Data Science in Finance and Economics

https://www.aimspress.com/journal/dsfe

DSFE, 6(2): 336–358.
DOI:10.3934/DSFE.2026012
Received: 17 November 2025
Revised: 02 April 2026
Accepted: 16 April 2026
Published: 01 June 2026

Research article

Monotone-constrained functional component analysis for forward risk-accumulation curves with Leak-Free temporal validation and regime classification

Çağlar SÖZEN*

Department of Finance and Banking, Giresun University, Giresun, Turkey

* **Correspondence:** Email: caglar.sozen@giresun.edu.tr.

Abstract: I studied whether a monotone-constrained functional representation could improve regime classification for forward risk-accumulation curves under a time-respecting empirical design. The empirical application focused on Apple Inc. (AAPL) and used a supportive tech-growth panel consisting of AAPL, MSFT, NVDA, AMZN, QQQ, XLK, SPY, XLC, XLY, and XLI for descriptive context. I constructed forward curves over a 60-day horizon and compared cumulative forward variance with root-cumulative forward volatility, taking the latter as the main empirical specification. To represent these structurally monotone curves, I implemented a monotone-constrained component extraction procedure, denoted mFPCA for convenience, based on isotonic projection and sequential deflation, and evaluated it against an unconstrained FPCA benchmark under identical classifiers and validation schemes. All transformations, thresholds, and tuning steps were estimated on the training slice only and then transferred unchanged to the evaluation slice. Under blocked validation, the preferred AAPL specification attained Accuracy 0.862 and Macro-F1 0.863; under rolling walk-forward evaluation, the corresponding values were 0.815 and 0.822. Relative to FPCA, the gains from mFPCA were positive but modest. Moreover, ablation results showed that cone depth did not improve the final AAPL specification, while sensitivity analysis supported $K = 3$ as the main component choice. Permutation importance further indicated that predictive information was concentrated primarily in the early-to-middle portion of the forward curve. Simulation evidence showed that mFPCA was more closely aligned with latent monotone structure than FPCA when judged by matched component angles, and that random K-fold validation remained mildly optimistic relative to blocked validation under temporally dependent regime-shift designs. Overall, the results supported monotone-constrained representation as a credible and structurally coherent approach to forward risk classification when combined with leak-free temporal evaluation.

Keywords: functional data analysis; monotone-constrained representation; forward risk-accumulation curves; temporal cross-validation; leak-free evaluation; regime classification

1. Introduction

Understanding how market risk accumulates over future horizons is central to portfolio allocation, stress monitoring, volatility-sensitive risk control, and early-warning systems in financial markets. Although much of the volatility literature concentrates on one-step-ahead conditional variance or on realized measures at a single horizon, many practical decisions depend not only on the level of future risk but also on the *path* by which that risk builds up across subsequent trading days. Forward risk-accumulation curves provide a natural representation for this purpose: For each origin date, they summarize the cumulative accumulation of future squared returns over a sequence of horizons. By construction, these curves are monotone nondecreasing in the horizon. This structural property makes them natural objects for functional data analysis (FDA), where the aim is to summarize high-dimensional curves through a small number of interpretable latent directions (Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012; Hsing and Eubank, 2015).

A difficulty arises, however, because classical functional principal component analysis (FPCA) does not explicitly incorporate monotonicity into the representation step. In noisy financial environments, especially under heavy tails, volatility clustering, and changing market states, unconstrained FPCA may produce oscillatory component functions that explain variation well in a statistical sense but are more difficult to interpret economically and may be less stable across samples (James et al., 2000; Rice and Wu, 2001; Chiou, 2012). For forward risk-accumulation curves, this issue is especially consequential: If the observed object is structurally monotone, then a representation built from unrestricted directions may preserve variance while diluting the economic geometry of risk accumulation. In this study, I am motivated by the view that, in this setting, monotonicity should enter the representation stage directly rather than being left to emerge indirectly.

Against this background, I study whether a monotone-constrained functional representation improves regime classification relative to an unconstrained FPCA benchmark under a fully time-respecting empirical design. The empirical application centers on AAPL and uses a reference panel comprising AAPL, MSFT, NVDA, AMZN, QQQ, XLK, SPY, XLC, XLY, and XLI for descriptive and contextual purposes. Two structurally monotone forward representations are considered over a 60-day horizon: Cumulative forward variance and root-cumulative forward volatility. The main empirical specification uses the root-cumulative representation, while the cumulative version is retained as a robustness comparison. Monotone component directions are extracted through a sequential projection-and-deflation procedure based on isotonic projection onto the monotone cone. These directions are then used to construct score-based features for three-regime classification (`calm`, `normal`, `stress`) using a class-weighted multinomial ridge classifier. All transformations, thresholds, and tuning steps are estimated on the training slice only and then transferred unchanged to the corresponding test slice.

The empirical evidence shows that representational choices matter materially. In the AAPL application, the root-cumulative representation dominates the cumulative alternative, and the monotone-constrained specification improves on the unconstrained FPCA benchmark under both blocked and rolling evaluations, although the gains are not large. The ablation results further show that adding cone depth is not beneficial in this particular design, whereas horizon-segment importance analysis indicates that discriminative information is concentrated primarily in the early-to-middle portion of the forward curve rather than at the far end of the horizon.

The simulation evidence complements the empirical results in three directions. First, under synthetic

monotone signal structures with heavy-tailed noise, monotone-constrained extraction tracks the latent monotone directions more closely than unconstrained FPCA when assessed by matched principal angles. Second, repeated multi-start estimation yields virtually identical fitted directions in the baseline simulation design, indicating strong practical stability of the extraction procedure. Third, in Markov-regime simulations, random K-fold validation remains optimistic relative to blocked validation, although the magnitude of this optimism is modest in the final data-generating design. Taken together, these findings suggest that the main value of the monotone-constrained representation lies less in dramatic predictive gains than in a more coherent, stable, and auditable low-dimensional summary of forward risk accumulation.

The contribution of the paper is fourfold. First, I introduce a monotone-constrained component-extraction procedure for forward risk-accumulation curves that is directly aligned with the structural monotonicity of the object of interest. Second, I provide a fully auditable empirical design in which all representation learning, labeling, scaling, and tuning steps are kept strictly within the training slice, thereby minimizing temporal leakage. Third, I evaluate the constrained representation against an unconstrained FPCA benchmark under identical classifiers and validation schemes, isolating the incremental role of the monotonicity restriction. Fourth, I show empirically that representational choices, such as root-cumulative versus cumulative construction and the inclusion of auxiliary depth features, can materially affect downstream performance. In this sense, the paper contributes to the broader literature showing that, in financial prediction problems, disciplined representation design and disciplined evaluation are inseparable.

2. Literature review and positioning

The study lies at the intersection of three literatures: Functional data analysis, volatility measurement and modeling, and time-respecting predictive evaluation. From the FDA side, the idea is to treat each observation as a function defined on a dense grid and to summarize cross-sectional variation through a small number of latent directions. Ramsay and Silverman (Ramsay and Silverman, 2005) provided the canonical foundation for this perspective, while Horváth and Kokoszka (Horváth and Kokoszka, 2012) and Hsing and Eubank (Hsing and Eubank, 2015) developed the broader theoretical and inferential framework for functional data. Within this tradition, FPCA is the standard dimension-reduction device: It extracts orthogonal directions that explain the largest share of variation in the sample covariance structure (James et al., 2000; Rice and Wu, 2001). Yet, the object studied here is not a generic functional observation. Forward risk-accumulation curves are structurally monotone in the horizon because they arise from cumulative future variation, and the FDA literature has paid much less direct attention to representation learning for this type of structurally monotone future-risk curve.

The second relevant literature concerns realized volatility and multi-horizon risk accumulation. Realized-volatility measures are widely used because they provide observable ex post proxies for latent variation and connect naturally to volatility dynamics, persistence, and forecasting (Barndorff-Nielsen and Shephard, 2002; Andersen et al., 2003). Within this literature, Corsi's HAR-RV model is especially influential because it formalizes the idea that volatility evolves across heterogeneous horizons and that short-, medium-, and long-run components jointly matter for persistence (Corsi, 2009). Related observation-driven frameworks, such as Realized-GARCH, likewise emphasize the informational role of realized measures in volatility modeling (Hansen et al., 2012). However, most of this literature

represents future risk through scalar summaries or parametric dynamics rather than through curve-level representations that preserve the shape of risk accumulation across horizons.

