

Adaptive Functional Principal Component Forecasting of Cross-Sectional ETF Return Distributions in Wasserstein Space

Çağlar Sözen

caglar.sozen@giresun.edu.tr

Giresun University

Mervener SÖZEN

Ondokuz Mayıs University

Research Article

Keywords: functional time series, distribution forecasting, quantile functions, Wasserstein distance, functional principal component analysis, adaptive forecasting, forecast evaluation

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9813865/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Adaptive Functional Principal Component Forecasting of Cross-Sectional ETF Return Distributions in Wasserstein Space

Çağlar Sözen^{a,*}, Mervener Sözen^b

^a*Department of Finance and Banking, Giresun University, Giresun, Turkey*

^b*Department of Statistics, Ondokuz Mayıs University, Samsun, Turkey*

Abstract

This paper studies one-step-ahead forecasting of daily cross-sectional ETF return distributions represented by quantile functions. The forecasting problem is formulated in one-dimensional Wasserstein space, where distributional accuracy is evaluated through distances between quantile functions. We propose an adaptive functional principal component forecasting framework centered on a strong FPCA autoregressive anchor. Two controlled refinements are examined: FPCA-QC, which adds a selected-quantile residual correction, and FPCA-GSD, which combines the FPCA anchor with global score dynamics through train-only state-adaptive weights. FPCA bases, score models, selected-quantile regressions, and score-dynamics components are re-estimated using only information available at each forecast origin, while state thresholds and adaptive weights are learned from the training period and kept fixed out-of-sample. Empirically, FPCA-GSD performs best in the main holdout, whereas FPCA-QC delivers the strongest rolling stability. The gains are modest relative to FPCA-AR but clearer against naive and unstable alternatives.

Keywords: functional time series, distribution forecasting, quantile functions, Wasserstein distance, functional principal component analysis, adaptive forecasting, forecast evaluation

*Corresponding author: Çağlar Sözen. E-mail: caglar.sozen@giresun.edu.tr
Email address: caglar.sozen@giresun.edu.tr (Çağlar Sözen)
URL: <https://orcid.org/0000-0002-3732-5058> (Çağlar Sözen),
<https://orcid.org/0000-0001-5603-5382> (Mervener Sözen)

1. Introduction

Forecasting financial returns is often approached through scalar objects such as conditional means, volatilities, downside-risk measures, or selected quantiles. These summaries are useful, but they necessarily compress the distributional information available in a cross-section of assets. For many portfolio, risk-management, and market-monitoring problems, the shape of the cross-sectional return distribution is itself informative. Its location, dispersion, asymmetry, and tail behavior can change rapidly across market conditions, and a scalar target may not reveal how these changes occur across the entire distribution.

This paper studies one-step-ahead forecasting of daily cross-sectional exchange-traded fund (ETF) return distributions. On each trading day, the available ETF returns are represented by an empirical quantile function observed on a fixed probability grid. This transforms the forecasting target into a functional object. The approach is natural because, for one-dimensional distributions, Wasserstein distances can be expressed through distances between quantile functions (Villani, 2009; Panaretos and Zemel, 2019). The resulting evaluation criterion therefore assesses the full predicted distribution rather than an isolated point summary.

The methodological starting point is a smooth functional anchor: functional principal component analysis autoregression (FPCA-AR). Functional principal component analysis (FPCA) has become a standard tool for functional time-series forecasting because it represents smooth curves through a low-dimensional set of scores that can be forecast by time-series models (Ramsay and Silverman, 2005; Hyndman and Ullah, 2007; Hyndman and Shang, 2009; Aue et al., 2015). In the present setting, FPCA-AR is not treated as a weak benchmark. It is the central baseline around which the adaptive specifications are constructed.

The paper introduces two adaptive refinements of this anchor. The first, FPCA with selected-quantile residual correction (FPCA-QC), uses selected-quantile autoregressive forecasts to construct a sparse residual correction relative to the FPCA forecast. This correction is interpolated across the full quantile grid and applied through a train-only state-adaptive weight. The second, FPCA with global score-dynamics combination (FPCA-GSD), combines the FPCA anchor with a global score-dynamics forecast, again

using train-only state-adaptive weights. Both refinements are intentionally restricted: they adjust the FPCA forecast locally or dynamically rather than replacing the functional representation.

A central design principle is leakage discipline. The target quantile function Q_t is constructed from the cross-sectional ETF returns on day t , so the forecast of Q_t must be formed before observing those returns. The dynamic forecasting components are therefore estimated using only information available at the forecast origin. This includes re-estimating the FPCA basis, mean function, score models, and selected-quantile regressions from the available history. State thresholds and adaptive weights are learned from the training portion of each evaluation scheme and then held fixed during the corresponding out-of-sample period. The same chronological convention is imposed in the main expanding holdout and in all rolling evaluation splits.

The empirical analysis uses a broad ETF panel covering U.S. equity, sector, fixed-income, commodity, real-estate, and international exposures. The main evaluation criterion is squared Wasserstein loss, supplemented by weighted left-tail Wasserstein loss, quantile mean absolute error, and errors in distributional summaries such as the 5 percent quantile, interquartile range, and 90 percent spread. Formal comparisons are based on Diebold–Mariano tests with Newey–West standard errors, lag sensitivity, and multiple-testing adjustments (Diebold and Mariano, 1995; Newey and West, 1987; Holm, 1979; Benjamini and Hochberg, 1995).

The results are interpreted in a deliberately measured way. The adaptive FPCA family delivers modest but systematic improvements, rather than a dramatic separation from the FPCA anchor. In the main holdout, FPCA-GSD achieves the lowest average squared Wasserstein loss. In rolling evaluation, FPCA-QC provides the strongest stability, with the largest number of rolling wins and top-three placements. At the same time, the formal Diebold–Mariano evidence does not support a claim of decisive superiority over FPCA-AR. The clearest empirical separation is instead between the FPCA-centered family and weaker or less stable alternatives such as naive persistence and transport-based benchmarks.

This paper makes three contributions. First, it formulates daily ETF panel forecasting as a distribution-valued problem by targeting the full cross-sectional return quantile function. Second, it proposes an adaptive FPCA-centered forecast family in which selected-quantile residual information and score-dynamics information are incorporated as controlled refinements of a strong functional anchor. Third, it evaluates the models un-

der a leakage-disciplined empirical design with both expanding and rolling out-of-sample schemes, Wasserstein-based loss functions, and conservative predictive-accuracy comparisons.

The contribution is therefore incremental but meaningful. The paper does not claim that adaptive FPCA forecasting uniformly dominates existing distributional time-series methods. Rather, it shows that carefully designed adaptive corrections can improve and stabilize a strong FPCA forecast for cross-sectional ETF return quantile functions, while preserving transparent forecast timing and avoiding test-period information leakage.

The rest of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 presents the methodological framework. Section 4 describes the data and empirical design. Section 5 reports the main results. Section 6 discusses their implications, and Section 7 concludes.

2. Related Literature

2.1. Functional time-series forecasting

This paper is first related to the literature on functional data analysis and functional time-series forecasting. Functional data analysis treats each observation as a curve or function rather than as a finite-dimensional vector. The foundational work of Ramsay and Silverman (2005) provides the statistical basis for representing, smoothing, and analyzing such functional observations. For dependent functional data, Bosq (2000) develops linear processes in function spaces, while Horváth and Kokoszka (2012) provide a broad treatment of inference for functional data and functional time series.

A central tool in this literature is FPCA. It provides a low-dimensional representation of functional variation by decomposing curves into orthogonal basis functions and associated scores. Hyndman and Ullah (2007) use a functional data approach for forecasting mortality and fertility rates, combining smooth functional representations with time-series forecasts of the resulting scores. Hyndman and Shang (2009) extend functional time-series forecasting with weighted FPCA and related methods, emphasizing the usefulness of functional representations when the object of interest is an entire curve rather than a scalar. Aue et al. (2015) further show how stationary functional time series can be predicted through finite-dimensional score representations and multivariate time-series techniques.

The present paper builds on this forecasting logic. The daily cross-sectional ETF return distribution is represented as a quantile function, and FPCA is

used to obtain a smooth low-dimensional anchor forecast. The distinctive feature is that FPCA is not used merely as one benchmark among many. Instead, it is treated as the central forecasting anchor, and the adaptive models are designed as restricted corrections around it.

2.2. Quantile functions, Wasserstein geometry, and distributional time series

The second strand of literature concerns distributional data analysis and Wasserstein geometry. For one-dimensional probability distributions, quantile functions provide a convenient representation because the squared 2-Wasserstein distance can be written as an L^2 distance between quantile functions. This connection follows from the optimal transport framework developed in Villani (2009) and has become central in statistical work on distribution-valued data. Panaretos and Zemel (2019) review statistical aspects of Wasserstein distances and emphasize their usefulness for comparing probability distributions as complex objects.

Recent statistical work has extended regression and time-series methods to non-Euclidean and distribution-valued responses. Petersen and Müller (2019) develop Fréchet regression for random objects in metric spaces, including probability distributions. Chen et al. (2023) propose Wasserstein regression for univariate probability measures, providing a framework for regression in the space of distributions. These contributions highlight that distributional objects require methods that respect their geometry rather than treating them as ordinary Euclidean vectors.

The paper is also related to distributional time-series models. Zhang et al. (2022) propose Wasserstein autoregressive models for density time series and explicitly motivate their framework with time-indexed distributions such as cross-sectional or intraday financial return distributions. Zhu and Müller (2023) develop autoregressive optimal transport models that operate on transport maps and provide an intrinsic time-series approach in Wasserstein space. These papers show that forecasting sequences of distributions is an active research area and that Wasserstein geometry is a natural language for such problems.

Relative to this literature, the present paper adopts a practical forecasting perspective for daily ETF panel distributions. It uses Wasserstein loss for evaluation and includes Wasserstein and transport autoregressive models as functional distributional alternatives. Its main contribution, however, is not to replace existing Wasserstein time-series models. Rather, it proposes a

leakage-disciplined adaptive FPCA family for cross-sectional return quantile functions and evaluates whether selected-quantile and score-dynamics corrections improve a strong FPCA anchor.

2.3. Cross-sectional ETF panels as distribution-valued forecasting targets

The paper also connects to empirical finance work that uses cross-sectional information to summarize market conditions. Cross-sectional return dispersion has long been used as an informative market-wide object because it reflects heterogeneity across assets and can be interpreted as a summary of common conditions, idiosyncratic risk, and changes in cross-sectional comovement (Solnik and Roulet, 2000; Goyal and Santa-Clara, 2003; Verousis and Voukelatos, 2018). A cross-section of asset returns therefore contains information about common shocks, dispersion, sector rotation, and tail behavior that may not be visible in a single index return.

ETF panels provide a natural empirical setting for this perspective. ETFs offer liquid and standardized exposure to broad market, sector, fixed-income, commodity, real-estate, and international asset classes, and the ETF literature emphasizes their role in price discovery, comovement, and market-wide information transmission (Elton et al., 2002; Ben-David et al., 2018). Observed on a common daily frequency, an ETF panel gives a compact cross-sectional representation of market conditions across multiple asset segments.

