# *Skepthical* review: *The Conditional Predictive Power of Sectoral Volatility Dispersion for VIX Innovations*

## Summary

The paper asks whether cross-sectional dispersion in sector realized volatility (Sectoral Volatility Dispersion, SVD) predicts future VIX changes and/or transitions into high-volatility regimes. Using daily data (2015–2026) for US sector ETFs, the S&P 500, and the VIX, it constructs 21-day realized volatilities by sector and defines SVD as the cross-sectional standard deviation, then normalizes it via a 252-day rolling z-score (Sec. 2.1). Market regimes are defined using a 3-state Gaussian Hidden Markov Model fit to daily VIX log-returns (Sec. 2.2, Sec. 3.1). Predictability is assessed with OLS regressions for $k$-day ahead VIX log-returns ($k = 5, 21$) and logistic regressions for transitions into the "High" HMM state (Sec. 2.3, Sec. 3.2–3.3). Empirically, SVD alone is not a robust predictor of future VIX innovations nor of regime transitions; classification performance is near chance (AUROC $\approx$ **0.53** with very low PR-AUC) (Sec. 3.3). The main positive result is conditional: at the 21-day horizon, the interaction **SVD** $\times$ **AvgCorr** is significant and negative, consistent with SVD being associated with higher future VIX primarily when average cross-sector correlation is low (Sec. 3.2). The contribution is best framed as an exploratory finding about the joint role of dispersion and comovement, but the manuscript needs stronger out-of-sample validation, clearer theoretical/related-work positioning, fuller methodological specification (data processing, correlation measure, HMM and logit design), and a more rigorous exposition of the interaction's marginal effects and economic magnitude to support the "fragility" interpretation (Sec. 1–4).

## Strengths

- Clear and focused research question on whether sector volatility dispersion contains predictive information for VIX dynamics and volatility regimes (Sec. 1).
- Feature construction (21-day realized volatility, SVD as cross-sectional dispersion, rolling z-score normalization) is intuitive and based on widely available data (Sec. 2.1; Eq. (1)–(2)).
- Interaction-based specification explicitly tests conditional predictability (dispersion matters depending on comovement), which is economically plausible and more nuanced than "SVD predicts VIX" (Sec. 2.3; Eq. (3); Sec. 3.2).
- Use of PR-AUC alongside AUROC acknowledges class imbalance in the regime-transition prediction problem (Sec. 3.3).
- The paper reports negative results candidly (SVD not a standalone timing signal) and includes a leave-one-sector sensitivity check (Sec. 3.4).
- Writing is generally clear and the empirical workflow (construction $\rightarrow$ regimes $\rightarrow$ regressions/classification $\rightarrow$ robustness) is easy to follow (Sec. 1–4).

## Major issues

1. **Predictive claims are evaluated essentially in-sample; there is no explicit, time-respecting separation between estimation and evaluation for the OLS forecasting regressions or the logistic transition models (Sec. 2.3; Sec. 3.2–3.3).** This makes it difficult to distinguish a stable predictive relationship from an in-sample pattern, especially given the limited sample and major episodes (e.g., COVID).

   *Recommendation:* Add an explicit out-of-sample forecasting design. For Sec. 3.2, implement rolling or expanding-window estimation that produces genuine $t \rightarrow t + k$ forecasts of $\Delta \ln(\text{VIX})_{t+k}$ using only information available at $t$, and report out-of-sample RMSE/MAE and out-of-sample $R^2$ (e.g., Campbell–Thompson) relative to benchmarks. For Sec. 3.3, compute AUROC/PR-AUC strictly out-of-sample using chronological splits (or rolling evaluation), and report uncertainty (e.g., block bootstrap over time). State the exact training window, update frequency, and evaluation period(s).

