Article

Not peer-reviewed version

Beyond Volatility: A Leakage-Safe Residual-Stress Signal for Drawdown Risk Monitoring

Ting Liu *

Posted Date: 16 March 2026

doi: 10.20944/preprints202603.0395.v2

Keywords: drawdown risk; early warning; residual stress; principal component analysis; sector ETFs; leakage-safe backtesting; risk monitoring; volatility

Article

Beyond Volatility: A Leakage-Safe Residual-Stress Signal for Drawdown Risk Monitoring

Ting Liu

Northwestern University, Evanston, IL, USA; tingliu2027@u.northwestern.edu

Abstract

Monitoring equity drawdown risk requires real-time indicators that can be implemented without look-ahead bias and that may add information beyond standard volatility measures. This study develops a leakage-safe residual-stress indicator from cross-sectional PCA reconstruction errors in U.S. sector excess returns and examines whether it contributes to drawdown-risk monitoring. Using daily adjusted prices for SPY and 11 U.S. sector ETFs over 2020–2025, this study computes sector excess returns relative to SPY, estimate the common component with principal component analysis (PCA), and define residual stress as the cross-sectional root-mean-square magnitude of out-of-sample reconstruction residuals. The PCA mapping is estimated using information available only through $t-1$, stress is computed at t, and high-stress regimes are defined using rolling train-only quantile thresholds shifted forward by one trading day. Performance is evaluated using drawdown-onset events and early-warning metrics including ROC-AUC, PR-AUC, and horizon-H precision and recall. The results indicate that although residual stress is not a superior standalone alternative to realized volatility, it remains the stronger benchmark in overall classification performance. Residual stress is most useful as a complementary indicator of cross-sectional market dislocation rather than as a replacement for volatility. In the baseline sample, residual-stress spikes cluster around drawdown onsets, and conditional regime analysis shows that when volatility is low, high residual stress is associated with a materially higher probability of a drawdown onset within the next $H = 21$ trading days than the low-stress/low-volatility regime. Event-overlap and lead-time diagnostics further suggest that residual stress can flag a subset of onset episodes not captured by a simple volatility-threshold rule, although its primary incremental value lies in conditional risk stratification rather than in systematically earlier triggering. A longer-history proxy-sector analysis yields similar evidence of conditional complementarity while also confirming the stronger standalone performance of volatility. The paper's contribution is to develop a leakage-safe, interpretable, cross-sectional residual-stress diagnostic that improves conditional drawdown-risk stratification, especially in otherwise low-volatility states, rather than a standalone replacement for realized volatility. This interpretation is supported by both a modern ETF baseline sample and a longer-history proxy-sector sample, with the broader sample providing more stable evidence of complementarity across a larger set of drawdown onsets.

Keywords: drawdown risk; early warning; residual stress; principal component analysis; sector ETFs; leakage-safe backtesting; risk monitoring; volatility

---

1. Introduction

Large equity drawdowns are economically costly and difficult to anticipate using stable linear predictors alone. In practice, the objective is not precise return forecasting, but *risk-state detection*: identifying periods in which downside risk is elevated so that investors and risk managers can adjust exposure, hedging intensity, and monitoring decisions in real time. This paper studies whether a leakage-safe measure of *cross-sectional residual stress* contains complementary information about drawdown risk beyond standard volatility-based signals.

The starting point is that sector returns often share a low-dimensional common structure driven by broad market and macroeconomic forces. During periods of market stress, however, sector-level repricing may become less well described by that dominant common component. We interpret this residual fragmentation as a potentially informative dimension of market stress. To measure it, we apply principal component analysis (PCA) [4] to *sector excess returns* (sector minus SPY), extract a K-dimensional common component, and define residual stress as the cross-sectional root-mean-square magnitude of out-of-sample PCA reconstruction residuals. Intuitively, the signal rises when sector movements become increasingly misaligned with the dominant low-dimensional market structure [3,5–7].

The central question is whether this residual-stress signal captures information not already summarized by conventional volatility measures. This distinction matters because volatility measures the aggregate time-series amplitude of market fluctuations, whereas residual stress is intended to capture *cross-sectional dislocation* after removal of the dominant common component. The contribution of the paper is not to propose residual stress as a replacement for volatility, but to test whether it provides *complementary, state-dependent, and event-level* information for drawdown-risk monitoring under a leakage-safe design. Empirically, the paper evaluates the signal in two complementary samples: a contemporary investable U.S. sector-ETF panel used as the main real-time monitoring setting, and a longer-history proxy-sector panel used to assess whether the same conditional and event-level patterns remain stable across a larger number of drawdown episodes.

A further contribution is implementation discipline. At each date t, the PCA mapping (including centering and loadings) is estimated using information available only through $t-1$, and the residual-stress score is then computed out-of-sample for date t. High-stress regimes are identified using a rolling *train-only* quantile threshold constructed only from historical stress scores and shifted forward by one trading day, so that the regime label used for monitoring at t depends only on information available through $t-1$. This timing discipline is designed to avoid look-ahead bias and to permit a clean evaluation of real-time warning performance.

Drawdown-warning performance is evaluated using drawdown-onset events, event-study visualizations, and early-warning classification metrics including ROC-AUC, PR-AUC, and horizon-H precision and recall. Because volatility-based indicators are the natural benchmark for market stress and risk timing [8,9], all evaluations compare residual stress against volatility-based measures under the same leakage-safe timing protocol. The analysis also examines event-overlap and lead-time diagnostics to assess whether residual stress and volatility provide distinct warning information. To illustrate implementation implications, the paper further reports transaction-cost-aware applications, including a stress-managed SPY overlay and a residual-ranked sector long–short portfolio, while explicitly distinguishing monitoring value from standalone trading profitability.

This study differs from volatility-forecasting and machine-learning early-warning papers in both object and mechanism [22]. Rather than forecasting volatility itself or benchmarking flexible classifiers on broad cross-asset predictors, it asks whether cross-sectional residual dislocation across sector ETFs contains useful information for equity drawdown-risk monitoring after the dominant common market structure has been removed. Therefore, its contribution lies in residual market-structure diagnostics and their leakage-safe evaluation, rather than in volatility prediction or generic predictive-model comparison.

This paper develops a residual-stress indicator from out-of-sample PCA residuals in U.S. sector excess returns and evaluates its usefulness for equity drawdown-risk monitoring against standard volatility benchmarks. The contribution is not a claim of predictive dominance over realized volatility, which remains stronger in the baseline classification results. Instead, the paper shows that residual stress captures cross-sectional market fragmentation after removal of the dominant common component and can therefore add useful conditional information, especially in otherwise low-volatility states. The analysis is implemented throughout under a lagged train-only design, linking the evidence to a feasible real-time monitoring framework.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the data, leakage-safe signal construction, and drawdown-onset definition. Section 4 reports early-warning performance, conditional-risk evidence, robustness checks, and practical applications. Section 5 discusses interpretation and limitations. Section 6 concludes.

2. Literature Review

This paper relates to three strands of research: (i) low-dimensional factor structure and statistical decompositions of asset returns, (ii) market stress measurement using cross-sectional information, and (iii) early-warning evaluation and risk-managed overlays for drawdown control.

2.1. Low-Dimensional Return Structure and PCA-Based Decompositions

A large literature models asset returns using a small number of common factors. Multi-factor asset-pricing frameworks provide an economic motivation for decomposing returns into common and idiosyncratic components, while statistical factor models and dimensionality-reduction methods estimate latent low-rank structure directly from return panels when the number or identity of underlying factors is uncertain. Within this literature, principal component analysis (PCA) is a standard tool for extracting dominant co-movement and isolating an orthogonal residual component [4,16].

In this paper, PCA is used as a parsimonious statistical filter rather than as a structural asset-pricing model. Applied to sector *excess* returns (sector minus SPY), PCA removes the dominant common component and produces a residual panel that captures cross-sectional movements not explained by the leading low-dimensional market structure [16,17]. This residual component provides the basis for the paper's stress measure.

2.2. Cross-Sectional Residual Dispersion and Market Stress

Beyond aggregate co-movement, cross-sectional return *dispersion* and residual variation can contain information about market conditions. Elevated residual dispersion may reflect disagreement, dislocation, sector rotation, or fragmentation that is not fully summarized by broad market moves [5–7]. Related ideas also appear in work on cross-sectional dispersion, herding, disagreement, and uncertainty measures constructed from return panels and other financial data.

The approach in this study operationalizes this intuition through a transparent residual-stress score defined as the cross-sectional root-mean-square magnitude of PCA reconstruction residuals from sector excess returns. The empirical question is not whether residual stress should replace conventional stress proxies, but whether it contains complementary information relative to them. Because realized and related volatility measures are widely used as benchmark indicators of market stress and risk timing [8,9], they provide the natural baseline against which to evaluate the additional informational content of residual dispersion.