Most forecasting exercises in empirical finance reduce this information to scalar summaries, such as the cross-sectional mean, dispersion, realized volatility, or selected quantiles. Such summaries are informative but partial. By contrast, the present paper treats the entire daily cross-sectional return distribution as the object to be forecast. This distribution-valued target allows the evaluation to account jointly for location, dispersion, asymmetry, and tail behavior. The ETF setting therefore provides a natural empirical environment for assessing whether functional distributional forecasts can add value beyond scalar or selected-quantile representations.

2.4. Quantile forecasting, probabilistic evaluation, and forecast comparison

A further relevant literature studies quantile forecasting, probabilistic forecast evaluation, and predictive-accuracy comparison. Quantile methods have a long history in econometrics, beginning with the regression quantile framework of Koenker and Bassett (1978). In financial applications, quantiles are often used to summarize downside risk, but a small number of quantiles cannot fully describe the shape of a distribution. The present paper therefore

uses selected-quantile autoregressions as an interpretable benchmark and as a source of residual correction, while the main target remains the full quantile function.

The evaluation of distributional forecasts is closely related to the broader probabilistic forecasting literature. [Gneiting and Raftery \(2007\)](#) formalize proper scoring rules for probabilistic forecasts, emphasizing that forecast evaluation should depend on the predictive distribution and the realized outcome. [Gneiting and Katzfuss \(2014\)](#) define probabilistic forecasts as predictive probability distributions over future quantities or events and stress the role of calibration and sharpness. Although the present paper uses Wasserstein-based losses rather than a full scoring-rule framework, it follows the same principle that the object being evaluated should be the predicted distribution rather than only a point forecast.

The paper also connects to the literature on forecast comparison and out-of-sample evaluation. [Diebold and Mariano \(1995\)](#) introduce a widely used test for comparing predictive accuracy between competing forecasts. [Newey and West \(1987\)](#) provide heteroskedasticity- and autocorrelation-consistent covariance estimation, which is useful when loss differentials are serially dependent. Since multiple pairwise comparisons can inflate false positives, the empirical analysis also reports Holm and Benjamini–Hochberg adjustments ([Holm, 1979](#); [Benjamini and Hochberg, 1995](#)). This conservative testing design is important for the interpretation of the results: lower average losses are reported separately from claims of statistically decisive superiority.

Overall, the related literature suggests a clear positioning for the present paper. Functional forecasting provides the FPCA anchor, Wasserstein geometry provides the distributional representation and loss, ETF panel data provide the empirical cross-section, and predictive-accuracy testing provides the evaluation discipline. The contribution lies at the intersection of these strands: a leakage-disciplined, FPCA-centered adaptive forecast family for daily cross-sectional ETF return quantile functions.

3. Methodology

This section develops the forecasting framework used to predict the daily cross-sectional distribution of ETF returns. The target is not a scalar return, volatility, or tail-risk measure, but the full cross-sectional return distribution represented by its quantile function. This places the problem in a one-dimensional Wasserstein setting, where distributional discrepancies can be

evaluated through distances between quantile functions.

The proposed framework is centered on a smooth functional benchmark, FPCA-AR. Two adaptive refinements are constructed around this anchor. The first, FPCA-QC, adds a selected-quantile residual correction to the FPCA forecast. The second, FPCA-GSD, combines the FPCA anchor with a global score-dynamics forecast. Both specifications are deliberately restricted: they adjust a strong FPCA forecast rather than replacing it with an unrestricted adaptive model.

The timing convention is strictly chronological. The forecast of Q_t is formed at the end of day $t - 1$, before the returns on day t are observed. Dynamic forecasting components such as the mean function, FPCA basis, score models, selected-quantile regressions, and global score-dynamics model are estimated using only the historical information available at the forecast origin. State thresholds and adaptive weights are selected using only the training portion of each evaluation scheme and are then kept fixed during the corresponding out-of-sample period. Thus, same-day and future observations affect neither the forecasts nor the state cutoffs or adaptive weights.

3.1. Forecast target and Wasserstein quantile representation

Let $r_{i,t}$ denote the close-to-close log return of ETF i on trading day t , where $i = 1, \dots, N_t$ indexes the ETFs available on that day. The cross-sectional empirical distribution of returns is denoted by F_t . The target object is the corresponding empirical quantile function

$$Q_t(u) = F_t^{-1}(u), \quad u \in (0, 1). \quad (1)$$

In the implementation, Q_t is evaluated on the fixed grid

$$\mathcal{U} = \{0.01, 0.02, \dots, 0.99\}, \quad (2)$$

so that the observed functional object is represented by

$$\mathbf{q}_t = \left(Q_t(u_1), \dots, Q_t(u_m) \right)^\top, \quad m = 99. \quad (3)$$

For one-dimensional distributions, the squared 2-Wasserstein distance can be written as the squared L^2 distance between quantile functions (Villani, 2009; Panaretos and Zemel, 2019; Petersen and Müller, 2019). The primary loss used in the paper is therefore the discrete approximation

$$L_{W_2}(\hat{Q}_t, Q_t) = \frac{1}{m} \sum_{j=1}^m \left\{ \hat{Q}_t(u_j) - Q_t(u_j) \right\}^2. \quad (4)$$

This criterion evaluates the full predicted distribution rather than only a conditional mean, variance, or single tail quantile.

The one-step-ahead information set is

$$\mathcal{H}_{t-1} = \{Q_1, \dots, Q_{t-1}\}. \quad (5)$$

The forecast $\hat{Q}_t$ is formed using only $\mathcal{H}_{t-1}$. No return observation from date t enters the predictor set for $\hat{Q}_t$. This convention is imposed in the main expanding holdout and in every rolling split.

3.2. FPCA autoregressive anchor

The central functional anchor is an FPCA autoregressive model. FPCA provides a low-dimensional representation of smooth functional variation and is widely used in functional data analysis and functional time-series forecasting (Ramsay and Silverman, 2005; Bosq, 2000; Hyndman and Ullah, 2007; Horváth and Kokoszka, 2012). At forecast origin $t - 1$, let $\mathcal{I}_{t-1}$ denote the historical index set available for estimation. This set is expanding in the main holdout design and rolling in the rolling backtests. The historical quantile functions are centered as

$$X_s(u) = Q_s(u) - \bar{Q}_{t-1}(u), \quad s \in \mathcal{I}_{t-1}, \quad (6)$$

where

$$\bar{Q}_{t-1}(u) = \frac{1}{|\mathcal{I}_{t-1}|} \sum_{s \in \mathcal{I}_{t-1}} Q_s(u). \quad (7)$$

The centered functions are approximated by

$$X_s(u) \approx \sum_{k=1}^K \xi_{s,k} \phi_k(u), \quad (8)$$

where $\phi_k(u)$ are empirical functional principal components and $\xi_{s,k}$ are the corresponding scores.

The number of retained components is selected at each forecast origin by a cumulative proportion-of-variance rule with target 0.95 and an upper cap of six components:

$$K = \min \left\{ k : \frac{\sum_{\ell=1}^k \lambda_\ell}{\sum_{\ell \geq 1} \lambda_\ell} \geq 0.95 \right\} \wedge 6. \quad (9)$$

For each retained score series, a univariate AR(1) model is fitted:

$$\xi_{s,k} = a_k + b_k \xi_{s-1,k} + \varepsilon_{s,k}. \quad (10)$$

Let $\hat{\xi}_{t,k}$ denote the resulting one-step-ahead score forecast. The raw FPCA-AR forecast is

$$\tilde{Q}_t^{\text{FPCA}}(u) = \bar{Q}_{t-1}(u) + \sum_{k=1}^K \hat{\xi}_{t,k} \phi_k(u). \quad (11)$$

Because a finite-grid reconstruction may produce minor monotonicity violations, the forecast vector is projected onto the monotone cone:

$$\hat{\mathbf{q}}_t^{\text{FPCA}} = \Pi_{\mathcal{M}}(\tilde{\mathbf{q}}_t^{\text{FPCA}}), \quad \mathcal{M} = \{\mathbf{x} \in \mathbb{R}^m : x_1 \leq x_2 \leq \cdots \leq x_m\}. \quad (12)$$

The projection is applied after forecasting and uses no future information. It is therefore a numerical admissibility adjustment rather than a forecasting input.

The FPCA-AR model is deliberately treated as a demanding functional anchor. The adaptive variants are evaluated as refinements around this anchor.

3.3. Selected-quantile autoregressive benchmark

A direct quantile-level benchmark is also included. Let

$$\mathcal{U}_S = \{0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.99\} \quad (13)$$

denote the selected quantile grid. For each $v \in \mathcal{U}_S$, an AR(1) model is fitted:

$$Q_s(v) = \alpha_v + \beta_v Q_{s-1}(v) + e_{s,v}. \quad (14)$$

The selected forecasts are linearly interpolated over $\mathcal{U}$ and projected onto the monotone cone. This model is interpretable because it directly forecasts tail, median, and interquartile locations. However, it does not exploit the full smooth low-dimensional structure of the quantile curve.

3.4. FPCA-QC residual anchor

The first adaptive extension is the FPCA-QC residual anchor. Its purpose is to preserve the smooth global structure of FPCA-AR while allowing selected quantile dynamics to correct local deviations. The construction is residual-based and is not a simple level mixture.

Let $\hat{Q}_t^{\text{FPCA}}(u)$ denote the projected FPCA forecast and let $\hat{Q}_t^S(v)$ denote the selected-quantile AR forecast at $v \in \mathcal{U}_S$. The selected-quantile residual is

$$\hat{e}_t^S(v) = \hat{Q}_t^S(v) - \hat{Q}_t^{\text{FPCA}}(v), \quad v \in \mathcal{U}_S. \quad (15)$$

This sparse residual function is linearly interpolated over the full grid:

$$\hat{e}_t^S(u) = \mathcal{I} \left[\{\hat{e}_t^S(v) : v \in \mathcal{U}_S\} \right] (u), \quad u \in \mathcal{U}, \quad (16)$$

where $\mathcal{I}[\cdot]$ denotes linear interpolation.

The raw FPCA-QC forecast is

$$\tilde{Q}_t^{\text{QC}}(u) = \hat{Q}_t^{\text{FPCA}}(u) + \omega_{s(t)}^{\text{QC}} \hat{e}_t^S(u), \quad (17)$$

where $s(t) \in \{\text{Low}, \text{Medium}, \text{High}\}$ is the state assigned at the forecast origin and $\omega_{s(t)}^{\text{QC}} \in [0, 1]$ is the corresponding train-only state-specific residual weight. The final forecast is

$$\hat{\mathbf{q}}_t^{\text{QC}} = \Pi_{\mathcal{M}} \left(\tilde{\mathbf{q}}_t^{\text{QC}} \right). \quad (18)$$

When $\omega_{s(t)}^{\text{QC}} = 0$, the model reduces to FPCA-AR. Positive weights allow selected-quantile dynamics to adjust the anchor in a controlled way.