2. **The manuscript lacks benchmark comparisons, so it is unclear whether SVD and especially SVD$\times$AvgCorr add incremental information beyond standard VIX dynamics (mean reversion, persistence) and common volatility predictors (Sec. 1; Sec. 2.3; Sec. 3.2; Sec. 4).**

   *Recommendation:* In Sec. 3.2, add benchmark models: (i) AR-type models for $\Delta \ln(\text{VIX})$ or changes in VIX; (ii) HAR-style models if using realized measures; (iii) controls-only specifications (e.g., lagged $\Delta \ln(\text{VIX})$, VIX level, realized SPX volatility). Compare in-sample and out-of-sample performance and test whether adding SVD and $\text{SVD} \times \text{AvgCorr}$ improves forecasts (nested model tests or $\Delta$OOS-$R^2$). Summarize incremental value clearly in Sec. 4.

3. **Data/replicability details are under-specified and there is an internal inconsistency in the sector universe: the Abstract says ten sector ETFs, but Sec. 2.1 lists nine and Eq. (2) uses N=9.** Additionally, the "balanced panel by removing missing days" approach can induce sample selection around missingness and inception/holiday alignment, affecting realized vol and correlations (Sec. 2.1).

   *Recommendation:* Make the sector universe consistent everywhere (Abstract, Sec. 2.1, Eq. (2), figures/tables): either correct to $N = 9$ or add the missing sector and update all computations. In Sec. 2.1, provide: data sources; whether prices are adjusted for splits/dividends; exact start/end dates after filtering; number of observations before/after balancing; what drives missingness; and robustness to alternative alignment rules (e.g., restricting to common inception date; forward-fill vs listwise deletion).

4. **Key model components are not defined with enough precision to be reproducible, especially AvgCorr, the HMM estimation choices, and the logistic-regression specification (Sec. 2.1–2.3; Sec. 3.1; Sec. 3.3).**

*Recommendation:* Add explicit definitions and implementation details: (i) define $\mathbf{AvgCorr}_t$ formally (e.g., mean over $i < j$ of rolling-window Pearson correlations of daily sector log-returns; specify the window length and any de-meaning); (ii) clarify whether $\mathbf{SVD}_t$ denotes raw SVD or the z-scored SVD, and introduce distinct notation if both appear; (iii) describe HMM estimation (algorithm, initialization, number of random restarts, convergence criteria, software/library) and whether regimes are based on filtered vs smoothed probabilities; (iv) write the exact logistic model equation (event definition, conditioning sample such as "not High at $t-1$", predictors, standardization, class weighting/resampling if any).

5. **The interpretation of the core interaction effect is underdeveloped: the text qualitatively claims "SVD predicts higher future VIX only when AvgCorr is low," but does not show marginal effects, confidence intervals, or economic magnitude.** This is crucial because the main SVD coefficient is not significant while $\mathbf{SVD} \times \mathbf{AvgCorr}$ is negative and significant (Sec. 3.2; Sec. 4).

   *Recommendation:* In Sec. 3.2, explicitly derive and report the marginal effect: $\partial\mathbb{E}[\Delta\ln(\mathrm{VIX})_{t+k}/\partial \mathrm{SVD}_t = \beta_1 + \beta_2 \cdot \mathrm{AvgCorr}_t$ and the positivity condition when $\beta_2 < 0$). Report distributions (mean/std/quantiles) of SVD and AvgCorr and compute marginal effects at representative correlation quantiles (10

   $\Delta\ln(\mathrm{VIX})$ (or marginal effect) over a grid of (SVD, AvgCorr), and translate effects into economically interpretable changes (percent VIX or VIX points at typical levels).

6. **SVD and AvgCorr may be mechanically linked through common-factor structure (changes in market factor variance can affect both cross-sectional volatility dispersion and pairwise correlations).** Without addressing this, the interaction may reflect broader market volatility dynamics rather than "decoupling/fragility" per se (Sec. 2.1–2.3; Sec. 3.2).

   *Recommendation:* Quantify and discuss the relation between SVD and AvgCorr: report their correlation and partial correlations. Re-estimate Sec. 3.2 adding market volatility controls (e.g., realized SPX volatility; VIX level; market variance proxy) to test whether $\mathbf{SVD} \times \mathbf{AvgCorr}$ remains significant. Consider alternative "decoupling" measures that more directly isolate comovement structure (e.g., first principal component explained variance; average $R^2$ from sector-on-market regressions; AvgCorr of residual returns after removing the market factor).