2.3. Drawdown Risk, Early-Warning Evaluation, and Risk-Managed Overlays

Predicting drawdowns is difficult because downside events are infrequent, nonlinear, and often regime-dependent [12,20]. A common response is to frame the problem as one of *risk-state detection* or *early warning*, where the goal is not precise return forecasting but timely identification of elevated downside risk. In this setting, event-based definitions and classification-oriented metrics are often more informative than average-return predictability alone.

In portfolio applications, such regime signals are sometimes mapped into overlays that reduce exposure during stress periods, trading off protection against turnover, tracking error, and transaction costs [12]. A related literature studies volatility-managed and risk-controlled exposures as implementable tools for drawdown mitigation [8,9]. Consistent with this perspective, we evaluate residual stress primarily as a *drawdown-risk monitoring* signal, using drawdown-onset early-warning metrics (ROC-AUC, PR-AUC, and horizon-H precision/recall), event-level diagnostics, and illustrative transaction-cost-aware applications under lagged and implementable timing [10,11,20].

2.4. Positioning of This Work

This paper sits at the intersection of four related literatures, but its objective differs from each of them in a specific way. Unlike volatility-forecasting studies, it does not seek to improve forecasts of market variance; instead, it asks whether cross-sectional residual fragmentation contains information about future drawdown onsets beyond realized volatility. Unlike PCA and statistical factor-model studies, it does not treat latent factors as the main empirical object. PCA is used here as a filtering step to remove the dominant common sector structure, after which the residual component is interpreted as a measure of cross-sectional dislocation. Relative to the broader literature on cross-sectional dispersion, disagreement, and herding, the proposed measure is more targeted because it is constructed from residual rather than raw sector excess returns. And unlike general machine-learning early-warning studies, the paper emphasizes an interpretable signal, onset-based evaluation, and an implementable train-only timing protocol rather than classifier complexity. Therefore, the contribution is not a new forecasting architecture or a claim of predictive dominance over volatility, but evidence that residual cross-sectional stress can improve conditional drawdown-risk monitoring within a leakage-safe framework [1].

3. Methodology

3.1. Data, Return Construction, and Reproducibility

The empirical design uses two complementary sector-based panels. The first is a contemporary U.S. sector-ETF sample consisting of the S&P 500 ETF (SPY) and 11 sector ETFs: XLC, XLY, XLP, XLE, XLF, XLV, XLI, XLK, XLB, XLU, and XLRE. This ETF sample is downloaded from Yahoo Finance, runs from 2020–01–01 to 2025–12–31, and serves as the contemporary, directly investable benchmark used in the main real-time monitoring analysis. The second is a longer-history proxy-sector panel spanning 1999–01–01 to 2025–12–31, used to assess whether the same conditional and event-level patterns remain stable across a larger number of drawdown episodes. For the ETF sample, daily adjusted close prices are used to account for dividends and stock splits. After download, all series are aligned to a common trading calendar, and any date with missing observations in any ETF is removed to ensure a balanced panel. The same alignment and reproducibility rules are applied within each sample so that signal construction and evaluation remain comparable across the contemporary ETF panel and the longer-history proxy panel.

Let $P_{i,t}$ denote the adjusted close price of sector ETF i on day t, and let $P_{m,t}$ denote the adjusted close price of SPY. Simple daily returns are computed as

$$r_{i,t} = \frac{P_{i,t}}{P_{i,t-1}} - 1, \qquad r_{m,t} = \frac{P_{m,t}}{P_{m,t-1}} - 1. \tag{1}$$

To isolate sector-specific movements beyond the broad market component, we define sector excess returns as

$$x_{i,t} = r_{i,t} - r_{m,t}. \tag{2}$$

Stacking these across the $N = 11$ sectors yields the cross-sectional excess-return vector

$$\mathbf{x}_t = (x_{1,t}, \dots, x_{N,t})^\top \in \mathbb{R}^N, \tag{3}$$

and collecting observations over $t = 1, \dots, T$ gives the return panel

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_T^\top \end{bmatrix} \in \mathbb{R}^{T \times N}. \tag{4}$$

All data used in this study are publicly available. Daily adjusted close prices are retrieved from Yahoo Finance via `quantmod`. A complete replication package (R scripts and configuration files)

 doi:10.20944/preprints202603.0395.v2

is provided as Supplementary Materials accompanying this paper. The configuration file freezes the sample period and preserves the leakage-safe timing protocol used throughout the paper. The replication package does not redistribute raw price data; instead, the download script retrieves them directly from the public source.

3.2. Leakage-Safe PCA Residual-Stress Construction

The paper measures cross-sectional market stress using the residual component of sector excess returns after removing a low-dimensional common structure. Let $\boldsymbol{\mu}_{t-1} \in \mathbb{R}^N$ denote the training-sample mean of sector excess returns, estimated using only information available through $t-1$. In the baseline specification, we use an expanding-window PCA with an initial burn-in window of $W_{\text{PCA}} = 252$ trading days. The centered excess-return vector at date t is

$$\tilde{\mathbf{x}}_t = \mathbf{x}_t - \boldsymbol{\mu}_{t-1}. \tag{5}$$

In the reported baseline specification, the common component is restricted to the first two principal components ($K = 2$), corresponding to PC1 and PC2 throughout the main analysis and robustness exercises.

Out-of-sample PCA projection. Let $\mathbf{L}_{t-1} = [\boldsymbol{\ell}_{1,t-1}, \boldsymbol{\ell}_{2,t-1}] \in \mathbb{R}^{N\times 2}$ denote the orthonormal loading matrix obtained by applying PCA to the centered training sample through $t-1$, so that

$$\mathbf{L}_{t-1}^{\top}\mathbf{L}_{t-1} = \mathbf{I}_2. \tag{6}$$

The corresponding out-of-sample factor score for day t is

$$\mathbf{f}_t = \mathbf{L}_{t-1}^{\top}\tilde{\mathbf{x}}_t \in \mathbb{R}^2. \tag{7}$$

The rank-2 common-component reconstruction is then

$$\widehat{\mathbf{x}}_t = \mathbf{L}_{t-1}\mathbf{f}_t = \mathbf{L}_{t-1}\mathbf{L}_{t-1}^{\top}\tilde{\mathbf{x}}_t. \tag{8}$$

Residual vector and stress score. The residual vector is defined as the reconstruction error

$$\mathbf{e}_t = \tilde{\mathbf{x}}_t - \widehat{\mathbf{x}}_t = \left(\mathbf{I}_N - \mathbf{L}_{t-1}\mathbf{L}_{t-1}^{\top}\right)\tilde{\mathbf{x}}_t \in \mathbb{R}^N. \tag{9}$$

By construction, $\mathbf{e}_t$ is orthogonal to the estimated factor space:

$$\mathbf{L}_{t-1}^{\top}\mathbf{e}_t = \mathbf{0}. \tag{10}$$

This study summarizes the cross-sectional magnitude of these residuals using the root-mean-square (RMS) residual-stress score

$$\text{Stress}_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N} e_{i,t}^2}. \tag{11}$$

Intuitively, Stress_t rises when sector movements are less well captured by the dominant low-rank common structure.

Throughout, the PCA mapping and the stress score at date t are constructed using only information available through $t-1$. This timing discipline ensures that Stress_t is implementable in real time and free of look-ahead bias. PCA provides the standard rank-2 least-squares approximation to the centered training panel, so the residual vector can be interpreted as the component of sector excess returns not explained by the leading low-dimensional common structure.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.0395.v2

3.3. Train-Only Regime Construction and Implementable Timing

To maintain a leakage-safe design, any quantity used for labeling, benchmarking, or trading at date t is constructed using information available no later than $t-1$.

Train-only stress threshold. Using a lookback window of length W_{thr} and quantile level q, we compute the rolling train-only threshold

$$Q_{t-1}^{(q)} = \text{Quantile}_q(\text{Stress}_{t-W_{\text{thr}}}, \ldots, \text{Stress}_{t-1}), \tag{12}$$

and define the stress-regime indicator as

$$\mathbb{I}_t^{\text{stress}} = \mathbb{1}\left\{\text{Stress}_t > Q_{t-1}^{(q)}\right\}. \tag{13}$$

In implementation, the rolling quantile in Equation (12) is evaluated with a right-aligned window and shifted forward by one trading day, so that the threshold applied at date t depends only on historical stress values observed through $t-1$.

Lagged, implementable signals. Any strategy or overlay that depends on the stress regime uses the lagged label $\mathbb{I}_{t-1}^{\text{stress}}$ to set exposure for day t. Likewise, sector-ranking signals for the long–short application are formed from the lagged residual vector $\mathbf{e}_{t-1}$, portfolio weights are set at the close of $t-1$, and realized portfolio returns are computed using contemporaneous excess returns $\mathbf{x}_t$. This convention ensures that both regime identification and portfolio formation are implementable and free of look-ahead bias.