3.5. Global score dynamics and FPCA-GSD combination

The second adaptive extension combines the FPCA anchor with a global score-dynamics forecast. At each forecast origin, a recent window of up to 250 historical observations is used to estimate a local FPCA representation. Let $\boldsymbol{\xi}_s = (\xi_{s,1}, \dots, \xi_{s,K})^\top$ denote the score vector. The score-dynamics model uses lagged and smoothed score features:

$$\mathbf{z}_s = \left(\boldsymbol{\xi}_{s-1}^\top, \boldsymbol{\xi}_{s-2}^\top, \bar{\boldsymbol{\xi}}_{s-1}^{(5)\top}, \bar{\boldsymbol{\xi}}_{s-1}^{(22)\top} \right)^\top, \quad (19)$$

where $\bar{\boldsymbol{\xi}}_{s-1}^{(5)}$ and $\bar{\boldsymbol{\xi}}_{s-1}^{(22)}$ are trailing moving averages over 5 and 22 observations. The score vector is forecast using multivariate ridge regression:

$$\boldsymbol{\xi}_s = \mathbf{A} + \mathbf{B}\mathbf{z}_s + \boldsymbol{\eta}_s, \quad (20)$$

with ridge penalty $\lambda = 10^{-3}$ in the main implementation. The reconstructed global score-dynamics forecast is denoted by $\hat{Q}_t^{\text{GSD}}(u)$.

The FPCA-GSD combination moves the FPCA forecast toward the global score-dynamics forecast:

$$\tilde{Q}_t^{\text{GSDC}}(u) = \hat{Q}_t^{\text{FPCA}}(u) + \omega_{s(t)}^{\text{GSD}} \left[\hat{Q}_t^{\text{GSD}}(u) - \hat{Q}_t^{\text{FPCA}}(u) \right]. \quad (21)$$

The final forecast is projected onto the monotone cone:

$$\hat{\mathbf{q}}_t^{\text{GSDC}} = \Pi_{\mathcal{M}} \left(\tilde{\mathbf{q}}_t^{\text{GSDC}} \right). \quad (22)$$

Thus FPCA-GSD is not a stand-alone replacement for the FPCA representation. It is an adaptive combination that can improve the anchor when global score-dynamics information is useful.

3.6. Train-only state construction and adaptive weights

The adaptive weights in (17) and (21) are state-specific. The state assignment is based on two recent distributional signals: dispersion and tail asymmetry. For a quantile function Q_t , define

$$\text{IQR}_t = Q_t(0.75) - Q_t(0.25), \quad (23)$$

and

$$\text{ASYM}_t = \{Q_t(0.95) - Q_t(0.50)\} - \{Q_t(0.50) - Q_t(0.05)\}. \quad (24)$$

At forecast origin $t - 1$, recent signals are computed from the trailing five observations:

$$\overline{\text{IQR}}_{t-1}^{(5)} = \frac{1}{5} \sum_{\ell=1}^5 \text{IQR}_{t-\ell}, \quad \overline{|\text{ASYM}|}_{t-1}^{(5)} = \frac{1}{5} \sum_{\ell=1}^5 |\text{ASYM}_{t-\ell}|. \quad (25)$$

These signals are standardized using means and standard deviations computed only from the training portion of the relevant evaluation scheme:

$$Z_{t-1}^{\text{IQR}} = \frac{\overline{\text{IQR}}_{t-1}^{(5)} - \mu_{\text{IQR}}}{\sigma_{\text{IQR}}}, \quad Z_{t-1}^{\text{ASYM}} = \frac{\overline{|\text{ASYM}|}_{t-1}^{(5)} - \mu_{\text{ASYM}}}{\sigma_{\text{ASYM}}}. \quad (26)$$

The composite state score is

$$S_{t-1} = 0.5Z_{t-1}^{\text{IQR}} + 0.5Z_{t-1}^{\text{ASYM}}. \quad (27)$$

Low, Medium, and High states are obtained from the training-sample tertiles of S_{t-1} . These thresholds are fixed within each evaluation scheme and are not

estimated from the test period. Thus, test-period observations affect neither the state cutoffs nor the state-specific adaptive weights.

For each scheme and adaptive model, global and state-specific weights are selected by a train-only pseudo-out-of-sample exercise. The candidate grid is

$$\Omega = \{0, 0.05, 0.10, \dots, 0.95, 1\}. \quad (28)$$

For each candidate ω , the corresponding adaptive forecast is evaluated over the last 80 train-only pseudo-out-of-sample observations, subject to a minimum history requirement of 80 observations. The raw state-specific weight minimizes the mean pseudo-out-of-sample Wasserstein loss within that state.

Because state-specific samples can be small, the raw state weight is shrunk toward the global weight:

$$\omega_s = \alpha_s \omega_s^{\text{raw}} + (1 - \alpha_s) \omega^{\text{global}}, \quad \alpha_s = \frac{n_s}{n_s + \kappa}, \quad (29)$$

where n_s is the number of train-only pseudo-out-of-sample observations in state s and $\kappa = 20$. This shrinkage prevents the adaptive layer from becoming an unstable state-specific refit and keeps the forecast family anchored to FPCA-AR.

3.7. Benchmark models

The benchmark set includes both simple distributional baselines and structured functional alternatives. The naive benchmark uses the most recent observed quantile function:

$$\hat{Q}_t^{\text{Naive}}(u) = Q_{t-1}(u). \quad (30)$$

The rolling barycenter benchmark averages recent historical quantile functions:

$$\hat{Q}_t^{\text{Bar}}(u) = \frac{1}{L} \sum_{\ell=1}^L Q_{t-\ell}(u), \quad L = 22. \quad (31)$$

The remaining competitors are selected-quantile AR, Wasserstein AR, transport AR, global score dynamics, and the FPCA-AR anchor. When reported as a standalone benchmark, global score dynamics is evaluated without FPCA anchoring. When used inside FPCA-GSD, it enters only through the adaptive combination with the FPCA forecast.

3.8. Evaluation design and statistical comparison

The primary ranking metric is the squared Wasserstein loss in (4). We also report its square root, a weighted left-tail Wasserstein loss, quantile mean absolute error, and summary errors for the left-tail 5 percent quantile, interquartile range, and 90 percent spread. The weighted loss is

$$L_{\text{w}W_2}(\hat{Q}_t, Q_t) = \sum_{j=1}^m w_j \left\{ \hat{Q}_t(u_j) - Q_t(u_j) \right\}^2, \quad \sum_{j=1}^m w_j = 1, \quad (32)$$

where the weights place greater emphasis on the left tail.

The empirical design has two complementary components. The main holdout uses an expanding recursive evaluation after an initial training fraction of 0.60. The rolling design uses 1000 training observations and 120 test observations, with the window advanced by 120 observations across splits. In both designs, the dynamic forecasting components are recomputed using only information available at the forecast origin, while state thresholds and adaptive weights are learned from the training portion of the corresponding scheme and held fixed during its out-of-sample evaluation.

Pairwise predictive-accuracy comparisons are based on Diebold–Mariano tests (Diebold and Mariano, 1995). For model a and benchmark b , define

$$d_t = L(\hat{Q}_t^a, Q_t) - L(\hat{Q}_t^b, Q_t). \quad (33)$$

Negative mean differences favor model a . The long-run variance of the loss differential is estimated with Newey–West HAC standard errors (Newey and West, 1987). Sensitivity is reported for Newey–West lag values 3, 5, and 10. Multiplicity-adjusted p -values are computed using Holm and Benjamini–Hochberg corrections (Holm, 1979; Benjamini and Hochberg, 1995).

The interpretation of the statistical tests is intentionally cautious. The adaptive FPCA variants may deliver lower average losses and better rolling stability without producing decisive superiority over FPCA-AR in every pairwise comparison. The main empirical claim is therefore not uniform dominance, but measured adaptive improvement around a strong functional baseline, together with clear gains over weaker and unstable alternatives.

3.9. Algorithmic summary

Algorithm 1 summarizes the one-step-ahead forecasting procedure.

Algorithm 1 Leakage-disciplined adaptive FPCA distribution forecast

Require: Historical quantile functions $\{Q_1, \dots, Q_{t-1}\}$, full grid $\mathcal{U}$, selected grid $\mathcal{U}_S$

Ensure: One-step-ahead forecasts $\hat{Q}_t^{\text{FPCA}}$, $\hat{Q}_t^{\text{QC}}$, and $\hat{Q}_t^{\text{GSDC}}$

- 1: Select the available historical index set $\mathcal{I}_{t-1}$ according to the expanding or rolling design.
 - 2: Estimate the mean quantile function and FPCA basis using only $\{Q_s : s \in \mathcal{I}_{t-1}\}$.
 - 3: Choose the number of FPCA components using the 0.95 PVE rule with an upper cap of six.
 - 4: Fit AR(1) models to the retained FPCA scores and construct the FPCA-AR forecast.
 - 5: Fit selected-quantile AR models on $\mathcal{U}_S$ and compute residuals relative to the FPCA forecast.
 - 6: Estimate the global score-dynamics model using the recent historical score window and ridge regression.
 - 7: Compute recent IQR and tail-asymmetry signals using only lagged quantile functions.
 - 8: Assign the current Low, Medium, or High state using thresholds learned from the training portion of the scheme.
 - 9: Retrieve the corresponding shrunk state-specific weights for FPCA-QC and FPCA-GSD.
 - 10: Construct the FPCA-QC residual forecast.
 - 11: Construct the FPCA-GSD combination forecast.
 - 12: Project all reconstructed quantile functions onto the monotone cone.
 - 13: Evaluate the forecasts only after Q_t is observed.
-

4. Empirical Design and Data

This section describes the data, the out-of-sample evaluation design, and the benchmark structure. The empirical design follows the timing convention introduced in Section 3: the cross-sectional quantile function Q_t is forecast at the end of day $t - 1$ using only quantile functions observed up to $t - 1$. Dynamic forecasting components are rebuilt from the information available at the forecast origin, while state thresholds and adaptive weights are selected within the training portion of each evaluation scheme.

4.1. Data

The empirical analysis uses daily closing prices downloaded from Yahoo Finance ([Yahoo Finance, 2026](#)). The requested sample period runs from January 2010 to December 2025. After aligning trading days and applying the minimum cross-sectional availability requirement, the effective sample contains 4022 daily cross-sectional quantile functions.

The ETF universe is designed to represent a broad cross-section of market exposures, including U.S. equity, sector, fixed-income, commodity, real-estate, and international funds. The panel contains SPY, QQQ, IWM, DIA, GLD, SLV, TLT, IEF, SHY, LQD, HYG, XLF, XLK, XLE, XLV, XLP, XLU, XLY, XLI, XLB, VNQ, EEM, and EFA. For compactness, ticker symbols are used throughout the remainder of the paper.

Daily close-to-close log returns are computed from the closing-price panel. On each trading day t , the available ETF returns form a cross-sectional return sample. The empirical quantile function of this cross-section is evaluated on the fixed grid $\mathcal{U} = \{0.01, 0.02, \dots, 0.99\}$. Days with fewer than 15 available ETF returns are removed. The resulting target is therefore a daily distribution-valued object rather than a scalar return, volatility, or tail-risk series. Because the empirical target is constructed from the closing-price panel used in the forecasting experiment, the results should be interpreted as forecasts of close-to-close ETF return cross-sections rather than dividend-adjusted total-return distributions.

Table 1 summarizes the sample, target construction, forecasting design, and main tuning parameters.

4.2. Out-of-sample evaluation design

The out-of-sample design has two complementary components. The main holdout uses an expanding recursive design. The first 60 percent of the effective sample is used as the initial training period, and one-step-ahead forecasts are generated recursively for the remaining observations. This design provides the primary long-horizon assessment of average out-of-sample performance.

The rolling evaluation uses 25 rolling splits. Each split uses 1000 historical observations for estimation and 120 observations for testing, and the rolling window is advanced by 120 observations at each split. This design examines whether relative model performance is stable across different market episodes, including calm periods, high-dispersion periods, and the stress episode around early 2020.

