PERFORMANCE OF PAIRS TRADING STRATEGIES BASED ON PRINCIPAL COMPONENT ANALYSIS METHODS

YUFEI SUN

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES

WORKING PAPERS

Performance of Pairs Trading Strategies Based on Principal Component Analysis Methods

Yufei Sun[1]*

[1]Faculty of Economic Sciences, University of Warsaw
*Corresponding authors: y.sun2@uw.edu.pl

Abstract: This thesis examines market-neutral, mean-reversion-based statistical arbitrage strategies in the Chinese equity market, using two factor decomposition methods: Principal Component Analysis (PCA) and sector-based Exchange-Traded Funds (ETFs). Residual returns are modeled as mean-reverting Ornstein–Uhlenbeck (OU) processes, generating contrarian signals. A 60-day rolling window ensures out-of-sample estimation. Realistic frictions are included via a 10-basis-point round-trip cost. Backtests from 2005 to 2024 compare four configurations: synthetic ETFs, fixed PCA, dynamic PCA, and trading-time volume adjustments. Both PCA- and ETF-based strategies deliver robust Sharpe ratios near 0.90–0.95. PCA portfolios perform better under high cross-sectional volatility, while ETF-based models remain stable during structural shifts. Incorporating trading volume enhances returns, especially for ETF models. Sensitivity analysis highlights the importance of threshold tuning and rolling-window lengths. These findings stress the critical role of factor construction and signal design in market-neutral strategies, suggesting further improvement via adaptive PCA and volume-weighted signals.

1. Introduction

Statistical arbitrage encompasses a wide range of investment strategies that share several defining characteristics. First, these strategies rely on systematic, rule-based trading signals that do not primarily depend on fundamental analysis. Second, portfolios constructed under this framework are typically designed to be market-neutral, with an overall beta that remains close to zero relative to broader market movements. Third, excess returns are sought through statistical methods by exploiting a large number of positions, each with a positive expected value. Such diversification is intended to yield relatively stable, low-volatility performance with limited correlation to market indices. Finally, the holding periods for statistical arbitrage strategies vary widely, ranging from ultra-short-term trades lasting only seconds to positions maintained over days, weeks, or even longer horizons.

Pairs trading is widely regarded as the foundational strategy within statistical arbitrage. It typically involves identifying two stocks—often from the same industry or with similar characteristics—that have historically exhibited closely correlated price movements. Let P_t and Q_t denote the price series of stocks P and Q, respectively. Their long-term relationship can be modeled as

Pairs trading is widely regarded as the foundational strategy within statistical arbitrage. It typically involves identifying two stocks—often from the same industry or with similar characteristics—that have historically exhibited closely correlated price movements. Let P_t and Q_t denote the price series of stocks P and Q, respectively. Their relationship can be modeled as:

$$ln\,(P_t) = \mu(t - t_0) + \beta ln\,(Q_t) + \varepsilon_t \tag{1}$$

or in differential form:

$$dP_t = \mu d_t + \beta dQ_t + d\varepsilon_t \tag{2}$$

where the residual term ε_t—known as the cointegration residual—captures deviations from the equilibrium between the two stocks and is assumed to be stationary or mean-reverting. In practice, the drift component is often negligible relative to fluctuations in ε_t and is therefore omitted. A typical pairs trading strategy involves going long stock P and shorting β units of stock Q when $\varepsilon_t < 0$ (indicating undervaluation of P relative to Q), and reversing the positions when $\varepsilon_t > 0$. Profits are realized as prices revert toward equilibrium. This mean-reverting

behavior is frequently interpreted as a market correction of temporary mispricings or overreactions, as discussed in Lo and MacKinlay (1990) and Pole (2007).

An alternative scenario arises when one stock persistently outperforms the other over a prolonged period. In such cases, the spread (or cointegration residual) fails to exhibit stationarity, and the price relationship may reflect a structural divergence rather than a transitory mispricing. As this thesis primarily focuses on mean-reverting relationships, these non-stationary cases are not examined in detail.

Beyond standard pairs trading, an extension—commonly referred to as generalized pairs trading—involves comparing individual stocks to benchmarks or indices rather than to other single securities. For example, within a specific sector such as biotechnology, each stock can be evaluated against a relevant sector index or ETF. By applying regression or cointegration techniques similar to those used in classical pairs trading, it is possible to identify stocks that are undervalued or overvalued relative to the sector benchmark.

Portfolios constructed under this framework typically involve long positions in stocks deemed undervalued and short positions in those considered overvalued, with each position hedged by its estimated exposure to the benchmark ETF or sector factor. Due to this hedging structure, the net exposure to the benchmark is minimal, making the resulting portfolio economically equivalent to a long–short equity strategy.

This thesis focuses on the construction, implementation, and performance evaluation of such statistical arbitrage strategies. Particular attention is given to the mean-reversion dynamics of residuals and the implications of factor selection for trading outcomes. Scenarios involving non-stationary residuals, which suggest persistent divergence rather than temporary mispricing, are deliberately excluded from the analysis.

Residual analysis forms the cornerstone of this research. Trading signals are generated based on relative value pricing within a sector or among groups of comparable firms, measured against benchmarks or latent factors. This is achieved by decomposing individual stock returns into systematic and idiosyncratic components, with the modeling effort focusing specifically on the idiosyncratic part, which captures deviations from broader market or sector trends. Formally, the decomposition can be written as:

$$r_{i,t} = u_i + \sum_{j=1}^{n} \beta_{ij} f_{j,t} + \varepsilon_{i,t} \qquad (3)$$

where $f_{j,t}$ denotes the returns of the $j-th$ systematic factor, β_{ij} is the exposure of stock i to that factor, and $\varepsilon_{i,t}$ is the idiosyncratic residual unexplained by the factor model.

A key practical challenge lies in determining the appropriate factor specification. While a similar issue arises in classical portfolio theory—typically in the context of selecting risk factors for asset pricing or risk management—the objective here is different. Rather than optimizing exposures to systematic factors, this study emphasizes removing them, as the dynamics of the residuals are what generate relative-value signals and ultimately determine profitability in statistical arbitrage.

The core contribution of this research lies in examining how different choices of factor sets affect the behavior of residuals and, by extension, the performance of the trading strategy. Alternative factor specifications induce distinct residual dynamics, which in turn generate different trading opportunities and PnL outcomes.

Recent research on mean-reversion and contrarian strategies has expanded beyond early foundational studies (e.g., Poterba and Summers, 1988; Lehmann, 1990; Lo and MacKinlay, 1990), incorporating more sophisticated models of return dynamics and factor structures. For example, Khandani and Lo (2009) examined market-neutral contrarian strategies during the 2007 liquidity crisis by constructing dollar-neutral portfolios based on past return quantiles, effectively trading "winners versus losers." While their approach relies on fixed trading intervals, more recent studies have shifted toward strategies that adapt trading frequency and rely on residual returns rather than total returns.

In particular, Kelly et al. (2020) introduce instrumented principal component analysis, while Gu et al. (2021) develop deep learning–based factor models—both of which enhance the explanatory power of systematic components and reduce residual noise. Building on these advances, this study designs mean-reversion strategies that explicitly remove factor-driven variation in returns. Rather than modeling raw price behavior, we focus on the dynamics of residuals after accounting for a chosen set of risk factors, thereby identifying relative mispricings that are more robust to market-wide movements.

This paper is organized as follows. Section 2 reviews the literature on mean-reversion strategies, market-neutral portfolio construction, and factor-based statistical arbitrage. It begins with foundational work on contrarian and mean-reverting behavior in asset prices, including early empirical evidence of return predictability. The review then turns to more recent studies that emphasize the role of risk factors in generating residual-based trading signals and the progression from simple linear models to more advanced approaches such as PCA, machine learning, and deep factor models. Special attention is given to research comparing different factor specifications—such as fundamental-based factors (e.g., Fama–French), statistical factors (e.g., PCA), and industry ETF proxies—in the context of designing market-neutral strategies. This discussion highlights both the theoretical rationale and empirical findings that motivate the methodologies adopted in this study.

Section 3 introduces the methodology. We explore the concept of market neutrality through two alternative approaches. The first extracts latent risk factors using PCA, following the framework introduced by Jolliffe (2011). The second constructs risk factors based on a set of industry-sector ETFs that serve as proxies for systematic market influences. Consistent with prior literature, our results show that applying PCA to the correlation matrix of a broad cross-section of Chinese equities yields statistically significant components that can often be interpreted as long–short portfolios across industry groups. Interestingly, the individual stocks with the highest loadings on each PCA-derived factor are not necessarily the largest firms by market capitalization within their sectors. This suggests that PCA-based factor construction may be less biased toward large-cap stocks compared to ETF-based methods, which typically follow capitalization-weighted schemes. Moreover, we observe that the proportion of total variance explained by a fixed number of PCA components fluctuates over time. This temporal variation implies that the dimensionality required to capture systematic return variation is itself dynamic, potentially reflecting changes in macroeconomic conditions, market structure, or investor risk appetite. These dynamics may help explain the performance differences observed between PCA-based and ETF-based strategies.

Sections 4 and 5 describe the construction of trading signals based on estimated residual processes. For each stock, residuals are computed at the close of each trading day using a rolling window of 60 trading days, roughly corresponding to one earnings cycle. This estimation is performed using only data prior to the trading day, thus replicating a real-time investment setting

and avoiding look-ahead bias. Trading signals are generated when the standardized residual (s-score) deviates significantly from its recent mean. Specifically, a position is opened when the s-score exceeds ±1.25, and is closed when the residual reverts within asymmetric thresholds (0.5 for long positions and 0.75 for short positions). These thresholds are applied uniformly across all stocks to ensure consistency and reduce overfitting.

The analysis uses daily end-of-day (EOD) data. The backtest period begins in 2005 for PCA-based strategies and in 2004 for ETF-based strategies, reflecting differences in data availability. The investment universe is restricted to Chinese stocks with a market capitalization above 2 billion RMB on the trading date, a filter that helps mitigate survivorship bias by excluding firms that fall below this threshold at any point in the historical sample. To preserve model transparency and reduce the risk of data mining, the residual estimation procedure and trading rules are deliberately kept simple and consistent across assets and time periods.

Section 6 presents a series of backtests evaluating the performance of various trading strategies, each employing a distinct method of factor construction for residual estimation. Specifically, we examine three configurations: (i) synthetic ETFs constructed from capitalization-weighted indices, (ii) a fixed number of PCA components, and (iii) a dynamic number of PCA components selected to exceed a specified explained-variance threshold. For all strategies, the residuals used to generate signals are estimated strictly out-of-sample, using only data from the preceding 60 trading days. To enhance the realism of execution modeling, we account for transaction costs by applying a 30-basis-point per-trip cost. This adjustment ensures that the reported performance metrics incorporate trading frictions commonly encountered in practice.

Section 7 provides a sensitivity analysis, exploring how changes in key parameters—such as the number of PCA components retained, the explained-variance threshold, the thresholds for opening and closing positions, and the rolling window length—affect mean-reversion performance and profitability. The results show that while the default settings (e.g., a 60-day rolling window with entry at ±1.25 standard deviations and asymmetric exit thresholds of 0.5 for longs and 0.75 for shorts) are generally robust, alternative parameter choices can yield different risk–return tradeoffs.

Finally, Section 8 summarizes the key findings and outlines directions for future research. This chapter highlights the practical relevance of PCA-based pairs trading, the value of sector ETF hedging, and the role of factor dimensionality—where sensitivity analysis indicates that dynamic PCA can be effective when explained-variance thresholds are appropriately calibrated (around 60–70%). It also points to the potential for further improvements using advanced machine learning techniques, alternative data sources, and adaptive thresholding mechanisms.

Overall, the thesis demonstrates that the choice of factor construction method—whether based on PCA eigenportfolios or industry-specific ETF proxies—materially influences residual-based signals, risk exposures, and empirical strategy performance. By systematically comparing and refining these approaches, the research contributes to a deeper understanding of how to design resilient, mean-reverting trading strategies that adapt to changing market conditions while maintaining market neutrality, thereby extending the literature on statistical arbitrage in the context of the Chinese equity market.

2. Literature Review

The study of mean-reversion and contrarian strategies has a long-standing history in financial economics. Foundational works by Poterba and Summers (1988), Lehmann (1990), and Lo and MacKinlay (1990) document statistically significant return reversals—at long horizons in the case of Poterba and Summers, and at short horizons in the case of Lehmann and Lo–MacKinlay. These findings laid the groundwork for market-neutral long–short equity strategies and inspired the development of systematic statistical arbitrage frameworks, such as the large-scale contrarian portfolios analyzed by Khandani and Lo (2009). In recent years, this research area has evolved substantially, incorporating residual-based signal construction, advanced factor models—including those derived via PCA—and machine learning techniques. This section reviews relevant literature across six interconnected areas: residual-based trading, PCA-based factor modeling and pair selection, machine learning applications, theoretical developments in contrarian strategies, reinforcement learning methods, and the role of liquidity and trading frictions.

PCA has been widely adopted for factor extraction and pair selection due to its ability to reduce dimensionality and identify latent risk factors embedded in asset return data. In early

applications (e.g., Avellaneda and Lee, 2010), PCA was used to construct market-neutral portfolios by regressing stock returns on principal components. More recent studies have extended this idea with richer models, broader datasets, and integration with machine learning.

Avellaneda and Lee (2010) present a seminal study on model-driven statistical arbitrage in the U.S. equities market, introducing a systematic approach based on the mean reversion of idiosyncratic returns. They propose two frameworks for extracting systematic risk factors: PCA and regressions on sector ETFs. In both cases, residual returns—after controlling for systematic components—are modeled as OU processes, enabling contrarian trading strategies that exploit temporary mispricings. Their backtests over 1997–2007 show that PCA-based approaches yield higher Sharpe ratios on average, particularly before 2003, while ETF-based models remain more stable in the later years. Notably, they also demonstrate that incorporating trading volume into signal generation improves performance, especially for ETF-based strategies. Furthermore, they link strategy effectiveness to broader market conditions, documenting deteriorating performance during periods of low volatility and systemic stress, such as the 2007 liquidity crisis. Their work not only provides robust empirical evidence but also offers a flexible framework for market-neutral strategies grounded in cross-sectional factor structures.