3.4. Drawdown-Onset Events and Horizon-H Early-Warning Labels

SPY drawdown and drawdown-onset events. Let $r_{m,t}$ denote the simple daily return of SPY from Equation (1). Define the cumulative equity curve by

$$E_t = \prod_{\tau \le t}(1 + r_{m,\tau}), \tag{14}$$

and the drawdown series by

$$\text{DD}_t = \frac{E_t}{\max_{\tau \le t} E_\tau} - 1. \tag{15}$$

Fix a drawdown threshold $\delta < 0$ (e.g., $\delta = -10\%$). The drawdown-region indicator is

$$\mathbb{I}_t^{\text{DD}} = \mathbb{1}\{\text{DD}_t \le \delta\}, \tag{16}$$

and the drawdown-onset event is defined as the first entry into the drawdown region:

$$\mathbb{I}_t^{\text{onset}} = \mathbb{I}_t^{\text{DD}}\left(1 - \mathbb{I}_{t-1}^{\text{DD}}\right). \tag{17}$$

Thus, $\mathbb{I}_t^{\text{onset}} = 1$ only on dates when the drawdown crosses below δ from above, avoiding repeated labels during the same drawdown episode.

Horizon-H early-warning label. For classification-style early warning, we define the binary target

$$y_t^{(H)} = \mathbb{1}\left\{\sum_{h=1}^{H} \mathbb{I}_{t+h}^{\text{onset}} \ge 1\right\}, \tag{18}$$

which equals one if at least one drawdown-onset event occurs within the next H trading days. The corresponding unconditional event rate is

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t^{(H)}. \tag{19}$$

 doi:10.20944/preprints202603.0395.v2

This study reports sensitivity checks over threshold, drawdown, and horizon definitions while preserving the same leakage-safe timing protocol throughout.

3.5. Benchmark Signals

Main volatility benchmark. Let $r_{m,t}$ denote the simple daily return of SPY from Equation (1). We define realized volatility as the annualized rolling standard deviation of daily SPY returns over a window of length W_{vol}:

$$\mathrm{Vol}_t = \sqrt{252} \cdot \mathrm{SD}\big(r_{m,t-W_{\mathrm{vol}}+1}, \ldots, r_{m,t}\big). \tag{20}$$

The continuous score Vol_t serves as the main benchmark in ROC/PR evaluation.

Train-only volatility regime. Analogous to the stress-regime construction, we compute the rolling train-only quantile threshold

$$Q^{(q_{\mathrm{vol}})}_{t-1,\mathrm{vol}} = \mathrm{Quantile}_{q_{\mathrm{vol}}}\Big(\mathrm{Vol}_{t-W_{\mathrm{thr,vol}}}, \ldots, \mathrm{Vol}_{t-1}\Big), \tag{21}$$

and define the volatility-regime indicator as

$$\mathbb{I}^{\mathrm{vol}}_t = \mathbb{1}\Big\{\mathrm{Vol}_t > Q^{(q_{\mathrm{vol}})}_{t-1,\mathrm{vol}}\Big\}. \tag{22}$$

As with the stress threshold, the rolling quantile in Equation (21) is right-aligned and shifted forward by one trading day so that the regime label at date t depends only on information available through $t-1$.

Realized volatility serves as the main benchmark in the baseline comparison. To assess whether residual stress contains complementary information relative to other standard stress proxies, appendix robustness analysis also considers additional benchmark signals, including downside semivolatility, the VIX when available, cross-sectional return dispersion, average pairwise sector correlation, and a sector breadth measure based on the share of negative excess returns. These auxiliary benchmark signals are constructed under the same train-only timing discipline as the main stress and volatility measures.

3.6. Bootstrap Inference

To assess sampling uncertainty in early-warning metrics and conditional-regime comparisons, the paper uses a moving block bootstrap that resamples contiguous time-series blocks of daily observations. This procedure preserves short-run serial dependence in signal realizations, early-warning labels, and regime indicators. Bootstrap confidence intervals are reported for selected ROC-AUC, PR-AUC, precision, recall, F1, and conditional event-rate differences. In addition, bootstrap comparisons are used to assess selected differences in ROC-AUC across benchmark signals. All bootstrap exercises preserve the same leakage-safe timing protocol as the main analysis.

3.7. Overlay Implementation and Transaction Costs

These portfolio rules are treated as illustrative applications rather than as the primary basis for evaluating the signal.

Baseline portfolio return. Let $\mathbf{w}_{t-1}$ denote the portfolio weights set at the close of day $t-1$ and held over day t. More generally, let $\mathbf{r}^*_t$ denote the contemporaneous vector of traded returns, where $\mathbf{r}^*_t$ corresponds to raw asset returns for the SPY overlay and to sector excess returns for the residual-ranked long–short application. The gross portfolio return is

$$r^{\mathrm{gross}}_{p,t} = \mathbf{w}^\top_{t-1}\mathbf{r}^*_t. \tag{23}$$

Overlay rule. Let $\mathbb{I}^{\text{ov}}_{t-1} \in \{0,1\}$ denote an implementable overlay trigger known at time $t-1$ (e.g., $\mathbb{I}^{\text{stress}}_{t-1}$ or $\mathbb{I}^{\text{vol}}_{t-1}$). For overlay intensity $\lambda \in (0,1)$, we scale risky exposure by

$$s_{t-1} = 1 - \lambda\, \mathbb{I}^{\text{ov}}_{t-1}. \tag{24}$$

The overlay-adjusted weights are

$$\tilde{\mathbf{w}}_{t-1} = s_{t-1}\mathbf{w}_{t-1}, \tag{25}$$

so that a "30% overlay" corresponds to $\lambda = 0.30$ and reduces risky exposure to 70% of the baseline allocation when the trigger is active. The remaining fraction $1 - s_{t-1}$ is allocated to cash with zero return, implying

$$\tilde{r}^{\text{gross}}_{p,t} = \tilde{\mathbf{w}}^{\top}_{t-1}\mathbf{r}^*_t = s_{t-1}\,\mathbf{w}^{\top}_{t-1}\mathbf{r}^*_t. \tag{26}$$

Transaction costs and net returns. This paper applies proportional transaction costs to portfolio turnover. Let

$$\text{TO}_t = \sum_i |\tilde{w}_{i,t} - \tilde{w}_{i,t-1}| \tag{27}$$

denote one-way turnover in weights at the rebalance from $t-1$ to t after the overlay is applied. Given a one-way cost rate c, the net return is

$$r^{\text{net}}_{p,t} = \tilde{r}^{\text{gross}}_{p,t} - c \cdot \text{TO}_t. \tag{28}$$

Unless otherwise stated, this study sets $c = 5$ bps per unit of one-way turnover and report performance statistics using the net return series $\{r^{\text{net}}_{p,t}\}$.

4. Results

4.1. Residual Stress Dynamics Around Drawdown Onsets

This study begins by examining whether the leakage-safe residual-stress signal is visibly related to subsequent drawdown episodes. Figure 1 overlays the residual-stress score with its train-only rolling quantile threshold in the top panel and compares it with the SPY drawdown series in the bottom panel. Rug marks indicate drawdown-onset events, defined as the first entry below the pre-specified drawdown threshold. Residual-stress spikes cluster around many onset dates, indicating that elevated cross-sectional residual dispersion tends to coincide with, and in some episodes precede, transitions into drawdown states.

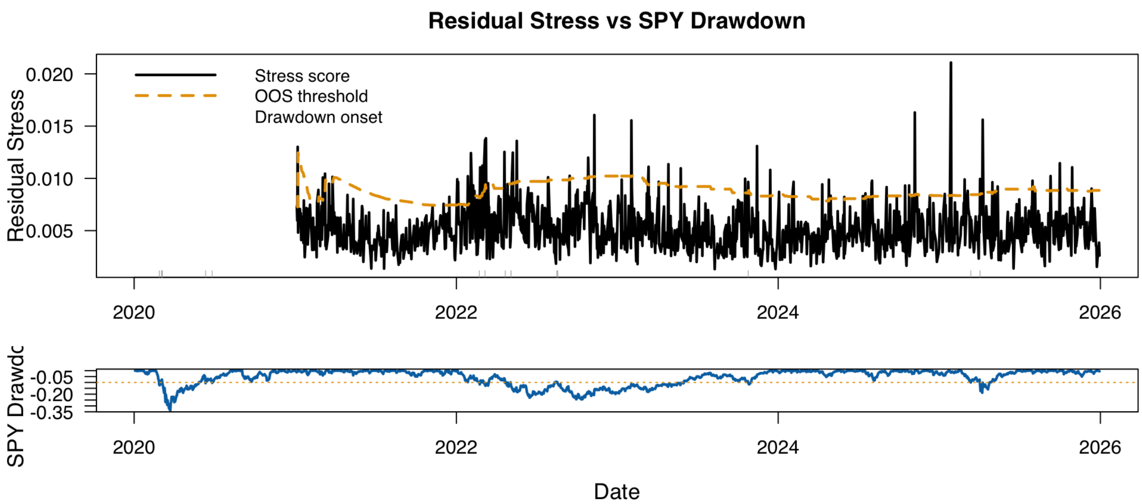

Figure 1. Residual stress (top) and SPY drawdown (bottom) with out-of-sample thresholds. Rug marks denote drawdown-onset events.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.0395.v2

Figure 2 reports an event-study view of the residual-stress score around drawdown onsets. For each onset, days are aligned by event time τ (with $\tau = 0$ at the onset), and the cross-event mean stress is computed over the ± 30 trading-day window. The shaded band summarizes cross-event dispersion. Average stress rises around onset episodes, consistent with the interpretation that residual dispersion captures deterioration in cross-sectional market structure during transitions into higher-risk states.