7. **The HMM regime model is insufficiently characterized and may not yield substantively distinct "Low/Moderate/High" regimes.** The reported average VIX levels by state are relatively close, and key regime diagnostics (transition matrix, durations, state variances) are not provided; the choice of three states is not justified (Sec. 2.2; Sec. 3.1). This matters because the logistic task depends entirely on the regime labels.

   *Recommendation:* Expand Sec. 3.1 with a regime diagnostics table: state-specific mean/variance of $\mathbf{\Delta \ln(VIX)}$ (the modeled series), implied distributions of VIX levels by state (not only means), the transition probability matrix, and expected spell lengths. Justify the 3-state choice (AIC/BIC or interpretability) and briefly check robustness to 2- and 4-state models. Validate that the "High" state captures known stress episodes (e.g., 2020) via time-series plots of smoothed state probabilities overlaid with VIX.

8. **The logistic-regression exercise is not well-aligned with the paper's main conditional finding: if predictability is conditional on AvgCorr, a logit model using only lagged SVD is unlikely to perform and does not test the core hypothesis.** Additionally, severe class imbalance calls for more diagnostics than AUROC alone (Sec. 3.3).

   *Recommendation:* Revise Sec. 3.3 to include AvgCorr and $\mathbf{SVD} \times \mathbf{AvgCorr}$ (and possibly key controls like lagged VIX level/regime duration) in the transition model, and test incremental value of SVD terms versus a baseline. Report event rate, confusion matrices at selected thresholds, Brier score, and calibration (reliability) plots; keep PR-AUC and add a no-skill PR baseline tied to prevalence. Consider forecasting the HMM's next-period high-state probability (or using ordered/multinomial models) rather than a hard transition indicator.

9. **Time-series dependence and horizon construction are not fully addressed.** For $k = 5$ and $k = 21$, $\Delta\ln(\mathrm{VIX})_{t+k}$ likely uses overlapping horizons, inducing serial correlation; Newey–West is mentioned but lag length and sensitivity are not documented. Al

   $\Delta\ln(\mathrm{VIX})_t$ and r$) should be clarified (Sec. 2.3; Sec. 3.2).},t

   *Recommendation:* In Sec. 2.3, define $\Delta\ln(\mathrm{VIX})_{t+k}$ explicitly (e.g., $\ln(\mathrm{VIX}_t)$) and state whether horizons overlap. Report the Newey–West lag choice (and rationale, e.g.,k-1) and sensitivity to alternative lags. Clarify timing assumptions: if regressors are observed at close of day t and the dependent variable spans t\rightarrow t+k, state this explicitly; consider robustness using lagged controls (t-1$) to avoid ambiguity.} / \mathrm{VIX

10. **Structural stability and subperiod robustness are not assessed despite the sample covering major regime changes (COVID, post-2020 volatility/correlation dynamics).** Coefficient stability is important for the interaction result and for any practical "monitoring" interpretation (Sec. 2.1; Sec. 3.2–3.3; Sec. 4).

    *Recommendation:* Add stability checks: estimate Sec. 3.2 and Sec. 3.3 models in subsamples (e.g., 2015–2019 vs 2020–2026; pre-/during-/post-COVID) and/or rolling-window regressions with coefficient paths for SVD, AvgCorr, and $\mathbf{SVD} \times \mathbf{AvgCorr}$. Include a brief Limitations subsection in Sec. 4 discussing sample dependence and how window choices (21-day vol; 252-day z-score) affect results.

11. **The paper is under-situated in prior literature on VIX forecasting, dispersion, correlation risk/breakdowns, and systemic fragility measures, making it hard to assess novelty and interpret the conditional finding (Sec. 1; Sec. 4).**

*Recommendation:* Add a Related Work section (e.g., Sec. 1.1) covering: (i) VIX forecasting (AR/HAR, implied vs realized predictors); (ii) dispersion measures (cross-sectional volatility/variance dispersion, sector dispersion); (iii) correlation risk/connectedness/systemic risk indicators; (iv) regime modeling with HMMs. Then sharpen Sec. 4's contribution statement to emphasize what is new (the conditional dispersion–correlation interaction in this sector-ETF setup) versus what is consistent with existing results.

