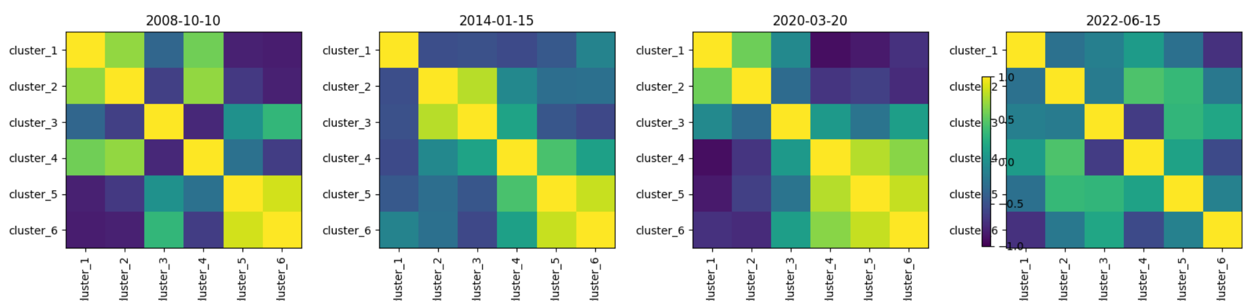


Figure 8: Correlation Snapshots

Figure 7 and 8 above show that network connectivity increases sharply during stress episodes. Nodes that are relatively independent in calm periods become tightly synchronized, and the system begins to behave as a more unified object. In quieter periods, structure is looser, connections weaken, and interactions become more heterogeneous.

### Systemic stress

The network layer provides a concise, time-varying summary of system-wide coupling. But a collection of statistics like average correlation, density, dominant eigenvalue is still just a set of correlated signals. To distill them into a single measure of systemic stress, we need a model that captures their shared latent structure.

An intuitive approach is to treat stress as a discrete regime such as calm versus crisis. But as the next section shows, this framing does not hold empirically. Instead, we model stress as a continuous latent process, inferred directly from the network statistics via a simple state-space model. This yields a single index that tracks the gradual accumulation and release of systemic pressure, aligning with the smooth evolution observed throughout the analysis.

### Discrete Regimes: Why they fail

A natural alternative to a continuous stress measure is to model markets as switching between discrete regimes. For example, “calm” and “crisis.” This is appealing conceptually as it matches the language of headlines and offers clear, interpretable states.

To test this, we applied hidden Markov models (HMMs) to the same network summary statistics (average absolute correlation, density, largest eigenvalue). If systemic stress were fundamentally discrete, an HMM should cleanly partition the timeline into distinct, persistent states with clearly separated statistical profiles.

The results were unconvincing. Estimated regimes were unstable across time, sensitive to initialization and parameter choices, and yielded states that overlapped substantially in their emission distributions. Rather than uncovering a small set of well-defined market environments, the models produced noisy, weakly separated clusters that shifted with minor variations in the training setup.

The absence of clean separation suggests that systemic risk does not evolve through discrete jumps. The empirical patterns observed throughout our analysis such as gradual rises in correlation, smooth compression of dimensionality, persistent but fluid network coupling are more consistent with a continuous approach. Stress accumulates and dissipates like pressure, not like a light switch.

Thus, rather than imposing artificial boundaries, we model stress as a continuous latent process. This aligns not only with the statistical failure of discrete models but with the geometric intuition that underlies the entire study: structure compresses and expands smoothly, not in jumps.

### Latent Systemic Stress Model

The network layer produced several time-varying statistics that summarize different aspects of system-wide coupling, including average absolute correlation $\bar{\rho}(t)$, network density $D(t)$ and the largest eigenvalue of the node level correlation matrix $\lambda_1(t)$.

While each metric is individually informative, they are highly correlated and appear to reflect a shared underlying phenomenon. Thus we introduce a latent state $S_t$ which is interpreted as the intensity of systemic coupling at time $t$. This state is not directly observed but is inferred through a simple state-space model:

$$\begin{aligned} S_t &= S_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \\ y_t &= \alpha + \beta S_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma) \end{aligned}$$

Here $y_t = [\bar{\rho}(t), D(t), \lambda_1(t)]$ is the vector of observed network statistics, treated as noisy measurements of the latent state $S_t$. The model is estimated via the Kalman filter yielding a smoothed time series $\hat{S}_t$ which we refer to as the Systemic Stress Index (SSI).

The standardized SSI is shown in figure 9. The index exhibits intuitive behavior: it rises during known stress episodes, remains elevated during sustained periods of instability, and decays gradually rather than exhibiting sharp regime-like jumps.

