Rolling PCA Eigenportfolios and Factor-Neutral Minimum-Variance Portfolios in a Multi-Asset ETF Universe (2014–2025)

Jacob Hahn

January 6, 2026

Contents

Abstract

This study explores whether linear-algebraic structure in ETF returns can be used for portfolio construction. Using daily log returns for a diversified ETF universe (2014–2025) and a walk-forward monthly rebalancing protocol, rolling PCA is implemented on the correlation matrix to extract statistical factors and construct (i) dollar-neutral eigenportfolios and (ii) factor-neutral minimum-variance portfolios. We find strong low-dimensional structure (top five PCs explain ∼86% of variance on average), but naïve factor-neutral optimization is unstable without explicit regularization. Shrinkage covariance estimation and gross-leverage constraints materially improve implementability net of transaction costs.

1 Introduction

Principal Component Analysis (PCA) is a core linear algebra tool for representing high-dimensional data within a lower-dimensional subspace. In quantitative finance, PCA applied to asset returns provides a natural way to extract *statistical factors* (eigenvectors of a covariance/correlation matrix) and to construct portfolios aligned with those factors. However, financial return distributions are unstable: factor structures drift across regimes, covariance matrices become ill-conditioned, and unconstrained optimization may produce extreme long/short positions that are ineffective out-of-sample.

This paper develops an end-to-end, reproducible pipeline: (1) build a clean ETF return dataset; (2) quantify rolling PCA factor structure and stability; (3) translate factors into eigenportfolios and factor-neutral minimum-variance portfolios; and (4) perform robustness checks with transaction costs and explicit leverage control.

2 Background

2.1 Eigenportfolios from PCA

Zhou and Luan formalize eigenportfolios as portfolios constructed from principal components of the asset return correlation matrix, emphasizing standardized returns, eigen-decomposition, and treating eigenvectors as portfolio weight vectors [4]. They also highlight a practical risk: selecting a single eigenportfolio purely by in-sample Sharpe ratio can overfit, motivating robustness checks and (in their setting) ensemble ideas that combine multiple top components.

2.2 Rolling PCA in Non-Stationary Markets

In financial time series, covariance structure changes over time. Hirsa et al. propose a robust rolling PCA framework (R2-PCA) aimed at improving stability in non-stationary settings, with particular focus on eigenvector sign indeterminacy (sign "flips") and component matching across time [2]. Their research emphasizes using cosine similarity between eigenvectors across consecutive windows to identify flips and realign signs, improving the continuity of projected factor time series.

3 Data

3.1 Universe and Data Source

This study analyzes daily adjusted close prices for a diversified universe of ETFs spanning equities (U.S. and international), fixed income (Treasuries, aggregate bonds, inflation-protected bonds), and other exposures. Prices were obtained programmatically via `yfinance` [1] and converted to daily log returns:

$$r_{t,i} = \log\left(\frac{P_{t,i}}{P_{t-1,i}}\right). \tag{1}$$

After filtering for data completeness, the final panel contains $N = 49$ ETFs and $T = 3008$ trading days from 2014-01-03 to 2025-12-18.

3.2 Dataset Construction and Missing Data

To ensure a complete matrix for PCA, assets were retained only if they met a coverage threshold and reasonable start/end-date requirements; after filtering, any remaining dates with missing values were dropped to produce a complete panel for rolling PCA and backtesting.

4 Methods

4.1 Walk-Forward Backtest Protocol

This study uses a walk-forward design with monthly rebalancing. A walk-forward analysis is algorithmic trading that tests a strategy's reliability by iteratively optimizing parameters on in-sample (past) data and then testing those settings on the next block of out-of-sample (unseen) data, rolling forward through history to simulate real-world performance and avoid overfitting Let T index rebalance dates (the first trading day of each month). At each rebalance date T, all quantities are estimated using a trailing window of length $W = 252$ trading days, i.e. using $\{r_{T-W}, \dots, r_{T-1}\}$ only. Portfolio weights chosen at date T are held until the next rebalance date.

4.2 Baseline Strategy

As a control, the study implements an equal-weight portfolio rebalanced monthly:

$$w_{T,i} = \frac{1}{N}, \quad i = 1, \dots, N.$$

This benchmark validates the data and backtest implementation and provides a reference risk profile for more complex strategies.