Table 1: Empirical design summary for the distribution-valued ETF panel forecasting experiment.

Design element	Specification
Sample window	2010-01-01 to 2025-12-31
Price field	Close
Effective sample size	4022 daily cross-sectional quantile functions
Panel size	23 ETFs
Minimum daily availability	15 ETFs
Forecast target	Daily cross-sectional ETF return quantile function Q_t
Forecast timing	Forecast Q_t at the end of day $t - 1$ using only $Q_1, \dots, Q_{t-1}$
Quantile grid	$u = 0.01, 0.02, \dots, 0.99$
Selected quantiles	0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.99
Main holdout	Expanding recursive one-step-ahead evaluation; initial train fraction = 0.60
Rolling evaluation	25 rolling splits; 1000 training observations, 120 test observations, step size = 120
Primary loss	Squared Wasserstein loss
Model family	FPCA-AR anchor; FPCA-QC anchor; FPCA-GSD combo
FPCA selection	PVE target = 0.95; maximum components = 6
Adaptive states	Low, Medium, High states based on recent IQR and absolute tail asymmetry
Adaptive weighting	Weight grid 0, 0.05, $\dots$, 0.95, 1; 80 train-only pseudo-OOS points
DM inference	Newey–West lags 3, 5, and 10

Notes: The ETF panel contains SPY, QQQ, IWM, DIA, GLD, SLV, TLT, IEF, SHY, LQD, HYG, XLF, XLK, XLE, XLV, XLP, XLU, XLY, XLI, XLB, VNQ, EEM, and EFA. Forecasts are generated under the chronological convention that Q_t is predicted before observing the ETF returns on date t .

For every forecast date in both designs, the dynamic forecasting components are estimated from the information set available at the forecast origin. In particular, the mean quantile function, FPCA basis, score dynamics, and selected-quantile regressions are rebuilt using only historical quantile functions. State thresholds, global adaptive weights, state-specific weights, and shrinkage weights are selected within the training portion of each evaluation scheme and are then held fixed during the corresponding out-of-sample evaluation. This distinction is important: the adaptive layer is trained without test-period information, while the curve-level forecasting components remain recursive

and date-specific. Since Q_t is constructed from ETF returns on date t , no return observation from date t enters the predictor set for $\hat{Q}_t$.

Predictive accuracy is evaluated primarily by squared Wasserstein loss. The reported supplementary metrics include the square-rooted Wasserstein loss, weighted left-tail Wasserstein loss, quantile mean absolute error, the absolute error of the 5 percent quantile, the absolute error of the interquartile range, and the absolute error of the 90 percent spread. Formal pairwise comparisons are based on Diebold–Mariano tests with Newey–West standard errors and lag sensitivity over 3, 5, and 10 lags.

4.3. Competing models and benchmark structure

The benchmark structure is organized around a demanding functional comparison set. The main anchor is FPCA-AR, which represents the full quantile curve through a low-dimensional FPCA basis and forecasts the retained scores autoregressively. The adaptive family contains two FPCA-centered extensions. FPCA-QC applies a selected-quantile residual correction to the FPCA forecast, whereas FPCA-GSD combines the FPCA forecast with a global score-dynamics forecast through train-only state-adaptive weights.

The remaining models provide alternative reference points. The naive benchmark uses the most recent observed quantile function. The rolling barycenter uses a recent average of historical quantile functions. The selected-quantile AR model forecasts a sparse set of quantile levels and interpolates the resulting curve. Wasserstein AR and transport AR provide additional functional distributional dynamics. Global score dynamics is included both as a standalone comparator and as the component used inside the FPCA-GSD combination.

This benchmark design avoids comparing the adaptive models only with weak persistence baselines. The proposed specifications are also evaluated against the FPCA-AR anchor and several functional distributional alternatives. The empirical question is therefore whether the adaptive FPCA family delivers measured improvements around a strong smooth functional baseline, rather than whether a single model uniformly dominates in every split and metric.

5. Results

This section reports the out-of-sample forecasting results. The presentation is descriptive first and interpretive second. The broader implications are discussed in Section 6. The central empirical pattern is that the leading

methods are concentrated around the FPCA representation: FPCA-GSD delivers the best main-holdout performance, FPCA-QC provides the strongest rolling stability, and FPCA-AR remains a highly competitive anchor. Expanded split diagnostics, adaptive-weight summaries, the full split-by-split scheme-wise loss table, and full-resolution diagnostic figures are reported in Appendices [Appendix A](#) and [Appendix B](#).

5.1. Main holdout results

The main holdout identifies FPCA-GSD combo as the best-performing specification under squared Wasserstein loss. Its average loss is 0.000109, compared with 0.000111 for FPCA-QC anchor and 0.000111 for FPCA-AR. The magnitude of the gain is modest, but it is directionally consistent with the role of FPCA-GSD as an adaptive score-dynamics correction around the FPCA anchor. The full split-by-split scheme-wise comparison is reported in Appendix Table [A4](#).

The detailed main-holdout ranking in Table [2](#) confirms this pattern. The three leading models are FPCA-GSD combo, FPCA-QC anchor, and FPCA-AR. Their losses are very close, indicating that the dominant empirical structure is the FPCA representation itself. The selected-quantile AR and rolling barycenter benchmarks are competitive but slightly weaker, whereas transport AR and the naive past-distribution benchmark are clearly less accurate. Figure [1](#) further shows that the adaptive gains should be interpreted dynamically. The adaptive variants improve on the FPCA anchor during parts of the holdout period, but the cumulative differences remain modest. This visual evidence supports the main interpretation: the proposed models are useful refinements of a strong FPCA benchmark rather than uniformly dominant replacements.

5.2. Rolling evaluation and stability

The rolling evaluation provides a stricter assessment of stability across market conditions. Table [3](#) summarizes rolling win counts, top-three counts, average ranks, and average losses across the 25 rolling splits. FPCA-QC anchor has the strongest rolling stability: it wins 10 rolling splits and appears in the top three in 21 out of 25 splits. FPCA-GSD combo wins 8 splits and appears in the top three in 17 splits, while FPCA-AR wins 4 splits and appears in the top three in 18 splits.

This pattern reinforces the main interpretation. FPCA-QC does not dominate the main holdout, but it is the most stable rolling specification.

Table 2: Main-holdout ranking under squared Wasserstein loss.

Rank	Model	Mean W_2^2	Mean W_2	Weighted W_2^2	Quantile MAE
1	FPCA-GSD combo	0.000109	0.007459	0.000131	0.006618
2	FPCA-QC anchor	0.000111	0.007489	0.000134	0.006616
3	FPCA-AR	0.000111	0.007493	0.000135	0.006621
4	Selected-quantile AR	0.000115	0.007573	0.000140	0.006641
5	Rolling barycenter	0.000115	0.007625	0.000140	0.006766
6	Wasserstein AR	0.000119	0.007714	0.000145	0.006855
7	Global score dynamics	0.000122	0.007799	0.000146	0.006974
8	Transport AR	0.000153	0.009167	0.000175	0.008143
9	Naive past distribution	0.000228	0.010510	0.000252	0.009401

Notes: The table reports average out-of-sample losses over the main holdout period. The ranking is based on mean squared Wasserstein loss (W_2^2). All models are evaluated on the same 1609 one-step-ahead forecasts. Weighted W_2^2 places greater emphasis on the left tail of the cross-sectional return distribution.

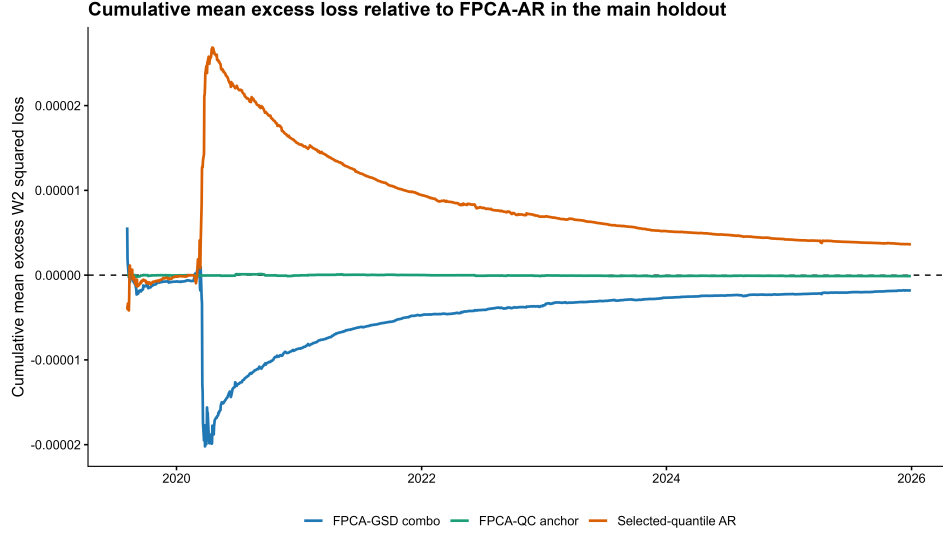


Figure 1: Cumulative mean excess loss relative to FPCA-AR in the main holdout. Negative values indicate periods in which the corresponding model improves on the FPCA-AR anchor.

FPCA-GSD is the best main-holdout model and remains highly competitive in rolling evaluation. FPCA-AR remains a strong anchor throughout. The adaptive models therefore add value, but the gain is best described as measured

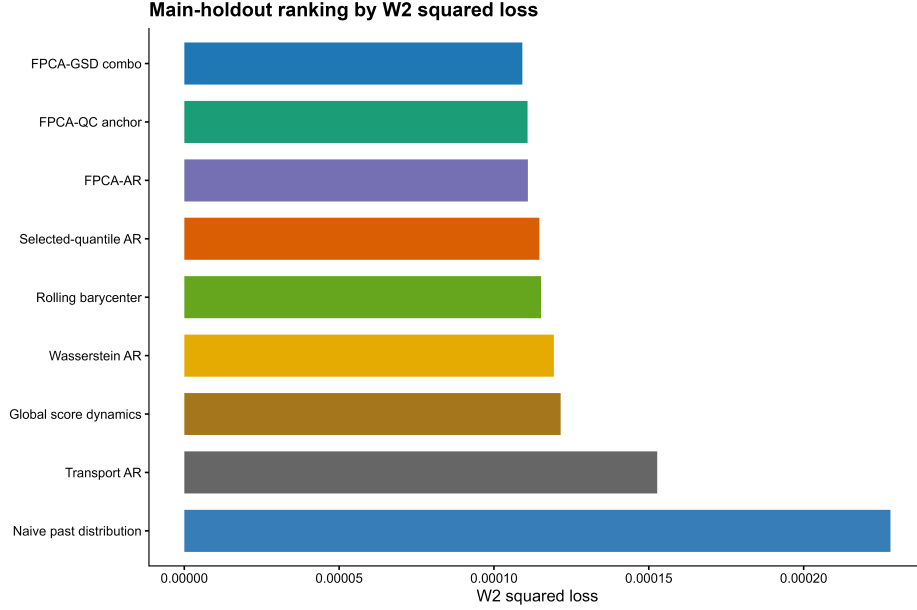


Figure 2: Main-holdout ranking by squared Wasserstein loss. The leading models are concentrated within the FPCA-based forecast family, while the naive and transport-based alternatives are visibly weaker.

rather than dramatic. The complete split diagnostics are reported in Appendix Table A1, the full split-wise loss comparison is reported in Appendix Table A4, and the full untrimmed rolling boxplot is shown in Appendix Figure A1.