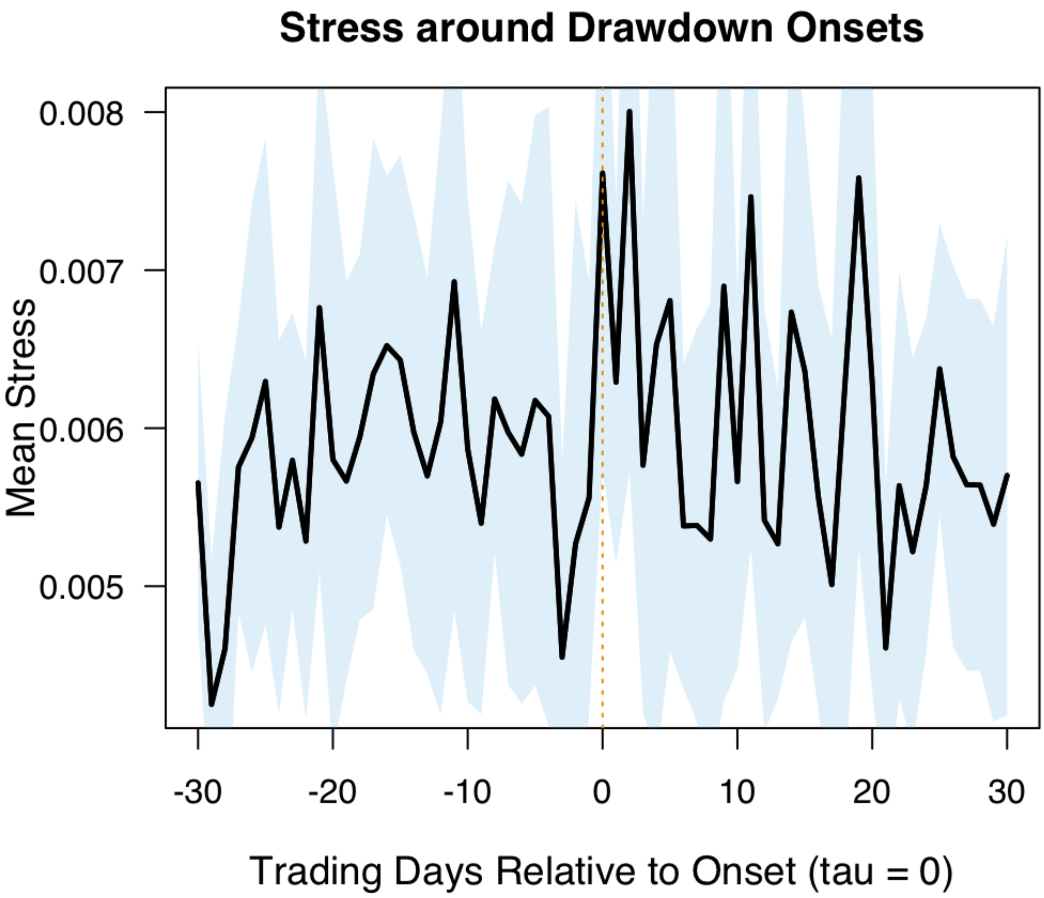

Figure 2. Event-study of residual stress around drawdown onsets (± 30 trading days; $\tau = 0$ at onset).

4.2. Early-Warning Evaluation for Drawdown Onsets

Next, this study evaluates whether the residual-stress score contains early-warning information for predicting whether a drawdown onset occurs within the next $H = 21$ trading days. Figure 3 reports the ROC and precision–recall curves for the residual-stress score. The ROC curve summarizes the trade-off between true- and false-positive rates across score thresholds, while the precision–recall curve is especially informative under class imbalance. In the baseline specification, the unconditional positive-label rate is $\bar{y} = \Pr(y_t^{(H)} = 1) \approx 13.3\%$ (Equation (19)).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.0395.v2

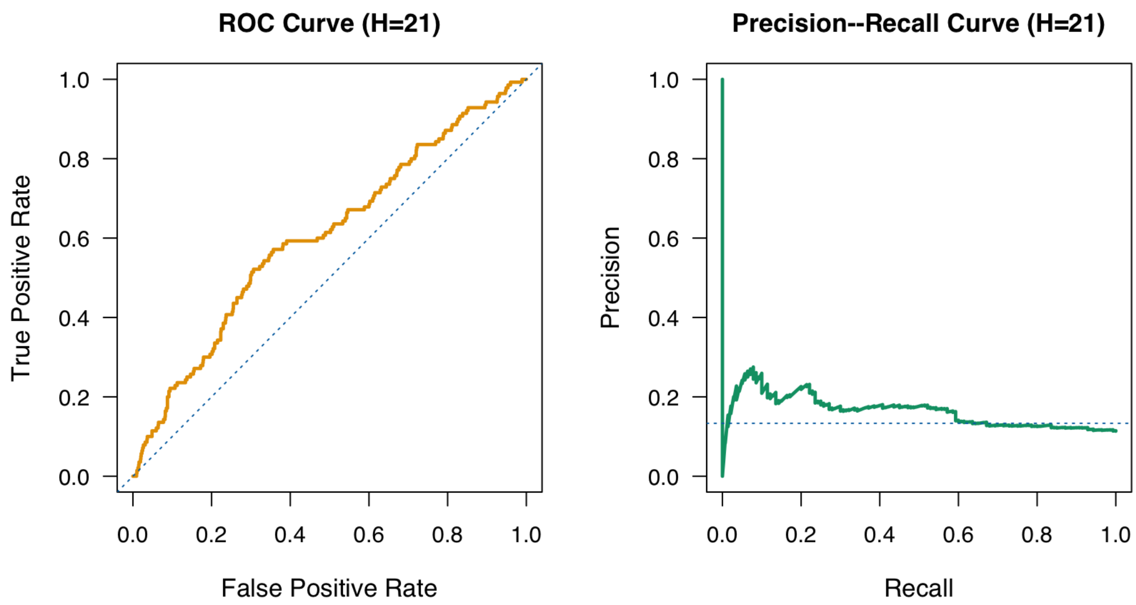

Figure 3. Early-warning performance ($H = 21$): ROC curve (left) and precision–recall curve (right) for the residual-stress score.

Table 1 (Panel A) reports aggregate early-warning metrics for residual stress and the SPY-volatility benchmark. Residual stress exhibits nontrivial warning content in the main sample, with ROC-AUC = 0.604 and PR-AUC = 0.162, both above a no-skill baseline. However, realized volatility remains the stronger standalone benchmark, with ROC-AUC = 0.685 and PR-AUC = 0.186, as well as higher precision, recall, and F1. Accordingly, the main purpose of the subsequent analysis is not to interpret residual stress as a superior standalone classifier, but to examine whether it contributes complementary information for drawdown-risk monitoring in settings where volatility alone may be incomplete.

4.3. Complementarity Relative to Volatility

The central empirical question is whether residual stress captures drawdown-risk information not already summarized by standard volatility measures. Table 1 separates baseline and complementarity evidence. Panel A reports standalone early-warning metrics for residual stress and SPY volatility, while Panels B–D evaluate whether residual stress adds information through joint-regime conditioning, event-level overlap diagnostics, and lead-time comparisons under the same train-only timing protocol used throughout the paper. In this setting, the role of residual stress is not to replace volatility as a standalone warning score, but to improve risk stratification by contributing conditional and event-level information beyond what is captured by volatility alone. The main-sample evidence is necessarily limited by the small number of drawdown-onset events and the infrequency of high-stress cells. For this reason, the longer-history proxy-universe robustness exercise is an important complement to the main-sample analysis. In that broader sample, the number of onset events increases substantially and the conditional complementarity pattern becomes more stable, supporting the interpretation that residual stress contributes useful information for drawdown-risk stratification beyond volatility.

Joint-regime conditioning.

Table 1 (Panel B) reports the conditional probability of a drawdown onset within the next $H = 21$ trading days under the joint regimes defined by residual stress and SPY volatility. Conditional on low volatility, high residual stress increases the onset probability from 7.78% in the low-stress/low-volatility regime to 17.02% in the high-stress/low-volatility regime. The joint high-stress/high-volatility regime exhibits the highest onset probability (36.0%), while the low-stress/high-volatility regime also carries elevated risk (28.66%). These patterns are consistent with the interpretation that residual stress provides complementary conditional information for drawdown-risk stratification, particularly in states where volatility alone remains subdued. At the same time, the high-stress regimes occur infrequently in the main sample ($N = 47$ and $N = 25$ days), and the bootstrap confidence interval for the Q10–Q00 difference includes zero. The main-sample evidence should therefore be interpreted as suggestive rather than definitive.

Event-level overlap and lead-time.