The third literature concerns shape-constrained estimation. When prior structure is given, shape restrictions can improve interpretability and stabilize estimation. For monotonicity, isotonic regression and cone projection provide one of the cleanest such frameworks. Barlow et al. (Barlow et al., 1972) and Robertson, Wright, and Dykstra (Robertson et al., 1988) established the classical theory of estimation under order restrictions and the algorithmic role of the pool-adjacent-violators algorithm. In functional settings, monotonicity and related shape restrictions have also been studied through smooth constrained regression and shape-constrained spline methods (Ramsay, 1988; Pya and Wood, 2015). These contributions strongly support the central methodological intuition of this paper: When the observed curves are structurally monotone, the extracted latent directions should be constrained accordingly rather than estimated in a fully unrestricted manner. Furthermore, the existing literature does not directly deliver the specific combination considered here, namely finance-oriented forward risk curves, monotone component extraction, and leak-free regime classification within a unified empirical framework.

A related but conceptually distinct literature concerns functional depth and curve centrality. Depth measures provide scalar summaries of how central or outlying a curve is relative to a reference sample and have been used for ranking, visualization, outlier detection, and classification in functional settings (Cuevas et al., 2007; López-Pintado and Romo, 2009; Sun and Genton, 2011; Zuo and Serfling, 2000). In this setting, a cone-based depth is attractive because it measures the distance of a centered curve to the monotone cone, thereby quantifying shape conformity. This makes it a natural candidate feature once monotonicity is treated as a first-order structural property. Even so, the empirical results in this paper show that the predictive contribution of such a feature is context-dependent rather than automatic: Cone depth is theoretically well motivated, but it does not improve performance in the final AAPL specification.

The final literature concerns evaluation under dependence. For time-indexed financial data, random K-fold cross-validation can yield overly optimistic estimates because it ignores temporal ordering and may inadvertently mix information across training and evaluation slices (Tashman, 2000; Arlot and Celisse, 2010; Bergmeir and Benítez, 2012). More appropriate alternatives include blocked validation and rolling-origin or walk-forward designs, which preserve temporal structure and offer a more realistic approximation to deployment conditions (Roberts et al., 2017; Cerqueira et al., 2020). This issue becomes even more important when labels themselves are constructed from future windows, as in forward risk-accumulation-based regime classification. In such cases, leakage can arise not only through the classifier, but also through centering, scaling, feature construction, thresholding, and internal model selection. Despite this, the predictive-evaluation literature rarely treats the representation step itself as leakage-sensitive, even though representation learning can be just as exposed to temporal contamination as the final classifier.

This positioning clarifies the paper's niche. The contribution is not to propose a universal replacement for FPCA in all financial functional-data settings, nor to claim that monotonicity alone yields large predictive gains. Rather, it is to show that when the target object is itself structurally monotone, a constrained representation can yield a more faithful and slightly more effective low-dimensional summary, provided that the entire empirical design is evaluated under a strict time-respecting protocol. In this sense, I address a narrower but important gap in the literature: The combination of forward

risk-accumulation curves, monotone-constrained representation learning, and fully auditable blocked and rolling validation within a single empirical framework.

3. Materials and methods

In this section, I present the empirical design, the monotone-constrained component-extraction procedure, the feature-engineering pipeline, the validation protocol, and the simulation framework used in the study. The empirical application centers on Apple Inc. (AAPL) and examines whether a monotone-constrained functional representation improves regime classification relative to unconstrained FPCA under a time-ordered validation design. Throughout the paper, the object of interest is referred to as *forward risk-accumulation curves*. In the empirical implementation, these curves are operationalized either as cumulative forward variance or as root-cumulative forward volatility. The main empirical specification uses the root-cumulative form, whereas the cumulative alternative is retained as a robustness check. All transformations, tuning steps, and label thresholds are estimated on the training slice only and then transferred unchanged to the corresponding evaluation slice.

3.1. Data, target asset, and forward risk-accumulation curves

The empirical analysis uses daily adjusted-close data from Yahoo Finance for a reference panel comprising Apple Inc. (AAPL), Microsoft Corporation (MSFT), NVIDIA Corporation (NVDA), Amazon.com, Inc. (AMZN), Invesco QQQ Trust (QQQ), the Technology Select Sector SPDR Fund (XLK), the SPDR S&P 500 ETF Trust (SPY), the Communication Services Select Sector SPDR Fund (XLC), the Consumer Discretionary Select Sector SPDR Fund (XLY), and the Industrial Select Sector SPDR Fund (XLI) over the period from 1 January 2019 to 31 December 2025 (Yahoo Finance, 2025). AAPL is the target asset in the classification analysis, while the remaining series are used for descriptive and contextual purposes. Let

$$r_t = \log(P_t/P_{t-1})$$

denote the daily log-return computed from the adjusted close P_t. For a fixed forecasting horizon H (here $H = 60$), define the forward curve

$$X_t(h) = \begin{cases} \sum_{j=1}^{h} r_{t+j}^2, & \text{cumulative forward variance,} \\ \left(\sum_{j=1}^{h} r_{t+j}^2\right)^{1/2}, & \text{root-cumulative forward volatility,} \end{cases} \qquad h = 1, \ldots, H. \tag{1}$$

Both representations are monotone nondecreasing in the horizon h by construction. The main analysis uses the second form, whereas the cumulative representation is considered only as a robustness comparison.

Stacking the curves for the target asset yields a matrix $X \in \mathbb{R}^{n\times H}$, where each row corresponds to one origin date and each column to one horizon point. Denote the training-sample column mean by $\mu \in \mathbb{R}^H$ and the centered matrix by

$$X_c = X - \mathbf{1}\mu^\top.$$

All centering and subsequent feature transformations are estimated on the training slice only and then applied unchanged to the test slice.

3.2. Monotonicity constraint and the isotonic cone

Let $\mathcal{H} = \mathbb{R}^H$ be equipped with the Euclidean inner product. The monotonicity restriction is encoded through the isotonic cone

$$\mathcal{K} = \left\{\phi \in \mathbb{R}^H : \phi(1) \leq \phi(2) \leq \cdots \leq \phi(H)\right\},$$

which is a nonempty closed convex cone. For any $y \in \mathbb{R}^H$, let $\Pi_{\mathcal{K}}(y)$ denote the Euclidean projection onto $\mathcal{K}$, computed via the pool-adjacent-violators algorithm (PAV) (Barlow et al., 1972; Robertson et al., 1988). Since $\mathcal{K}$ is closed and convex, $\Pi_{\mathcal{K}}(y)$ exists, is unique, and is nonexpansive.

3.3. Monotone-constrained component extraction procedure

Throughout the paper, I denote the implemented monotone-constrained component extraction procedure by mFPCA for convenience. The procedure is designed as a projection-and-deflation representation scheme for structurally monotone curves, building on the broader FPCA and shape-constrained estimation literature (James et al., 2000; Rice and Wu, 2001; Barlow et al., 1972; Robertson et al., 1988). Let K denote the number of retained components. The main empirical specification uses $K = 3$, and $K \in \{2, 3, 4, 5\}$ is examined in a separate sensitivity analysis. Components are extracted sequentially from the training matrix X_c through monotone projection and rank-one deflation. The resulting directions are used as low-dimensional monotone components in the downstream classification stage.

For the k-th component, let R_{k-1} denote the current residual matrix, with $R_0 = X_c$. Starting from an initial vector $\phi^{(0)}$, the algorithm iterates

$$s^{(m)} = R_{k-1}\phi^{(m)}, \qquad v^{(m+1)} = R_{k-1}^{\top} s^{(m)}, \qquad \phi^{(m+1)} = \frac{\Pi_{\mathcal{K}}(v^{(m+1)})}{\|\Pi_{\mathcal{K}}(v^{(m+1)})\|_2}. \tag{2}$$

The iteration is terminated when

$$\|\phi^{(m+1)} - \phi^{(m)}\|_2 < 10^{-8}, \tag{3}$$

or when a maximum of 200 iterations is reached. After convergence, the score vector for the k-th component is

$$z_k = R_{k-1}\phi_k,$$

and the residual is updated by rank-one deflation,

$$R_k = R_{k-1} - z_k\phi_k^{\top}. \tag{4}$$

This procedure is repeated for $k = 1, \ldots, K$.