## Minor issues

1. Terminology/notation is sometimes informal or inconsistent: "VIX innovations" is used while the dependent variable is $\mathbf{\Delta \ln(VIX)}_{t+k}$; AvgCorr is discussed without a symbol/formula; "market fragility" remains qualitative rather than operational (Sec. 1; Sec. 2.1–2.3).

   *Recommendation:* Tighten definitions in Sec. 2.1–2.3: define "VIX innovations" as $\mathbf{\Delta \ln(VIX)}_{t+k}$; introduce notation for $\mathbf{AvgCorr}_t$ and define what "low" vs "high" correlation means (e.g., quantiles); provide an operational definition of "fragility" as a region in (SVD, AvgCorr) space (e.g., high SVD and low AvgCorr).

2. Descriptive statistics for the key predictors are limited, which makes it harder to gauge how extreme the "high SVD / low correlation" episodes are and how often they occur (Sec. 2.1; Sec. 3.2).

   *Recommendation:* Add a small table with mean/std/quantiles of SVD (raw and z-scored), AvgCorr, and their correlation; optionally include counts of days in key regions (e.g., SVD z-score $> 1$ and AvgCorr below 20th percentile).

3. The logistic-regression results are summarized primarily via AUROC/PR-AUC, but additional diagnostics would help interpret failure modes (e.g., whether the model is poorly calibrated or simply has no separation) (Sec. 3.3).

   *Recommendation:* Add Brier score and a calibration plot; report precision/recall at one or two operating points (e.g., fixed recall), and include confidence intervals for AUROC/PR-AUC (block bootstrap).

4. The discussion of why predictability appears at 21 days but not 5 days is brief, and the roles of the control variables ($\Delta\ln(\mathrm{VIX})_{t,\ r}$) are not interpreted in much depth (Sec. 3.2).},t

   *Recommendation:* Add a short interpretation paragraph in Sec. 3.2 linking coefficient signs to standard VIX dynamics (mean reversion, leverage effects) and proposing a mechanism for horizon dependence (e.g., dispersion/correlation information being incorporated into implied vol more slowly).

5. Figures have several clarity and scientific-communication issues (caption mismatches, accessibility, and potentially overstated visual narratives), and some figure evaluation details (in-sample vs out-of-sample) are not stated (Fig. 1–3; Sec. 3.1–3.4).

   *Recommendation:* Update figures after adding out-of-sample evaluation: ensure captions match panels (e.g., ROC+PR both present), state evaluation protocol in captions, add uncertainty bands where feasible, and adopt colorblind-safe palettes with less-obscuring regime shading (e.g., top-strip shading). Ensure Figure 3 is labeled as "leave-one-sector sensitivity" (not LOOCV) unless time-series CV is actually implemented.

6. ADF stationarity test for normalized SVD is mentioned without specifying the regression form (constant/trend) and lag selection; this reduces interpretability (Sec. 3 / wherever ADF is reported).

   *Recommendation:* Report the ADF specification (constant and/or trend), lag selection rule, test statistic, and p-value; if space is limited, move details to an appendix.

## Very minor issues

1. Formatting/polish issues: inconsistent heading styles (e.g., stray "# 3.2"), unresolved LaTeX placeholders (e.g., "Figure ??", "Table ??"), and inconsistent capitalization/quotation for regime labels (Sec. 2.2; Sec. 3.1–3.4).

   *Recommendation:* Standardize headings, resolve all cross-references with \label/\ref, and use consistent regime label formatting ("Low/Moderate/High") throughout.

2. Notation could be clearer for raw vs normalized SVD: Eq. (2) defines raw SVD, but later "SVD_t" appears to mean the z-scored version without an explicit switch (Sec. 2.1–2.3).

   *Recommendation:* Introduce separate symbols (e.g., $\mathrm{SVD}{{\rm raw}, t}**and**\mathrm{SVD}_t henceforth denotes the normalized z-score.}, t}) or explicitly state at the start of Sec. 2.3 that $\mathrm{SVD}

3. Figure labeling and annotation could be improved (panel labels, axis units/definitions, legend placement) to make plots self-contained (Fig. 1–3).