5.3. Statistical comparison

Because FPCA-QC is the most stable adaptive specification across rolling splits, the compact Diebold–Mariano table focuses on FPCA-QC. The main-holdout ranking separately identifies FPCA-GSD as the lowest-loss specification. Table 4 reports the main-holdout Diebold–Mariano comparisons for FPCA-QC anchor against selected benchmarks. Negative mean differences favor FPCA-QC.

The results show clear improvements over the naive past-distribution benchmark and transport AR. The comparison against selected-quantile AR is also favorable under squared Wasserstein loss for the shorter Newey–West lag choices, although the strength of significance weakens as the lag length increases. The comparison with FPCA-AR is more cautious: the mean differences are negative, but the adjusted p -values in the full diagnostics do

Table 3: Rolling-split stability summary under squared Wasserstein loss.

Model	Rolling wins	Top-three count	Average rank	Average W_2^2 loss
FPCA-QC anchor	10	21	2.32	0.000081
FPCA-GSD combo	8	17	2.56	0.000081
FPCA-AR	4	18	2.88	0.000081
Rolling barycenter	2	6	4.32	0.000083
Selected-quantile AR	1	6	4.24	0.000082
Global score dynamics	2	4	5.76	0.000087
Wasserstein AR	1	3	5.68	0.000087
Transport AR	0	0	8.00	0.000112
Naive past distribution	0	0	9.00	0.000160

Notes: The table summarizes model stability across 25 rolling one-step-ahead evaluation splits. Rolling wins count the number of splits in which a model attains the lowest average W_2^2 loss. Top-three count gives the number of splits in which a model ranks among the three best-performing methods. Lower average ranks and lower average W_2^2 losses indicate stronger rolling performance.

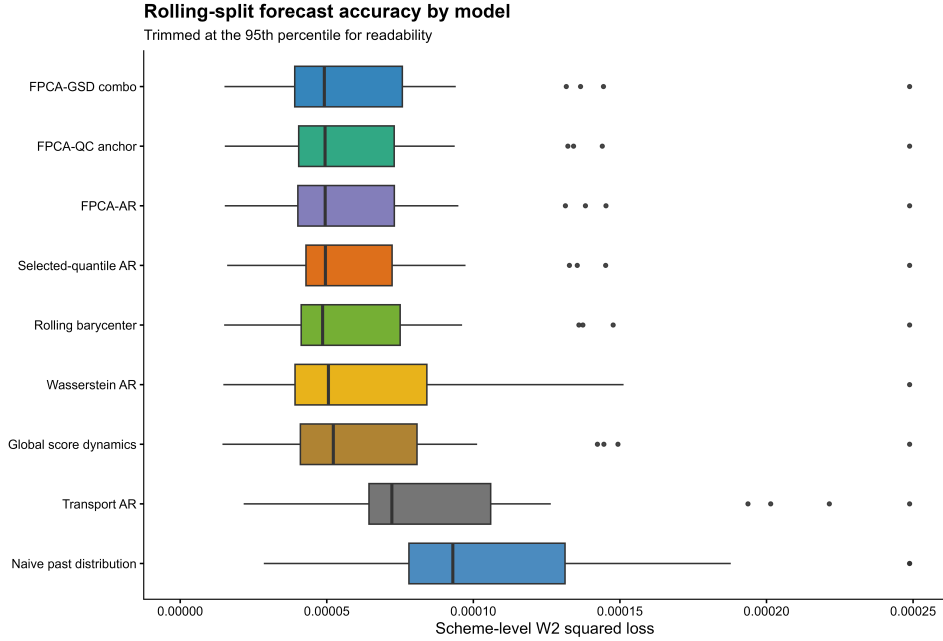


Figure 3: Rolling-split forecast accuracy by model. The trimmed display highlights the lower-loss functional forecast family and the weaker performance of unstable alternatives.

Table 4: Main-holdout Diebold–Mariano comparisons for FPCA-QC anchor.

Loss	Benchmark	Mean difference	DM statistic	Two-sided p -value
W_2^2	Naive past distribution	-1.172×10^{-4}	-3.8518	0.0001
W_2^2	Selected-quantile AR	-3.787×10^{-6}	-2.3541	0.0186
W_2^2	FPCA-AR	-1.103×10^{-7}	-1.2576	0.2085
W_2^2	Transport AR	-4.186×10^{-5}	-7.7044	< 0.0001
W_2^2	Global score dynamics	-1.064×10^{-5}	-2.4839	0.0130
Weighted W_2^2	Naive past distribution	-1.175×10^{-4}	-5.0146	< 0.0001
Weighted W_2^2	Selected-quantile AR	-6.093×10^{-6}	-1.7450	0.0810
Weighted W_2^2	FPCA-AR	-3.255×10^{-7}	-1.3440	0.1789
Weighted W_2^2	Transport AR	-4.085×10^{-5}	-5.9401	< 0.0001
Weighted W_2^2	Global score dynamics	-1.118×10^{-5}	-1.8847	0.0595

Notes: The table reports compact main-holdout Diebold–Mariano comparisons for FPCA-QC anchor using Newey–West lag 5. Mean differences are computed as the loss of FPCA-QC minus the loss of the benchmark; negative values therefore favor FPCA-QC. The full lag-sensitivity and multiplicity-adjusted results are reported in the appendix.

not indicate decisive superiority over the FPCA anchor.

Additional lag-sensitivity and multiplicity-adjusted Diebold–Mariano diagnostics lead to the same substantive conclusion: FPCA-QC performs reliably better than weak or unstable alternatives, but the evidence does not support a claim of decisive superiority over FPCA-AR.

5.4. Stress dates and full quantile-function diagnostics

The final diagnostic analysis examines the full predicted quantile functions on representative dates. Figure 4 compares observed and forecasted summaries of the quantile functions in the main holdout, including the left-tail 5 percent quantile and the interquartile range. The smoothed display shows that the leading models track the main time variation in cross-sectional dispersion and left-tail behavior, while differences among the FPCA-based models remain moderate. The raw version of this display is reported in Appendix Figure A2.

Figure 5 reports full quantile functions for a typical date and a stress date. The leading models capture the broad shape of the cross-sectional distribution, but discrepancies remain visible in difficult episodes, especially in the tails and in the full dispersion of the quantile function. Appendix Figure A3 reports the corresponding full model display.

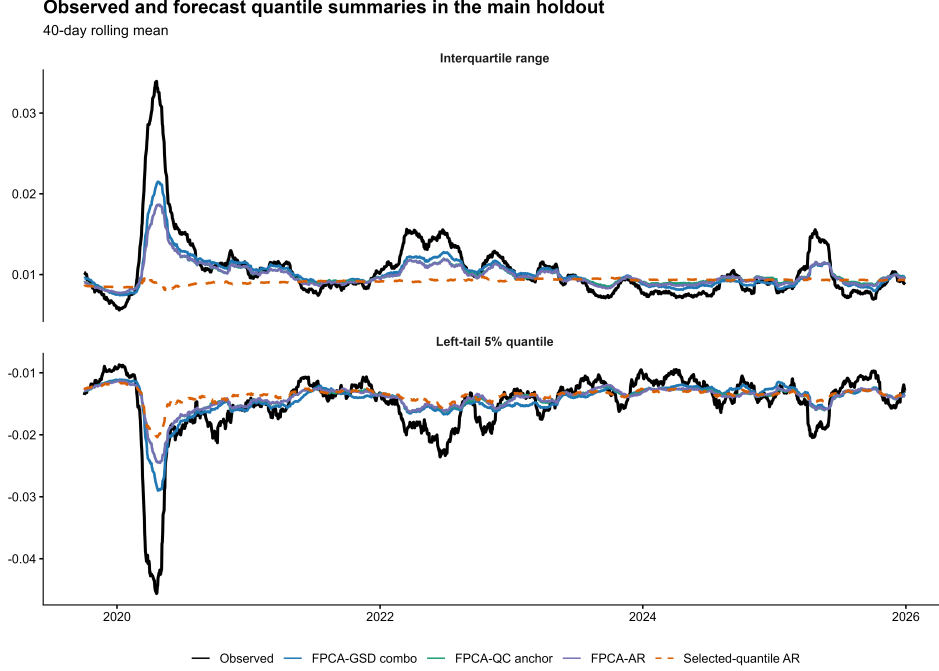


Figure 4: Observed and forecast quantile summaries in the main holdout. The display uses a rolling mean to emphasize medium-run movements in the left-tail quantile and interquartile range.

Overall, the results support three conclusions. First, the best main-holdout performance is achieved by FPCA-GSD combo. Second, the best rolling stability is achieved by FPCA-QC anchor. Third, the adaptive FPCA family improves reliably over naive and less stable alternatives, while the evidence against the FPCA-AR anchor is more measured. The proposed contribution is therefore a leakage-disciplined adaptive distributional forecasting framework that delivers modest but systematic gains around a strong FPCA-based functional representation.

6. Discussion

The empirical results indicate that most of the predictable structure in the daily cross-sectional ETF return distribution is captured by the FPCA representation itself. The close performance of FPCA-GSD, FPCA-QC, and FPCA-AR in the main holdout and rolling evaluations suggests that the low-dimensional score structure provides a demanding benchmark. The adaptive

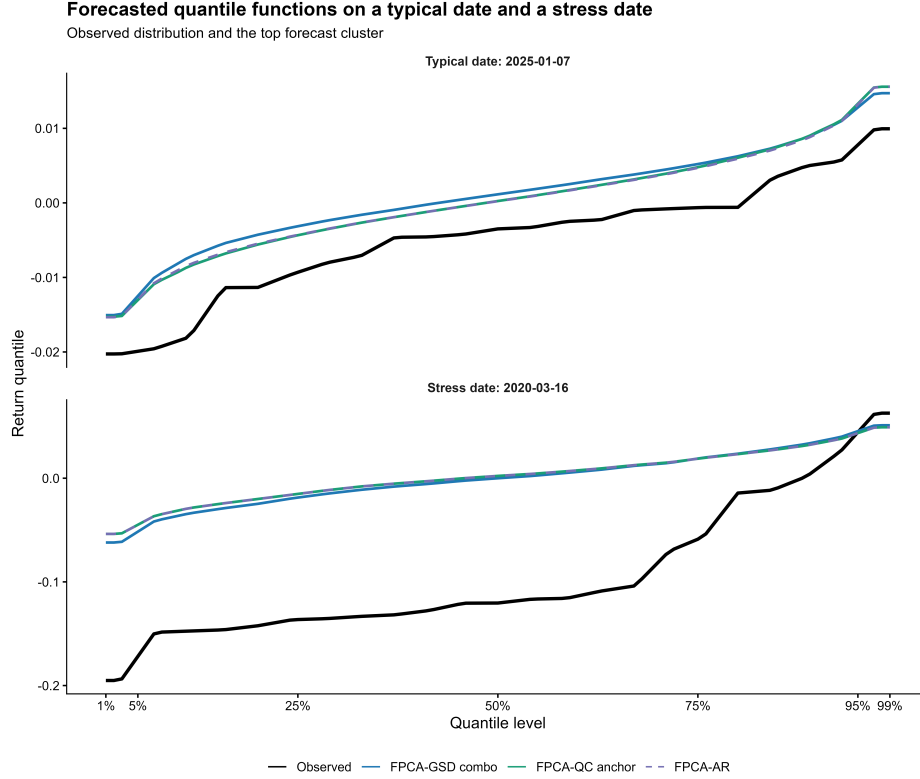


Figure 5: Forecasted quantile functions on a typical date and a stress date. The figure highlights both the strength of the FPCA-based forecast family and the remaining limitations during difficult market episodes.

extensions improve forecasts mainly by adjusting this representation, rather than by introducing a fundamentally different source of predictive power.