4.3 Rolling PCA on Correlation

Because ETFs differ in volatility and scale, PCA is performed on the *correlation* matrix. Within each window at date T, returns are de-meaned and standardized, yielding a standardized return matrix $Z_T \in R^{W \times N}$. The sample correlation matrix is:

$$C_T = \frac{1}{W-1} Z_T^\top Z_T.$$

Then compute the eigen-decomposition:

$$C_T v_{T,k} = \lambda_{T,k} v_{T,k}, \quad k = 1, \dots, N,$$

with eigenvalues sorted in descending order. The explained variance ratio (EVR) is:

$$\mathrm{EVR}_{T,k} = \frac{\lambda_{T,k}}{\sum_{j=1}^{N} \lambda_{T,j}}.$$

4.4 Factor Stability and Sign Alignment

Eigenvectors are defined only up to sign, so we align signs across consecutive rebalance dates by flipping $v_{T,k}$ whenever $\langle v_{T,k}, v_{T-\Delta,k} \rangle < 0$, where Δ denotes one rebalance step. This approach follows the cosine-similarity sign-fix described in the R2-PCA research[2]. Stability is measured by absolute cosine similarity:

$$\mathrm{Stab}_{T,k} = \left| \langle v_{T,k}, v_{T-\Delta,k} \rangle \right|.$$

4.5 Eigenportfolios

Following the eigenportfolio perspective in Zhou and Luan [4], for each component k we construct a dollar-neutral eigenportfolio by demeaning and normalizing to unit gross exposure:

$$\tilde{v}_{T,k} = v_{T,k} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top v_{T,k}, \quad w_T^{(k)} = \frac{\tilde{v}_{T,k}}{\|\tilde{v}_{T,k}\|_1}.$$

Weights are held constant until the next rebalance.

4.6 Minimum-Variance and Factor-Neutral Minimum-Variance

Let Σ_T denote a covariance estimator using the trailing window. The minimum-variance portfolio solves:

$$\min_w \frac{1}{2} w^\top \Sigma_T w \quad \text{s.t.} \quad \mathbf{1}^\top w = 1.$$

To remove exposure to the top PCA directions, define $V_{T,K} = [v_{T,1}, \ldots, v_{T,K}]$ and solve a factor-neutral problem:

$$\min_w \frac{1}{2} w^\top \Sigma_T w \quad \text{s.t.} \quad \mathbf{1}^\top w = 1,\ V_{T,K}^\top w = 0.$$

Empirically, this formulation can produce extreme offsetting positions. Therefore stabilize it with:

1. **Shrinkage covariance estimation (Ledoit–Wolf).** We use a shrinkage estimator to reduce noise and improve conditioning [3].
2. **Explicit gross-leverage constraint.** We solve a convex quadratic program with an ℓ_1 constraint:

$$\min_w \frac{1}{2} w^\top \Sigma_T w \quad \text{s.t.} \quad \mathbf{1}^\top w = 1,\ V_{T,K}^\top w = 0,\ \|w\|_1 \le L_{\max}.$$

4.7 Transaction Costs and Metrics

Turnover at rebalance date T is:

$$\mathrm{TO}_T = \sum_{i=1}^{N} |w_{T,i} - w_{T^-,i}|.$$

Net wealth is reduced at rebalance dates by a proportional cost $c \cdot \mathrm{TO}_T$ with $c = 0.001$ (10 bps per 100% turnover). Report CAGR, annualized volatility, Sharpe ratio (risk-free rate set to zero), and maximum drawdown.

5 Results

5.1 PCA Diagnostics

Rolling PCA indicates strong low-dimensional structure: the top five principal components explain approximately 85.6% of total variance on average. Factor stability is stratified by component order: PC1 and PC2 remain highly stable over time, while higher-order PCs show greater regime sensitivity.