Sarmento and Horta (2020) propose a novel two-step machine learning framework to enhance traditional pairs trading. First, they apply PCA for dimensionality reduction, extracting latent risk factors, and then employ the OPTICS clustering algorithm to identify statistically robust candidate pairs. This unsupervised selection method significantly outperforms standard approaches, yielding an average portfolio Sharpe ratio of 3.79, compared with 3.58 for the best conventional method and 2.59 for a simpler baseline. Second, to mitigate episodes of persistent divergence and prolonged losses, they introduce a forecasting-based trading model using ARMA, LSTM, and LSTM encoder–decoder architectures. This predictive step reduces the number of days with portfolio drawdowns by approximately 75 %, albeit at the cost of some reduction in overall profitability. Their framework is backtested on 208 commodity-linked ETFs using 5-minute bar data from 2009 to 2018, with transaction costs duly incorporated.

Caneo and Kristjanpoller (2021) investigate statistical arbitrage in six Latin American equity markets (2013–2017), covering 338 stocks, by applying a PCA-based multifactor model to construct mean-reverting spreads. They propose two adaptive threshold selection criteria using

moving training–trading windows, and show that their strategy achieves an average Sharpe ratio about 1.55 higher than that of the corresponding regional market benchmarks. Their analysis also reveals that the largest eigenvalue is statistically dominant, with the associated eigenportfolio closely co-moving with the overall market. Moreover, the number of significant components varies inversely with market volatility, highlighting the dynamic structure of risk decomposition.

Xiang and He (2022) propose an innovative asset pricing model in which arbitrage factors derived from pairs trading are employed to explain the equity risk premium in the Chinese stock market. Their approach begins with the construction of mean-reverting spreads using pairs trading strategies across a large sample of A-share stocks. These spreads, modeled as OU processes, capture idiosyncratic deviations between paired securities. To extract common variation in these residuals, the authors apply PCA and interpret the leading principal components as systematic pricing factors. Using Fama–MacBeth two-pass regressions and double-sorting portfolio analysis, they demonstrate that these PCA-derived arbitrage factors are strongly associated with cross-sectional variation in future stock returns, with explanatory power that remains robust after controlling for traditional firm characteristics. The study suggests that market inefficiencies captured via statistical arbitrage can constitute systematic risk sources within a rational asset-pricing framework.

Gatta et al. (2023) introduce a multi-horizon clustering framework for statistical arbitrage that leverages PCA-derived risk factors computed at both daily and monthly frequencies. After extracting these factors, they apply an Adaptive Lasso procedure for feature selection to identify those most relevant in explaining each stock's returns. Stocks are then clustered based on their exposure profiles, and within each cluster, common-factor exposures are neutralized to isolate idiosyncratic spread components. An optimal market-neutral portfolio is subsequently constructed using constrained optimization via Sequential Least Squares Programming. Thorough backtests across six major equity markets—the Italian, German, U.S., Japanese, Brazilian, and Indian—demonstrate robust outperformance relative to minimum-variance, mean-variance, and Exponential Gradient benchmarks, confirming the effectiveness and generality of the approach.

Han et al. (2023) propose a novel pairs trading framework that combines unsupervised learning techniques with PCA. Rather than relying solely on cointegration- or distance-based

criteria, they construct a multidimensional feature set comprising both return-based signals and firm-specific fundamental variables. PCA is employed to reduce dimensionality and extract dominant variations, which are then input into clustering algorithms such as k-means, DBSCAN, and agglomerative clustering to form groups of similar stocks. Within each cluster, trading pairs are selected based on short-term (one-month) return divergence or momentum rankings. Empirical tests across U.S., U.K., and Hong Kong equity markets show that the agglomerative clustering strategy delivers an annualized return of 24.8% and a Sharpe ratio of 2.69, with drawdowns limited to 12.3%. In contrast, clustering based solely on price-based features yields a Sharpe ratio of only ~1.44 and a drawdown close to 50%, underscoring the critical role of incorporating firm-specific information in machine learning–driven statistical arbitrage strategies.

Rotondi and Russo (2024) present a clustering-based pairs trading strategy that incorporates PCA-derived distance metrics for stock pair selection. They evaluate three distance measures— (1) Euclidean distance, (2) PCA-transformed Euclidean distance, and (3) a novel partial correlation–based distance—using S&P 500 constituents over the period 2000–2023. After clustering, a simple long–short strategy is implemented exclusively on the identified pairs, yielding statistically and economically significant excess returns net of transaction costs. Specifically, average monthly excess returns range from 36 to 41 basis points, with Sharpe ratios between 0.20 and 0.30, corresponding to annualized Sharpe ratios of 0.72 to 0.99. The partial correlation–based distance provides the highest risk-adjusted performance, attributed to superior clustering accuracy as measured by industry-sector purity. These results remain robust across sensitivity checks, suggesting that PCA-enabled clustering offers a meaningful enhancement for pairs trading.

Krause and Calliess (2024) propose a deep learning–based statistical arbitrage framework that generalizes traditional PCA residual models through the use of an autoencoder architecture. While classical statistical arbitrage methods typically rely on PCA to extract linear mean-reverting components from asset returns, the authors design an end-to-end policy network in which the autoencoder simultaneously learns nonlinear latent structures and optimizes trading performance. Empirically, PCA-based models using OU residuals achieve Sharpe ratios of 0.87 to 0.96 and annualized returns of 3.9% to 4.6%, with volatility around 4.4% to 4.7%. In contrast, the autoencoder-based policy network delivers Sharpe ratios between 1.42 and 1.81, mean returns of 5% to 6%, and substantially lower volatility (~3.4% to 3.7%). These findings

demonstrate that nonlinear, end-to-end learning with autoencoders can outperform PCA in capturing profitable mean-reversion signals and improving risk-adjusted performance in statistical arbitrage.

In summary, the recent literature provides strong empirical and methodological support for residual-based trading strategies and PCA-derived factor models in statistical arbitrage. PCA has proven effective not only for extracting systematic components from stock returns but also for identifying mean-reverting spreads through clustering and distance-based approaches. Collectively, these studies provide robust evidence that isolating idiosyncratic return components is essential for designing market-neutral strategies, and that PCA remains a valuable tool in both residual estimation and pair selection. These insights directly motivate the PCA-based residual-driven trading framework developed in the next section.

3. Quantitative Construction of Risk Factors and Market-Neutral Residuals

Consider a universe of N stocks, each with returns denoted by R_i over a given trading period (e.g., daily log returns). We begin with a single-factor specification, where the market portfolio—typically proxied by a capitalization-weighted index such as the S&P 500—has returns denoted by F. Each stock's return can then be expressed as:

$$R_i = \beta_i F + \tilde{R}_i \tag{4}$$

where β_i measures the sensitivity of stock i to the market factor F, and $\tilde{R}_i$ represents the idiosyncratic component, assumed to be uncorrelated with F.

Extending this to a multi-factor framework, the generalized form is:

$$R_i = \sum_{j=1}^{m} \beta_{ij} F_j + \tilde{R}_i \tag{5}$$

where m denotes the number of systematic factors, F_j represents the return of factor j, and β_{ij} is the factor loading of stock i on factor j.

A portfolio composed of stock weights $\{Q_i\}_{i=1}^{N}$ is market-neutral if it has no net exposure to any of the systematic risk factors. Formally, this condition is met when the portfolio's aggregate factor exposures (i.e., factor betas) satisfy:

$$\beta_j^{portfolio} = \sum_{i=1}^{N} \beta_{ij} Q_i = 0, \ j = 1,2,\dots,m \tag{6}$$

Under this condition, portfolio returns are uncorrelated with factor movements and are entirely driven by idiosyncratic components:

$$\sum_{i=1}^{N} Q_i R_i = \sum_{i=1}^{N} Q_i \left(\sum_{j=1}^{m} \beta_{ij} F_j + \tilde{R}_i\right) = \sum_{j=1}^{m}\left(\sum_{i=1}^{N} \beta_{ij} Q_i\right) F_j + \sum_{i=1}^{N} Q_i \tilde{R}_i = \sum_{i=1}^{N} Q_i \tilde{R}_i \quad (7)$$

Empirical evidence from developed markets—particularly the G8 economies—suggests that approximately 10 to 20 factors are generally sufficient to capture the majority of systematic risk, with around 15 being most commonly used. Studies such as Laloux et al. (2000) and Plerou et al. (2000) report that these systematic factors can explain roughly 50% of the total variance in stock returns. The central challenge, therefore, lies in identifying and extracting these latent systematic components effectively in order to construct robust market-neutral portfolios.

3.1 The PCA Method for Factor Extraction

PCA is a widely adopted statistical technique for extracting latent risk factors from financial datasets. It is particularly well-suited for high-dimensional settings, such as equity returns, where the number of assets N may be large relative to the number of time observations M. In the context of pairs trading and statistical arbitrage, PCA facilitates the identification of systematic structures and the isolation of residual (idiosyncratic) components that may exhibit mean-reverting behavior.

Standardized Returns and Correlation Matrix Construction

Let $S_i(t)$ denote the adjusted price of stock i at time t, accounting for dividends and splits. The daily log return of stock i over a window of M trading days prior to a reference date t_0 is computed as:

$$R_{ik} = \ln\left(\frac{S_i(t_0 - (k-1)\Delta t)}{S_i(t_0 - k\Delta t)}\right), \ k = 1, \dots, M, i = 1, \dots, N \quad (8)$$

where $\Delta t = \frac{1}{252}$ corresponds to one trading day. To control for heteroskedasticity across assets, the returns are standardized:

$$Y_{ik} = \frac{R_{ik} - \bar{R}_i}{\sigma_i} \quad (9)$$

with mean and standard deviation defined by:

$$\bar{R}_i = \frac{1}{M}\sum_{k=1}^{M} R_{ik}\,,\sigma_i^2 = \frac{1}{M-1}\sum_{k=1}^{M}(R_{ik} - \bar{R}_i)^2$$

From the matrix of standardized returns $Y \in \mathbb{R}^{N \times M}$, the empirical correlation matrix $\rho \in \mathbb{R}^{N \times N}$ is defined as:

$$\rho_{ij} = \frac{1}{M-1}\sum_{k=1}^{M} Y_{ik}Y_{jk} \tag{10}$$

By which is symmetric and positive semi-definite. By construction, the diagonal elements satisfy:

$$\rho_{ii} = \frac{1}{M-1}\sum_{k=1}^{M} Y_{ik}^2 = \frac{1}{M-1}\frac{\sum_{k=1}^{M}(R_{ik}-\bar{R}_i)^2}{\sigma_i} = 1 \tag{11}$$

In practical applications, the correlation matrix ρ is often high-dimensional—typically 500×500 or larger—while the available data is relatively limited. Using an excessively long historical window ($M \gg N$) incorporates outdated information and may fail to reflect current market dynamics. Conversely, restricting the estimation to a short recent window ($M < N$) produces a noisy and potentially ill-conditioned correlation matrix, a well-documented issue in the literature on Random Matrix Theory

To balance this trade-off, we adopt a rolling one-year (252 trading days) window for correlation estimation. This captures recent market structure while avoiding excessive lookback bias. PCA is then applied to extract the dominant eigenportfolios, filtering out noise and providing a stable low-rank approximation even in high-dimensional settings.

Eigen-Decomposition and Spectral Structure

The correlation matrix ρ is then subjected to eigen-decomposition. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$ denote the ordered eigenvalues, and $v^{(j)} \in \mathbb{R}^N$ the corresponding eigenvectors. The spectral decomposition is given by:

$$\rho = \sum_{j=1}^{N} \lambda_j v^{(j)}(v^{(j)})^T \tag{12}$$

Empirically, the spectrum of ρ typically displays a few large eigenvalues that are well separated from the rest of the spectrum, which is referred to as the bulk or noise spectrum (see Figure 1A and 1B). These leading eigenvalues correspond to systematic factors, such as market-

wide or sector-specific components, whereas the bulk is mainly attributed to idiosyncratic fluctuations.

To characterize the distribution of eigenvalues more formally, we define the empirical spectral density as:

$$D(x, y) = \frac{\#\{\lambda_j \in (x, y)\}}{N} \tag{13}$$

which measures the proportion of eigenvalues lying within the interval (x, y). This function highlights the dominance of low-variance (noisy) components within the spectrum (see Figure 2).

Selecting Significant Components and Eigenportfolios

To construct a low-rank approximation of the correlation structure, two main approaches are commonly employed to determine the number of principal components m to retain. The first approach is fixed-rank truncation, in which a predetermined number of leading eigenvectors are selected—often based on economic intuition, such as the number of major industry sectors in the market. The second approach is variance thresholding, where the smallest number of components m is retained such that the cumulative explained variance exceeds a predefined threshold α, typically set between 80% and 90%. This method ensures that the selected components capture the majority of systematic variation present in the data, while discarding lower-variance components likely associated with noise.

$$\frac{\sum_{j=1}^{m} \lambda_j}{\sum_{j=1}^{N} \lambda_j} \geq \alpha$$

Once the top m components are selected, eigenportfolios are defined by rescaling the eigenvectors using the inverse volatility of each stock:

$$Q_i^{(j)} = \frac{v_i^{(j)}}{\sigma_i}, \qquad j = 1, \dots, m$$

where $v_i^{(j)}$ is the i-th element of the j-th eigenvector and σ_i is the volatility of stock i. The return of the j-th eigenportfolio at time k is then given by:

$$F_{jk} = \sum_{i=1}^{N} \frac{v_i^{(j)}}{\sigma_i} R_{ik} \tag{14}$$

By construction, these factor returns $\{F_{jk}\}$ are mutually uncorrelated and represent latent systematic drivers of asset returns. The return of each asset can therefore be expressed as:

$$R_{ik} = \sum_{j=1}^{m} \beta_{ij} F_{jk} + \tilde{R}_{ik} \tag{15}$$

where $\tilde{R}_{ik}$ denotes the idiosyncratic residual, uncorrelated with the extracted factors.