Table 1 (Panel C) evaluates whether residual stress flags onset events not captured by volatility thresholding within a fixed 63-trading-day lookback window. In the main sample, both alarms occur before 57.1% of onset events, residual stress uniquely flags one onset (7.1%), volatility uniquely flags none, and neither alarm occurs before 35.7% of onsets. Table 1 (Panel D) reports lead-time diagnostics for paired events in which both alarms occur. Median lead times are similar for stress and volatility (57.5 versus 60.0 trading days), and the stress alarm arrives earlier in 37.5% of paired cases. Taken together, these results suggest that the main-sample value of residual stress lies more in conditional risk stratification and partial event differentiation than in systematically earlier warning signals. Residual stress therefore appears most useful as a complementary, state-dependent signal rather than as a standalone substitute for volatility.

Table 1. Early-warning performance and volatility-complementarity diagnostics in the main sample ($H = 21$, $\delta = -10\%$). Panel A reports standalone early-warning metrics for residual stress and SPY volatility. Panel B reports regime-conditional onset probabilities under the joint residual-stress and volatility regimes. Panel C summarizes event-level overlap within a 63-day lookback window. Panel D reports lead-time diagnostics for paired events.

Panel A: Early-warning metrics (higher is better).

Signal	ROC-AUC	PR-AUC	Precision	Recall	F1	Event rate
Residual stress	0.604	0.162	0.236	0.121	0.160	0.133
SPY volatility	0.685	0.186	0.297	0.386	0.335	0.133

Panel B: Conditional onset probability by joint regimes.

Regime (Stress, Vol)	N days	$\mathbb{P}(\text{onset within } H)$
Low, Low (Q00)	1003	0.0778
High, Low (Q10)	47	0.1702
Low, High (Q01)	157	0.2866
High, High (Q11)	25	0.3600

Panel C: Event-level overlap within a 63-day lookback window.

Category	N events	Share
Both	8	0.571
Stress only	1	0.071
Vol only	0	0.000
Neither	5	0.357

Panel D: Lead-time to onset (paired events; lookback = 63).

Metric	Value
n_{paired}	8
Median lead (stress)	57.5
Median lead (vol)	60.0
Mean lead (stress)	51.75
Mean lead (vol)	57.75
Pr(stress earlier)	0.375

4.4. Broader-Sample Evidence and Robustness

The paper's central interpretation is evaluated from two complementary empirical views: a modern sector-ETF baseline sample and a longer-history proxy-sector sample. The modern sector-ETF

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.0395.v2

sample (2020–2025) provides the primary contemporary baseline, but it is short and contains only 14 drawdown onsets. We therefore pair it with a longer-history proxy sector universe spanning 1999–2025, which raises the onset count to 55 and allows a more informative assessment of whether the conditional and event-level role of residual stress is stable across a broader set of drawdown episodes. We also report moving-block-bootstrap inference and parameter-sensitivity checks to quantify sampling uncertainty and assess the robustness of the main interpretation.

Long-history proxy-universe evidence.

The longer-history proxy-sector sample provides a broader-event view of the paper's main interpretation by expanding the analysis to 1999–2025 and increasing the number of drawdown onsets from 14 to 55. In this sample, residual stress remains weaker than SPY volatility as a standalone classifier (ROC-AUC = 0.569 versus 0.615; PR-AUC = 0.108 versus 0.131), consistent with the baseline sample. However, the complementary role of residual stress becomes more stable. Conditional on low volatility, high residual stress raises the drawdown-onset probability from 7.72% in the low-stress/low-volatility regime to 15.06% in the high-stress/low-volatility regime, and the bootstrap interval for the $Q10 - Q00$ contrast is strictly positive: 0.0735 with 95% interval $[0.0244, 0.1288]$. Event-level diagnostics also strengthen in the broader sample: residual stress uniquely flags seven onset events, volatility uniquely flags none, and the stress alarm arrives earlier in 66.7% of paired cases, with median lead times of 52 trading days for stress versus 40 trading days for volatility. Taken together, the longer-history evidence supports the same substantive conclusion as the modern ETF sample, but with greater event depth and more stable conditional evidence.

Bootstrap comparison across multiple baselines.

Moving-block-bootstrap inference in the modern ETF sample quantifies the uncertainty around the baseline comparisons. The intervals confirm that residual stress contains nontrivial warning content but do not support a claim of standalone dominance over volatility-based benchmarks. Relative to SPY volatility, the bootstrap difference in ROC-AUC is negative (-0.102), with a 95% interval of $[-0.214, 0.028]$. Relative to the VIX, the difference is also negative (-0.126), with a 95% interval of $[-0.198, -0.046]$. Residual stress nevertheless compares favorably to several weaker auxiliary baselines, including average pairwise correlation and breadth. These results reinforce the paper's interpretation that the value of residual stress lies in conditional and event-level complementarity rather than in unconditional predictive dominance.

Table 2. Summary of baseline and broader-sample evidence for residual-stress complementarity.

Sample	Onsets	ROC-AUC S	ROC-AUC V	PR-AUC S	PR-AUC V	$Q10-Q00$	95% CI	S-only	Pr(S earlier)
Modern ETF sample	14	0.604	0.685	0.162	0.186	0.092	[-0.006, 0.228]	1	0.375
Long-history proxy sample	55	0.569	0.615	0.108	0.131	0.074	[0.024, 0.129]	7	0.667

Notes: S denotes residual stress and V denotes SPY realized volatility. The modern ETF sample spans 2020–2025; the long-history proxy sample spans 1999–2025. $Q10 - Q00$ is the difference in drawdown-onset probability between the high-stress/low-volatility and low-stress/low-volatility regimes. S-only denotes onset episodes uniquely flagged within the event-matching window; V-only equals zero in both samples. Pr(S earlier) is the share of paired events for which residual stress signaled before volatility.

Parameter sensitivity.

Additional sensitivity checks over the train-only threshold quantile, drawdown definition, and early-warning horizon leave the broad interpretation unchanged. Across these alternatives, residual stress continues to exhibit nontrivial warning content, while volatility remains the stronger standalone benchmark on average. The contribution of residual stress therefore continues to lie in leakage-safe conditional and event-level complementarity rather than in unconditional predictive dominance.

4.5. Economic Performance and Implementation Frictions

As a secondary implementation check, we examine whether the warning signals can be mapped into simple overlay and trading rules after transaction costs. These exercises are not part of the paper's primary contribution in drawdown-risk monitoring. Instead, they are included to illustrate implementation frictions and to clarify whether information useful for risk-state identification also translates into portfolio decisions.

Figure 4 reports cumulative wealth for SPY buy-and-hold returns, two stress-managed SPY overlay variants (residual-stress and volatility-based), and the residual-ranked long–short portfolio reported as a net-return series. Figure 5 shows the corresponding drawdown profiles.

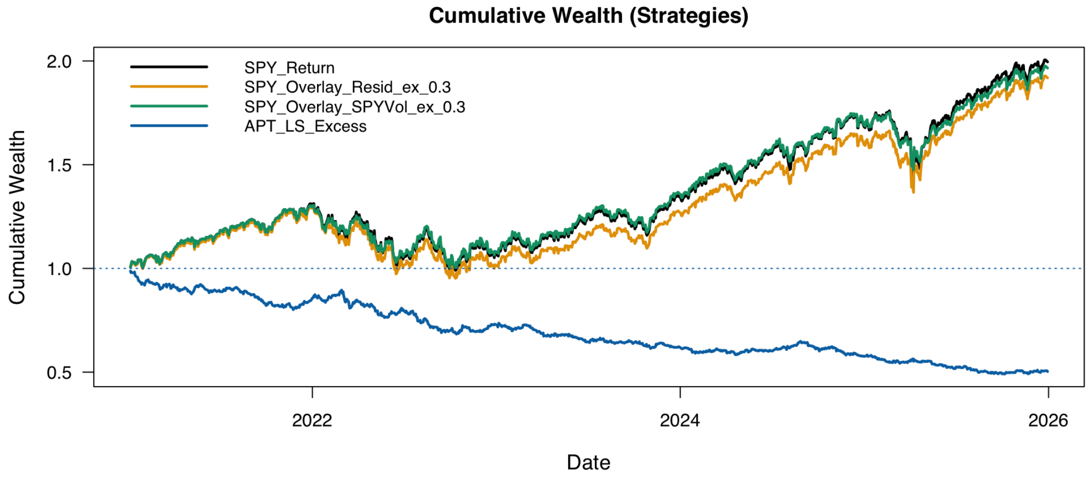

Figure 4. Cumulative wealth for SPY buy-and-hold, two stress-managed SPY overlays, and the residual-ranked long–short strategy (net returns).