   *Recommendation:* Add panel labels (a/b/c), define axes (e.g., "$\mathbf{\Delta \ln(VIX)}_{t+21}$"), ensure legends do not overlap data, and reduce dense date ticks (e.g., annual ticks).

4. Minor language issues (overlong sentences, occasional dramatic phrasing) slightly reduce clarity (Sec. 1; Sec. 4).

   *Recommendation:* Edit for concision and precision; remove repetitive phrasing between Introduction and Conclusions and replace vague claims with references to specific coefficients/results (e.g., the 21-day interaction term in Sec. 3.2).

## Key statements and references

- • **The only statistically significant predictor in the 5-day ahead VIX OLS regression is the concurrent market return $r_{\mathrm{MKT},t}$, whose positive coefficient reflects the well-documented contemporaneous leverage effect documented in prior literature.**
- *Reference(s):* (none)
- • **At the 21-day forecast horizon, the interaction term between Sectoral Volatility Dispersion (SVD) and average sector correlation in the OLS regression is negative and statistically significant ($\beta = -0.1267$, $p = 0.013$), implying that increases in SVD are associated with higher future VIX innovations only when average cross-sector correlation is low, whereas this effect is dampened or reversed when correlation is high.**
- *Reference(s):* (none)
- • **The logistic regression model using lagged normalized SVD to predict transitions into the High VIX regime exhibits poor discriminatory performance, with an AUROC of 0.526 and a Precision-Recall AUC of 0.0195, indicating that SVD in isolation has very low precision for timing rare regime shifts in line with prior findings on the difficulty of forecasting such transitions.**
- *Reference(s):* (none)

## Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

**Maths relevance:** light

The paper contains a small number of central mathematical definitions (realized volatility, cross-sectional dispersion, and a regression with an interaction term). There are no multi-step derivations, but internal consistency of definitions, notation, and implied marginal effects can be audited. The main internal inconsistency is the stated number of sector ETFs (ten vs nine), which affects the definition of SVD and AvgCorr.

**Checked items**

1. ✔ **Annualized 21-day realized volatility definition** (Eq. (1), Sec. 2.1, p.3)
    - **Claim:** Defines sector $i$ annualized 21-day realized volatility as $\sqrt{252}$ times the standard deviation of the prior 21 daily log returns.
    - **Checks:** dimensional/units, notation/definition consistency
    - **Verdict:** PASS; confidence: high; impact: moderate
    - **Assumptions/inputs:** $r_{i,t}$ are daily log returns, $\mathbf{StDev}(r_{i,t-20:t})$ is the sample standard deviation over a 21-day window, $\mathbf{252}$ is the annualization factor (trading days/year)
    - **Notes:** The formula is dimensionally coherent: returns are dimensionless, standard deviation is dimensionless, annualization via $\sqrt{252}$ yields annualized volatility. No internal algebra issues.
2. ✔ **Cross-sectional dispersion (SVD) formula** (Eq. (2), Sec. 2.1, p.3)
    - **Claim:** Defines $\mathbf{SVD}_t$ as the cross-sectional standard deviation of the contemporaneous sector realized volatilities.
    - **Checks:** algebraic form, definition consistency
    - **Verdict:** PASS; confidence: high; impact: moderate
    - **Assumptions/inputs:** $\sigma_{i,t}^{(21)}$ are defined by Eq. (1), $\overline{\sigma}_t^{(21)}$ is the cross-sectional mean at time $t$, $N$ is the number of sectors included
    - **Notes:** Eq. (2) matches the standard sample cross-sectional standard deviation with denominator $N-1$ and outer square root.
3. ✖ **Sector count consistency (N and listed ETFs)** (Abstract p.1; Sec. 2.1 p.2–3; Eq. (2) p.3)
    - **Claim:** The paper uses a consistent sector universe for SVD construction.
    - **Checks:** symbol/definition consistency
    - **Verdict:** FAIL; confidence: high; impact: critical
    - **Assumptions/inputs:** Abstract statement 'ten US sector ETFs', Sec. 2.1 lists nine ETFs, Eq. (2) uses $N = 9$
    - **Notes:** Internal contradiction: the abstract and narrative claim ten sector ETFs, but the methods enumerate nine and set $N = 9$. This affects the defined SVD and AvgCorr universes and therefore the interpretation of all models using SVD/AvgCorr.
4. ⚠ **Rolling Z-score normalization of SVD** (Sec. 2.1, p.3)
    - **Claim:** Raw SVD is transformed into a rolling 252-day Z-score using rolling mean and standard deviation.
    - **Checks:** definition completeness, notation consistency
    - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
    - **Assumptions/inputs:** A rolling window is used for both mean and standard deviation, The normalized series is then used in regressions