This formulation implies a low-rank approximation of the correlation matrix:

$$\rho_{ab} = \sum_{j=1}^{m} \lambda_j v_a^{(j)} v_b^{(j)} + \eta_a^2 \delta_{ab} \tag{16}$$

where $a, b = 1, \dots, N$ are asset indices, and

$$\eta_a^2 = 1 - \sum_{j=1}^{m} \lambda_j (v_a^{(j)})^2$$

ensures that the diagonal elements remain normalized to unity. The Kronecker delta δ_{ab} captures the idiosyncratic noise component.

In this way, PCA separates the systematic cross-sectional structure of returns from residual noise, yielding a set of orthogonal factors suitable for risk modeling, market-neutral portfolio construction, and residual-based trading signal generation.

Figure 1A. Top 50 eigenvalues of the correlation matrix of A-share stock returns
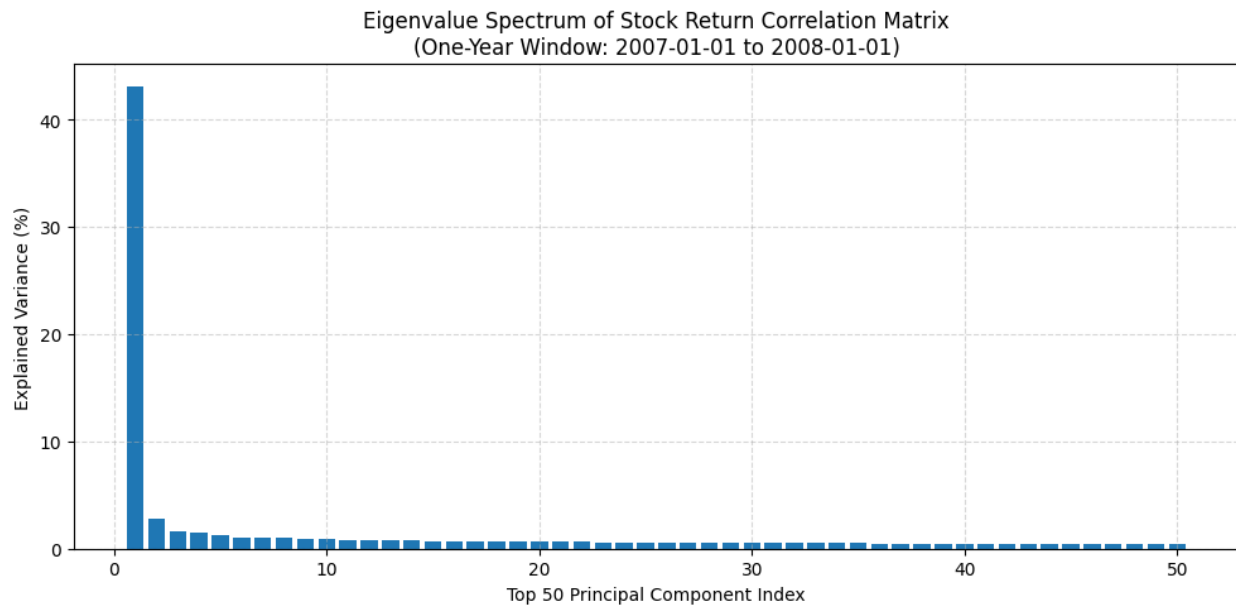
Note: Top 50 eigenvalues of the correlation matrix of A-share stock returns computed using a one-year window ending on 1 January 2008. The eigenvalues are derived from daily return data for a filtered universe of 303 stocks

with imputed values, and are shown as percentages of explained variance. The first principal component (market eigenvalue) dominates the spectrum, followed by a steep decay characteristic of common sector or idiosyncratic effects.
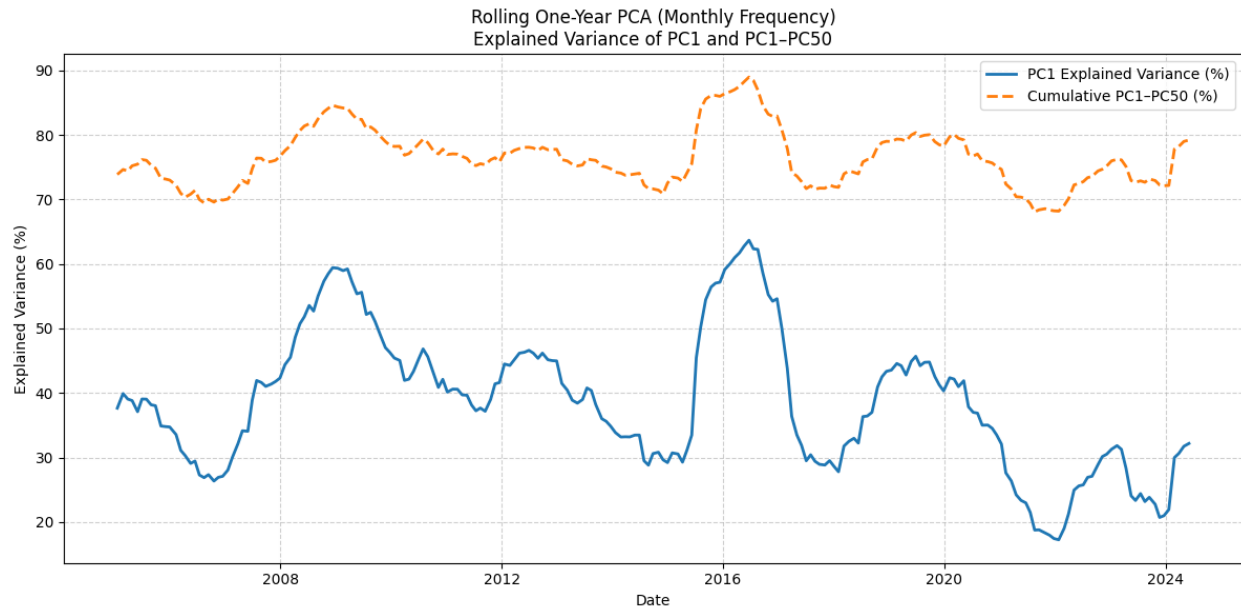
Figure 1B. Time series of the PC1 explained variance and the cumulative PC1–PC50

Note: Time series of the explained variance of the first principal component (PC1) and the cumulative variance explained by the top 50 components (PC1–PC50), based on a rolling one-year PCA using daily returns from 303 A-share stocks. The window rolls monthly from 2007 to 2024. The plot reveals time-varying strength of the market component and fluctuations in the overall explanatory power of the principal subspace, reflecting evolving cross-sectional structure in return correlations.

Figure 2. The distribution of PCA eigenvalues
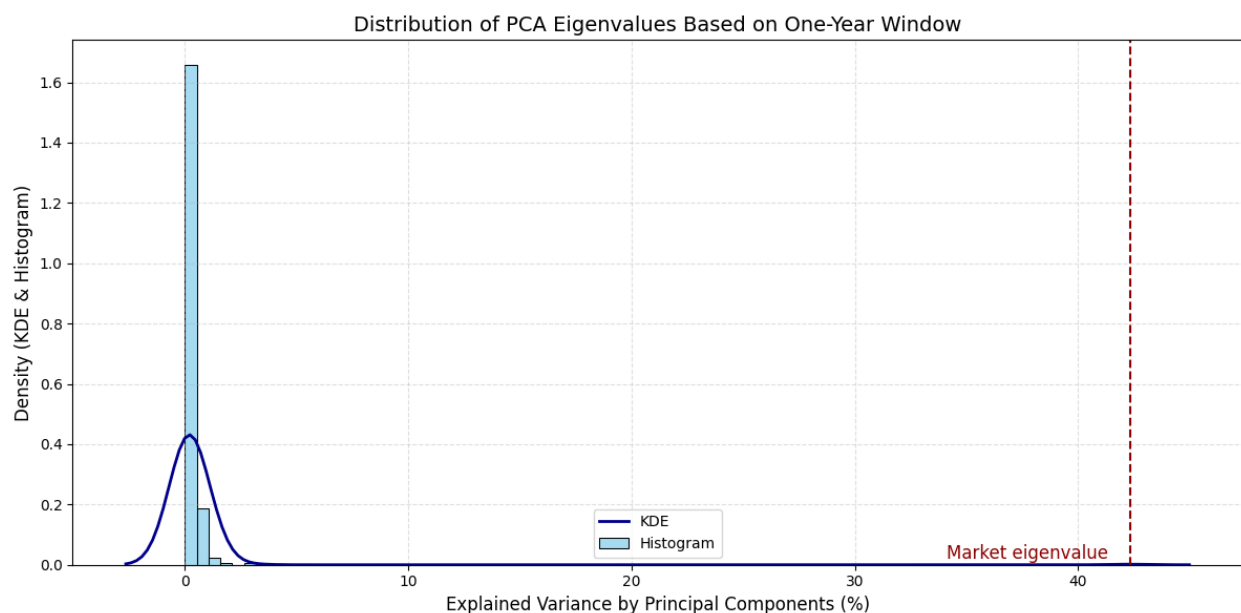
Note: The distribution of PCA eigenvalues estimated using a one-year window ending on 1 January 2008, based on daily returns of 303 A-share stocks. The histogram and KDE curve show a clear separation between a large 'market eigenvalue' and the rest of the bulk spectrum. The bulk of the eigenvalues are densely clustered near zero, while the dominant eigenvalue—highlighted in red—captures a significant proportion of the total variance, consistent with the presence of a strong market-wide factor.

3.2 Interpretation of Eigenvectors and Eigenportfolios

Previous studies (e.g., Laloux et al., 2000) show that the leading eigenvector of the return correlation matrix typically corresponds to the market mode. This interpretation arises from the fact that all components of the first eigenvector $v_i^{(1)}$, for $i = 1, 2, \ldots, N$, are positive, implying that the associated eigenportfolio assigns positive weights to all assets. Formally, the portfolio weights are defined as:

$$Q_i^{(1)} = \frac{v_i^{(1)}}{\sigma_i}$$

where σ_i denotes the volatility of asset i. This weighting scheme produces allocations inversely proportional to asset volatilities, which empirically tends to approximate capitalization-weighted market portfolios. Although not identical, the market-mode eigenportfolio and the capitalization-weighted benchmark exhibit a strong resemblance, as illustrated in Figure 3.

In contrast, the interpretation of the remaining eigenvectors is less straightforward. These higher-order eigenvectors necessarily contain both positive and negative components to ensure orthogonality with the leading eigenvector. However, due to the absence of an intrinsic ordering among stocks—unlike in applications such as interest-rate term structure PCA (Litterman, 1991) or volatility surface decomposition (Cont and Da Fonseca, 2002)—a direct analogy to "shape modes" is not applicable.

Instead, following Scherer and Avellaneda (2002) and Plerou et al. (2000), we rank the components of each eigenvector $v^{(k)}$ in descending order:

$$v_{n_1}^{(k)} \geq v_{n_2}^{(k)} \geq \cdots \geq v_{n_N}^{(k)}$$

where $\{n_1, n_2, \ldots, n_N\}$ is a permutation of asset indices based on coefficient magnitude. This reordering often reveals coherence: firms with adjacent ranks tend to belong to the same industry group. Empirical results from Table 1 and 2 along with Figure 4, 5 and 6 confirm this phenomenon, particularly for the second and third eigenvectors. However, coherence weakens for deeper eigenvectors and eventually disappears within the bulk of the spectrum, where components predominantly capture unstructured, idiosyncratic noise.

Accordingly, eigenportfolios constructed from intermediate eigenvectors can be interpreted as long–short portfolios contrasting industries or sectors. This aligns with the view that the eigenvalue spectrum reflects a hierarchy of market influences: from broad market-wide movements, through sectoral dynamics, down to residual noise.

Figure 3. Comparative evolution of the principal eigenportfolio and the capitalization-weighted portfolio
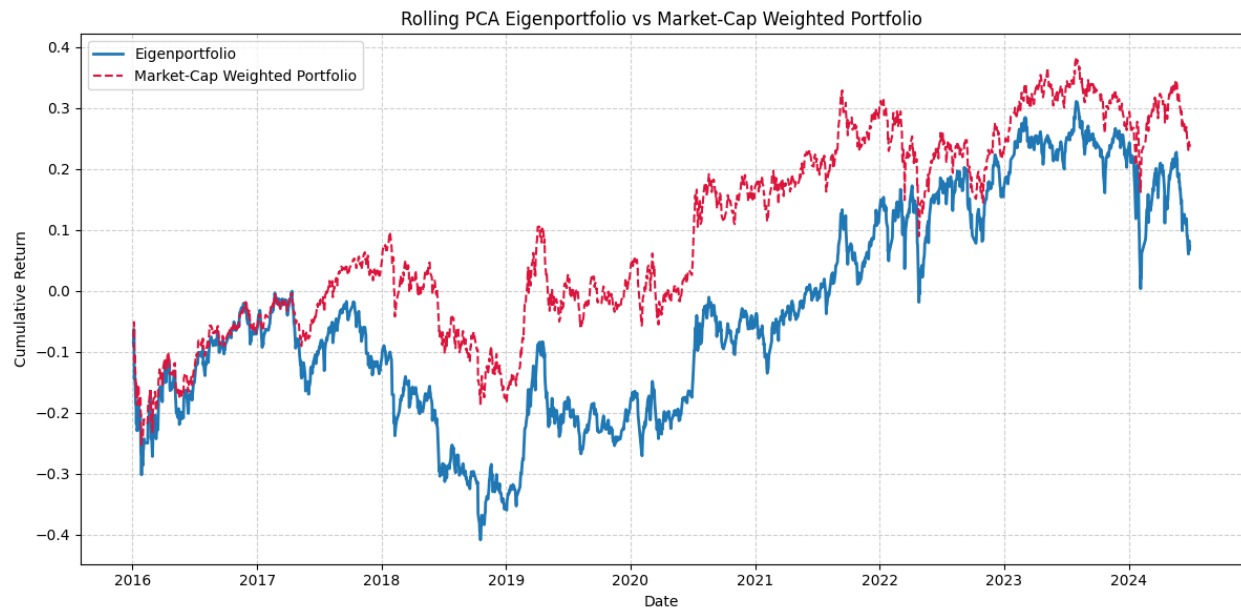
Note: Comparative evolution of the principal eigenportfolio and the capitalization-weighted portfolio from January 2016 to January 2024. The eigenportfolio is constructed using rolling PCA with a one-year window and updated monthly. Both portfolios exhibit broadly similar trends over time, with the eigenportfolio capturing the main dynamics of the market.