This interpretation is consistent with the nature of the target object. A full cross-sectional quantile function reflects movements in location, dispersion, asymmetry, and tail shape. Many of these movements are smooth and persistent enough to be represented by a small number of functional principal components. The role of the adaptive layer is therefore necessarily limited: it can correct local selected-quantile deviations or incorporate additional score-dynamics information, but it operates around an already competitive forecast. The empirical contribution is consequently best viewed as a disciplined improvement to a strong functional baseline.

The contrast between FPCA-GSD and FPCA-QC is informative. FPCA-

GSD achieves the lowest main-holdout loss, which suggests that global score dynamics can add useful information over a long post-training evaluation period. FPCA-QC, by contrast, provides the strongest rolling stability. This difference is consistent with their construction. The GSD combination can move the FPCA forecast toward broader score-level dynamics, whereas the QC correction is tied more closely to selected quantile residuals. The latter mechanism is less aggressive and may therefore be more stable across shorter rolling windows and changing market conditions.

The Diebold–Mariano evidence further supports a cautious reading of the results. The adaptive models clearly improve over weak or unstable alternatives such as naive persistence and transport AR. However, the tests do not provide decisive evidence that FPCA-QC is superior to FPCA-AR. This distinction is important. Lower average losses and stronger rolling ranks indicate useful forecasting gains, but they do not justify a broad claim of uniform dominance over the FPCA anchor.

The stress-date diagnostics highlight the remaining difficulty of the problem. Average Wasserstein losses summarize overall forecast accuracy, but they do not fully describe performance during rapidly changing distributional episodes. In such periods, discrepancies remain visible in the tails and in the overall dispersion of the predicted quantile functions. This suggests that future improvements may need to focus less on average curve fit and more on the early identification of abrupt changes in cross-sectional dispersion and tail asymmetry.

Overall, the findings support a restrained interpretation. A carefully constrained FPCA-centered family can improve upon weak benchmarks, remain competitive with a strong FPCA anchor, and provide a transparent framework for forecasting full cross-sectional return distributions. The main value of the approach lies in combining smooth functional dimension reduction with train-only adaptive corrections while preserving a clear out-of-sample forecasting protocol.

7. Conclusion

This paper studied one-step-ahead forecasting of daily cross-sectional ETF return distributions represented as quantile functions. The forecasting problem was formulated in a one-dimensional Wasserstein representation, and the empirical design ensured that forecasts of Q_t used only information available up to the end of day $t - 1$.

The proposed framework used FPCA-AR as a smooth functional anchor and added two controlled adaptive refinements: FPCA-QC, based on selected-quantile residual correction, and FPCA-GSD, based on an adaptive combination with global score dynamics. In the empirical results, FPCA-GSD delivered the best main-holdout performance, while FPCA-QC provided the strongest rolling stability. The adaptive specifications improved clearly over naive and unstable alternatives, but their advantage over the FPCA anchor was measured rather than decisive.

These findings support adaptive FPCA forecasting as a useful refinement for distribution-valued financial forecasting. At the same time, stress-period diagnostics show that rapid changes in tail behavior and cross-sectional dispersion remain difficult to anticipate. Future work could extend the framework with stress-sensitive state variables, nonlinear score dynamics, or tail-focused adaptive components, provided that the same chronological discipline is maintained.

Declarations

- Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
- Conflict of interest/Competing interests: The authors declare no competing interests.
- Ethics approval and consent to participate: Not applicable, as this study did not involve human participants or animals.
- Consent for publication: All authors have approved the final manuscript for publication.
- Data availability: The data used in this study are publicly available from Yahoo Finance. Processed data and code are available from the corresponding author upon reasonable request.
- Author contributions: Çağlar Sözen: Conceptualization, methodology, software, formal analysis, investigation, writing–original draft, writing–review and editing. Mervener Sözen: Methodology, validation, formal analysis, writing–review and editing.

Appendix A. Additional empirical results

This appendix reports expanded empirical results that support the main-text findings. Table A1 gives the full split diagnostics for the main holdout and rolling designs. Table A2 reports the train-only global state thresholds and global adaptive weights by evaluation scheme. Table A3 reports the raw and shrunk state-adaptive weights used by the FPCA-QC and FPCA-GSD adaptive specifications. Table A4 reports the full split-by-split scheme-wise comparison under squared Wasserstein loss. The expanded tables are reported for auditability rather than for substantive model ranking. Together, they document the chronological construction of the evaluation design, the adaptive weighting mechanism, and the split-level loss comparisons.

Table A1: Split diagnostics for the main holdout and rolling evaluation designs.

Scheme	History	Train rows	Test rows	First test date	Last test date
Main holdout	Expanding	2413	1609	2019-08-07	2025-12-30
Roll. 1	Rolling	1000	120	2013-12-24	2014-06-17
Roll. 2	Rolling	1000	120	2014-06-18	2014-12-05
Roll. 3	Rolling	1000	120	2014-12-08	2015-06-01
Roll. 4	Rolling	1000	120	2015-06-02	2015-11-18
Roll. 5	Rolling	1000	120	2015-11-19	2016-05-12
Roll. 6	Rolling	1000	120	2016-05-13	2016-11-01
Roll. 7	Rolling	1000	120	2016-11-02	2017-04-26
Roll. 8	Rolling	1000	120	2017-04-27	2017-10-16
Roll. 9	Rolling	1000	120	2017-10-17	2018-04-10
Roll. 10	Rolling	1000	120	2018-04-11	2018-09-28
Roll. 11	Rolling	1000	120	2018-10-01	2019-03-25
Roll. 12	Rolling	1000	120	2019-03-26	2019-09-13
Roll. 13	Rolling	1000	120	2019-09-16	2020-03-06
Roll. 14	Rolling	1000	120	2020-03-09	2020-08-26
Roll. 15	Rolling	1000	120	2020-08-27	2021-02-18
Roll. 16	Rolling	1000	120	2021-02-19	2021-08-10
Roll. 17	Rolling	1000	120	2021-08-11	2022-01-31
Roll. 18	Rolling	1000	120	2022-02-01	2022-07-25
Roll. 19	Rolling	1000	120	2022-07-26	2023-01-13
Roll. 20	Rolling	1000	120	2023-01-17	2023-07-10
Roll. 21	Rolling	1000	120	2023-07-11	2023-12-28
Roll. 22	Rolling	1000	120	2023-12-29	2024-06-21
Roll. 23	Rolling	1000	120	2024-06-24	2024-12-11
Roll. 24	Rolling	1000	120	2024-12-12	2025-06-06
Roll. 25	Rolling	1000	120	2025-06-09	2025-11-26

Notes: The main holdout uses an expanding recursive evaluation design. Rolling splits use a fixed 1000-observation training window and a 120-observation test window. Dates refer to the first and last forecasted observations in each test block.

Table A2: Train-only state-threshold and global-weight summary by evaluation scheme.

Scheme	History	μ_{IQR}	σ_{IQR}	μ_{ASYM}	σ_{ASYM}	$q_{1,S}$	$q_{2,S}$	QC w_{glob}	QC n	GSD w_{glob}	GSD n
Main	Exp.	0.0085	0.0036	0.0068	0.0044	-0.4523	0.0834	0.0000	80	0.3500	80
Roll. 1	Roll.	0.0091	0.0040	0.0084	0.0055	-0.4323	0.0788	0.3500	80	0.7000	80
Roll. 2	Roll.	0.0088	0.0038	0.0082	0.0054	-0.4384	0.0917	0.0000	80	0.6500	80
Roll. 3	Roll.	0.0085	0.0038	0.0080	0.0055	-0.4385	0.0821	0.0000	80	0.4000	80
Roll. 4	Roll.	0.0085	0.0038	0.0078	0.0052	-0.4254	0.0745	0.0000	80	0.3500	80
Roll. 5	Roll.	0.0079	0.0025	0.0068	0.0037	-0.4164	0.1630	0.3500	80	0.0000	80
Roll. 6	Roll.	0.0081	0.0027	0.0070	0.0037	-0.4328	0.1784	0.0000	80	0.2000	80
Roll. 7	Roll.	0.0080	0.0028	0.0070	0.0037	-0.4365	0.1618	0.0000	80	0.0000	80
Roll. 8	Roll.	0.0081	0.0029	0.0067	0.0035	-0.4478	0.1651	0.0000	80	0.0500	80
Roll. 9	Roll.	0.0078	0.0030	0.0061	0.0031	-0.4643	0.1355	0.0000	80	0.8000	80
Roll. 10	Roll.	0.0080	0.0033	0.0059	0.0032	-0.4723	0.1047	0.3000	80	0.0000	80
Roll. 11	Roll.	0.0081	0.0033	0.0058	0.0031	-0.4676	0.1028	0.0000	80	0.0000	80
Roll. 12	Roll.	0.0083	0.0036	0.0055	0.0027	-0.4580	0.0858	0.5000	80	0.0000	80
Roll. 13	Roll.	0.0083	0.0035	0.0054	0.0027	-0.4716	0.1131	0.4500	80	0.7000	80
Roll. 14	Roll.	0.0080	0.0037	0.0052	0.0026	-0.4523	0.0822	0.0000	80	0.0000	80
Roll. 15	Roll.	0.0094	0.0065	0.0061	0.0057	-0.4056	-0.0560	0.2000	80	0.1000	80
Roll. 16	Roll.	0.0098	0.0065	0.0066	0.0059	-0.4326	0.0135	0.4500	80	0.3000	80
Roll. 17	Roll.	0.0102	0.0064	0.0069	0.0058	-0.4282	0.0341	0.5000	80	0.2000	80
Roll. 18	Roll.	0.0106	0.0062	0.0071	0.0058	-0.4117	0.0389	0.6500	80	0.2500	80
Roll. 19	Roll.	0.0113	0.0063	0.0074	0.0058	-0.4162	0.0952	0.6500	80	0.1500	80
Roll. 20	Roll.	0.0113	0.0061	0.0076	0.0057	-0.4026	0.0816	0.5500	80	0.4000	80
Roll. 21	Roll.	0.0116	0.0060	0.0078	0.0057	-0.3818	0.0633	0.5500	80	0.5000	80
Roll. 22	Roll.	0.0116	0.0059	0.0077	0.0056	-0.3865	0.0499	0.5500	80	0.3500	80
Roll. 23	Roll.	0.0101	0.0031	0.0071	0.0035	-0.4625	0.2097	0.0000	80	0.5500	80
Roll. 24	Roll.	0.0099	0.0031	0.0067	0.0029	-0.4675	0.2091	0.7000	80	0.0500	80
Roll. 25	Roll.	0.0100	0.0034	0.0065	0.0026	-0.4841	0.2124	0.0000	80	0.0000	80

Notes: Exp. denotes the expanding main-holdout design and Roll. denotes the rolling evaluation design. $q_{1,S}$ and $q_{2,S}$ are the train-only tertile cutoffs of the composite state score S_{t-1} . QC denotes FPCA-QC anchor and GSD denotes FPCA-GSD combo. Global weights are selected by train-only pseudo-out-of-sample loss minimization; n is the number of pseudo-out-of-sample observations used for global-weight selection.