- **Notes:** No explicit equation is provided for the Z-score (e.g., $(\mathbf{SVD}_{\text{raw},t} - \mu_t)/s_t$). This is likely standard but cannot be verified from the text alone; also the same symbol $\mathbf{SVD}_t$ is later used for the normalized value without a formal redefinition.

5. ✔ **Average cross-sector correlation definition** (Sec. 2.1, p.3)
   - **Claim:** $\mathbf{AvgCorr}_t$ is the mean of all unique pairwise correlations across sector ETF returns over a rolling 21-day window.
   - **Checks:** definition clarity, indexing/normalization sanity
   - **Verdict:** PASS; confidence: medium; impact: minor
   - **Assumptions/inputs:** Correlations are computed pairwise over the same rolling window, Averaging is over unique $i < j$ pairs
   - **Notes:** Conceptually consistent, but would benefit from an explicit formula to remove ambiguity about normalization (e.g., division by $N(N-1)/2$) and return window indexing.
6. ✔ **HMM input transformation for VIX** (Sec. 2.2, p.3)
   - **Claim:** Fits the HMM to daily VIX log returns $\Delta\ln(\mathrm{VIX}_t) = \ln(\mathrm{VIX}_t / \mathrm{VIX}_{t-1})$.
   - **Checks:** algebra/notation
   - **Verdict:** PASS; confidence: high; impact: minor
   - **Assumptions/inputs:** $\mathbf{VIX}_t > 0$ so logs are defined
   - **Notes:** The transformation is correctly stated and self-consistent.
7. ⚠ **OLS regression specification with interaction** (Eq. (3), Sec. 2.3, p.3–4)
   - **Claim:** Models $k$-day-ahead VIX log return as a linear function of $\mathrm{SVD}_t, \mathrm{SVD}_t \times \mathrm{AvgCorr}_t, \mathrm{contemporaneous}\ \Delta\ln(\mathrm{VIX})_t, \mathrm{and contemporaneous\ market\ return}_t$.},t
   - **Checks:** symbol/definition consistency, dimensional/units, time-index consistency
   - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
   - **Assumptions/inputs:** $\mathrm{SVD}_t$ refers to the normalized $(Z-\mathrm{scored})\ SVD$, $\mathrm{AvgCorr}_t$ is contemporaneous, $\Delta\ln(\mathrm{VIX})$ is the next$-$k$-days VIX log return
   - **Notes:** The model is algebraically well-formed, but $\Delta\ln(\mathrm{VIX})_{t+k}$ is not explicitly defined for $k>1$ (e.g., $\ln(\mathrm{VIX}_t)$). Without an explicit definition, the exact meaning of the horizon and overlapping-return structure cannot be verified from the text alone.}/\mathrm{VIX
8. ⚠ **Interaction-term interpretation via marginal effects** (Sec. 3.2, p.5–7 (discussion of $\beta_2 < 0$))
   - **Claim:** A negative coefficient on $\mathbf{SVD}_t \times \mathbf{AvgCorr}_t$ implies SVD predicts higher future VIX innovations only when AvgCorr is low.
   - **Checks:** algebraic implication check, logic from model to narrative
   - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
   - **Assumptions/inputs:** Linear model: $E[y|\cdot] = \alpha + \beta_1 \mathrm{SVD} + \beta_2 \mathrm{SVD}{\cdot}\mathrm{AvgCorr} + \ldots$, Marginal effect of SVD is $\beta_1 + \beta_2{\cdot}\mathrm{AvgCorr}$
   - **Notes:** From the model, $\beta_2 < 0$ implies the marginal effect of SVD decreases as AvgCorr increases ($\partial/\partial\mathbf{AvgCorr}$ of the marginal effect is $\beta_2$). However, the claim that the effect becomes positive 'only when AvgCorr is low' additionally requires that $\beta_1 + \beta_2{\cdot}\mathbf{AvgCorr} > 0$ for low AvgCorr and $< 0$ otherwise. The paper does not state this threshold condition explicitly.
9. ⚠ **Logistic transition model specification completeness** (Sec. 2.3 and Sec. 3.3, p.4 and p.7)
   - **Claim:** Predicts probability of transitioning into High regime on day $t$ (conditional on not High at $t-1$) using lagged SVD.
   - **Checks:** model specification completeness, symbol/definition consistency
   - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
   - **Assumptions/inputs:** Binary response variable defined as an HMM regime transition indicator, Logistic link is used
   - **Notes:** No explicit equation is given for the logit model, the exact binary event definition is not formalized, and the conditioning ('not in High at $t-1$') is described but not written mathematically.