Figure 4. First principal component loadings
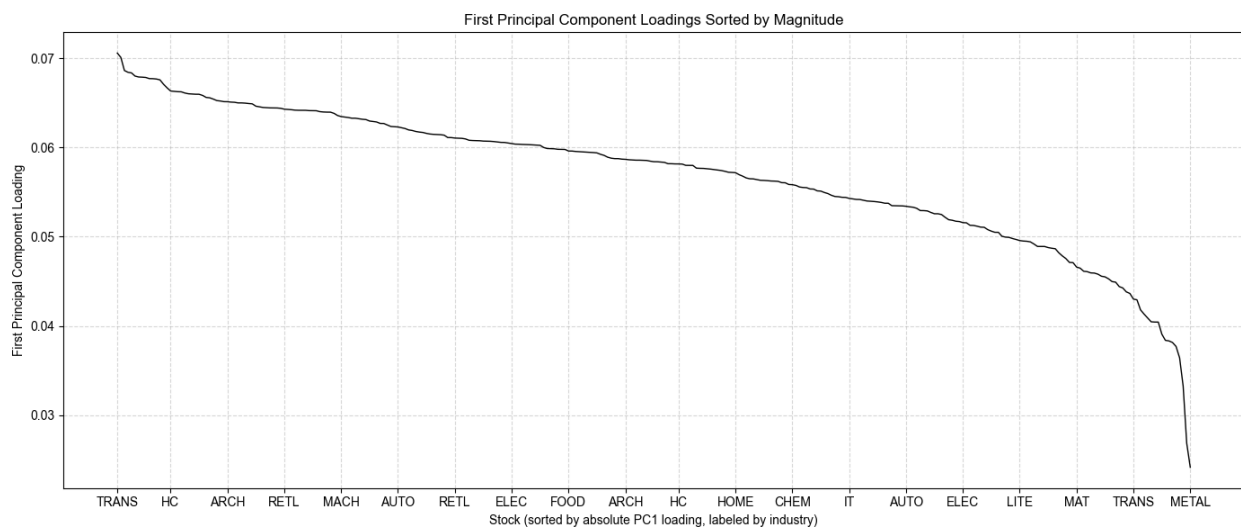
Note: presents the loadings of the first principal component (i.e., the leading eigenvector) derived from the PCA of the stock return correlation matrix. The loadings are sorted by their absolute magnitude, highlighting the stocks that contribute most strongly to the market-wide common component. Each stock is annotated along the x-axis with the abbreviation of its corresponding industry sector. The higher the absolute loading, the more the stock moves in sync with the market's dominant co-movement structure captured by the first principal component. The industry sector abbreviations are as follows: AGRI (Agriculture, Forestry, Animal Husbandry and Fishery), CHEM (Basic Chemicals), STEEL (Steel), METAL (Non-ferrous Metals), ELEC (Electronics), HOME (Household Appliances), FOOD (Food and Beverage), TEXT (Textile and Apparel), LITE (Light Industry Manufacturing), HC (Healthcare – Pharmaceuticals and Biotechnology), UTIL (Utilities), TRANS (Transportation), RE (Real Estate), RETL (Commerce and Retail), SERV (Social Services), CONS (Conglomerates), AUTO (Automotive), BANK (Banking), FIN (Non-bank Financials), MAT (Building Materials), ARCH (Construction and Decoration), POWER (Electrical Equipment), MACH (Machinery), DEF (National Defense and Military Industry), IT (Information Technology), MEDIA (Media), COMM (Telecommunications), COAL (Coal), OIL (Oil and Petrochemicals), ENV (Environmental Protection), and BEAU (Personal Care and Beauty).

Figure 5. Second principal component loadings
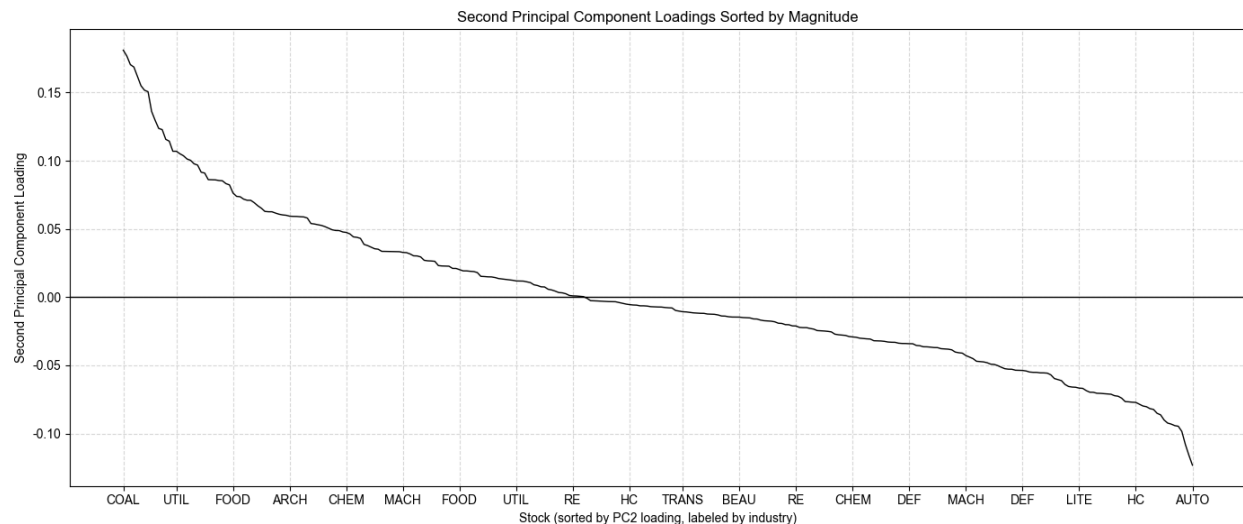
Note: This figure plots the loadings of the second eigenvector obtained from the PCA of the stock return correlation matrix. The stocks are sorted in descending order of their PC2 coefficients (not by absolute value), and each is labeled with the abbreviation of its corresponding industry sector. The second principal component often reflects an industry-specific or style-based long-short structure, in contrast to the market-wide component captured by the first principal component. Industry abbreviations follow the same convention as in Figure 4, where, for example, "COAL" refers to coal industry, "UTIL" to utilities, "FOOD" to food & beverage, and so on.

Figure 6. Third principal component loadings
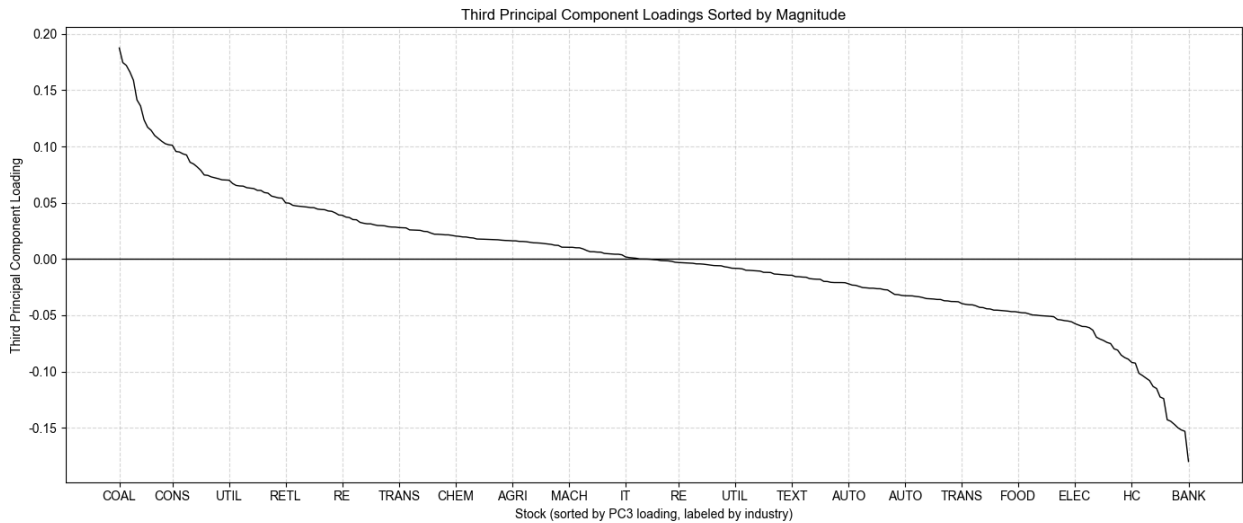
Note: This figure displays the loadings of the third principal component (PC3), obtained from the PCA of the stock return correlation matrix. Stocks are ordered from highest to lowest based on their PC3 loadings, and each is labeled by the abbreviation of its corresponding industry sector. Unlike the first component, which captures overall market direction, and the second component, which often reflects a sector-level long–short structure, the third component may highlight more subtle themes such as industry rotations or factor-based segmentation. The horizontal black line at zero marks the sign change in component contributions. Industry abbreviations follow the same notation as in [Figure 4](#).

Table 1. Top and bottom 10 stocks sorted by second principal component loading

Top 10 stocks			Bottom 10 stocks		
Code	**Name**	**Industry**	**Code**	**Name**	**Industry**
000937	China Coal Energy Co., Ltd.	Coal Mining	600303	Vantone Industrial Co., Ltd.	Automobile
600585	Anhui Conch Cement Company	Building Materials	000909	Digital China Information Service Co., Ltd.	Real Estate
600016	China Minsheng Banking Corp.	Banking	600884	Ningbo Thermal Power Co., Ltd.	Power Equipment
600348	Yangquan Coal Industry	Coal Mining	600644	Leshan Electric Power Co., Ltd.	Utilities
000898	Baotou Steel Union Co., Ltd.	Steel	600130	Infovision Optoelectronics	Electronics
600028	China Petroleum & Chemical Corp. (Sinopec)	Oil & Petrochemicals	000705	Wuhan Humanwell Healthcare Group	Healthcare & Biotech
600188	Yanzhou Coal Mining Company	Coal Mining	600283	Qianjiang Water Resources Dev Co.	Environmental Protection
600036	China Merchants Bank	Banking	600883	Inner Mongolia Yili Industrial	Agriculture & Animal

Top 10 stocks			Bottom 10 stocks		
Code	Name	Industry	Code	Name	Industry
000983	Shanxi Coking Coal Group Co., Ltd.	Coal Mining	600360	Hoshine Silicon Industry Co., Ltd.	Electronics
600015	Industrial and Commercial Bank of China (ICBC)	Banking	600235	Minfeng Special Paper Co., Ltd.	Light Manufacturing

Note: The top stocks are primarily from energy, coal, and banking sectors, while the bottom stocks are associated with real estate, utilities, and consumer industries. Stock codes, company names, and industry classifications are shown.

Table 2. Top and bottom 10 stocks sorted by third principal component loading.

Top 10 stocks			Bottom 10 stocks		
Code	Name	Industry	Code	Name	Industry
000937	China Coal Energy Co., Ltd.	Coal Mining	600016	China Minsheng Banking Corp.	Banking
600123	Shanxi Lanhua Sci-Tech Innovation Co., Ltd.	Coal Mining	600036	China Merchants Bank	Banking
000983	Shanxi Coking Coal Group Co., Ltd.	Coal Mining	600535	Tasly Pharmaceutical Group	Healthcare & Biotech
600348	Yangquan Coal Industry Group Co., Ltd.	Coal Mining	600518	Kangmei Pharmaceutical Co., Ltd.	Healthcare & Biotech
600188	Yanzhou Coal Mining Company	Coal Mining	600011	Huadian Power International Corp.	Utilities
000060	Minmetals Development Co., Ltd.	Non-ferrous Metals	600015	Industrial and Commercial Bank of China (ICBC)	Banking
600456	Baotou Huazi Industry Co., Ltd.	Non-ferrous Metals	600066	BAIC BluePark New Energy Technology Co., Ltd.	Automobile
600362	Jiangxi Copper Corporation Limited	Non-ferrous Metals	600085	Tonghua Dongbao Pharmaceutical Co., Ltd.	Healthcare & Biotech
600282	Nanjing Iron and Steel Company	Steel	600887	Yantai Changyu Pioneer Wine Co., Ltd.	Food & Beverage
600508	Shaanxi Coal and Chemical Industry Co., Ltd.	Coal Mining	600276	Jiangsu Hengrui Pharmaceuticals Co., Ltd.	Healthcare & Biotech

Note: Top stocks are predominantly from energy and metals sectors, including coal mining and non-ferrous metals, while the bottom stocks are concentrated in banking, pharmaceuticals, and consumer-related industries. Stock codes, company names, and industry classifications are shown.

3.3 The ETF-Based Industry Approach for Residual Extraction

An alternative to PCA for extracting residuals is to use sector-based ETFs as proxies for risk factors. In this approach, the returns of industry-representative ETFs serve as explanatory variables in stock-level single-factor regressions.