Initialization is multi-start. The first start is obtained by projecting the leading right singular vector of the current residual matrix onto the isotonic cone,

$$\phi_{\mathrm{SVD}}^{(0)} = \frac{\Pi_{\mathcal{K}}(v_1)}{\|\Pi_{\mathcal{K}}(v_1)\|_2},$$

where v_1 is the leading right singular vector of R_{k-1}. Additional starts are generated as random monotone vectors by cumulative summation of absolute Gaussian draws followed by normalization. In the empirical application, six starts are used in the main fit, while a larger number is employed in the

initialization-stability analysis. For each component, the selected start is the one that maximizes the training objective

$$Q(\phi) = \|R_{k-1}\phi\|_2^2. \tag{5}$$

The extracted directions should be interpreted as ordered monotone component directions. Because the procedure combines isotonic projection with sequential deflation, the resulting components are not imposed to form an exact orthonormal eigenbasis. Their role is representational rather than spectral in the strict classical sense.

3.4. Cone depth as an auxiliary feature

A cone-depth measure is used to quantify the proximity of a centered curve to the isotonic cone. This construction is motivated by the functional depth literature, which studies centrality and outlyingness for curves in infinite- and finite-dimensional settings (Cuevas et al., 2007; López-Pintado and Romo, 2009; Sun and Genton, 2011; Zuo and Serfling, 2000). For a centered curve $x \in \mathbb{R}^H$, define

$$\mathrm{CD}(x) = \frac{1}{1 + \|x - \Pi_{\mathcal{K}}(x)\|_2 / s}, \tag{6}$$

where

$$s = \mathrm{median}(\|X_c^{(\mathrm{tr})}(i, \cdot)\|_2)$$

is a robust scale estimated on the training slice only. Raw cone-depth values are standardized using the training-sample mean and standard deviation and then applied unchanged to the test slice. This avoids joint scaling of training and test observations and preserves the chronological integrity of the evaluation.

Cone depth is retained as an auxiliary, structurally motivated feature and is evaluated in ablation analysis; however, it is not included in the final AAPL specification because it does not improve predictive performance under either blocked or rolling validation.

3.5. Feature construction

Using the training-fitted mean function μ and component directions $\{\phi_k\}_{k=1}^K$, the score features are

$$z_{t,k} = (X_t - \mu)^\top \phi_k, \qquad k = 1, \ldots, K.$$

In the main specification, $K = 3$. From these scores and from the original target curve $X_t(\cdot)$, the following numerical features are constructed:

1. Component scores $z_{t,1}, z_{t,2}, z_{t,3}$.
2. Score ratios:
$$\frac{z_{t,2}}{z_{t,1} + \varepsilon}, \qquad \frac{z_{t,3}}{z_{t,1} + \varepsilon},$$
with $\varepsilon = 10^{-8}$.
3. Early, middle, and late horizon summaries:
$$E_t = \frac{1}{|I_E|} \sum_{h \in I_E} X_t(h), \qquad M_t = \frac{1}{|I_M|} \sum_{h \in I_M} X_t(h), \qquad L_t = \frac{1}{|I_L|} \sum_{h \in I_L} X_t(h),$$
where I_E, I_M, and I_L partition the horizon grid into three equal blocks.

4. Contrasts $E_t - L_t$ and $M_t - L_t$.
5. Lag-1 and moving-average features of the component scores.
6. Optionally, the standardized cone-depth feature and its lag.

The constant $\varepsilon = 10^{-8}$ is included only for numerical stabilization. If $z_{t,1}$ is very close to zero, the ratio may become unstable or mechanically inflated. The constant is chosen to be negligible relative to the empirical scale of the scores, so that it has essentially no effect away from near-zero denominators while preventing spurious blow-ups in rare nearly singular cases.

All lagged and moving-average features are built in the true chronological order of the target series. Training and test rows are transformed using the training-fitted representation without reordering the underlying time index. This prevents contamination that may arise when lagged features are constructed after concatenating train and test segments in artificial order.

After feature construction, numerical columns are filtered using the training slice only, and standardization is carried out using training-sample means and standard deviations prior to classifier fitting; the same transformations are then applied unchanged to the test slice.

3.6. Labeling rule, classifier, and validation design

The regime labels are defined from the target asset itself. Let

$$Y_t^{(H_0)} = \sum_{j=1}^{H_0} r_{t+j}^2$$

denote the H_0-day cumulative forward variance of the target asset, with $H_0 = 20$. On each training slice, the empirical 30th and 70th percentiles $(q_{0.3}, q_{0.7})$ of $Y_t^{(H_0)}$ are computed and used to define

$$\texttt{calm}: \; Y_t^{(H_0)} \le q_{0.3}, \qquad \texttt{stress}: \; Y_t^{(H_0)} \ge q_{0.7}, \qquad \texttt{normal}: \; q_{0.3} < Y_t^{(H_0)} < q_{0.7}.$$

These thresholds are recomputed within each training fold and then applied unchanged to the corresponding test slice.

The classifier is a class-weighted multinomial ridge model implemented through `glmnet` with $\alpha = 0$. Hyperparameter selection is performed through blocked inner cross-validation whenever the class structure permits a feasible blocked split; otherwise, a stable one-vs-rest ridge fallback is used. This preserves chronological ordering during tuning as well (Friedman et al., 2010).

Performance is evaluated under two time-ordered schemes:

1. **Blocked K-fold cross-validation:** $K = 5$ contiguous date blocks.
2. **Rolling walk-forward evaluation:** Minimum training length $T_{\min} = 200$ observations and test span $\Delta T = 30$.

In both schemes, all representational objects—means, monotone directions, feature standardizers, near-zero-variance filtering, and label thresholds—are estimated on the training slice only and then frozen on the test slice.

The reported performance measures are Accuracy, Macro-F1, class-wise F1, one-vs-rest precision–recall (PR) curves, and class-wise area under the PR curve (AUPRC) (Davis and Goadrich, 2006; He and Garcia, 2009; Saito and Rehmsmeier, 2015).

3.7. Comparator models and robustness checks

All main classification results are paired with an unconstrained FPCA benchmark using the same number of retained components, the same classifier, and the same blocked and rolling evaluation protocol. This comparison isolates the incremental contribution of the monotonicity constraint itself.

The empirical study includes four major layers of robustness analysis:

1. **Representation robustness:** Root-cumulative (main) versus cumulative forward curves.
2. **Ablation analysis:** mFPCA with and without cone depth, and FPCA with and without cone depth.
3. **Component sensitivity:** $K \in \{2, 3, 4, 5\}$.
4. **Initialization stability:** Repeated fits across multiple random starts and seeds.

3.8. Permutation importance for horizon segments

For interpretability, the horizon $H = 60$ is divided into 12 nonoverlapping five-step bins. Under the blocked-CV specification, let F^{base} denote the baseline Macro-F1 on the held-out fold and let $F^{(j,b)}$ denote the Macro-F1 after randomly permuting the j-th horizon-bin feature in the test fold on repetition $b = 1, \ldots, B$. The raw importance of bin j is

$$\mathrm{PI}_j = \frac{1}{B} \sum_{b=1}^{B} \left(F^{\text{base}} - F^{(j,b)} \right). \tag{7}$$

Negative values are truncated at zero and the resulting nonnegative importances are normalized to sum to one:

$$\widetilde{\mathrm{PI}}_j = \frac{\max\{\mathrm{PI}_j, 0\}}{\sum_{\ell=1}^{12} \max\{\mathrm{PI}_\ell, 0\}}. \tag{8}$$

Thus, a larger value indicates a larger deterioration in held-out Macro-F1 when the corresponding horizon segment is destroyed by permutation.