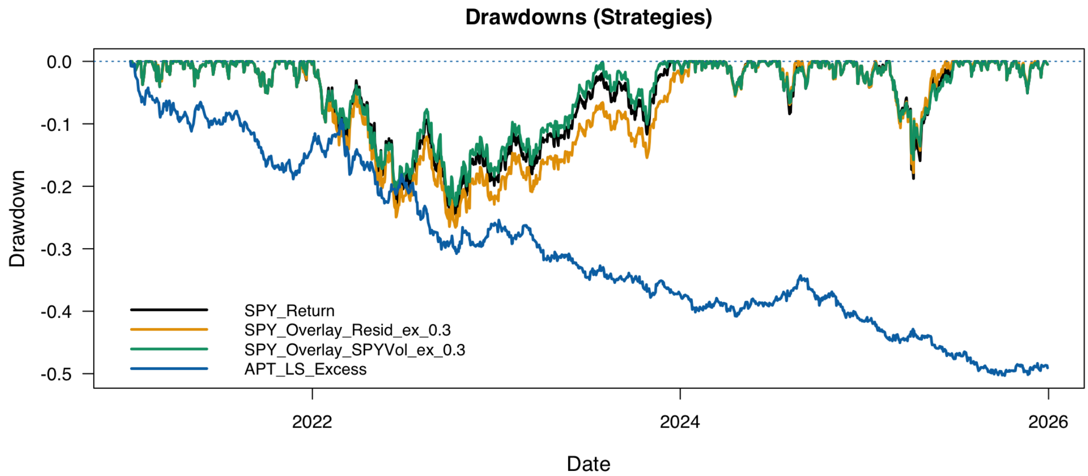

Figure 5. Drawdowns for SPY buy-and-hold and the stress-managed SPY overlay strategies (net returns).

Table 3 summarizes annualized performance over the main sample. The volatility-based overlay attains the highest Sharpe ratio and the smallest maximum drawdown among the SPY-based strategies. The residual-stress overlay is economically interpretable as a simple risk-state rule, but in this sample it slightly underperforms the SPY buy-and-hold benchmark and exhibits a somewhat deeper drawdown. The residual-ranked long–short strategy performs poorly after transaction costs, indicating that information useful for aggregate drawdown-risk monitoring does not necessarily translate into robust short-horizon cross-sectional alpha. Accordingly, these implementation results reinforce the interpretation of residual stress as a monitoring and risk-management variable rather than as a source of standalone trading alpha.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.0395.v2

Table 3. Main-sample strategy performance summary (annualized).

Strategy	Ann. Return	Ann. Vol	Sharpe	MaxDD
SPY buy-and-hold	0.1491	0.1711	0.8975	-0.2450
SPY overlay (residual stress, 30%)	0.1400	0.1665	0.8702	-0.2658
SPY overlay (volatility, 30%)	0.1456	0.1541	0.9592	-0.2308
Residual-ranked long–short	-0.1292	0.0999	-1.3344	-0.5031

Note: "Overlay (30%)" denotes scaling risky portfolio weights by 0.7 when the specified trigger is active (the remaining 30% is held in cash at zero return). "Net returns" subtract proportional transaction costs of $c = 5$ bps per unit of one-way turnover.

Figure 6 reports the 21-day rolling mean of the residual-ranked long–short portfolio's one-way turnover, defined as $\frac{1}{2}\sum_i |w_{t,i} - w_{t-1,i}|$. The persistently high turnover implies frequent rebalancing, so transaction costs can materially reduce net performance.

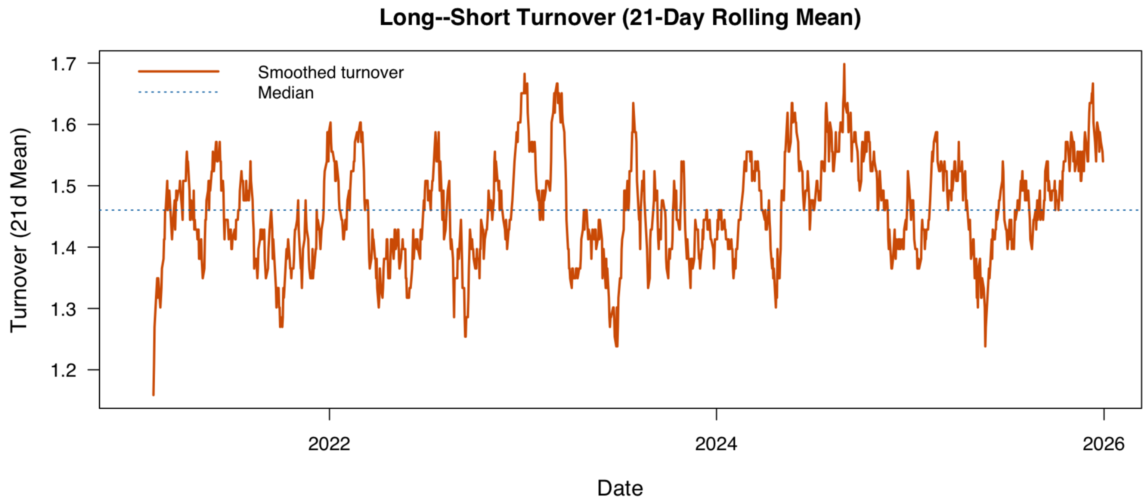

Figure 6. Residual-ranked long–short turnover (21-day rolling mean).

4.6. Descriptive PCA Structure in Sector Excess Returns

For completeness, we summarize the descriptive PCA structure of the sector excess-return panel. Figure 7 reports the PCA scree plot. In the main sample, the first two principal components explain approximately 64.1% of the cross-sectional variation, supporting the parsimonious baseline choice $K = 2$ used throughout the analysis.

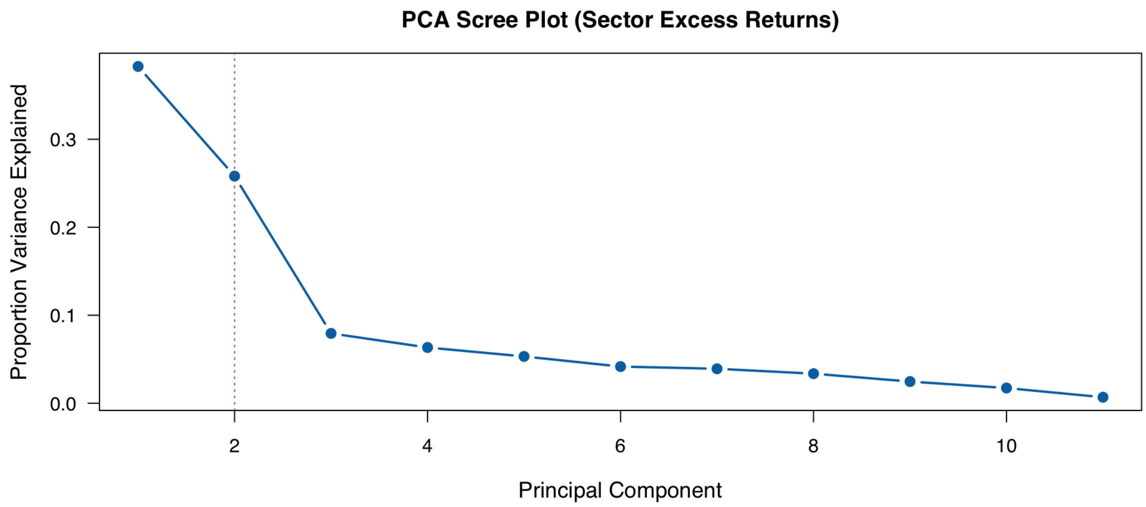

Figure 7. PCA scree plot of sector excess returns (sector minus SPY).

doi:10.20944/preprints202603.0395.v2

Table 4. Variance explained by PC1 and PC2 in the main-sample descriptive PCA on sector excess returns.

PC	Explained variance (%)	Cumulative (%)
PC1	38.27	38.27
PC2	25.81	64.08

To interpret the economic content of the leading components, Figure 8 displays the PC1–PC2 loading heatmap across the 11 sector ETFs, and Table 5 reports the largest contributors by absolute loading. The heatmap indicates that a low-dimensional factor structure captures much of the common variation in sector excess returns. Differences in sign reflect opposite directional co-movement with the corresponding latent component.

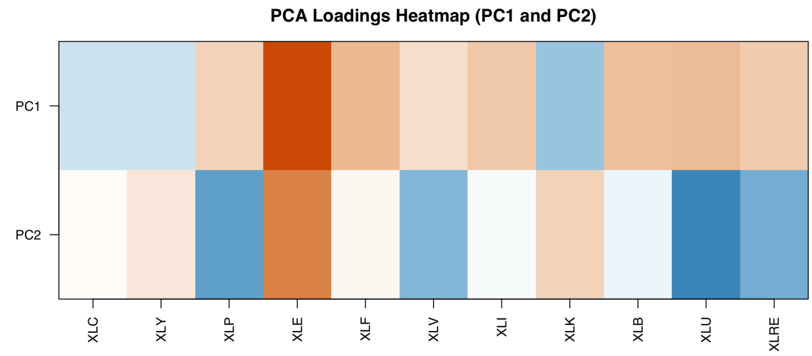

Figure 8. PC1–PC2 loading heatmap across the 11 sector ETFs.

Table 5. Largest absolute PCA loadings for PC1 and PC2 in the main-sample descriptive PCA.