Table 3 reports industry classifications and the number of stocks with market capitalizations exceeding RMB 2 billion as of January 4, 2016. This breakdown contextualizes the relative size and importance of each sector in the equity universe. Each industry is linked to a corresponding sector ETF, which is used as a simplified factor in return modeling.

Unlike eigenportfolios—which are orthogonal by construction and hence produce uncorrelated factors—sector ETFs often exhibit substantial cross-correlation. Such correlation can introduce redundancy and multicollinearity: two related ETFs may both load significantly on a stock's return (sometimes with opposite signs) if the firm spans multiple activities, thereby hurting interpretability and inflating estimation variance in residuals.

To address these concerns, various regression techniques have been explored in the literature. One promising class includes sparse regression methods, which reduce overfitting by selecting only the most relevant explanatory variables. A well-known example is the Matching Pursuit algorithm ((Davis et al., 1997), which iteratively selects predictors that maximize explained variance while promoting parsimony. Another approach is Ridge Regression (Jolliffe, 2011), which imposes shrinkage penalties to mitigate multicollinearity among predictors.

In this study, we adopt a simple and interpretable one-to-one matching: each stock i is assigned to a single ETF based on its primary industry classification (see Table 3). Residual returns are then estimated via a univariate regression,

$$R_{i,t} = \alpha_i + \beta_i F_{g(i),t} + \tilde{R}_{i,t} \tag{17}$$

where $R_{i,t}$ is the return of stock i, $F_{g(i),t}$ is the return of the matched sector ETF for industry $g(i)$, β_i is the factor loading, and $\tilde{R}_{i,t}$ is the residual component not captured by the sector factor. If returns are demeaned in preprocessing, α_i can be set to zero by construction. This one-to-one design avoids the instability arising from correlated predictors and yields a transparent residual suitable for subsequent mean-reversion/statistical-arbitrage strategies.

Table 3. Summary statistics by industry based on market capitalization as of January 4, 2016

Index	Industry	Num of Stocks	Average MktCap	Max MktCap	Min MktCap
			Market cap unit: RMB millions		
1	Transportation	20	15,458.00	54,496.00	4,944.00
2	Media	2	15,333.00	22,643.00	8,023.00
3	Utilities	16	17,671.00	86,520.00	2,827.00
4	Agriculture & Forestry	8	9,424.00	19,093.00	3,832.00
5	Healthcare	31	18,608.00	91,999.00	3,938.00
6	Retail	13	5,576.00	8,690.00	3,489.00
7	Defense & Military	5	19,216.00	46,342.00	10,008.00
8	Basic Chemicals	20	10,262.00	38,627.00	2,193.00
9	Home Appliances	6	12,019.00	24,018.00	3,896.00
10	Building Materials	12	12,830.00	64,275.00	3,807.00
11	Construction & Decoration	6	11,563.00	37,409.00	3,677.00
12	Real Estate	24	10,489.00	40,188.00	3,680.00
13	Non-ferrous Metals	12	14,441.00	29,738.00	3,457.00
14	Machinery	16	14,766.00	45,562.00	2,962.00
15	Automobile	20	13,693.00	46,645.00	4,044.00
16	Coal	6	14,453.00	25,337.00	6,859.00
17	Environmental Protection	8	14,187.00	30,968.00	4,714.00
18	Power Equipment	5	8,575.00	14,512.00	4,117.00
19	Electronics	7	17,316.00	64,923.00	5,632.00
20	Oil & Petrochemicals	6	89,298.00	456,766.00	2,730.00
21	Social Services	5	8,025.00	14,820.00	2,603.00
22	Textiles & Apparel	9	7,616.00	12,700.00	4,585.00
23	Conglomerates	4	8,748.00	13,535.00	4,269.00
24	Personal Care	1	3,762.00	3,762.00	3,762.00
25	Information Technology	7	22,968.00	40,128.00	3,620.00
26	Light Manufacturing	6	7,646.00	9,501.00	4,145.00
27	Telecommunications	4	17,519.00	29,441.00	4,271.00
28	Steel	6	14,561.00	28,224.00	4,880.00
29	Banking	3	237,435.00	356,468.00	88,985.00
30	Non-bank Financials	2	88,441.00	171,266.00	5,615.00
31	Food & Beverage	13	16,151.00	90,287.00	3,573.00
Total	**-**	**303**	**17,789.00**	**456,766.00**	**2,193.00**

Note: The table reports the number of stocks, average, maximum, and minimum market capitalization (in million RMB) for each industry sector. "Total" aggregates across all industries.

4. An Equity Valuation Approach Based on Relative Value

This section presents a quantitative approach to equity valuation based on relative value, which evaluates a stock's performance relative to its industry sector or factor benchmark. The

model uses either sector-based ETFs or PCA-derived synthetic factors as explanatory variables. An extension incorporating trading volume is introduced in Section 5. Although the current framework relies solely on price and volume data, it can be extended to include fundamental variables such as analyst revisions, earnings momentum, and other measurable financial indicators.

We adopt a continuous-time framework, where the price of stock $S_i(t)$, for $i = 1, \dots, N$, evolves according to a stochastic differential equation. Specifically, the return dynamics follow:

$$\frac{dS_i(t)}{S_i(t)} = \alpha_i dt + \sum_{j=1}^{N} \beta_{ij} \frac{dI_j(t)}{I_j(t)} + dX_i(t) \tag{18}$$

where $I_j(t)$ denotes the $j - th$ factor (either ETF or PCA-based), β_{ij} are the factor loadings, and $dX_i(t)$ represents the idiosyncratic return component not explained by systematic factors.

In the case of a single sector ETF, the model simplifies to:

$$\frac{dS_i(t)}{S_i(t)} = \alpha_i dt + \beta_{ij} \frac{dI(t)}{I(t)} + dX_i(t) \tag{19}$$

with $I(t)$ corresponding to the ETF of the stock's primary industry classification.

The idiosyncratic return process $\tilde{X}_i(t)$ is defined as:

$$d\tilde{X}_i(t) = \alpha_i dt + dX_i(t) \tag{20}$$

where the drift term $\alpha_i dt$ represents the expected excess return relative to the benchmark, and $dX_i(t)$ captures stock-specific shocks. This component is modeled as a mean-reverting OU process:

$$dX_i(t) = \kappa_i(\mu_i - X_i(t))dt + \sigma_i dW_i(t), \kappa_i > 0 \tag{21}$$

where κ_i is the speed of mean reversion, μ_i is the long-term equilibrium level, σ_i is the volatility of residual shocks, and $dW_i(t)$ is a standard Brownian motion. The OU process is stationary and can be regarded as the continuous-time analogue of the $AR(1)$ model.

The expected conditional increment is given by:

$$E[dX_i(t)|X_i(s), s \leq t] = \kappa_i(\mu_i - X_i(t))dt \tag{22}$$

implying that the forecasted return depends on the deviation of $X_i(t)$ from its equilibrium value μ_i.

The parameters $\alpha_i, \kappa_i, \mu_i, \sigma_i$ are stock-specific and are assumed to be locally constant within a moving estimation window. In practice, we estimate these parameters using a 60-day rolling window and retain only those stocks with sufficiently high mean-reversion speeds (i.e., large κ_i) to ensure both robustness of the stationarity assumption and profitability of trading signals. Details of the estimation procedure are provided in the next section and in Appendix A.

Solving the OU process yields the closed-form solution:

$$X_i(t_0 + \Delta t) = e^{-\kappa_i \Delta t} X_i(t_0) + (1 - e^{-\kappa_i \Delta t})\mu_i + \sigma_i \int_{t_0}^{t_0 + \Delta t} e^{-\kappa_i(t_0 + \Delta t - s)} dW_i(s) \qquad (23)$$

In the long-run limit ($\Delta t \to \infty$), the process converges to its stationary distribution:

$$E[X_i(t)] = \mu_i, Var[X_i(t)] = \frac{\sigma_i^2}{2\kappa_i} \qquad (24)$$

This formulation naturally lends itself to relative-value trading. A market-neutral position is constructed by going long the stock and short the corresponding sector ETF (or factor portfolio), scaled by the estimated factor loading. The expected instantaneous return of such a portfolio is:

$$\mathbb{E}[d\Pi_i(t)] = \alpha_i dt + \kappa_i(\mu_i - X_i(t))dt \qquad (25)$$

where the drift α_i captures any systematic excess return and the second term reflects the forecasted mean-reversion gain. When $X_i(t) > \mu_i$, the expected return is negative, generating a short signal; conversely, $X_i(t) < \mu_i$ produces a long signal.

The speed of mean reversion κ_i determines how quickly deviations from the mean are corrected. Its inverse,

$$\tau_i = 1/\kappa_i$$

defines the characteristic time scale. Assets with small τ_i (fast mean reversion) offer more reliable short-term opportunities, and the strategy therefore retains only those stocks with sufficiently large κ_i to ensure both estimation robustness and trading profitability.

5. Construction of Trading Signals

This section outlines the construction of trading signals derived from residual processes modeled under a mean-reversion framework. Residual returns are assumed to follow an OU process, with parameters estimated using a 60-day rolling window (i.e., $T_1 = 60/252$). This

window provides a balance between responsiveness and estimation stability, roughly corresponding to a single corporate earnings cycle.

As a filtering mechanism, only those stocks with estimated mean-reversion times shorter than one-thirtieth of a year ($\tau_i < 1/30 \approx 8.4$ trading days) are retained. This restriction ensures that the residual process exhibits sufficiently fast mean reversion within the estimation window, thereby supporting the local stationarity assumption and enhancing the statistical reliability of signals for short-horizon trading.

To provide a preliminary empirical assessment, Table 5 reports the historical Sharpe ratios achieved by strategies based on these signals, applied to portfolios sorted by sector ETFs over the full sample period. Comprehensive backtesting results will be presented in a later section. Technical details on the estimation of OU parameters and the construction of standardized signals (e.g., s-scores) are provided in the following subsection.

5.1 Signal Design Based on Pure Mean-Reversion Dynamics

In the baseline signal construction approach, we consider a simplified version of the OU process by omitting the drift term. In this specification, the residual process $X_i(t)$ is modeled as purely mean-reverting without directional bias. Under such an assumption, the equilibrium standard deviation of the process is:

$$\sigma_{eq,i} = \frac{\sigma_i}{\sqrt{2\theta_i}}$$

where σ_i denotes the volatility parameter of the residual process, and θ_i is the mean-reversion speed.

To standardize deviations from equilibrium across stocks with heterogeneous volatility profiles, we define a dimensionless statistic, the s-score, as:

$$s_i(t) = \frac{X_i(t) - \mu_i}{\sigma_{eq,i}} \tag{26}$$

where μ_i represents the long-term equilibrium level of the residual process. The s-score thus measures the normalized deviation of the residual from its equilibrium, serving as the central decision variable in our mean-reversion trading strategy. Figure 7 illustrates the evolution of the

s-score for the residuals of CMBC relative to the Banking sector. A higher absolute value of $s_i(t)$ indicates a greater disequilibrium, and therefore a stronger trading signal.

The trading rules based on the s-score are structured as threshold-based triggers:

- Open Long Position: if $s_i(t) < -s_{entry}$,
- Open Short Position: if $s_i(t) > +s_{entry}$,   (27)
- Close Long Position: if $s_i(t) > -s_{close}$,
- Close Short Position: if $s_i(t) < +s_{close}$.

Each trading signal is implemented as a market-neutral position. When a long signal is triggered (i.e., $s_i(t) < -s_{entry}$), the strategy buys one dollar of the corresponding stock while simultaneously selling β_i dollars of the associated sector ETF, thereby hedging systematic exposure. This construction ensures neutrality with respect to the sector-level factor used in residual estimation.

In the case of a multi-factor specification—such as regressions on multiple ETFs or on principal components extracted via PCA—the hedging leg consists of a weighted portfolio of factor exposures. Specifically, the investor sells β_{i1} dollars of Factor 1, β_{i2} dollars of Factor 2, …, up to β_{im} dollars of Factor mmm, where β_{im} is the loading of stock i on factor j. The reverse applies to short signals. All positions are closed by unwinding both the stock and the hedging portfolio using the same regression-derived weights.

This hedging procedure preserves dollar neutrality against the underlying systematic factors, ensuring that expected returns are driven solely by the behavior of the idiosyncratic (mean-reverting) component. Such design allows the strategy to remain agnostic to broad market movements while exploiting temporary mispricings at the stock level.

For implementation consistency and scalability across assets, uniform entry and exit thresholds are applied. These thresholds are determined empirically through backtesting on historical data from 2010 to 2014. The optimal values are:

$$s_{entry} = 1.25, s_{close,\,long} = 0.50, s_{close,\,short} = 0.75$$

The asymmetric closing rules—liquidating short positions at a higher s-score (0.75) than long positions (0.50)—were found to improve performance during the training period and are thus retained in subsequent backtests.

Overall, this rule-based framework capitalizes on statistically significant deviations of the residual from its mean, under the assumption that such excursions are temporary and will revert over a characteristic horizon determined by $1/\theta_i$. By focusing exclusively on substantial deviations, the strategy filters out noise and concentrates on high-conviction opportunities with favorable mean-reversion profiles.