3.9. Simulation design and evaluation criteria

The simulation study complements the empirical analysis in three ways. First, a recovery experiment generates synthetic curves from latent monotone directions under heavy-tailed noise and compares monotone-constrained extraction with unconstrained FPCA. Let $\{\phi_1^0, \phi_2^0, \phi_3^0\}$ denote the latent monotone directions used in the data-generating signal and let $\{\widehat{\phi}_1, \widehat{\phi}_2, \widehat{\phi}_3\}$ denote the estimated directions. After the best permutation and sign alignment, the component-wise principal angles are

$$\theta_k = \arccos\left(\left| \left\langle \phi_k^0, \widehat{\phi}_{\pi(k)} \right\rangle \right| \right) \times \frac{180}{\pi}, \qquad k = 1, 2, 3, \tag{9}$$

and the mean of these matched angles is reported. Smaller angles indicate that the estimated directions more closely track the latent monotone structure. The simulation study also reports the signal reconstruction error

$$\mathrm{MSE}_{\text{sig}} = \frac{1}{nH} \| \widehat{S} - S^0 \|_F^2,$$

where S^0 is the true signal matrix and $\widehat{S}$ is its low-rank reconstruction.

Second, an initialization-stability experiment repeats the monotone-constrained extraction across multiple seeds and reports angles to a reference fit, cumulative explained variance, and monotonicity violations.

Third, a Markov-regime simulation evaluates both (i) the optimism of random K-fold relative to blocked K-fold and (ii) the classification difference between monotone-constrained extraction and FPCA under stationary, moderate-shift, and strong-shift scenarios. These simulations are reported under both blocked and rolling evaluations and therefore parallel the empirical protocol.

3.10. Computational remarks

Each PAV projection is $O(H)$. For a given residual matrix $R \in \mathbb{R}^{n\times H}$, one iteration of the component update requires a matrix–vector multiplication $R^\top(R\phi)$, which is $O(nH)$, followed by projection and normalization. With $H = 60$, $K = 3$, and a moderate number of iterations, the constrained fit remains computationally lightweight in practice. Relative to low-rank FPCA score extraction, the additional cost is limited to isotonic projection and multi-start selection.

3.11. Scope and extension

The empirical design is intentionally focused on AAPL in order to provide a clean and auditable evaluation of the shape-constrained representation. The method itself is not restricted to this setting. The same projection-and-deflation strategy can, in principle, be embedded in multivariate or multi-asset functional settings, for example by fitting asset-specific monotone directions or by constructing shared cross-asset functional representations. Likewise, while this implementation operates directly on discretized forward curves, spline-based or integrated bases could be incorporated prior to monotone projection if additional smoothing were desired. The direct grid representation is retained here in order to keep the monotonicity constraint transparent and the leakage-control protocol auditable.

4. Empirical analysis

I now turn to the empirical evidence for the AAPL design. The analysis is organized around five questions. First, what are the unconditional return characteristics of the supportive tech-growth panel? Second, how strong is the contemporaneous dependence structure around AAPL? Third, how does the preferred monotone-constrained representation perform under blocked and rolling evaluations? Fourth, how sensitive are the results to auxiliary design choices such as cone depth, the number of retained components, and the curve representation itself? Fifth, which portions of the forward curve contribute most to predictive discrimination? Throughout, the raw return sample ran from 1 January 2019 to 31 December 2025, whereas the effective curve sample was shorter because the construction of forward risk-accumulation curves requires future returns up to horizon $H = 60$. In this application, the effective sample used in the classification analysis ran from 3 January 2019 to 3 October 2025 and contained $N_{\text{effective}} = 1698$ forward curves.

Table 1 reports descriptive statistics for daily log-returns in the supportive panel. Mean returns are small relative to dispersion across all assets, which is typical at the daily frequency. The panel is also heterogeneous in risk. NVDA is the most volatile series, with daily standard deviation around 3.2%, while SPY is the least volatile broad market reference in the set, with standard deviation around 1.2%. Skewness is generally negative, especially for SPY, XLC, XLI, and XLY, indicating heavier downside

moves than upside moves. Kurtosis is above the Gaussian benchmark for all series and is particularly high for SPY and XLI, confirming that the sample contains substantial tail risk and episodic turbulence. These stylized facts motivate two aspects of the design. First, forward risk-accumulation curves provide a natural way to summarize how turbulent episodes propagate across future horizons. Second, any low-dimensional representation used for classification must remain numerically stable under heavy tails and shifting market conditions.

Table 1. Descriptive statistics of daily log-returns for the supportive tech-growth panel.

Symbol	N	Mean	SD	Skewness	Kurtosis
AAPL	1758	0.001	0.020	−0.083	6.619
AMZN	1758	0.001	0.022	−0.077	4.603
MSFT	1758	0.001	0.018	−0.185	8.063
NVDA	1758	0.002	0.032	−0.004	4.232
QQQ	1758	0.001	0.015	−0.383	7.522
SPY	1758	0.001	0.012	−0.588	14.125
XLC	1758	0.001	0.014	−0.518	7.520
XLI	1758	0.001	0.014	−0.526	13.786
XLK	1758	0.001	0.017	−0.291	8.903
XLY	1758	0.001	0.015	−0.578	7.860

Notes: The table reports sample size, mean, standard deviation, skewness, and kurtosis for daily log-returns over the raw return sample. The supportive panel is used for descriptive and contextual purposes, while the classification analysis is conducted on AAPL.

Table 2. Correlation matrix of daily log-returns for the supportive tech-growth panel.

Asset	AAPL	AMZN	MSFT	NVDA	QQQ	SPY	XLC	XLI	XLK	XLY
AAPL	1.000	0.589	0.702	0.583	0.820	0.779	0.710	0.591	0.826	0.705
AMZN	0.589	1.000	0.670	0.581	0.773	0.666	0.698	0.437	0.696	0.758
MSFT	0.702	0.670	1.000	0.670	0.862	0.805	0.758	0.583	0.870	0.715
NVDA	0.583	0.581	0.670	1.000	0.803	0.703	0.630	0.503	0.816	0.652
QQQ	0.820	0.773	0.862	0.803	1.000	0.937	0.887	0.721	0.975	0.888
SPY	0.779	0.666	0.805	0.703	0.937	1.000	0.870	0.887	0.933	0.899
XLC	0.710	0.698	0.758	0.630	0.887	0.870	1.000	0.695	0.828	0.813
XLI	0.591	0.437	0.583	0.503	0.721	0.887	0.695	1.000	0.738	0.776
XLK	0.826	0.696	0.870	0.816	0.975	0.933	0.828	0.738	1.000	0.838
XLY	0.705	0.758	0.715	0.652	0.888	0.899	0.813	0.776	0.838	1.000

Table 2 reports the correlation matrix for the same panel. Dependence is strong and highly structured. AAPL is tightly connected to the technology and growth segment of the market, with especially high correlations with XLK (0.826), QQQ (0.820), and MSFT (0.702). More generally, QQQ, SPY, and XLK form a dense core with pairwise correlations above 0.93 in some cases, indicating a pronounced growth-market factor over the sample. These correlations are reported for contextual and descriptive purposes rather than as direct inputs to the classifier. Their relevance here is interpretive: The target

asset is embedded in a highly connected growth-oriented environment, which makes it plausible that shifts in the shape of its forward risk-accumulation curve reflect broader market conditions rather than purely idiosyncratic noise.

Table 3 reports the major classification results. Under blocked five-fold validation, the preferred specification attains Accuracy = 0.862 and Macro-F1 = 0.863. Under rolling walk-forward evaluation, these decline to Accuracy = 0.815 and Macro-F1 = 0.822. The difference between the two schemes is expected: Blocked folds preserve temporal order but remain less demanding than genuine rolling-origin prediction. Overall, the rolling results remain strong, indicating that the representation retains predictive value when transported across time. The close alignment between Accuracy and Macro-F1 also suggests that performance is not being driven solely by the largest class.

Table 3. AAPL: Overall classification performance under blocked and rolling evaluations.