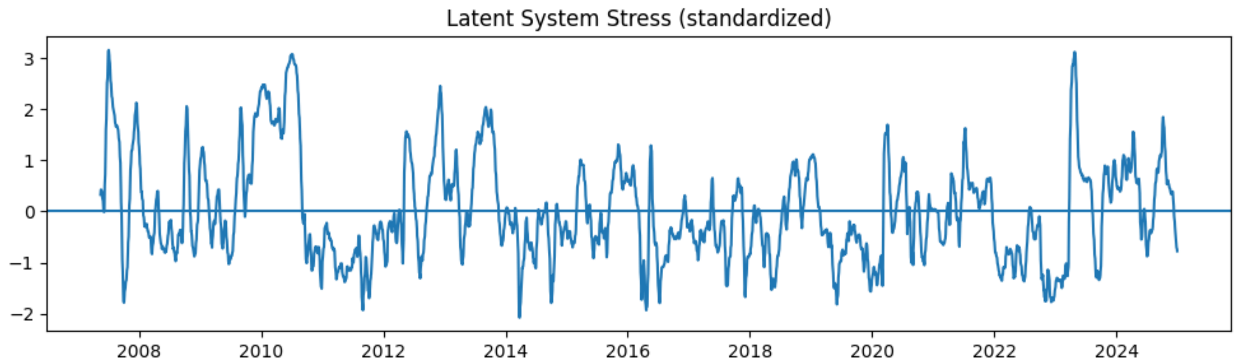


Figure 9: Standardized SSI

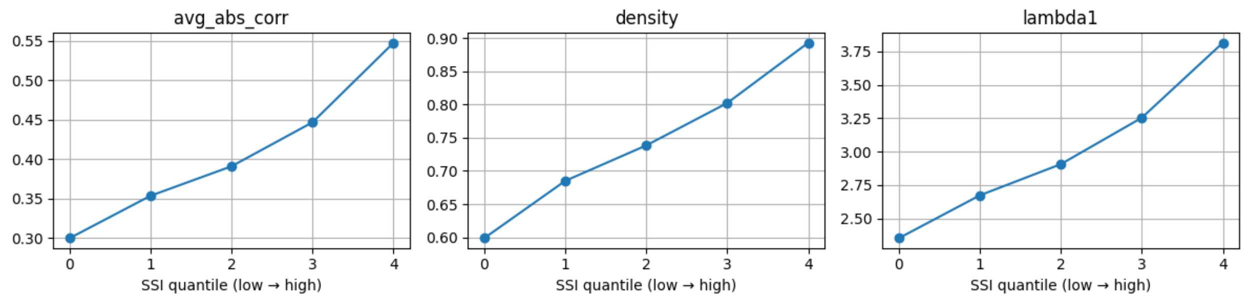


Figure 10: Standardized SSI

To assess whether the index meaningfully orders systemic conditions, we condition the original network statistics on SSI quantiles. Observations are grouped by SSI level (low to high), and within each group we compute the average of $\bar{\rho}(t), D(t)$ and $\lambda_1(t)$. The results are shown below in Figure 10. This provides internal validation that the latent state is not merely a smoothed combination of inputs, but captures a coherent ordering of systemic structure.

At this stage, the model is not interpreted as causal or structural. Its role is representational. It provides a compact, continuous summary of system-wide coordination that aligns with the geometric intuition developed throughout the analysis.

The SSI is intended to order systemic conditions, not to predict or classify them. The monotonic relationship with network metrics confirms that the index meaningfully ranks periods by their degree of stress-induced coupling, consistent with the structural compression observed in earlier stages. This does not constitute external validation, but it establishes internal consistency. The latent state orders market conditions in a way that reflects the underlying evolution of volatility co-movement.

While the latent state-space model provides a convenient compression of multiple network statistics into a single index, it should not be interpreted as a structural model of market dynamics. The inferred state is sensitive to modeling choices such as noise variances, scaling of inputs, and the assumed random walk evolution. Moreover, the Kalman filter inherently produces smoothed estimates, which may introduce a perception of continuity even if the underlying system contains sharper transitions. The model is therefore best understood as a representational tool rather than a causal mechanism.

## Discussion

This project began as an exploration rather than a solution. The goal was not to build a causal model of markets, propose a trading strategy, or claim discovery of a new factor. Instead, it asked a simpler and more open-ended question: what kinds of structure become visible in volatility when basic mathematical tools are applied carefully, and what might that structure suggest about the nature of systemic risk?

The methods used throughout are intentionally simple. Principal component analysis, clustering, correlation networks, and state-space models are all standard tools. None are optimized for this specific problem, and none are presented as optimal choices. The value of the exercise lies in transparency. Each step is interpretable, each transformation admits a geometric interpretation, and each result can be traced back to its construction rather than emerging from opaque modeling assumptions.

Several consistent patterns emerge. Volatility structure appears to be at least partly endogenous. Even after removing a market-wide component and avoiding any sector or factor information, the system does not become random. Assets organize into coherent communities, and those communities exhibit persistent interaction patterns. Structure is not imposed; it emerges.

Systemic risk also appears to evolve continuously rather than discretely. Attempts to impose regime boundaries using hidden Markov models produced unstable and weakly separated states. In contrast, treating stress as a continuous latent process yields a representation that aligns naturally with the gradual buildup and decay visible across all layers of the analysis.

Periods of stress are further characterized by a collapse in effective dimensionality. Whether viewed through eigenvalue concentration, network connectivity, or latent coupling intensity, instability is associated with increasing coordination and decreasing structural richness. As stress rises, the system becomes simpler, tighter, and more constrained.

These are not definitive conclusions. They are observations within the limits of the methods used, and those limits are substantial.

A recurring tension throughout the analysis is that the tools are not perfectly matched to the object being studied. Linear methods are applied to a system that is almost certainly nonlinear. Correlation-based geometry is used to approximate structure that may be curved, multi-scale, or topological in nature. In several places, the methods feel less like precise instruments and more like blunt approximations. But that limitation is also informative: it reveals both what these tools capture effectively and where they begin to fail.

There are clear directions where this framework could be extended. Nonlinear embeddings, manifold learning, kernel methods, or tools from algebraic topology may be more appropriate for capturing the true geometry of market structure. These directions are not pursued here, not because they are unimportant, but because the purpose of this work is exploratory rather than exhaustive. They represent natural continuations rather than omissions.

In that sense, this article is best read as a map of questions rather than a collection of answers. It documents what becomes visible when relatively simple tools are applied with care, and where those tools begin to break down. That boundary between what is captured and what is lost is arguably the most informative outcome of the entire exercise.