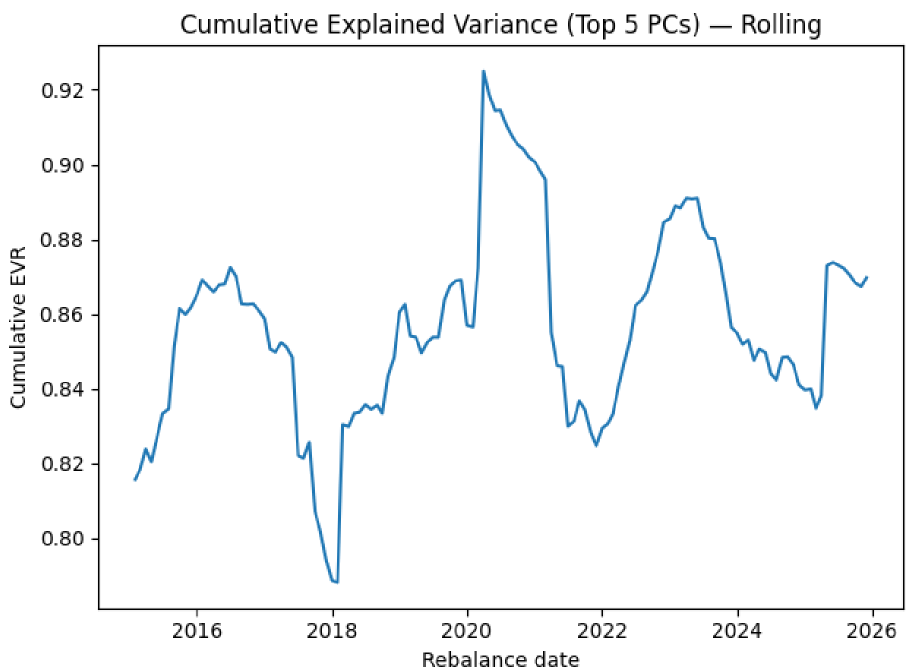

Figure 1: Rolling PCA cumulative explained variance (top K PCs) using a 252-day window and monthly re-estimation.

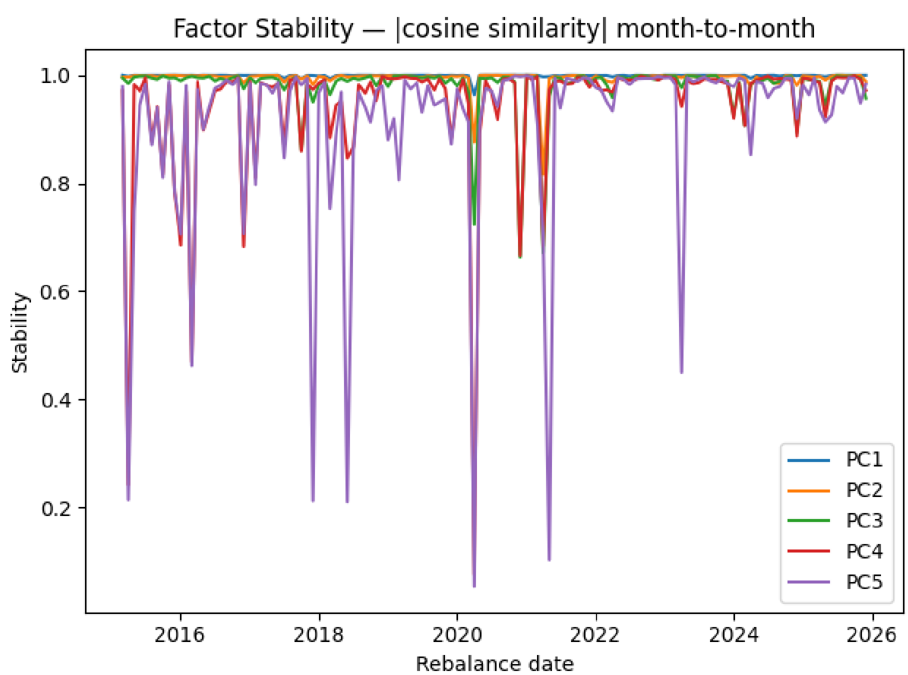

Figure 2: Month-to-month factor stability (absolute cosine similarity after sign alignment). PC1–PC2 are most stable; higher PCs show regime sensitivity.

5.2 Strategy Backtests

This study compares a monthly equal-weight benchmark to PCA-derived eigenportfolios and minimum-variance families. Eigenportfolios capture specific patterns in returns, but they don't necessarily make money; in particular, a purely in-sample selection approach can lead to instability, consistent with the overfitting caution discussed by Zhou and Luan [4]. Unconstrained factor-neutral minimum-variance produced high turnover and extreme gross leverage, motivating the stabilized formulation.

Strategy	CAGR_net	Sharpe_net	MaxDD_net	AvgTurnover	AvgGrossLev
Equal-Weight (monthly)	0.075969	0.629901	-0.298247	0.007634	1.000000
Eigenportfolio PC1 (L1=1, sum=0)	0.035322	0.465775	-0.198533	0.054419	1.000000
Eigenportfolio PC2 (L1=1, sum=0)	-0.034971	-0.408931	-0.386176	0.088639	1.000000
Eigenportfolio PC3 (L1=1, sum=0)	0.004235	0.107225	-0.153515	0.146277	1.000000
Min-Var (sum=1)	0.015673	0.724423	-0.084867	0.312136	1.843992
Factor-Neutral Min-Var (sum=1, neutral PCs 1–2)	0.008100	0.188669	-0.123380	1.360096	7.323483

Table 1: Week 3 performance summary (net of transaction costs, $c = 0.001$). Eigenportfolios are dollar-neutral (self-financing) with unit gross exposure; min-variance portfolios are fully invested with potential shorting (gross leverage may exceed 1).