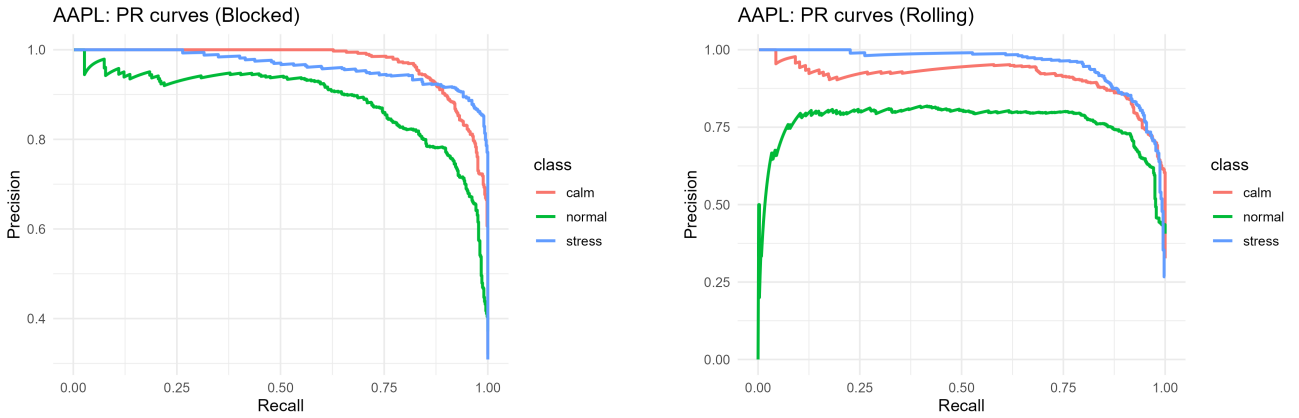
CV type	Accuracy	Macro-F1
Blocked	0.862	0.863
Rolling	0.815	0.822

The class-wise results in Table 4 clarify where the classification problem is easiest and hardest. Under blocked evaluation, the two extreme regimes are identified more cleanly than the middle regime: F1 equals 0.878 for `calm`, 0.919 for `stress`, and 0.791 for `normal`. The AUPRC values tell the same story, with especially strong blocked performance for `calm` (0.973) and `stress` (0.964). Under rolling evaluation, all classes deteriorate, but the decline is most pronounced for `normal`, whose AUPRC falls to 0.764. This pattern is substantively plausible. The central regime is intrinsically less distinct because it lies between the two extremes of unusually low and unusually high forward risk accumulation. Furthermore, the calm and stress states correspond to more clearly separated parts of the future-risk distribution and remain easier to detect even out of sample.

Table 4. AAPL: Class-wise performance under blocked and rolling evaluations.

Class	F1 (Blocked)	F1 (Rolling)	AUPRC (Blocked)	AUPRC (Rolling)
calm	0.878	0.830	0.973	0.912
normal	0.791	0.755	0.889	0.764
stress	0.919	0.881	0.964	0.953

Figure 1 shows the corresponding one-vs-rest precision–recall curves. The blocked curves remain uniformly above the rolling curves, which is consistent with the stronger blocked AUPRC values. In both panels, `normal` is the most challenging class, especially as recall approaches one. The `stress` and `calm` curves, by contrast, remain comparatively elevated across a broad recall range. In deployment terms, this means that the model is more reliable for detecting extreme future-risk conditions than for distinguishing observations that sit near the middle of the future-risk distribution.



(a) Blocked evaluation. **(b)** Rolling evaluation.

Figure 1. AAPL: One-vs-rest precision–recall curves under blocked and rolling evaluations.

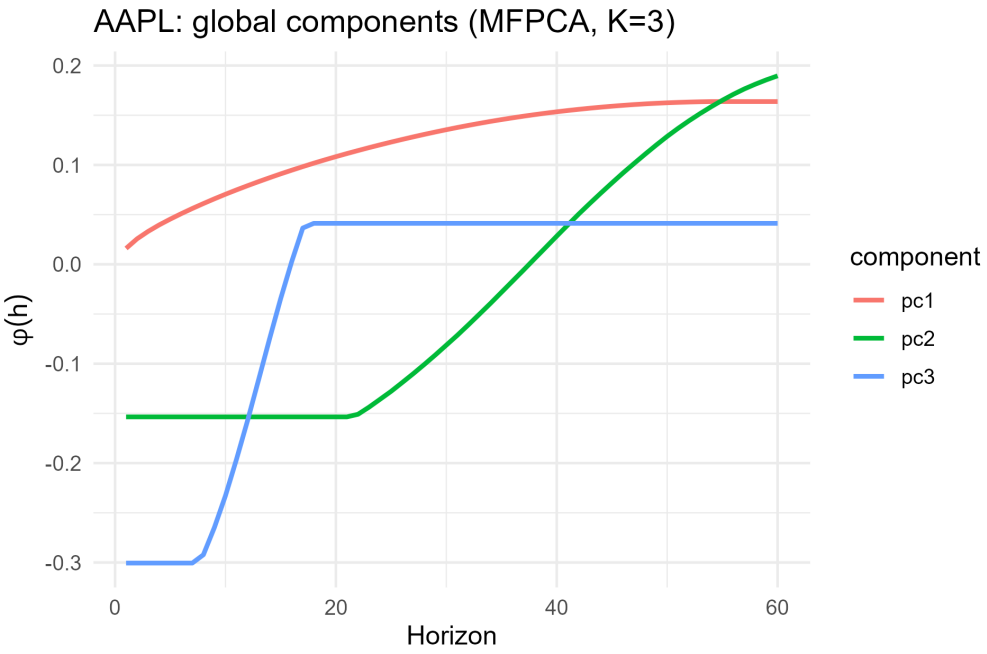


Figure 2. AAPL: Global monotone component directions for the preferred $K = 3$ specification.

The representation itself is summarized in Table 5 and Figure 2. The first reported monotone component explains 87.2% of total variation, while the first three jointly explain 96.1%. Thus, although the representation is not rank-one, it is still strongly concentrated in a small number of directions. The component figure also addresses an important interpretive issue: All three reported directions are visually nondecreasing over the horizon grid, which is consistent with the structural monotonicity built into the extraction procedure. At the same time, these directions should be interpreted as ordered monotone component directions rather than as an exact orthonormal eigenbasis. This distinction matters because the implemented projection-and-deflation procedure is designed primarily for low-dimensional representation, not for exact spectral orthogonality in the classical FPCA sense.

Table 5. Compact summary of the preferred AAPL specification.

Target	N_{eff}	Sample	Repr.	K	PC1 share	Cum. share	Acc. (B)	F1 (B)	Acc./F1 (R)
AAPL	1698	2019-01-03 to 2025-10-03	Root-cum.	3	0.872	0.961	0.862	0.863	0.815 / 0.822

Table 6 isolates the role of the shape constraint and the cone-depth feature. Three conclusions emerge. First, mFPCA outperforms FPCA when both are run without depth, under both blocked and rolling evaluations. Second, the gains are positive but limited, which is why the contribution is best interpreted as a disciplined representational improvement rather than a dramatic predictive breakthrough. Third, cone depth does not help in this design. Once depth is added, both mFPCA and FPCA deteriorate substantially and become nearly indistinguishable. Hence, the gains in the preferred specification are attributable primarily to the monotone representation itself, not to the auxiliary depth variable.

Table 6. AAPL: Ablation results comparing mFPCA/FPCA and depth/no-depth.

Method	Depth	CV	Accuracy	Macro-F1
MFPCA	Yes	Blocked	0.813	0.811
MFPCA	Yes	Rolling	0.790	0.797
MFPCA	No	Blocked	0.862	0.863
MFPCA	No	Rolling	0.815	0.822
FPCA	Yes	Blocked	0.813	0.810
FPCA	Yes	Rolling	0.790	0.797
FPCA	No	Blocked	0.845	0.844
FPCA	No	Rolling	0.810	0.817

Table 7 examines sensitivity to the number of retained components. The move from $K = 2$ to $K = 3$ produces a clear improvement in both blocked and rolling performance. Increasing the number of components beyond three raises the explained-variance share only marginally and yields at most negligible blocked gains, while slightly weakening rolling performance. Thus, $K = 3$ is a natural main specification: It already explains 96.1% of total variation and provides the best balance between representational richness and temporal generalization.

Table 7. AAPL: Sensitivity to the number of retained components.

K	CV	Variance explained	Accuracy	Macro-F1
2	Blocked	0.950	0.835	0.834
2	Rolling	0.950	0.795	0.801
3	Blocked	0.961	0.862	0.863
3	Rolling	0.961	0.815	0.822
4	Blocked	0.969	0.864	0.864
4	Rolling	0.969	0.812	0.819
5	Blocked	0.974	0.864	0.864
5	Rolling	0.974	0.810	0.817

Robustness across representations is reported in Table 8. The root-cumulative forward volatility representation dominates the cumulative forward variance alternative by a wide margin: Blocked Macro-F1 rises from 0.729 to 0.863, and rolling Macro-F1 rises from 0.735 to 0.822. Thus, the choice between cumulative and root-cumulative construction is not innocuous in this application. The root transformation is not merely a cosmetic re-expression of the same information; rather, it materially improves downstream classification. For this reason, the root-cumulative design is treated as the main empirical representation, while the cumulative version is retained only as a robustness benchmark.