PC	Ticker	Loading
PC1	XLE	0.7611
PC1	XLF	0.2766
PC1	XLU	0.2587
PC1	XLK	-0.2553
PC2	XLU	-0.5413
PC2	XLE	0.5033
PC2	XLP	-0.4233
PC2	XLRE	-0.3652

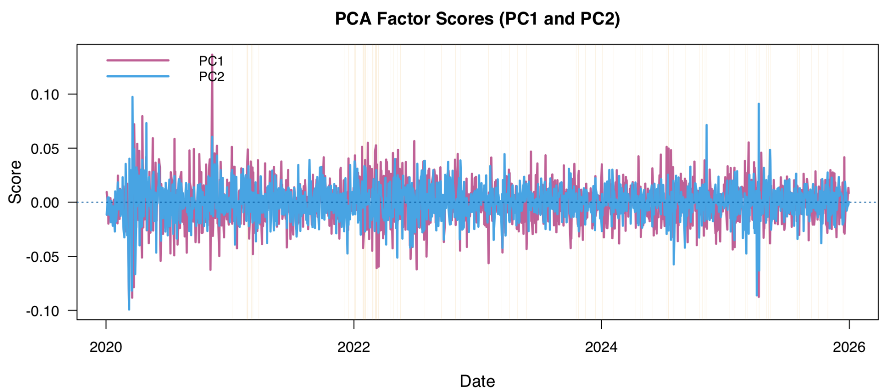

Figure 9. PC1–PC2 factor scores with stress periods shaded.

 doi:10.20944/preprints202603.0395.v2

5. Discussion

Realized volatility remains the stronger standalone benchmark in aggregate early-warning metrics, but the results indicate that residual stress contains economically meaningful complementary information in conditional and event-level settings. This study develops a leakage-safe PCA–APT residual-stress index from the cross section of sector excess-return residuals and evaluates its usefulness for monitoring SPY drawdown risk. The main empirical finding is that residual stress helps characterize cross-sectional market dislocation around drawdown-risk transitions, even though realized volatility remains stronger in unconditional average classification performance in both the main sample and the longer-history robustness exercise. Residual-stress spikes cluster around drawdown onsets (Figures 1 and 2), and the joint-regime evidence in Table 1 indicates that the signal can refine near-term drawdown-risk assessment even when realized volatility remains subdued.

This interpretation is consistent with the paper's core framing. Volatility captures aggregate time-series fluctuation, whereas residual stress captures fragmentation in the cross section after removal of the dominant common component. It is also useful to distinguish this paper from volatility-forecasting and general machine-learning early-warning studies. The aim here is not to improve volatility prediction or to compete primarily on classifier complexity, but to isolate a different information channel—cross-sectional residual fragmentation after removing the dominant sector common component—and to evaluate that channel under an implementable, leakage-safe design. The economic interpretation of the signal is therefore structural and cross-sectional rather than purely time-series or model-driven.

5.1. Residual Stress as a Drawdown-Risk Monitoring Signal

The first implication of the results is that cross-sectional residual dispersion appears meaningfully related to transitions into higher-risk market states. Figure 1 shows that large residual-stress spikes tend to cluster around drawdown-onset events, while Figure 2 indicates that average residual stress rises around onset windows. Taken together, these patterns support the interpretation that unusual dispersion in sector-level residual movements is associated with deterioration in market structure during risk-off transitions. Economically, this is plausible: when sectors reprice in a more fragmented or rotational manner, a low-dimensional common-factor structure explains less of the cross section, and the residual component becomes larger. The proposed stress index is designed precisely to capture this type of dislocation.

5.2. Complementary Information Relative to Volatility

The main value of residual stress is not stronger standalone discrimination than volatility, but finer risk stratification in conditional settings. Table 1, Panel A shows that realized volatility achieves stronger average discrimination than residual stress in the main sample. Panel B, however, indicates that residual stress can still refine drawdown-risk assessment within volatility states. In particular, when volatility is low, moving from the low-stress/low-volatility regime to the high-stress/low-volatility regime raises the probability of a drawdown onset within the next $H = 21$ trading days from approximately 7.8% to approximately 17.0%. The joint high-stress/high-volatility regime exhibits the highest onset probability, at approximately 36.0%. These patterns are consistent with the interpretation that residual stress captures a dimension of cross-sectional fragility not fully summarized by time-series volatility alone.

At the same time, the main-sample evidence should be interpreted with caution. High-stress regimes occur infrequently, and the bootstrap confidence interval for the Q10–Q00 difference includes zero, so the main-sample evidence is better viewed as suggestive rather than definitive. For this reason, the appendix robustness exercise based on a longer-history proxy sector universe is an important part of the empirical assessment. In that proxy-universe sample, the number of drawdown onsets increases from 14 to 55, the Q10–Q00 event-rate difference remains positive, and its bootstrap confidence interval excludes zero. Event-level diagnostics also become more supportive: residual stress uniquely flags

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.0395.v2

onset events not captured by volatility and arrives earlier than volatility in a majority of paired cases. Taken together, these results support a conservative interpretation: residual stress is best viewed as a complementary monitoring signal whose value is clearer in conditional analysis and longer-history robustness evidence than in standalone unconditional classification. The longer-history proxy-universe exercise is especially important because it materially increases the number of drawdown-onset events and yields more stable evidence of complementarity than the modern ETF sample alone.

5.3. Residual Stress as a Regime Indicator Rather than a Standalone Return Factor

The empirical results also suggest that residual stress should be interpreted primarily as a *regime-monitoring* variable rather than as a conventional return-prediction factor. The paper's main evidence comes from drawdown-onset events, regime-conditional probabilities, and classification-style warning metrics rather than from strong unconditional average return predictability. This interpretation is appropriate for the problem considered here. Drawdowns are infrequent and nonlinear events, so the value of a signal may lie less in forecasting average returns during normal periods and more in identifying states in which downside-risk management becomes especially relevant. In this sense, the residual-stress index is better viewed as a market-state indicator for monitoring and risk control than as a robust standalone source of alpha.

5.4. Implications for Risk-Management Overlays

The stress-managed SPY overlays illustrate how the signal can be mapped into practical risk-control decisions. In the reported sample, the volatility-based overlay delivers the strongest overall overlay performance, while the residual-stress overlay remains close to the buy-and-hold benchmark and does not improve downside protection to the same extent (Table 3; Figures 4 and 5). This result is informative rather than discouraging. It suggests that a simple binary residual-stress rule with a fixed 30% haircut is too coarse to produce a strong standalone overlay effect in this sample, even though the signal contains useful monitoring information.

More broadly, the overlay results highlight a practical trade-off between protection, turnover, and tracking error. Stronger drawdown mitigation would likely require either a more aggressive exposure reduction, a lower activation threshold, or a continuous scaling rule that maps stress intensity into portfolio exposure. Related extensions such as smoothing, hysteresis, or multi-day confirmation rules may also improve implementability by reducing unnecessary switching. The key takeaway is that residual stress appears more naturally suited to *risk monitoring and conditional exposure adjustment* than to a simple one-threshold mechanical de-risking rule.

5.5. Why the Residual-Ranked Long–Short Strategy Can Underperform

The residual-ranked long–short strategy performs poorly in the reported sample, with negative annualized return, weak risk-adjusted performance, and deep drawdown (Table 3). Figure 6 also shows persistently high turnover, which implies substantial implementation frictions after transaction costs. This underperformance is not inconsistent with the paper's main thesis. A signal that is informative about *aggregate drawdown risk* need not translate into profitable short-horizon cross-sectional alpha after costs. Cross-sectional residuals can be noisy, unstable, and costly to trade when portfolios are rebalanced frequently.

These results support a narrower interpretation of the long–short exercise. It is best viewed as a diagnostic of whether residual signals can be monetized directly, not as the central contribution of the paper. In practical terms, improving the tradability of the long–short strategy would likely require signal smoothing, less frequent rebalancing, stronger portfolio constraints, volatility scaling, or turnover penalties.

5.6. Economic Interpretation of the PCA Factor Space

The descriptive PCA results remain useful for interpretation even though the main contribution of the paper lies in the residual component rather than in the factor scores themselves. Table 4 and

Figure 7 show that the first two principal components explain approximately 64.1% of the cross-sectional variation in sector excess returns, supporting the parsimonious choice $K = 2$. Figure 8 and Table 5 indicate that a subset of sectors contributes disproportionately to the leading components, consistent with a low-dimensional common structure driven by broad macroeconomic or industry forces. This matters because the stress index is defined relative to that common structure: it measures the extent to which the cross section behaves abnormally *after* the dominant common component has been removed. Figure 9, which plots PC1 and PC2 with stress-period shading, provides an additional descriptive view of how the common factor space evolves across stress and non-stress periods, but the empirical signal of interest remains the residual dispersion orthogonal to that space.