Table A3: Train-only state-adaptive raw and shrunk mixture weights by evaluation scheme.

Scheme	History	State	QC raw	QC shrunk	QC n	GSD raw	GSD shrunk	GSD n
Main	Exp.	Low	0.0000	0.0000	46	0.0000	0.1061	46
Main	Exp.	Medium	1.0000	0.5122	21	1.0000	0.6829	21
Main	Exp.	High	0.0000	0.0000	13	0.0000	0.2121	13
Roll. 1	Roll.	Low	0.0000	0.1400	30	0.4000	0.5200	30
Roll. 1	Roll.	Medium	0.8000	0.6269	32	1.0000	0.8846	32
Roll. 1	Roll.	High	0.9500	0.6342	18	0.6000	0.6526	18
Roll. 2	Roll.	Low	0.0000	0.0000	58	0.5000	0.5385	58
Roll. 2	Roll.	Medium	0.0000	0.0000	20	0.6000	0.6250	20
Roll. 2	Roll.	High	0.0000	0.0000	2	0.6500	0.6500	2
Roll. 3	Roll.	Low	0.0000	0.0000	33	0.9000	0.7113	33
Roll. 3	Roll.	Medium	0.3000	0.1605	23	0.9000	0.6674	23
Roll. 3	Roll.	High	0.2000	0.1091	24	0.0000	0.1818	24
Roll. 4	Roll.	Low	0.0000	0.0000	31	0.0000	0.1373	31
Roll. 4	Roll.	Medium	0.1500	0.0992	39	0.2500	0.2839	39
Roll. 4	Roll.	High	0.0000	0.0000	10	1.0000	0.5667	10
Roll. 5	Roll.	Low	0.0000	0.2333	10	1.0000	0.3333	10
Roll. 5	Roll.	Medium	0.2500	0.2885	32	0.1000	0.0615	32
Roll. 5	Roll.	High	0.4000	0.3828	38	0.0000	0.0000	38
Roll. 6	Roll.	Low	0.0000	0.0000	8	0.2000	0.2000	8
Roll. 6	Roll.	Medium	0.5500	0.2679	19	0.1500	0.1756	19
Roll. 6	Roll.	High	0.0000	0.0000	53	0.2500	0.2363	53
Roll. 7	Roll.	Low	0.0000	0.0000	50	0.0000	0.0000	50
Roll. 7	Roll.	Medium	0.0000	0.0000	22	0.0000	0.0000	22
Roll. 7	Roll.	High	0.0000	0.0000	8	0.0000	0.0000	8
Roll. 8	Roll.	Low	0.0000	0.0000	63	0.0500	0.0500	63
Roll. 8	Roll.	Medium	0.1500	0.0689	17	0.0000	0.0270	17
Roll. 8	Roll.	High	0.0000	0.0000	0	0.0500	0.0500	0
Roll. 9	Roll.	Low	0.0000	0.0000	58	0.8500	0.8372	58
Roll. 9	Roll.	Medium	0.9500	0.4976	22	0.6500	0.7214	22
Roll. 9	Roll.	High	0.0000	0.0000	0	0.8000	0.8000	0
Roll. 10	Roll.	Low	0.9000	0.6073	21	0.2500	0.1280	21
Roll. 10	Roll.	Medium	1.0000	0.7500	36	0.0000	0.0000	36
Roll. 10	Roll.	High	0.0000	0.1395	23	0.0000	0.0000	23
Roll. 11	Roll.	Low	0.0000	0.0000	35	0.0000	0.0000	35
Roll. 11	Roll.	Medium	0.0000	0.0000	31	0.0500	0.0304	31
Roll. 11	Roll.	High	0.0000	0.0000	14	0.0000	0.0000	14
Roll. 12	Roll.	Low	0.6000	0.5474	18	0.4000	0.1895	18
Roll. 12	Roll.	Medium	0.7000	0.6216	31	0.0000	0.0000	31
Roll. 12	Roll.	High	0.5000	0.5000	31	0.0000	0.0000	31
Roll. 13	Roll.	Low	0.8500	0.6798	27	0.0500	0.3266	27
Roll. 13	Roll.	Medium	1.0000	0.7179	19	1.0000	0.8462	19
Roll. 13	Roll.	High	0.2500	0.3241	34	0.6000	0.6370	34
Roll. 14	Roll.	Low	1.0000	0.6000	30	0.4000	0.2400	30
Roll. 14	Roll.	Medium	1.0000	0.5556	25	0.0000	0.0000	25
Roll. 14	Roll.	High	0.0000	0.0000	25	0.0000	0.0000	25
Roll. 15	Roll.	Low	0.2000	0.2000	0	0.1000	0.1000	0
Roll. 15	Roll.	Medium	0.2000	0.2000	4	0.1000	0.1000	4
Roll. 15	Roll.	High	0.2500	0.2396	76	0.1000	0.1000	76

Continued on next page

Table A3: Train-only state-adaptive raw and shrunk mixture weights by evaluation scheme. Continued.

Scheme	History	State	QC raw	QC shrunk	QC n	GSD raw	GSD shrunk	GSD n
Roll. 16	Roll.	Low	0.4500	0.4500	4	0.3000	0.3000	4
Roll. 16	Roll.	Medium	0.4000	0.4217	26	0.6500	0.4978	26
Roll. 16	Roll.	High	0.4500	0.4500	50	0.2500	0.2643	50
Roll. 17	Roll.	Low	1.0000	0.7222	16	0.8000	0.4667	16
Roll. 17	Roll.	Medium	0.3500	0.3955	46	0.3500	0.3045	46
Roll. 17	Roll.	High	0.4500	0.4763	18	0.0000	0.1053	18
Roll. 18	Roll.	Low	0.3000	0.4795	19	0.3000	0.2744	19
Roll. 18	Roll.	Medium	0.9000	0.8206	43	0.2500	0.2500	43
Roll. 18	Roll.	High	0.4000	0.5316	18	0.2000	0.2263	18
Roll. 19	Roll.	Low	0.6500	0.6500	2	0.1500	0.1500	2
Roll. 19	Roll.	Medium	0.9500	0.8346	32	0.1000	0.1192	32
Roll. 19	Roll.	High	0.5000	0.5455	46	0.1500	0.1500	46
Roll. 20	Roll.	Low	1.0000	0.7429	15	0.9500	0.6357	15
Roll. 20	Roll.	Medium	0.2000	0.3296	34	0.0000	0.1481	34
Roll. 20	Roll.	High	0.6000	0.5804	31	0.4500	0.4304	31
Roll. 21	Roll.	Low	0.2000	0.3400	30	1.0000	0.8000	30
Roll. 21	Roll.	Medium	1.0000	0.8421	37	0.4000	0.4351	37
Roll. 21	Roll.	High	0.2500	0.4318	13	0.0500	0.3227	13
Roll. 22	Roll.	Low	0.0500	0.2038	45	0.6500	0.5577	45
Roll. 22	Roll.	Medium	1.0000	0.8235	31	0.3500	0.3500	31
Roll. 22	Roll.	High	0.5500	0.5500	4	0.3500	0.3500	4
Roll. 23	Roll.	Low	0.0000	0.0000	43	0.4000	0.4476	43
Roll. 23	Roll.	Medium	0.4000	0.2095	22	0.7000	0.6286	22
Roll. 23	Roll.	High	0.4500	0.1929	15	0.4000	0.4857	15
Roll. 24	Roll.	Low	0.2000	0.4326	23	0.0000	0.0233	23
Roll. 24	Roll.	Medium	0.7500	0.7353	48	0.3500	0.2618	48
Roll. 24	Roll.	High	0.7000	0.7000	9	0.0500	0.0500	9
Roll. 25	Roll.	Low	1.0000	0.3750	12	0.8500	0.3188	12
Roll. 25	Roll.	Medium	0.3000	0.1605	23	0.0000	0.0000	23
Roll. 25	Roll.	High	0.0000	0.0000	45	0.0000	0.0000	45

Notes: Raw weights are selected by train-only pseudo-out-of-sample loss minimization over the candidate grid $\{0, 0.05, \dots, 0.95, 1\}$. Shrunk weights are obtained by shrinking the state-specific raw weight toward the corresponding global weight using $\alpha_s = n_s/(n_s + 20)$. The reported counts are the train-only pseudo-out-of-sample observations available in each state. QC denotes FPCA-QC anchor and GSD denotes FPCA-GSD combo.

Table A4: Full split-by-split performance under squared Wasserstein loss.

Scheme	Leading adaptive	Adaptive loss	FPCA-AR loss	Leading external comparator	Comparative loss	Interpretation
Main	FPCA-GSD combo	0.000109	0.000111	Selected-quantile AR	0.000115	Adaptive family leads
Roll. 1	FPCA-GSD combo	0.000027	0.000028	Global score dynamics	0.000027	Empirically close

Continued on next page

Table A4: Full split-by-split performance under squared Wasserstein loss. Continued.

Scheme	Leading adaptive	Adaptive loss	FPCA- AR loss	Leading external comparator	Comparative loss	Interpretation
Roll. 2	FPCA-GSD combo	0.000033	0.000033	Rolling barycenter	0.000034	FPCA-centered cluster
Roll. 3	FPCA-QC anchor	0.000044	0.000044	Selected- quantile AR	0.000045	FPCA-centered cluster
Roll. 4	FPCA-QC anchor	0.000073	0.000073	Selected- quantile AR	0.000072	Comparator slightly leads
Roll. 5	FPCA-GSD combo	0.000066	0.000066	Rolling barycenter	0.000064	Comparator leads
Roll. 6	FPCA-GSD combo	0.000045	0.000045	Selected- quantile AR	0.000046	FPCA-centered cluster
Roll. 7	FPCA-QC / FPCA-GSD	0.000024	0.000024	Wasserstein AR	0.000025	FPCA-centered cluster
Roll. 8	FPCA-GSD combo	0.000015	0.000015	Global score dynamics	0.000014	Comparator slightly leads
Roll. 9	FPCA-QC anchor	0.000053	0.000053	Selected- quantile AR	0.000054	FPCA-centered cluster
Roll. 10	FPCA-QC anchor	0.000022	0.000022	Selected- quantile AR	0.000022	Empirically close
Roll. 11	FPCA-GSD combo	0.000079	0.000079	Selected- quantile AR	0.000079	Empirically close
Roll. 12	FPCA-QC anchor	0.000043	0.000043	Global score dynamics	0.000043	Empirically close
Roll. 13	FPCA-QC anchor	0.000066	0.000066	Rolling barycenter	0.000066	Empirically close
Roll. 14	FPCA-GSD combo	0.000596	0.000596	Rolling barycenter	0.000609	FPCA-centered cluster
Roll. 15	FPCA-QC anchor	0.000094	0.000095	Rolling barycenter	0.000096	Adaptive family leads
Roll. 16	FPCA-QC anchor	0.000046	0.000047	Selected- quantile AR	0.000047	Adaptive family leads
Roll. 17	FPCA-QC anchor	0.000047	0.000049	Selected- quantile AR	0.000048	Adaptive family leads
Roll. 18	FPCA-QC anchor	0.000134	0.000138	Selected- quantile AR	0.000135	Adaptive family leads
Roll. 19	FPCA-QC anchor	0.000144	0.000145	Selected- quantile AR	0.000145	Adaptive family leads
Roll. 20	FPCA-GSD combo	0.000051	0.000053	Global score dynamics	0.000052	Adaptive family leads
Roll. 21	FPCA-QC anchor	0.000051	0.000052	Rolling barycenter	0.000048	Comparator leads
Roll. 22	FPCA-GSD combo	0.000039	0.000040	Wasserstein AR	0.000039	Empirically close
Roll. 23	FPCA-GSD combo	0.000049	0.000049	Selected- quantile AR	0.000050	FPCA-centered cluster
Roll. 24	FPCA-GSD combo	0.000132	0.000131	Selected- quantile AR	0.000133	FPCA-AR slightly leads