Table 8. AAPL: Robustness across cumulative and root-cumulative representations.

Representation	Acc. (B)	F1 (B)	Acc. (R)	F1 (R)
Root-cumulative	0.862	0.863	0.815	0.822
Cumulative	0.736	0.729	0.726	0.735

Finally, Table 9 and Figure 3 report normalized conditional permutation importance over twelve nonoverlapping five-step horizon bins. The dominant predictive mass is concentrated in the early-to-middle portion of the forward curve, especially around segments 3–6. The single most important bin is seg4 (h = 16–20), followed by seg5 (21–25), seg3 (11–15), and seg6 (26–30). By contrast, the far end of the horizon carries very little incremental importance once the remaining representation is already in place. This pattern is notably different from a simple long-horizon persistence story. For this application, the main discriminative information appears to lie in how future risk begins and then consolidates over the first month of the forward window, rather than in the extreme tail end of the 60-day horizon.

Table 9. AAPL: Conditional permutation importance by horizon segment.

Segment	Importance
seg1	0.000
seg2	0.031
seg3	0.175
seg4	0.351
seg5	0.274
seg6	0.107
seg7	0.047
seg8	0.006
seg9	0.006
seg10	0.000
seg11	0.000
seg12	0.002

Taken together, the empirical results support four conclusions. First, the preferred specification is the root-cumulative monotone representation with $K = 3$ and no cone-depth augmentation. Second, mFPCA improves on unconstrained FPCA under identical classifiers and evaluation protocols, although the gap is not large. Third, these gains are strongest for the low- and high-risk regimes, while the middle regime remains the most difficult to separate under temporal drift. Fourth, the most informative part of the forward curve lies in the early-to-middle horizon range rather than at the far end of the 60-day window.

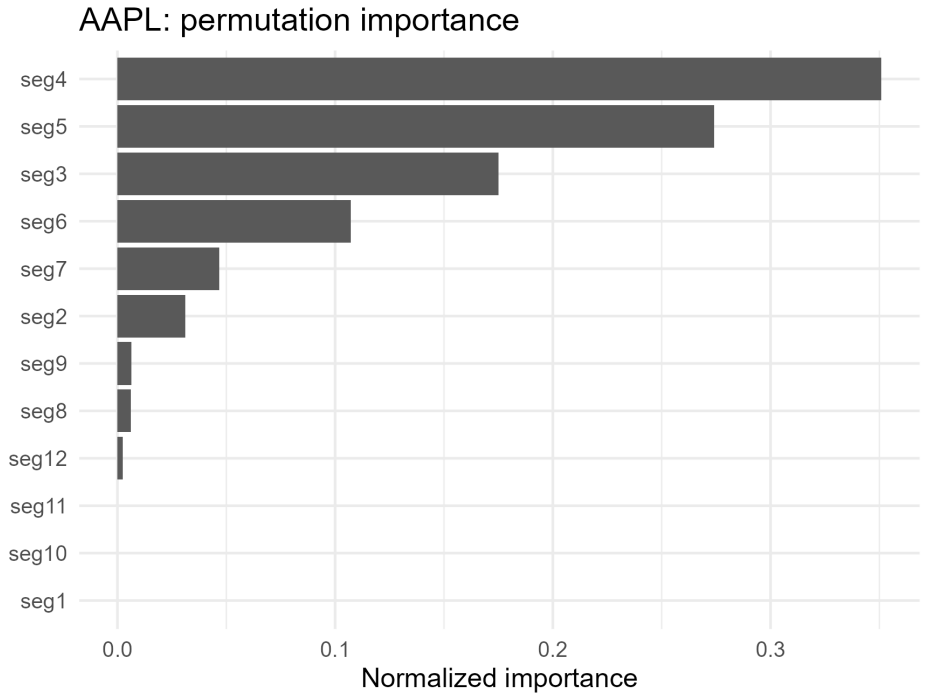


Figure 3. AAPL: Normalized conditional permutation importance by horizon segment.

5. Simulation

The simulation study complements the empirical analysis in three directions. First, it evaluates whether the monotone-constrained extraction procedure recovers latent monotone directions more closely than unconstrained FPCA in a controlled setting. Second, it assesses whether random K-fold validation remains optimistic relative to blocked validation under temporal dependence and regime shifts. Third, it compares mFPCA and FPCA as representational inputs to the same downstream classifier under blocked and rolling evaluations. In all cases, the emphasis is on directional alignment, temporal realism, and relative performance, rather than on claiming exact latent recovery or universally large predictive gains.

Table 10 reports the recovery experiment, and Figure 4 visualizes the corresponding matched-angle patterns. Across all six (n, SNR) combinations, mFPCA attains smaller matched mean component angles than FPCA, and the win rate is 1.00 throughout. The advantage becomes more pronounced as signal quality improves: For example, at $n = 300$ and $\mathrm{SNR} = 5$, the mean-angle gain is about 4.58°. At the same time, the absolute angles remain sizable, lying roughly between 48° and 53°. This is important for interpretation. The simulation supports the claim that mFPCA is *better aligned* with the latent monotone structure than FPCA, but it does not justify stronger language such as near-exact or high-precision recovery. The reconstruction-error differences are negligible, which is also informative: The constrained representation improves directional faithfulness without requiring a large sacrifice in low-rank signal reconstruction.

Table 10. Recovery study using matched angles to the latent monotone directions and signal reconstruction error.

n	SNR	Angle (mFPCA)	Angle (FPCA)	ΔAngle	MSE (mFPCA)	MSE (FPCA)	Win rate
150	2.5	53.215	53.270	0.056	0.000	0.001	1.000
150	3.5	51.267	53.117	1.850	0.000	0.000	1.000
150	5.0	48.450	53.005	4.555	0.000	0.000	1.000
300	2.5	52.404	53.186	0.782	0.000	0.001	1.000
300	3.5	50.808	53.067	2.259	0.000	0.000	1.000
300	5.0	48.398	52.977	4.579	0.000	0.000	1.000

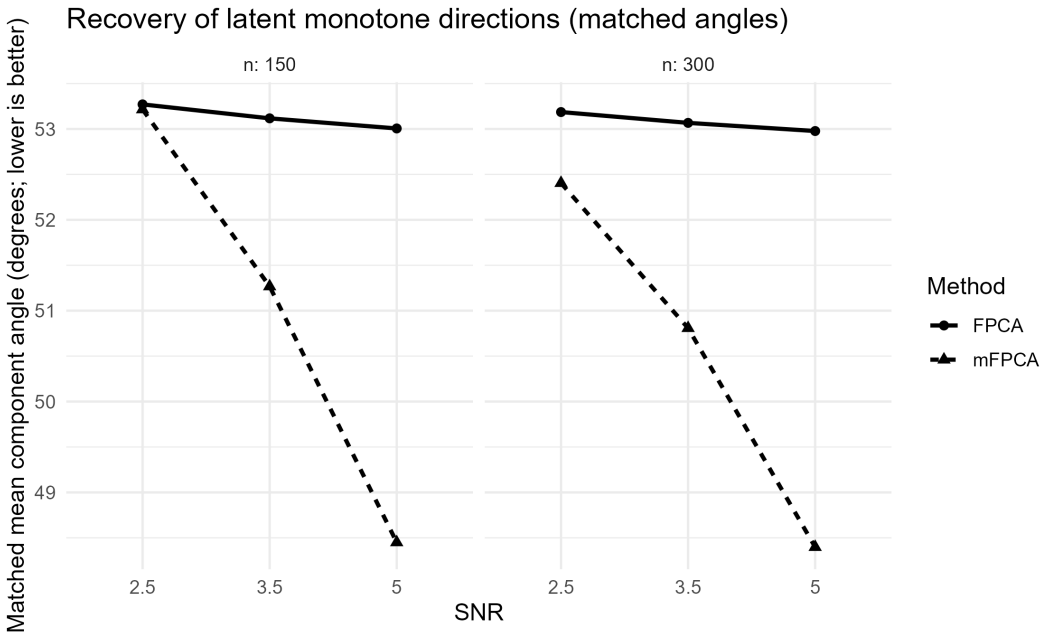


Figure 4. Recovery of latent monotone directions under varying sample size and signal-to-noise ratio. Lower matched mean angles indicate closer alignment with the latent monotone structure.