5.7. Robustness Considerations and Limitations

The paper reports robustness exercises over threshold, drawdown, and horizon definitions while preserving the same leakage-safe timing protocol throughout. Across these alternatives, the broad qualitative conclusion remains stable: residual stress contains nontrivial warning content and remains associated with drawdown risk, while the volatility benchmark remains more competitive in average standalone classification performance. At the same time, the conditional regime results and longer-history appendix evidence support the interpretation that residual stress contributes complementary information for monitoring drawdown risk.

Several limitations should nevertheless be acknowledged. First, the main text results are based on a relatively short modern sample concentrated on one U.S. market proxy and one sector-ETF cross section, even though the appendix expands the event count using longer-history proxies. Second, train-only thresholds and rolling windows involve unavoidable design choices; different choices may alter the frequency and timing of stress labels. Third, because the number of drawdown-onset events is limited in the main sample, some conditional estimates in Table 1 are based on relatively small cell counts and should therefore be interpreted with caution. Fourth, while the implementation is explicitly designed to be leakage-safe, public market data can still be affected by revisions, symbol changes, or vendor-specific quirks. These considerations do not overturn the main interpretation of residual stress as a complementary monitoring signal, but they do suggest that broader samples, additional markets, and alternative data sources would be useful for establishing external validity more firmly.

5.8. Future Work

Several extensions follow naturally from the present results. First, residual stress could be combined with complementary market-stress variables such as implied volatility, credit spreads, liquidity indicators, or macro-financial uncertainty measures in order to improve joint risk-state detection. Second, the current binary threshold framework could be replaced with a probabilistic mapping from residual stress into drawdown-risk probabilities, for example through walk-forward logistic or hazard-style models. Third, broader validation across other equity universes, international markets, and multi-asset settings would help determine whether cross-sectional residual dispersion is a general indicator of market fragmentation or is more specific to U.S. sector structure. Finally, deeper economic investigation of the mechanisms behind elevated residual stress—such as sector rotation, funding pressure, or disagreement shocks—would strengthen the interpretation of the signal and potentially guide more effective overlay design.

6. Conclusions

This paper develops a leakage-safe PCA–APT residual-stress index for equity-market drawdown monitoring using a cross-section of U.S. sector ETFs. The proposed measure is constructed by extracting a low-dimensional common structure from sector excess returns relative to SPY and defining residual stress as the cross-sectional root-mean-square magnitude of the out-of-sample residual component. A train-only rolling quantile rule is then used to convert the continuous stress score into an implementable regime label, allowing drawdown-warning evaluation and illustrative portfolio applications without look-ahead bias.

 doi:10.20944/preprints202603.0395.v2

The paper's contribution is to develop a leakage-safe, interpretable, cross-sectional residual-stress diagnostic that improves conditional drawdown-risk stratification, especially in otherwise low-volatility states, rather than a standalone replacement for realized volatility. The paper does not claim standalone predictive dominance over realized volatility. In both the modern ETF baseline sample and the longer-history proxy-sector sample, realized volatility remains the stronger unconditional benchmark on standard classification metrics. The contribution instead lies in showing that residual stress adds complementary conditional and event-level information for drawdown-risk monitoring under an implementable, leakage-safe design.

Empirically, residual-stress spikes cluster around drawdown-onset episodes, and the joint-regime evidence indicates that elevated residual stress is associated with higher near-term drawdown risk even when volatility remains relatively low. The broader-sample results are especially important because they materially increase the number of drawdown-onset events and provide more stable support for the paper's central interpretation. In the longer-history proxy-sector sample, the conditional low-volatility contrast remains positive with a strictly positive bootstrap interval, and the event-level diagnostics are more supportive than in the modern ETF sample alone. Taken together, these findings suggest that the main conclusion is not purely episodic, even though the modern sample remains limited.

The portfolio exercises are best interpreted as implementation diagnostics rather than as a separate contribution. The stress-managed SPY overlay illustrates how the signal can be translated into a simple risk-control rule, whereas the weak residual-ranked long–short results indicate that information useful for aggregate drawdown-risk monitoring does not necessarily produce robust short-horizon cross-sectional alpha after costs. This further supports interpreting residual stress primarily as a regime-monitoring and risk-management variable rather than as a standalone trading factor.

Several extensions follow naturally. Future work should evaluate the signal across broader equity universes, international markets, and multi-asset settings; combine residual stress with complementary indicators such as implied volatility, credit spreads, liquidity measures, and macro-financial uncertainty variables; and develop probabilistic drawdown-risk models that map residual stress into calibrated warning probabilities. Additional work on continuous exposure-scaling rules, turnover-aware overlay design, and external validation would further clarify the practical role of residual-stress indicators in systematic financial risk management.

Author Contributions: Conceptualization, T.L.; methodology, T.L.; software, T.L.; validation, T.L.; formal analysis, T.L.; investigation, T.L.; data curation, T.L.; writing—original draft, T.L.; writing—review & editing, T.L.; visualization, T.L. The author has read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data were obtained from Yahoo Finance. The datasets analyzed during the current study are publicly available from these sources.

Acknowledgments: The author used AI-assisted tools for language editing and formatting. The author reviewed and takes responsibility for the content.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. Exploratory Combined-Model Exercise

As an exploratory check, we also considered a walk-forward model combining residual stress and volatility signals. In this sample, the combined specification did not improve on the standalone volatility benchmark and was sensitive to specification choices. We therefore do not emphasize this exercise in the paper.

 doi:10.20944/preprints202603.0395.v2

References

1. Li, H.; Liu, T. Portfolio optimization based on the LSTM forecasting model. *Proceedings of the 2nd International Conference on Financial Technology and Business Analysis* **2023**, *48*(1), 97–106.
2. Pandey, P.N. ETF Risk–Return and Price Discovery in Crisis Regimes: Evidence from Indian Markets. *SSRN* **2025**, 5899923.
3. Ross, S.A. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* **1976**, *13*(3), 341–360.
4. Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer: New York, NY, USA, 2002.
5. Campbell, J.Y.; Lettau, M.; Malkiel, B.G.; Xu, Y. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The Journal of Finance* **2001**, *56*(1), 1–43.
6. Christie, W.G.; Huang, R.D. Following the Pied Piper: Do individual returns herd around the market? *Financial Analysts Journal* **1995**, *51*(4), 31–37.
7. Chang, E.C.; Cheng, J.W.; Khorana, A. An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance* **2000**, *24*(10), 1651–1679.
8. Moreira, A.; Muir, T. Volatility-managed portfolios. *The Journal of Finance* **2017**, *72*(4), 1611–1644.
9. Barroso, P.; Santa-Clara, P. Momentum has its moments. *Journal of Financial Economics* **2015**, *116*(1), 111–120.
10. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*(1), 29–36.
11. Davis, J.; Goadrich, M. The relationship between Precision–Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*; 2006; pp. 233–240.
12. Ciciretti, V.; Nandy, M.; Pallotta, A.; Lodh, S.; Senyo, P.K.; Kartasova, J. An early-warning risk signals framework to capture systematic risk in financial markets. *Quantitative Finance* **2025**, *25*, 757–771.
13. Febrian, E.; Herwany, A. CAPM and APT validation test before, during, and after financial crisis in emerging market: Evidence from Indonesia. *SSRN* **2010**.
14. Koutoulas, G.; Kryzanowski, L. Integration or segmentation of the Canadian stock market: Evidence based on the APT. *Canadian Journal of Economics* **1994**, 329–351.
15. Baghdadabad, M.R.T.; Glabadanidis, P. An extensile method on the arbitrage pricing theory based on downside risk (D-APT). *International Journal of Managerial Finance* **2014**, *10*, 54–72.
16. Yang, L.; Rea, W.; Rea, A. Financial insights from the last few components of a stock market PCA. *International Journal of Financial Studies* **2017**, *5*, 15.
17. Neves-Silva, R.; et al. Complex principal component analysis of dynamic correlations in financial markets. In *Intelligent Decision Technologies: Proceedings of the 5th KES International Conference on Intelligent Decision Technologies (KES-IDT 2013)*; Volume 255; p. 111; 2013.
18. Liu, T. A Comparative Study of Transformer-Based and Classical Models for Financial Time-Series Forecasting. *Journal of Risk and Financial Management* **2026**, *19*(3), 203. doi:10.3390/jrfm19030203.
19. Quiroga-Juárez, C.A.; Villalobos-Escobedo, A. Analysis of stock market behavior of the major financial exchanges worldwide using multivariate analysis (principal component analysis PCA) for the period 2011 to 2014. Working Paper, Instituto Tecnológico Metropolitano (ITM), **2025**.
20. Geboers, H.; Depaire, B.; Annaert, J. A review on drawdown risk measures and their implications for risk management. *Journal of Economic Surveys* **2023**, *37*, 865–889.
21. Acharya, V.; Steffen, S. Stress tests for banks as liquidity insurers in a time of COVID. *VoxEU.org* **2020** (March).
22. Liu, T. Volatility Forecasting and Early-Warning Market Stress Detection: A Leakage-Safe Evaluation with Tree Ensembles and Transformers. *Preprint* **2026**, 04 March; available at Research Square. doi:10.21203/rs.3.rs-9015347/v1.