**Limitations**

- Audit is based only on the provided PDF text; no supplementary appendices or full equation derivations were available to inspect.
- No numeric validation was performed (per scope), so any checks that would require verifying ranges/distributions of variables (e.g., whether AvgCorr is 'low' enough for a positive marginal effect) are limited to symbolic conditions.
- Figures and tables were not used for recalculation; only their stated model relationships were checked for algebraic/notation consistency.

## Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

13 internal numerical checks were performed: 12 PASS and 1 FAIL. The only failure is a cross-section count inconsistency for the number of sector ETFs (10 vs 9), which affects the definition of the core dataset/universe. Other checks (pair-count from $N = 9$, repeated constants, horizon labeling, p-value threshold logic, basic range/sanity checks, and simple performance-metric differences) are internally consistent.

**Checked items**

1. ✖ **C1_sector_count_internal_consistency** (Page 1 Abstract; Page 2 Data; Page 3 Eq. (2))
   - **Claim:** Paper inconsistently describes sector ETF count: Abstract says "ten US sector ETFs" but Methods list nine ETFs and Eq. (2) sets $N = 9$.
   - **Checks:** cross_section_count_consistency
   - **Verdict:** FAIL
   - **Notes:** Inconsistent counts (abstract vs equation, abstract vs methods).
2. ✔ **C2_pairwise_correlation_count_from_N** (Page 3, Section 2.1 (average cross-sector correlation) + Eq. (2) $N = 9$)
   - **Claim:** Average cross-sector correlation is computed as the mean of all unique pairwise correlations across the nine sector ETFs; with $N = 9$, unique pairs should be $N(N-1)/2 = 36$.
   - **Checks:** combinatorial_recalculation
   - **Verdict:** PASS
   - **Notes:** Computed $N(N-1)/2$.
3. ✔ **C3_annualization_factor_consistency** (Page 3 Eq. (1) and nearby text)
   - **Claim:** 21-day realized volatility is annualized using $\sqrt{252}$ in Eq. (1); check that $\sqrt{252}$ is applied consistently with stated 21-day window and that 252-day window is used for Z-score normalization.
   - **Checks:** constant_reuse_consistency
   - **Verdict:** PASS
   - **Notes:** Checked repeated constants (252) and realized vol window (21).
4. ✔ **C4_horizons_k_match_reported_models** (Page 4 Section 2.3; Page 5 Table 1)
   - **Claim:** OLS model is estimated for horizons $k = 5$ and $k = 21$ days; Table 1 should correspond exactly to 5-Day Ahead and 21-Day Ahead results.
   - **Checks:** label_to_parameter_consistency
   - **Verdict:** PASS
   - **Notes:** Compared set equality of horizons.
5. ✔ **C5_table1_significance_bold_rule_vs_pvalues** (Page 5 Table 1 note + p-values in rows)
   - **Claim:** Table note says bold indicates statistical significance at 5% level; verify which p-values are $< 0.05$ and therefore should be bold (even if formatting not visible in parsed text).
   - **Checks:** pvalue_threshold_classification
   - **Verdict:** PASS
   - **Notes:** Classified significance by strict $p < 0.05$.
6. ✔ **C6_table1_only_significant_predictor_claim_check** (Page 5 Section 3.2 narrative + Table 1 (5-day column))
   - **Claim:** Narrative says: "The only significant predictor is the concurrent market return" for 5-day horizon; verify from Table 1 p-values that only $r_{\mathrm{MKT},t}$ has $p < 0.05$ in 5-day model.
   - **Checks:** narrative_vs_table_pvalues
   - **Verdict:** PASS
   - **Notes:** Verified narrative 'only significant predictor' for 5-day model.
7. ✔ **C7_table1_r2_values_plausible_range** (Page 5 Table 1)
   - **Claim:** $R^2$ values reported as 0.016 and 0.047; verify they lie in $[0, 1]$.
   - **Checks:** range_check
   - **Verdict:** PASS
   - **Notes:** Checked bounds $[0, 1]$.
8. ✔ **C8_observations_match_across_models** (Page 5 Table 1)
   - **Claim:** Table 1 reports Observations = 2,828 for both 5-day and 21-day regressions; verify equality.
   - **Checks:** equality_check
   - **Verdict:** PASS
   - **Notes:** Compared observation counts.
9. ✔ **C9_hmm_regime_means_ordering** (Page 5 Section 3.1 (bulleted means))
   - **Claim:** Regime labels Low/Moderate/High should correspond to increasing mean VIX: $17.54 < 19.29 < 21.90$.
   - **Checks:** monotonic_order_check