In the initialization-stability experiment, repeated multi-start estimation yields virtually identical fitted directions in the baseline design. This is consistent with the empirical application, where the initialization-stability summary also collapses to zero angle variation. The interpretation should nevertheless remain practical rather than absolute: The simulation indicates that, in this design, the implemented monotone projection-and-deflation algorithm is numerically stable across starts.

Table 11 examines validation optimism under temporal dependence, and Figure 5 provides the corresponding graphical comparison. Random K-fold validation remains optimistic relative to blocked validation in all three scenarios, but the magnitude is modest in the final Markov-regime design. The average inflation is about 0.011 in the moderate-shift case, 0.012 in the stationary case, and 0.007 in the strong-shift case. Thus, the revised simulation no longer supports a claim of dramatic optimism inflation. What it does support is the more cautious statement that random reshuffling still tends to overstate performance when the data are temporally dependent, even if the effect size is not large in the final design.

Table 11. Leakage study comparing random and blocked K-fold validation under stationary and shifted Markov-regime DGPs.

Scenario	Accuracy (Random)	Accuracy (Blocked)	Δ
Moderate shift	0.720	0.709	0.011
Stationary	0.736	0.724	0.012
Strong shift	0.745	0.738	0.007

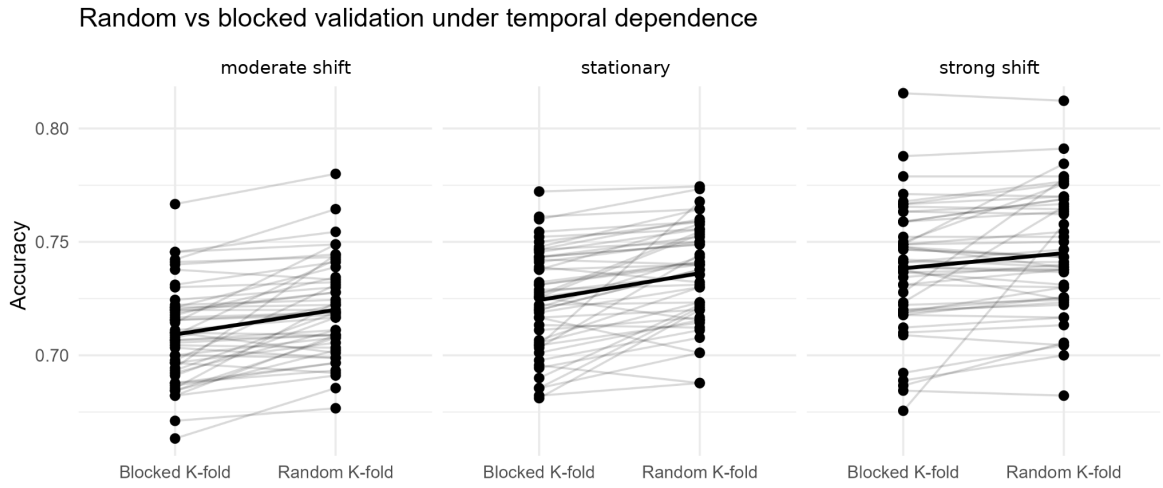


Figure 5. Random versus blocked validation under temporally dependent Markov-regime simulations.

The final simulation exercise compares mFPCA and FPCA as inputs to the same classifier under both blocked and rolling evaluations. Table 12 shows that the performance differences are generally positive but small, while Figure 6 displays the corresponding mFPCA-minus-FPCA differences. Under moderate shift and strong shift, mFPCA has slight advantages in both Accuracy and Macro-F1, especially under rolling evaluation. Under the stationary blocked scenario, however, FPCA is marginally better. Thus, the simulation evidence does not support a universal dominance claim. Instead, it suggests that the gains from monotone-constrained extraction are limited, scenario-dependent, and more naturally interpreted as improvements in structural regularization than as large standalone predictive gains.

Table 12. Classification comparison of monotone-deflation mFPCA and unconstrained FPCA under stationary and shifted Markov-regime DGPs.

Scenario	CV	Acc. (mFPCA)	Acc. (FPCA)	F1 (mFPCA)	F1 (FPCA)
Moderate shift	Blocked	0.757	0.756	0.675	0.675
Moderate shift	Rolling	0.791	0.786	0.719	0.707
Stationary	Blocked	0.749	0.752	0.682	0.686
Stationary	Rolling	0.762	0.763	0.682	0.681
Strong shift	Blocked	0.766	0.763	0.583	0.582
Strong shift	Rolling	0.828	0.825	0.700	0.690

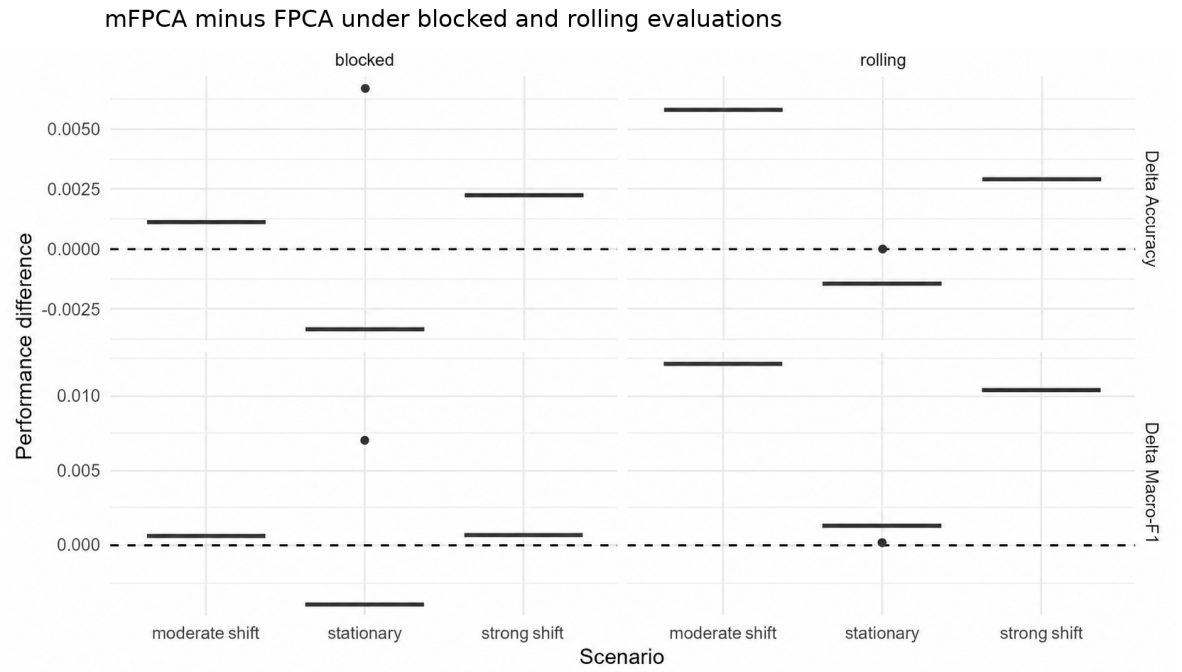


Figure 6. mFPCA minus FPCA under blocked and rolling evaluations in the Markov-regime simulation.

Overall, the simulation evidence supports three restrained conclusions. First, monotone-constrained extraction is more closely aligned with latent monotone structure than unconstrained FPCA when judged by matched component angles. Second, random K-fold validation remains mildly optimistic under temporal dependence, which justifies blocked or rolling evaluation even when the inflation is not dramatic. Third, the predictive gains from mFPCA over FPCA are mostly positive but small and scenario-dependent, which is consistent with the empirical application. In short, the value of the monotone-constrained representation lies primarily in interpretability, numerical stability, and evaluation discipline, rather than in large universal gains in raw predictive accuracy.

6. Conclusions and recommendations

In this paper, I examined whether a monotone-constrained functional representation could improve regime classification for forward risk-accumulation curves under a time-ordered empirical design. Using AAPL as the empirical focus and a supportive tech-growth panel for descriptive context, I compared a monotone projection-and-deflation representation against an unconstrained FPCA benchmark under identical classifiers and validation protocols. The empirical results showed that the preferred specification is the root-cumulative representation with $K = 3$ retained components and no cone-depth augmentation. Under this design, the monotone-constrained representation improved on FPCA under both blocked and rolling evaluations, while preserving visually monotone component directions and a highly concentrated low-dimensional summary of the forward curve.