Figure 7. Rolling S-score of Stock CMBC vs. Banking sector
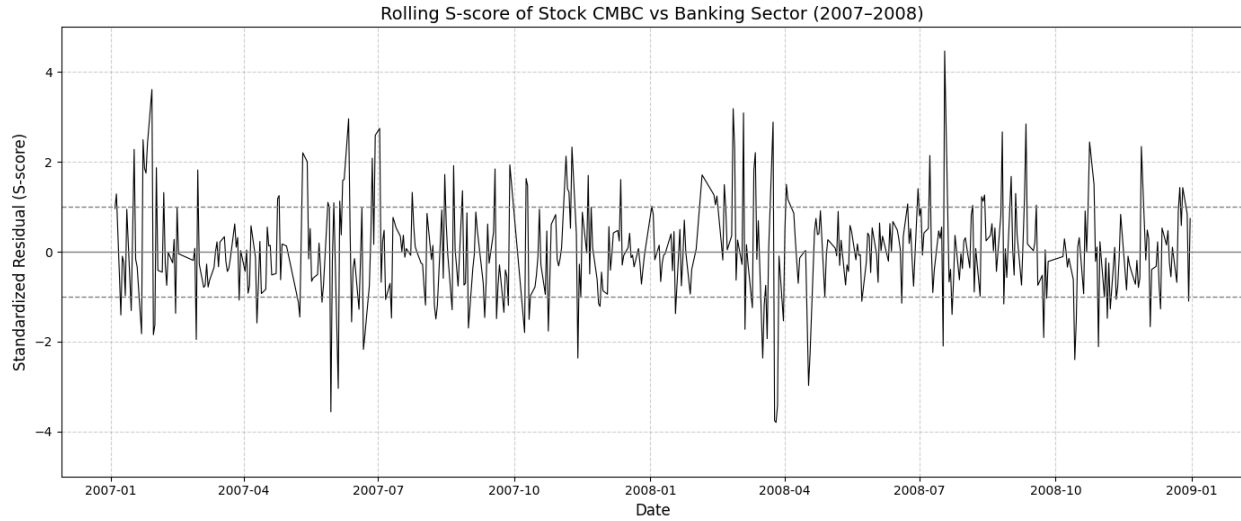
Note: Rolling S-score of China Minsheng Banking Corp. (CMBC) vs. the Banking sector index during 2007–2008. The S-score represents standardized residuals from a rolling regression of stock returns on sector returns, used to detect mean-reversion signals.

5.2 Signal Design Based on Mean-Reversion with Drift

In the pure mean-reversion model introduced in the previous section, the drift term of the residual process was omitted under the assumption that it is negligible relative to the equilibrium volatility, $\sigma_{eq,i} = \sigma_i/\sqrt{2\theta_i}$. However, it is possible to explicitly incorporate a drift component, which leads to a modified signal that accounts for potential directional bias in the residual dynamics.

Under the OU specification,

$$dX_i(t) = \theta_i\big(\mu_i - X_i(t)\big)\,dt + \sigma_i dW_i(t)$$

the conditional expectation over a short interval dt is

$$E[dX_i(t)] = \theta_i(\mu_i - X_i(t))\,dt \tag{28}$$

Rewriting this expression in terms of the standardized residual (s-score):

$$s_i(t) = \frac{X_i(t) - \mu_i}{\sigma_{eq,i}}$$

we obtain,

$$E[dX_i(t)] = \theta_i \left(\frac{\mu_i}{\theta_i \sigma_{eq,i}} - s_i(t) \right) \sigma_{eq,i} dt \tag{29}$$

This motivates the definition of a drift-adjusted s-score as:

$$s_i^{mod}(t) = s_i(t) - \frac{\mu_i}{\theta_i \sigma_{eq,i}} \tag{30}$$

The intuition is straightforward. If the residual process has a positive drift ($\mu_i > 0$), the effective short signal is weakened, since the stock tends to rise even after adjusting for factor exposures. Conversely, a negative drift strengthens the short signal. This adjustment therefore introduces a momentum-like correction into the symmetric mean-reversion framework.

For example, under the original strategy, a short position would be opened if $s_i > s_{entry}$. Under the modified formulation, the same threshold is harder to reach when $\mu_i > 0$, thereby reducing the risk of shorting a stock with persistent upward bias.

From a statistical perspective, μ_i can be interpreted as the equilibrium level of the residual process, which, when estimated over short windows (e.g., 60 trading days), resembles the slope of a local moving average. Incorporating this term thus introduces a momentum filter into the otherwise purely mean-reverting framework.

Empirical analysis over the 2010–2014 training period shows that the same entry and exit thresholds calibrated for the pure mean-reversion strategy also work well for the drift-adjusted case. In practice, the estimated values of μ_i are relatively small—typically on the order of 15 basis points—while the average reversion time is about 7 trading days and the equilibrium residual volatility is around 300 basis points. Hence, the correction term,

$$\frac{\mu_i}{\theta_i \sigma_{eq,i}} \approx \frac{0.0015 \times 7}{0.03} \approx 0.35$$

is modest and does not materially affect the decision variable.

Backtests confirm that this adjustment yields only marginal performance improvements at the daily trading frequency. For simplicity and robustness, we therefore adopt the original s-score formulation in the main implementation and omit further consideration of the drift-adjusted variant in subsequent sections.

In summary, while theoretically appealing, incorporating a constant drift term into the residual process does not materially enhance trading performance. This suggests that, once systematic risk exposures are controlled for, the idiosyncratic component of stock returns exhibits limited persistent momentum over the short horizons relevant to this strategy.

6. Results

To evaluate the effectiveness of the proposed strategy, we conducted a series of backtesting experiments using historical stock data. The central objective is to simulate daily trading across the entire stock universe based on the signal rules defined earlier (see Equation 27). Each trading day, we update model parameters—such as factor loadings (betas) and residuals—and compute trading signals, ensuring that all information used is strictly backward-looking and adheres to a rolling window framework.

During parameter estimation, we retain only those stocks with sufficiently fast mean reversion. Specifically, the mean-reversion time scale is defined as $\tau_i = 1/\kappa_i$. If τ_i exceeds 8.4 trading days (i.e., $\tau_i < 1/8.4$), the stock is excluded from trading. This filtering ensures that only stocks with strong and rapid mean-reversion dynamics are eligible for the strategy.

All trades are assumed to be executed at the day's closing price. To account for trading frictions—including transaction costs and price slippage—we impose a round-trip cost of 30 basis points per completed trade. In addition, we include a per-trade slippage parameter $\epsilon = 0.0005$ (5 bps) in the portfolio update equation to capture incremental execution costs.

Let E_t denote the total portfolio equity at time t. The evolution of the portfolio equity over a single trading interval Δt is given by:

$$E_{t+\Delta t} = E_t + E_t r \Delta t + \sum_{i=1}^{N} Q_{it} R_{it} - (\sum_{i=1}^{N} Q_{it}) r \Delta t + \sum_{i=1}^{N} \frac{Q_{it} D_{it}}{S_{it}} - \sum_{i=1}^{N} |Q_{i(t+\Delta t)} - Q_{it}| \varepsilon \quad (31)$$

Here,

- $Q_{it} = E_t \gamma_t$ represents the capital allocated to stock i at time t,

- R_{it} is the return of stock i over $(t, t + \Delta t)$,

- r is the risk-free interest rate,

- $\Delta t = 1/252$ corresponds to one trading day in a year,

- D_{it} is the dividend paid by stock i during the period,

- S_{it} is the stock price at time t,

- $\varepsilon = 0.0005$ reflects the per-trade slippage.

The allocation coefficient γ_t is a fixed proportion of portfolio equity, identical across stocks, and calibrated to maintain a target leverage. For instance, if we aim to hold 200 long and 200 short positions with a gross leverage of 4 (i.e., 2× equity on the long side and 2× on the short side), we set $\gamma_t = 2/200 = 0.01$. Thus, γ_t can be interpreted as the maximum fraction of equity allocated to any single stock.

This leverage calibration, validated using data from 2002–2004, produces a portfolio with approximately 10% annualized volatility. Importantly, leverage choices do not affect risk-adjusted metrics such as the Sharpe ratio. Comparable strategies—for example, the "1/2 + 1/2" contrarian portfolio of Khandani and Lo (2009)—yield similar Sharpe ratios once normalized for exposure.

Given the binary nature of the trading signals, the strategy follows a so-called bang-bang (all-or-nothing) implementation. Specifically, when a signal is triggered, the model takes a full long or short position in the corresponding stock, and the position is fully closed when the exit condition is met. Although this binary approach may appear simplistic, empirical evidence suggests that it consistently outperforms gradual or fractional rebalancing schemes. The performance advantage arises from two factors. First, by avoiding frequent incremental trades, the strategy reduces sensitivity to estimation noise in the signals. Second, the lower turnover reduces cumulative transaction costs, including slippage and round-trip fees, thereby improving net performance.

In summary, the simulation framework validates the practicality and robustness of the proposed mean-reversion strategy. The following sections provide a comprehensive empirical

evaluation across subsample periods and examine the comparative performance of different model specifications and residual extraction techniques.

6.1 Synthetic Industry Indices as Factors

To extend the back-testing period prior to the emergence of tradable sector ETFs, we constructed synthetic, capitalization-weighted sector indices to serve as risk factors in our statistical arbitrage model. The motivation was to allow for a direct comparison with the PCA-based approach and to address the absence of certain sector ETFs, as some industries did not yet have corresponding tradable instruments. Each synthetic index aggregates the returns of constituent stocks within a given industry, weighted by market capitalization.

Daily returns for these synthetic indices were computed directly from constituent stock returns, aggregated with capitalization weights. For parameter estimation, a rolling 60-day window preceding each estimation date was applied. Individual stock returns were then regressed on the returns of their respective sector indices, and the residuals from these regressions were modeled as OU processes. These residuals formed the basis for trading signals under the assumption of mean reversion.

To achieve market neutrality, portfolio exposure was hedged daily using the CSI 300 ETF, an ETF tracking the CSI 300 index. This ensured that the portfolio maintained a near-zero beta with respect to the broader market, especially given that synthetic ETFs, unlike actual ETFs, are not tradable instruments. In line with the view that market inefficiencies are transient and mean-reverting, we further centered each stock's residual series by subtracting the cross-sectional average residual mean across the universe. This adjustment was found to reduce model bias and enhance strategy performance, consistent with findings by Avellaneda and Lee (2010).

Table 4 reports the annual performance metrics of the synthetic ETF-based strategy from 2005 through 2024. Over the full sample, the strategy exhibited a mean annual return of 9.00% with a standard deviation of 9.82%, yielding an overall Sharpe ratio of 0.92. Notably, the strategy performed strongly in several isolated years, such as 2009, 2015, and 2019, where annual Sharpe ratios exceeded 2.4, indicating favorable risk-adjusted returns. However, performance deteriorated in certain years—most evidently in 2017 and 2020—with negative Sharpe and Calmar ratios.

The strategy's robustness was further evaluated using various risk and distributional metrics. For example, in 2009, the Sortino ratio reached 4.70, and the Calmar ratio peaked at 7.87, reflecting both limited drawdown and high returns. Additionally, the Conditional VaR (CvaR) at the 95% confidence level remained modest across most years, typically under 3%, indicating effective downside risk control. The Omega ratio, which captures the proportion of gains relative to losses, consistently exceeded 1.1 in strong-performing years and reached a maximum of 1.84 in 2019.

Table 4. Annual performance metrics of the synthetic ETF-based statistical arbitrage strategy with trading costs included (2005–2024)

Year	Mean	Std. Dev.	Sharpe	Sortino	Max Drawdown	Calmar	t-stat	z-stat	JB test (p-value)	CVaR (95%)	Omega
2005	0.1694	0.2317	0.7312	1.5110	-0.1964	0.8627	0.7031	0.7031	0.0000	-0.0241	1.1714
2006	0.1229	0.2499	0.4920	0.7823	-0.1631	0.7539	0.4811	0.4811	0.0000	-0.0288	1.1321
2007	0.0442	0.1110	0.3984	0.5505	-0.1608	0.2750	0.3904	0.3904	0.0000	-0.0158	1.0711
2008	0.1426	0.0887	1.6078	2.1284	-0.0577	2.4718	1.5885	1.5885	0.0000	-0.0124	1.3167
2009	0.1885	0.0677	2.7821	4.7027	-0.0239	7.8738	2.7376	2.7376	0.3413	-0.0078	1.5838
2010	0.0978	0.0675	1.4498	2.3811	-0.0405	2.4126	1.4207	1.4207	0.0000	-0.0082	1.2861
2011	0.0467	0.0505	0.9243	1.3723	-0.0372	1.2554	0.9095	0.9095	0.0000	-0.0070	1.1718
2012	0.1048	0.0446	2.3513	3.7359	-0.0190	5.5260	2.3089	2.3089	0.4876	-0.0057	1.4645
2013	0.0233	0.0595	0.3927	0.6748	-0.0476	0.4901	0.3816	0.3816	0.0031	-0.0077	1.0658
2014	0.0612	0.0465	1.3149	2.0348	-0.0272	2.2461	1.2965	1.2965	0.1470	-0.0060	1.2463
2015	0.2273	0.0922	2.4643	4.2807	-0.0523	4.3459	2.4249	2.4249	0.0000	-0.0105	1.5266
2016	0.1126	0.0454	2.4782	4.0652	-0.0201	5.6042	2.4386	2.4386	0.4631	-0.0055	1.4874
2017	-0.0466	0.0529	-0.8804	-1.1969	-0.0570	-0.8163	-0.8664	-0.8664	0.0000	-0.0083	0.8628
2018	0.0499	0.0527	0.9460	1.2858	-0.0624	0.7994	0.9290	0.9290	0.0000	-0.0073	1.1704
2019	0.1785	0.0506	3.5270	6.2941	-0.0252	7.0835	3.4705	3.4705	0.0000	-0.0056	1.8424
2020	-0.0124	0.0541	-0.2295	-0.3609	-0.0413	-0.3010	-0.2254	-0.2254	0.5567	-0.0071	0.9642
2021	0.0252	0.0719	0.3505	0.6629	-0.0871	0.2893	0.3442	0.3442	0.0005	-0.0083	1.0580
2022	0.1344	0.0531	2.5299	4.1298	-0.0304	4.4142	2.4792	2.4792	0.7352	-0.0066	1.5060
2023	0.0374	0.0414	0.9044	1.5572	-0.0398	0.9408	0.8863	0.8863	0.1487	-0.0050	1.1559
2024	0.0978	0.0833	1.1740	1.6487	-0.0464	2.1100	0.7965	0.7965	0.0000	-0.0119	1.2523
All	0.0900	0.0982	0.9172	1.3689	-0.2088	0.4312	3.9706	3.9706	0.0000	-0.0119	1.2276

Note: Performance includes a round-trip transaction cost of 30 basis points per trade. The final row ("All") summarizes the full-period statistics.