- **Verdict:** PASS
- **Notes:** Checked strict ordering.

10. ✔ **C10_adf_pvalue_vs_statistic_sign_consistency** (Page 4 Section 3.1)
    - **Claim:** ADF test reported as statistic = $-4.97$ with $p < 0.001$; check internal sign/plausibility: statistic is negative and p-value bound is between 0 and 1.
    - **Checks:** sign_and_range_check
    - **Verdict:** PASS
    - **Notes:** Basic numeric sanity checks.
11. ✔ **C11_logistic_auc_margin_over_random** (Page 7 Section 3.3; Page 8 Figure 2 caption)
    - **Claim:** AUROC reported as $0.526$ and described as marginally better than random guess ($0.500$); verify difference is $0.026$.
    - **Checks:** difference_recalculation
    - **Verdict:** PASS
    - **Notes:** Computed AUROC margin over random.
12. ✔ **C12_pr_auc_range_check** (Page 7 Section 3.3; Page 8 Figure 2 caption)
    - **Claim:** PR AUC reported as $0.0195$; verify it lies in $[0, 1]$.
    - **Checks:** range_check
    - **Verdict:** PASS
    - **Notes:** Checked bounds $[0, 1]$.
13. ✔ **C13_loocv_mean_sd_vs_baseline_auroc** (Page 7 Section 3.4; Page 8 Figure 3 caption)
    - **Claim:** LOOCV AUROC mean = $0.5272$ with SD = $0.0094$; baseline model performance is $0.526$. Check baseline is within a small number of SDs of mean and that mean is close to baseline (difference $0.0012$).
    - **Checks:** consistency_with_summary_stats
    - **Verdict:** PASS
    - **Notes:** Checked baseline proximity to LOOCV mean (diff and z-score).

**Limitations**

- Audit used only the provided PDF parsed text; no underlying data, code, or supplemental tables were available to recompute econometric results.
- No numerical extraction was performed from plots/figures (pixel reading), per instruction; figure-based values not explicitly printed in text were treated as unavailable.
- Several checks are limited to internal arithmetic/sanity/consistency (counts, ranges, simple differences) rather than validating statistical procedures or reproducing model estimates.
- Multiple key empirical results (OLS coefficients/p-values including Newey-West corrections, ADF test confirmation, HMM regime mean recomputation, AUROC/PR AUC and LOOCV distribution recomputation) could not be independently recomputed without the underlying time series, residuals, state assignments, and prediction outputs.