5.3 Stabilization and Final Specification

There was successful stabilized factor-neutral optimization with Ledoit–Wolf covariance shrinkage and an explicit gross-leverage constraint. A robustness grid over neutrality strength $K_{\text{neutral}} \in \{1, 2\}$ and leverage caps $L_{\max}$ supported a final implementable specification:

$$K_{\text{neutral}} = 1 \text{ (neutralize PC1 only)}, \quad L_{\max} = 3.0,$$

with monthly rebalancing and a 252-day estimation window. This choice avoided frequent feasibility-driven cap escalation while keeping leverage and turnover controlled net of costs.

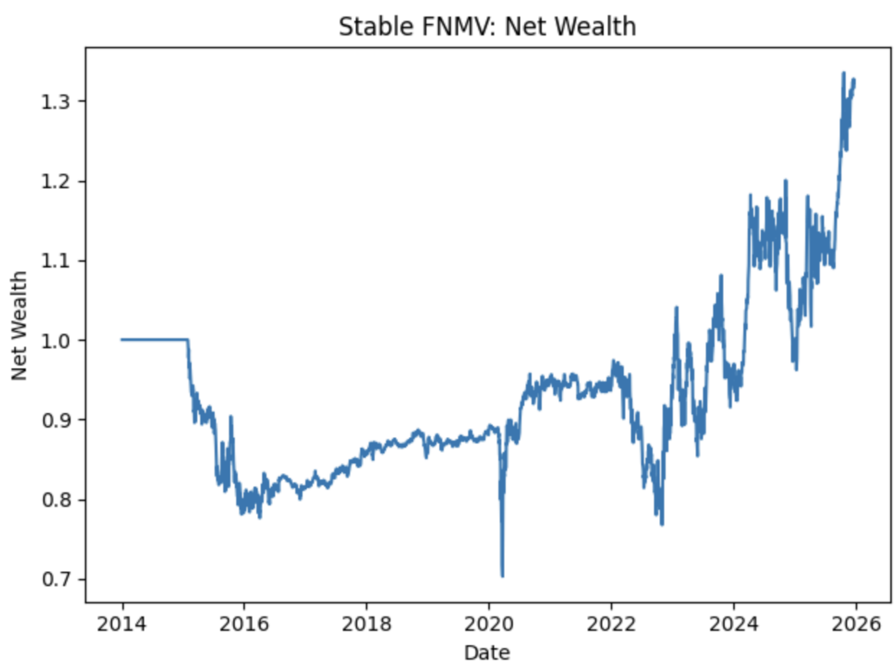

Figure 3: Stable factor-neutral minimum-variance (FNMV) strategy: net wealth (net of transaction costs).

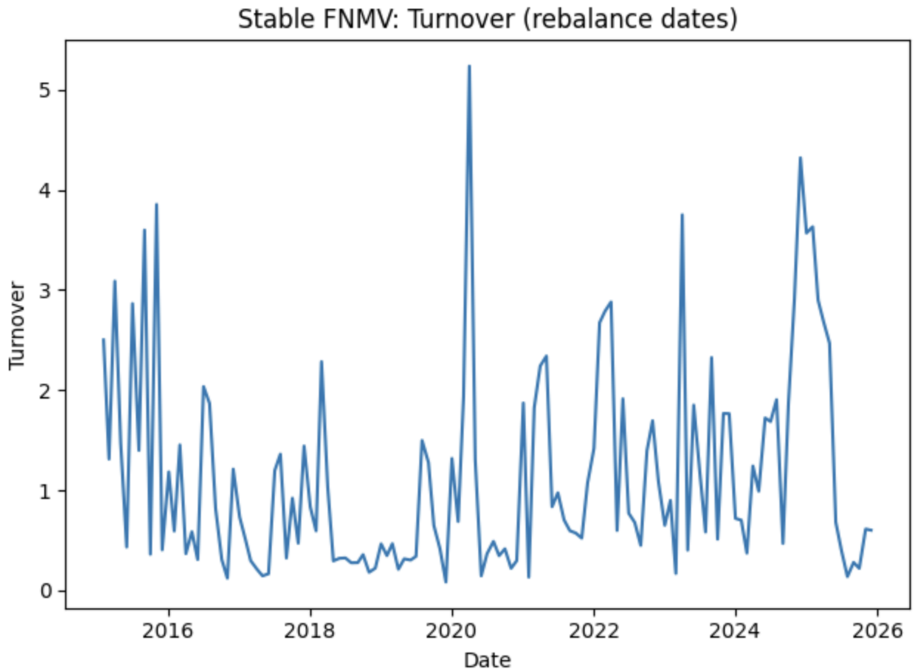

Figure 4: Stable FNMV strategy: turnover on rebalance dates. Spikes indicate regime shifts and/or changes in estimated covariance structure.