To validate the statistical significance of these returns, t-statistics and z-statistics were computed annually. Several years, including 2009, 2015, and 2019, showed t-statistics well above 2.0, supporting the rejection of the null hypothesis of zero mean return. Additionally, the

Jarque–Bera test was conducted to assess normality of returns, with p-values frequently indicating departures from Gaussianity, especially in years with extreme Sharpe ratios.

Figure 8 shows the cumulative PNL of the synthetic ETF-based strategy from 2005 to 2024. The strategy demonstrates a consistent upward trend over the full sample period, with particularly strong growth phases observed between 2015–2016 and again from 2019 onward. These periods are characterized by sustained momentum and limited drawdowns, contributing significantly to the long-term profitability of the strategy. Short-term stagnations and minor drawdowns are visible in certain intervals, such as 2007–2008 and 2016–2017, reflecting occasional market environments where the mean-reversion signals generated from synthetic ETF residuals were less effective or temporarily misaligned with broader market movements.

Figure 8. Historical PNL statistics for the strategy using synthetic ETFs as factors from 2005–2024
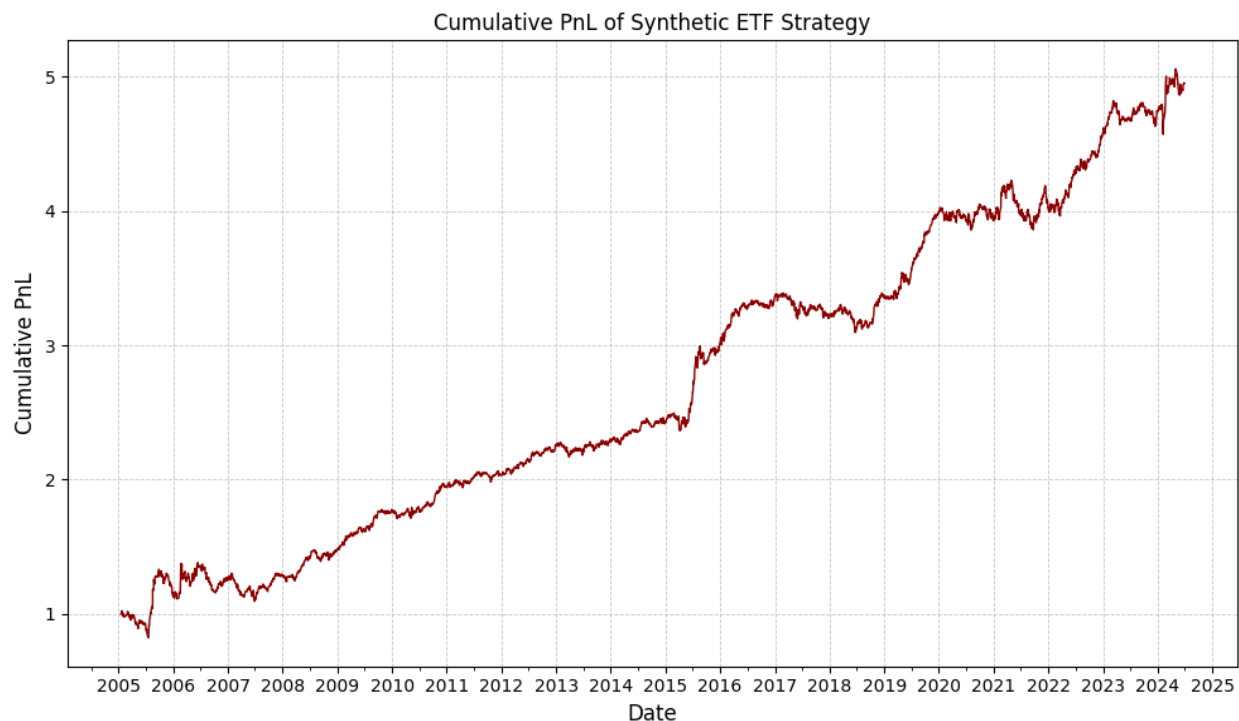
The strong performance post-2015 may be attributed to several factors. First, the improved signal quality due to more stable residual dynamics and faster mean reversion in low-volatility regimes enhances trade timing. Second, greater sector dispersion in certain years increases the relative-value opportunities that the strategy seeks to exploit. Third, tighter risk control, including SPY beta-neutral hedging and residual centering techniques, may have contributed to more robust downside protection and capital preservation. On the other hand, periods of

underperformance often coincide with compressed cross-sectional volatility or heightened market correlation, reducing the effectiveness of sector-based residual deviations as trading signals. These environments tend to obscure true mispricings and lead to weaker signal-to-noise ratios.

In summary, while the synthetic ETF-based approach offers a viable proxy for risk factor modeling in the absence of historical ETF data, its performance varies considerably across market cycles. It is most effective during periods of heightened volatility or market dislocations, consistent with the underlying mean-reversion assumption.

6.2 PCA as Factors

PCA implementation, summarized in Table 5 and contrasted with the synthetic-ETF variant in Figure 9, displays a markedly different life-cycle of effectiveness. Across 2005–2024 it delivered a compound annual return of 8.33 percent with 8.83 percent volatility, yielding a slightly higher full-sample Sharpe ratio (0.94) yet a lower Calmar ratio (0.51) than the synthetic-ETF benchmark. The smaller peak-to-trough drawdown (–16.5 percent versus –20.9 percent) suggests tighter downside containment, but the slower capital recovery implied by the Calmar figure indicates that large gains arrived in fewer, more concentrated bursts.

During the first half-decade (2005–2010) the PCA strategy outpaced its synthetic counterpart by a wide margin. Annual Sharpe ratios climbed steadily from 1.02 in 2005 to an exceptional 3.70 in 2009, when fat-tailed sector dislocations pushed the Sortino ratio above 5.6 and lifted the Calmar ratio past 8. The eigenportfolios extracted from rolling 60-day return matrices evidently captured latent common risks—growth versus value, liquidity, and leverage— that broad industry baskets only imperfectly proxied before liquid sector ETFs became prevalent. High t-statistics (above 3.6 in 2009) and Omega ratios near 1.9 confirm that the gains were statistically robust and skewed toward favorable outcomes.

Between 2011 and 2016 the two approaches converged as cross-sectional dispersion normalised after the post-crisis reset. Both models alternated quarters of out- and under-performance. The PCA engine retained an edge in responsiveness—evident in quicker recoveries during 2015 and 2016—yet the synthetic indices produced consistently higher Sortino ratios and smoother equity curves, reflecting greater stability despite similar Sharpe profiles. This phase

illustrates the trade-off between adaptivity, which favors PCA, and the interpretability and robustness that capitalization-weighted sector indices provide under moderate-volatility regimes.

From 2018 onward the synthetic-ETF strategy assumed clear leadership. The PCA variant registered two negative-Sharpe years (2019 and 2024) and failed to match the red-line ascent visible in Figure 9. Three structural shifts help explain the reversal. First, accelerated factor rotation—driven by retail flows, thematic trading and quantitative crowding—rendered rolling 60-day covariance estimates noisy; resulting eigenvector "flip-flops" diluted the explanatory power of the fifteen retained PCs and increased turnover costs. Second, sector realignment—owing to China's supply-side reform, global ESG adoption and a surge in semiconductor and AI capital expenditure—heightened the relevance of well-defined industry clusters. Capitalization-weighted synthetic indices adapted automatically, whereas the PCA basis struggled to anchor on a stable set of latent factors. Third, volatility regime shifts around the COVID-19 pandemic elevated the kurtosis of residuals; Jarque–Bera p-values collapsed toward zero in 2019, while results in 2024 suggested a temporary reversion toward Gaussianity, highlighting unstable tail behavior across regimes.

Years of notable success for the PCA strategy, such as 2009, 2010, 2015, 2017 and 2018, share three features: t- and z-statistics often around or above 2.0, Omega ratios exceeding 1.3, and daily CvaR 95% contained below 1.5 percent of equity. Loss years—2012, 2019 and 2024—display the mirror image: sub-unitary Sharpe ratios, negative Calmar figures and Omega below 1.0, underscoring the sensitivity of eigenportfolios to correlation noise.

Imposing eigenvector-stability constraints or applying shrinkage techniques such as Ledoit–Wolf or random-matrix filtering could curb factor turnover and reduce estimation error in later years. Allowing the number of retained components to fluctuate with an explained-variance threshold, rather than fixing it at fifteen, would align dimensionality with prevailing market complexity and lessen over-fitting. A regime-switching ensemble that blends PCA- and sector-based residuals—weighted by real-time dispersion and correlation metrics—promises smoother performance across cycles of fragmentation and convergence, while embedding volatility-jump or higher-moment hedges could mitigate the fat-tail drag observed in 2019 and 2024.

Taken together, the evidence implies that PCA factors excel when latent, non-industry commonalities dominate price behavior—such as the early post-crisis and style-rotation

phases—but underperform when investor attention re-anchors on clear-cut sector themes or when correlation regimes shift abruptly. Conversely, synthetic-industry indices prosper in environments of sector-specific dispersion and offer superior robustness at the cost of slower reaction speed. A carefully engineered hybrid framework, allowing market structure to determine whether adaptive or interpretable factors receive the risk budget, stands to capture the complementary advantages documented in Figure 9.

Table 5. Annual performance of the strategy using 15 PCA-based factors (2005–2024), with trading costs included.

Year	Mean	Std. Dev.	Sharpe	Sortino	Max Drawdown	Calmar	t-stat	z-stat	JB test (p-value)	CVaR (95%)	Omega
2005	0.1551	0.1518	1.0219	1.4039	-0.1095	1.4163	0.9826	0.9826	0.0000	-0.0208	1.2368
2006	0.0785	0.1068	0.7348	1.2215	-0.0750	1.0472	0.7186	0.7186	0.0000	-0.0131	1.1425
2007	0.0573	0.1208	0.4739	0.6210	-0.1647	0.3477	0.4644	0.4644	0.0000	-0.0171	1.0812
2008	0.1298	0.1157	1.1224	1.7345	-0.0594	2.1848	1.1090	1.1090	0.0001	-0.0149	1.2047
2009	0.3522	0.0953	3.6977	5.6140	-0.0430	8.1951	3.6385	3.6385	0.0002	-0.0116	1.8761
2010	0.1294	0.0756	1.7114	2.8199	-0.0452	2.8608	1.6771	1.6771	0.0013	-0.0094	1.3335
2011	0.0262	0.0658	0.3979	0.6247	-0.0715	0.3662	0.3916	0.3916	0.1852	-0.0088	1.0679
2012	-0.0085	0.0697	-0.1221	-0.1755	-0.0519	-0.1641	-0.1199	-0.1199	0.0000	-0.0101	0.9802
2013	0.0260	0.0798	0.3256	0.5406	-0.0521	0.4986	0.3164	0.3164	0.9766	-0.0101	1.0531
2014	0.0387	0.0632	0.6126	0.9188	-0.0526	0.7356	0.6040	0.6040	0.1428	-0.0083	1.1056
2015	0.1701	0.1162	1.4630	2.3896	-0.0854	1.9918	1.4396	1.4396	0.0000	-0.0145	1.2921
2016	0.0167	0.0668	0.2498	0.2981	-0.0700	0.2382	0.2458	0.2458	0.0000	-0.0106	1.0441
2017	0.1236	0.0578	2.1374	3.7191	-0.0350	3.5313	2.1032	2.1032	0.9353	-0.0069	1.4046
2018	0.1070	0.0652	1.6404	2.3339	-0.0464	2.3079	1.6109	1.6109	0.0000	-0.0092	1.3213
2019	-0.0216	0.0711	-0.3033	-0.4132	-0.1096	-0.1968	-0.2984	-0.2984	0.0000	-0.0108	0.9501
2020	0.0539	0.0842	0.6398	1.0735	-0.0508	1.0607	0.6282	0.6282	0.4600	-0.0108	1.1086
2021	0.0916	0.0779	1.1769	2.0863	-0.0413	2.2179	1.1557	1.1557	0.5679	-0.0093	1.2024
2022	0.0410	0.0792	0.5182	0.8431	-0.0530	0.7746	0.5078	0.5078	0.0010	-0.0100	1.0887
2023	0.0777	0.0580	1.3407	2.1632	-0.0425	1.8275	1.3138	1.3138	0.6199	-0.0075	1.2371
2024	-0.0441	0.0719	-0.6134	-0.9274	-0.0558	-0.7896	-0.4162	-0.4162	0.8901	-0.0097	0.9050
All	0.0833	0.0883	0.9434	1.3736	-0.1647	0.5060	4.0840	4.0840	0.0000	-0.0120	1.1806

Figure 9 presents a comparative view of the cumulative PnL between the synthetic ETF-based strategy and the PCA-based strategy using 15 eigenportfolios over the period 2005–2024. In the earlier years (2005–2010), the PCA-based strategy shows superior performance, suggesting that principal components extracted from stock returns captured latent common risk factors not fully represented by sector-based indices. Between 2010 and 2016, both strategies delivered broadly similar outcomes, with alternating phases of outperformance: the PCA approach displayed higher

responsiveness to short-term deviations, while the synthetic ETF strategy provided smoother and more stable return paths.

From 2017 onward, the synthetic ETF strategy began to outperform more consistently. This shift likely reflects structural changes in market organization that reinforced the importance of sector-based co-movements, making capitalization-weighted synthetic indices more reliable for residual estimation. By contrast, the PCA-based approach became more vulnerable to factor instability due to evolving correlation patterns and increased noise in rolling covariance estimates.