Several empirical findings are especially important. First, the root-cumulative representation clearly outperformed the cumulative alternative, showing that the representation choice itself materially affects downstream classification. Second, the gains from monotone-constrained extraction were positive

but not dramatic, which suggests that the main contribution lies in representational discipline and interpretability rather than in large improvements in raw predictive accuracy. Third, the auxiliary cone-depth feature did not improve performance in the final specification and therefore should be viewed as a context-dependent extension rather than as a universally beneficial ingredient. Fourth, predictive information was concentrated primarily in the early-to-middle portion of the forward curve, rather than at the far end of the 60-day horizon. Finally, the blocked and rolling evaluations jointly confirmed that the middle regime remains the hardest to separate, whereas the calm and stress regimes are more reliably identified.

The simulation evidence complements these conclusions in a restrained but useful way. In controlled monotone-signal settings, the proposed extraction procedure produced directions that were more closely aligned with the latent monotone structure than those obtained from unconstrained FPCA when judged by matched component angles. However, the absolute angle levels remained nontrivial, so the appropriate interpretation is relative improvement rather than near-exact recovery. In the validation experiment, random K-fold remained optimistic relative to blocked validation under temporally dependent Markov-regime designs, but the size of this optimism was modest in the final simulation setup. Likewise, classification gains of mFPCA over FPCA were mostly positive but small and scenario-dependent. Taken together, these results reinforce the empirical message of the paper: The main value of the monotone-constrained representation lies in interpretability, numerical stability, and evaluation discipline, rather than in universal predictive dominance.

From a practical standpoint, several recommendations emerge. First, when the target object is structurally monotone by construction, representation learning should respect that structure directly rather than rely entirely on unconstrained variance maximization. Second, all representational objects, including centering, component extraction, scaling, thresholding, and tuning, should be estimated on the training slice only and transferred unchanged to the test slice. Third, for temporally ordered financial classification tasks, blocked or rolling evaluations should be preferred over random reshuffling, even when the resulting optimism gap is only moderate. Fourth, in imbalanced regime problems, Accuracy should be reported together with Macro-F1, PR curves, and class-wise AUPRC. Fifth, empirical model building should treat design choices such as the curve representation, the number of retained components, and auxiliary features as first-order decisions rather than as secondary implementation details.

The study also has limitations. Empirically, the analysis is centered on AAPL, so the magnitude of the gains should not be extrapolated mechanically to other assets, markets, or sampling frequencies. Methodologically, the implemented monotone projection-and-deflation procedure is designed to produce ordered monotone component directions, not an exact orthonormal eigenbasis in the strict classical FPCA sense. The classifier is intentionally simple and transparent; other learners may interact differently with the same representation. In addition, the curve construction is based on daily returns and simple cumulative transformations, so alternative realized-variation measures or alternative regime-label definitions may lead to different levels of separability.

These limitations point to several directions for future research. One natural extension is to study richer shape-constrained representations, including combinations of monotonicity with convexity, smoothness, or sparsity. A second is to generalize the framework to multivariate or multi-asset functional settings in which shared or asset-specific monotone representations are learned jointly. A third is to investigate more flexible classifiers while preserving the same training-slice-only protocol. A fourth is to develop stronger theory for monotone projection-and-deflation procedures under dependence, including

stability, rates, and inferential properties. Finally, external validity should be assessed across other equities, asset classes, and higher-frequency environments.

Overall, the paper provides evidence that disciplined representation design and disciplined temporal evaluation must be treated jointly. In the empirical application studied here, a monotone-constrained root-cumulative representation provides a more interpretable and slightly more effective low-dimensional summary of forward risk accumulation than unconstrained FPCA, while a training-slice-only protocol ensures that the reported gains remain empirically credible.

Author contributions

Çağlar Sözen: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing—Original Draft, Writing—Review & Editing, Visualization. The author has approved the final manuscript for publication.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The author declares no competing interests.

Declarations

- Ethics approval and consent to participate: Not applicable, as this study did not involve human participants or animals.
- Data availability: Data will be made available on request.
- Materials availability: Not applicable, as the study did not involve physical materials.

References

Andersen TG, Bollerslev T, Diebold FX, et al. (2003) Modeling and forecasting realized volatility. *Econometrica* 71: 579–625. https://doi.org/10.1111/1468-0262.00418

Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. *Stat Surv* 4: 40–79. https://doi.org/10.1214/09-SS054

Barlow RE, Bartholomew DJ, Bremner JM, et al. (1972) Statistical inference under order restrictions. Wiley, New York.

Barndorff-Nielsen OE, Shephard N (2002) Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *J R Stat Soc B* 64: 253–280. https://doi.org/10.1111/1467-9868.00336

Bergmeir C, Benítez JM (2012) On the use of cross-validation for time series predictor evaluation. *Inf Sci* 191: 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Cerqueira V, Torgo L, Mozetič I (2020) Evaluating time series forecasting models: an empirical study on cross-validation and windows. *Int J Forecast* 36: 30–44. https://doi.org/10.1007/s10994-020-05910-7

Chiou JM (2012) Dynamical functional prediction and classification, with application to traffic flow prediction. *Ann Appl Stat* 6: 1588–1614. https://doi.org/10.1214/12-AOAS595

Corsi F (2009) A simple approximate long-memory model of realized volatility. *J Financ Econometrics* 7: 174–196. https://doi.org/10.1093/jjfinec/nbp001

Cuevas A, Febrero M, Fraiman R (2007) Robust estimation and classification for functional data via projection-based depth notions. *Comput Stat* 22: 481–496. https://doi.org/10.1007/s00180-007-0053-0

Davis J, Goadrich M (2006) The relationship between precision–recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, 233–240. https://doi.org/10.1145/1143844.1143874

Hansen PR, Huang Z, Shek HH (2012) Realized GARCH: a joint model for returns and realized measures of volatility. *J Appl Econometrics* 27: 877–906. https://doi.org/10.1002/jae.1234

He H, Garcia EA (2009) Learning from imbalanced data. *IEEE Trans Knowl Data Eng* 21: 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Horváth L, Kokoszka P (2012) *Inference for functional data with applications*. Springer, New York.

Hsing T, Eubank RL (2015) *Theoretical foundations of functional data analysis, with an introduction to linear operators.* Wiley, Chichester.

James GM, Hastie TJ, Sugar CA (2000) Principal component models for sparse functional data. *Biometrika* 87: 587–602. https://doi.org/10.1093/biomet/87.3.587

López-Pintado S, Romo J (2009) On the concept of depth for functional data. *J Amer Stat Assoc* 104: 718–734.

Pya N, Wood SN (2015) Shape constrained additive models. *Stat Comput* 25: 543–559. https://doi.org/10.1007/s11222-013-9448-7

Ramsay JO (1988) Monotone regression splines in action. *Stat Sci* 3: 425–441. https://doi.org/10.1214/ss/1177012761

Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd ed. Springer, New York.

Rice J, Wu CO (2001) Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* 57: 253–259. https://doi.org/10.1111/j.0006-341X.2001.00253.x

Roberts DR, Bahn V, Ciuti S, et al. (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40: 913–929. https://doi.org/10.1111/ecog.02881

Saito T, Rehmsmeier M (2015) The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE* 10: e0118432. https://doi.org/10.1371/journal.pone.0118432

Sun Y, Genton MG (2011) Functional boxplots. *J Comput Graph Stat* 20: 316–334. https://doi.org/10.1198/jcgs.2011.09224

Tashman LJ (2000) Out-of-sample tests of forecasting accuracy: an analysis and review. *Int J Forecast* 16: 437–450. https://doi.org/10.1016/S0169-2070(00)00065-0

Yahoo Finance (2025) Historical data for equities and ETFs. Available from: https://finance.yahoo.com/.

Robertson T, Wright FT, Dykstra RL (1988) Order restricted statistical inference. Wiley, New York.

Zuo Y, Serfling R (2000) General notions of statistical depth function. *Ann Stat* 28: 461–482. https://doi.org/10.1214/aos/1016218226

Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33: 1–22. https://doi.org/10.18637/jss.v033.i01

AIMS Press