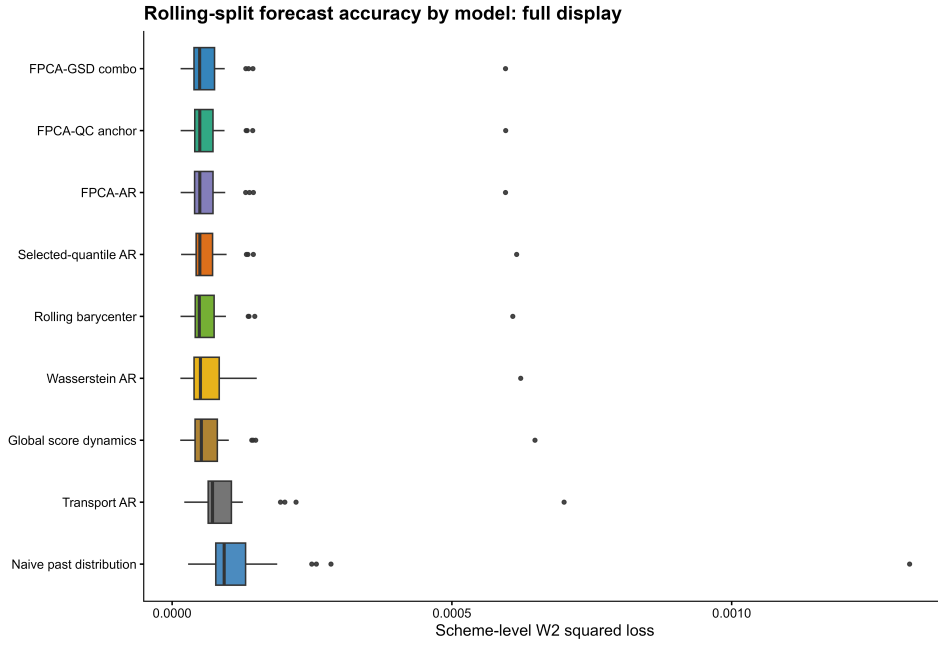


Figure A1: Rolling-split forecast accuracy by model: full untrimmed display.

Roll. 25	FPCA-GSD combo	0.000035	0.000036	Rolling barycenter	0.000036	Adaptive family leads
----------	-------------------	----------	----------	-----------------------	----------	--------------------------

Notes: Loss values are average squared Wasserstein losses within each evaluation scheme. The adaptive family consists of FPCA-QC anchor and FPCA-GSD combo. External comparators exclude FPCA-AR and the adaptive FPCA models. The table is reported for split-level auditability; the main-text interpretation is based on the compact main-holdout and rolling summaries.

Appendix B. Additional figures

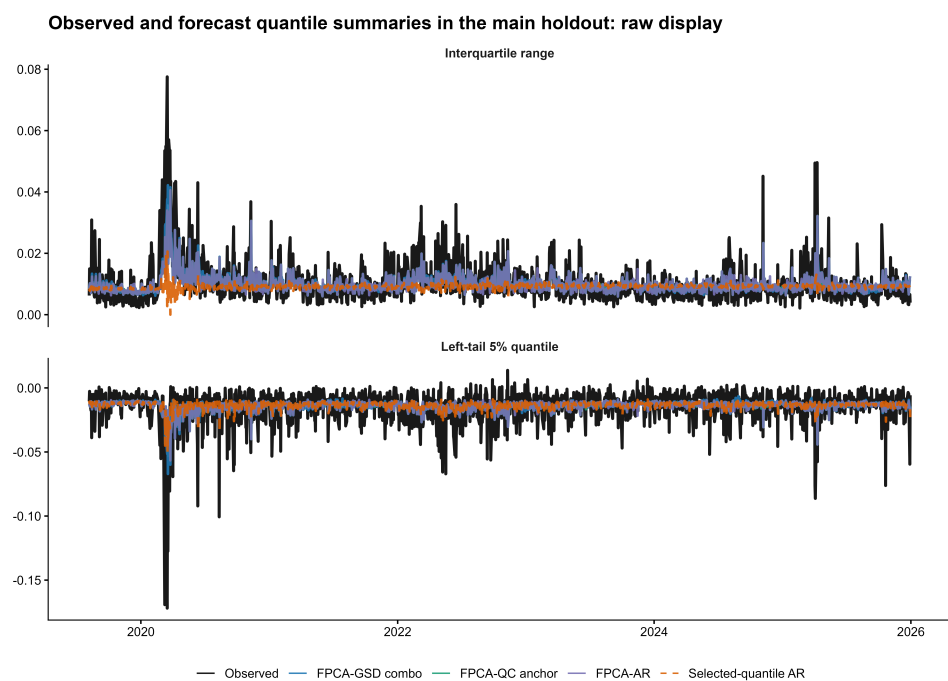


Figure A2: Observed and forecast quantile summaries in the main holdout: raw display.

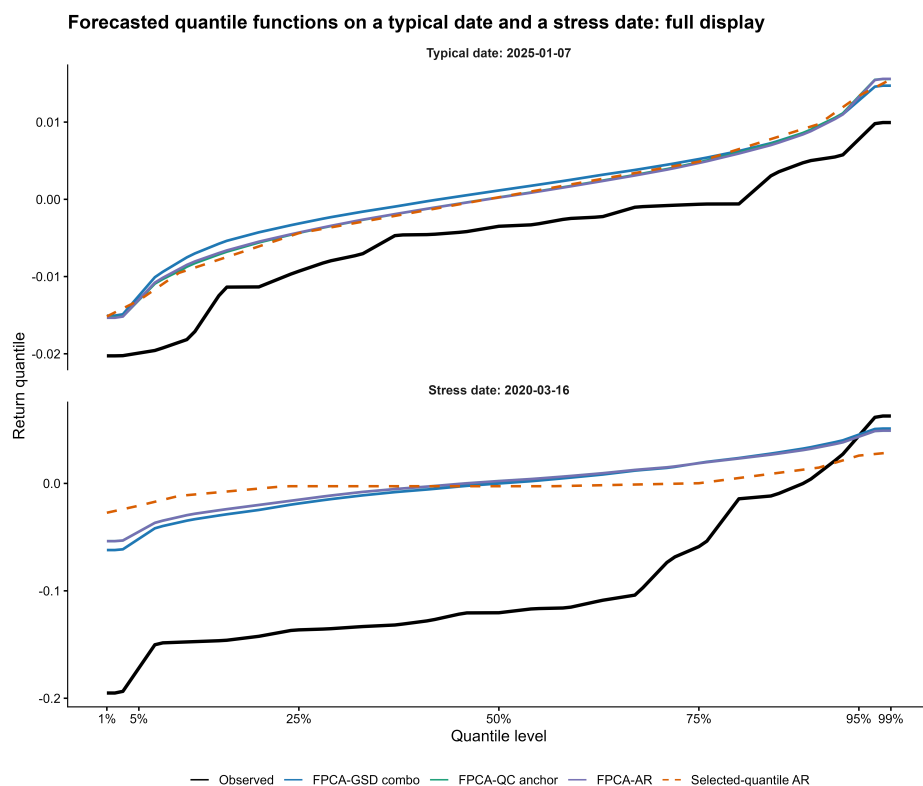


Figure A3: Forecasted quantile functions on a typical date and a stress date: full model display.

Appendix C. Implementation details

All forecasts are generated recursively using only information available at the forecast origin. For each forecast date, the historical quantile-function sample is rebuilt, the FPCA basis is re-estimated, the retained score models are refitted, and the selected-quantile and global score-dynamics components are reconstructed from the available history. Adaptive states are assigned using thresholds learned from the training portion of the relevant evaluation scheme.

The state-adaptive weights are estimated separately within each evaluation scheme using only train-period pseudo-out-of-sample comparisons. The recent state signal is constructed from trailing IQR and absolute tail-asymmetry summaries, standardized by train-only means and standard deviations, and discretized using train-only tertiles. Global and state-specific candidate weights are selected over the grid $\{0, 0.05, \dots, 0.95, 1\}$. State-specific raw weights are shrunk toward the global weight using the shrinkage rule described in Section 3.6. These thresholds and weights are then held fixed during the corresponding out-of-sample evaluation. This design prevents the adaptive layer from becoming an unconstrained state-specific refit and keeps the proposed models anchored to the FPCA-AR forecast. Final quantile functions are projected onto the monotone cone by isotonic regression after forecasting.

References

- Aue, A., Norinho, D. D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110(509), 378–392.
- Ben-David, I., Franzoni, F., and Moussawi, R. (2018). Do ETFs increase volatility? *The Journal of Finance*, 73(6), 2471–2535.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300.
- Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Springer, New York.
- Chen, Y., Lin, Z., and Müller, H.-G. (2023). Wasserstein regression. *Journal of the American Statistical Association*, 118(542), 869–882.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Elton, E. J., Gruber, M. J., Comer, G., and Li, K. (2002). Spiders: Where are the bugs? *Journal of Business*, 75(3), 453–472.
- Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1, 125–151.
- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Goyal, A. and Santa-Clara, P. (2003). Idiosyncratic risk matters! *The Journal of Finance*, 58(3), 975–1007.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, New York.

- Hyndman, R. J. and Shang, H. L. (2009). Forecasting functional time series. *Journal of the Korean Statistical Society*, 38(3), 199–211.
- Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51(10), 4942–4956.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Panaretos, V. M. and Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and Its Application*, 6, 405–431.
- Petersen, A. and Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics*, 47(2), 691–719.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. 2nd ed., Springer, New York.
- Solnik, B. and Roulet, J. (2000). Dispersion as cross-sectional correlation. *Financial Analysts Journal*, 56(1), 54–61.
- Verousis, T. and Voukelatos, N. (2018). Cross-sectional dispersion and expected returns. *Quantitative Finance*, 18(5), 813–826.
- Villani, C. (2009). *Optimal Transport: Old and New*. Springer, Berlin.
- Yahoo Finance. (2026). Historical daily market data for exchange-traded funds. Available at: <https://finance.yahoo.com/>. Accessed January 24, 2026.
- Zhang, C., Kokoszka, P., and Petersen, A. (2022). Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis*, 43(1), 30–52.
- Zhu, C. and Müller, H.-G. (2023). Autoregressive optimal transport models. *Journal of the Royal Statistical Society: Series B*, 85(3), 1012–1033.