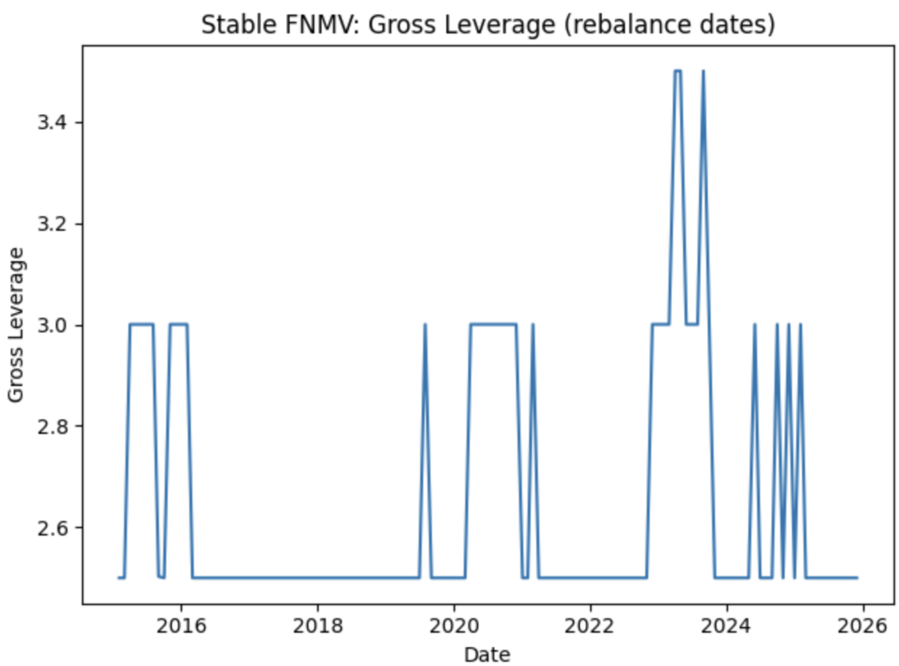

Figure 5: Stable FNMV strategy: gross leverage on rebalance dates. Leverage remains bounded near the imposed ℓ_1 cap, improving implementability relative to an unconstrained factor-neutral optimizer.

6 Discussion

There are two takeaways revealed in this study. First, the ETF return space exhibits strong low-dimensional structure: a small number of rolling principal components explain most cross-sectional variation, and the leading components remain stable through time. Second, translating statistical factors into *tradable* portfolios requires explicit attention to estimation error and constraints. The instability of selecting or optimizing directly on PCA components is consistent with the overfitting concerns emphasized by Zhou and Luan [4]. Moreover, non-stationarity and eigenvector indeterminacy motivate the sign-alignment and stability diagnostics highlighted in the R2-PCA research[2]. Finally, shrinkage covariance estimation and gross-leverage constraints are essential to keep factor-neutral optimization feasible and implementable net of transaction costs.

7 Conclusion

This study built a reproducible pipeline connecting linear algebra to portfolio engineering: rolling eigendecomposition to extract factors, eigenportfolios to isolate those factors, and constrained factor-neutral minimum-variance optimization to hedge dominant modes of co-movement. Empirically, PCA reveals persistent structure in ETF returns, but naïve factor-neutral optimization is unstable. Robust rolling factor alignment and explicit leverage control materially improve implementability and interpretability net of costs.

References

[1] Ran Aroussi and contributors. yfinance: Yahoo! finance market data downloader, 2025. Python package.

[2] Ali Hirsa, Federico Klinkert, Satyan Malhotra, and Ryan Holmes. Robust rolling pca (r2-pca), 2023. Poster, Columbia Data Science Institute Poster Session (Fall 2023). PDF hosted by Columbia Data Science Institute (uploaded Jan 2024).

[3] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2004.

[4] ZhengXiang Zhou and Yuqi Luan. Eigen portfolios: From single component models to ensemble approaches, 2025. arXiv preprint.