Overall, the evidence suggests that the synthetic ETF strategy offers greater robustness in later years, whereas the PCA framework was better suited to earlier periods of fragmented market structure. The contrasting results underscore the complementary nature of the two approaches: PCA excels in detecting latent, non-industry risk drivers, while synthetic indices provide more stability when sector clustering dominates.

Figure 9. Synthetic ETF-based strategy vs. PCA-based strategy
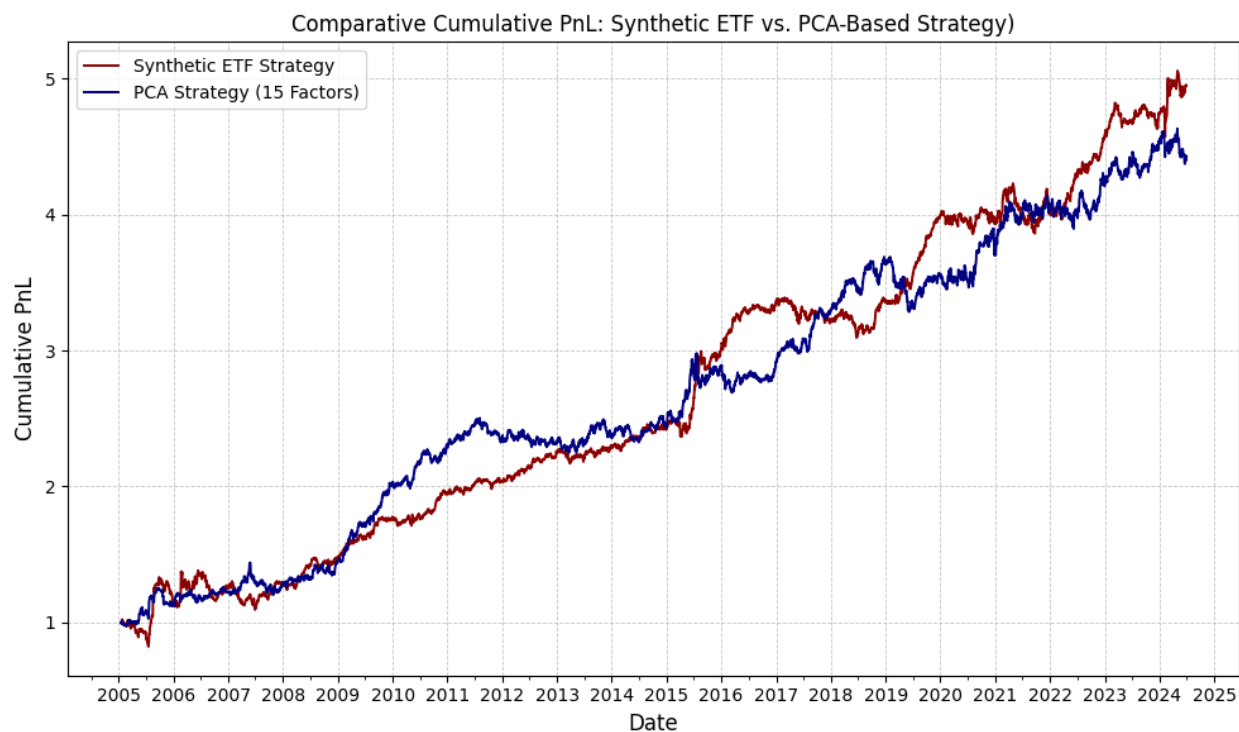
Note: Cumulative PNL comparison between the synthetic ETF-based strategy and the PCA-based strategy with 15 eigenportfolios (2005–2024). The PCA strategy outperforms in earlier years, while the synthetic ETF strategy delivers more stable and superior performance in recent periods.

Figure 10 plots the temporal evolution of two variables: the proportion of variance explained by the top 15 principal components extracted from stock return cross-sections (left axis), and the annualized 252-day rolling volatility of the CSI300 Index (right axis), spanning the

period from 2006 to 2024. A clear co-movement emerges: episodes of elevated market volatility—such as the 2008 global financial crisis, the 2015 Chinese equity market correction, and the COVID-19 outbreak in 2020—are associated with a sharp rise in the fraction of variance captured by the leading eigenvectors.

This relationship reflects a well-established phenomenon in empirical asset pricing: during turbulent markets, idiosyncratic risk declines in relative importance, while systematic shocks become more pervasive across assets. As correlations among securities strengthen, a smaller set of latent factors explains a larger share of total return variation. Conversely, in tranquil periods marked by low aggregate volatility, return dynamics are more dispersed and idiosyncratic, leading to a reduction in the explanatory power of the same set of principal components.

These findings are consistent with the theoretical predictions of factor models under regime-switching volatility, and highlight the importance of adapting factor estimation methodologies to the prevailing market environment.

Figure 10. Explained variance (top 15 principal components) & CSI300 rolling volatility
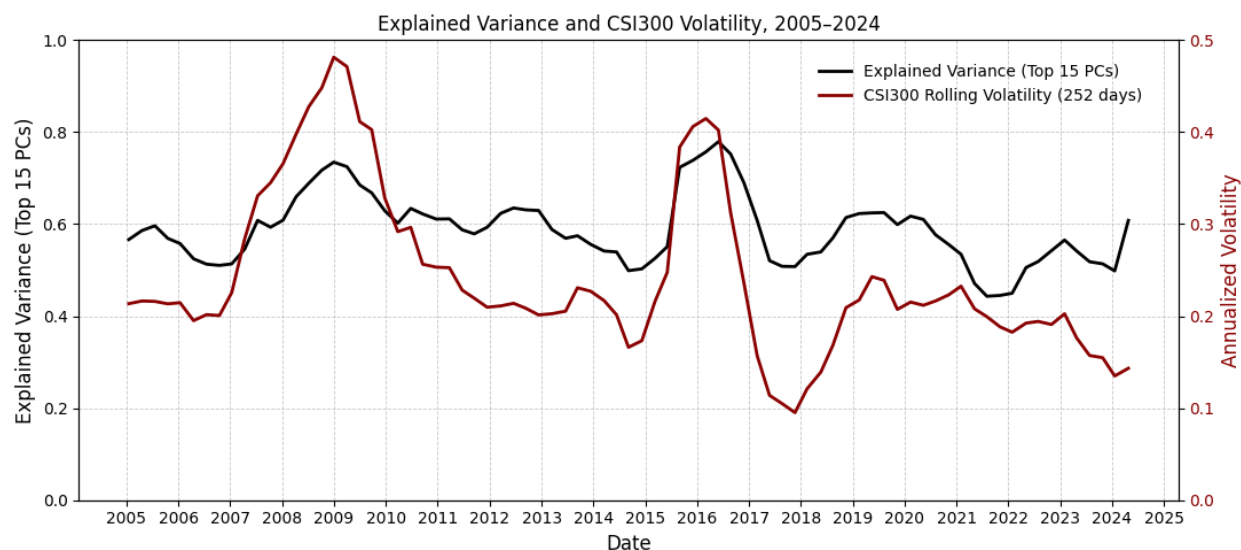
Note: Percentage of variance explained by the top 15 principal components (left axis) and the annualized rolling volatility of the CSI300 Index (right axis) from 2005 to 2024. The explained variance tends to rise during periods of elevated market volatility, suggesting increased common factor dominance during market stress.

6.3 PCA with Fixed Explained Variance

In this section, we evaluate the performance of a PCA-based strategy where the number of factors is chosen endogenously to explain a fixed proportion of return variance. Specifically, we retain the minimum number of principal components required to account for 55% of cross-sectional variance at each rebalancing period. Unlike the fixed-15-PCs approach, this method allows factor dimensionality to adapt dynamically to changing market conditions.

Table 6 summarizes the results. Over 2005–2024, the portfolio earned an annualized return of 6.3% at 12.7% volatility, producing a Sharpe ratio of 0.50—roughly half the risk-adjusted efficiency of the fixed-15-factor benchmark. A maximum drawdown of –23.7% in 2016 and a Calmar ratio of 0.27 highlight the fragility of this approach: although daily tail risk is modest (average CVaR95% of –1.7%), capital recovery after large losses is slow.

Performance is highly regime-dependent. In environments where a few dominant factors overwhelm idiosyncratic noise—most notably the 2008 global financial crisis and the 2021–2022 post-pandemic reopening—the adaptive model performs strongly. Its low dimensionality reduces estimation error, yielding Sharpe ratios above 1.2, Sortino ratios above 2.0, and Calmar values exceeding 3.5, indicating rapid recovery from shallow drawdowns. By contrast, in markets lacking clear directional themes but characterized by elevated, choppy volatility—such as 2016 and, to a lesser extent, 2020—the model struggles. In these cases, discarding lower-rank factors omits economically meaningful structure, leading to underfitting and weak signal extraction. Sharpe ratios fall to around –1.5, drawdowns deepen, and Omega drops below 1.0.

The full-sample statistics reflect a classic bias–variance trade-off. The 55% variance threshold curtails noise and turnover in turbulent regimes but sacrifices informational breadth in calmer periods, producing only half the profit per unit of risk relative to the 15-factor benchmark. Empirically, annual t- and z-statistics cluster near unity, while the Jarque–Bera test almost always rejects normality, indicating that residual payoffs remain fat-tailed even after dimension reduction.

These results suggest two refinements. First, applying a volatility-conditioned threshold— e.g., 55% during high-volatility regimes but 65–70% when volatility and correlations subside— could preserve robustness while recapturing alpha in tranquil markets. Second, blending the adaptive model with a small overlay of fixed PCs would mitigate underfitting and deliver

smoother performance across the full dispersion cycle. Without such refinements, the endogenous 55% rule should be regarded as a defensive, low-beta complement rather than a stand-alone alpha engine.

Table 6. Annual performance of the PCA-based strategy using a variable number of factors selected to explain 55% of total return variance (2005–2024), with trading costs included.

| Year	Mean	Std. Dev.	Sharpe	Sortino	Max Drawdown	Calmar	t-stat	z-stat	JB test (p-value)	CVaR (95%)	Omega
2005	0.1346	0.1742	0.7726	0.9305	-0.1165	1.1549	0.7429	0.7429	0.0000	-0.0265	1.1737
2006	0.0865	0.1012	0.8546	1.3465	-0.0783	1.1047	0.8357	0.8357	0.0000	-0.0133	1.1651
2007	-0.0005	0.1391	-0.0035	-0.0046	-0.2204	-0.0022	-0.0034	-0.0034	0.0000	-0.0194	0.9994
2008	0.4177	0.3109	1.3436	3.1614	-0.1170	3.5714	1.3275	1.3275	0.0000	-0.0278	1.3568
2009	0.0458	0.1805	0.2535	0.3799	-0.1968	0.2326	0.2494	0.2494	0.3494	-0.0247	1.0425
2010	0.1270	0.1081	1.1749	1.8498	-0.0762	1.6660	1.1514	1.1514	0.2686	-0.0143	1.2131
2011	0.0641	0.0744	0.8614	1.3507	-0.0534	1.1989	0.8476	0.8476	0.0011	-0.0102	1.1507
2012	0.0036	0.0893	0.0404	0.0552	-0.0547	0.0661	0.0397	0.0397	0.0000	-0.0134	1.0067
2013	-0.0467	0.0921	-0.5072	-0.7703	-0.0735	-0.6359	-0.4929	-0.4929	0.4326	-0.0123	0.9218
2014	0.0068	0.0614	0.1106	0.1600	-0.0498	0.1366	0.1091	0.1091	0.0077	-0.0088	1.0186
2015	0.1550	0.1457	1.0640	1.4285	-0.1138	1.3625	1.0469	1.0469	0.0000	-0.0212	1.2056
2016	-0.2074	0.1391	-1.4906	-1.6057	-0.2199	-0.9431	-1.4667	-1.4667	0.0000	-0.0257	0.7558
2017	0.0865	0.0776	1.1146	1.8345	-0.0725	1.1926	1.0967	1.0967	0.0000	-0.0098	1.2089
2018	0.0919	0.0699	1.3152	2.0743	-0.0455	2.0184	1.2915	1.2915	0.0000	-0.0094	1.2526
2019	0.1112	0.1004	1.1070	1.6360	-0.1071	1.0379	1.0893	1.0893	0.0000	-0.0141	1.2061
2020	-0.0246	0.0984	-0.2498	-0.4039	-0.0743	-0.3308	-0.2453	-0.2453	0.8391	-0.0130	0.9610
2021	0.0843	0.0693	1.2159	2.2353	-0.0424	1.9877	1.1940	1.1940	0.5050	-0.0081	1.2125
2022	0.0451	0.0671	0.6724	1.1195	-0.0463	0.9739	0.6589	0.6589	0.2969	-0.0084	1.1145
2023	0.0133	0.0527	0.2528	0.4181	-0.0580	0.2294	0.2478	0.2478	0.9532	-0.0068	1.0407
2024	0.0699	0.0731	0.9562	1.3678	-0.0336	2.0816	0.6487	0.6487	0.0000	-0.0101	1.1734
All	0.0632	0.1268	0.4983	0.7213	-0.2368	0.2667	2.1572	2.1572	0.0000	-0.0173	1.1048

Figure 11 contrasts the cumulative performance of two beta-hedged PCA mean-reversion strategies in China's equity market from 2005 to 2024. The fixed-dimension model, which always retains the top 15 principal components, outperforms for most of the sample, particularly during post-2010 expansion phases and other periods of elevated market integration. By

including a broader set of systematic co-movements, this specification delivers smoother equity growth and superior risk-adjusted returns.

In contrast, the adaptive specification retains only the minimum number of principal components required to explain 55 percent of cross-sectional variance. This defensive design reduces factor dimensionality in low-volatility regimes, thereby limiting noise but also discarding potentially relevant lower-rank factors. The resulting under-fitting dampens alpha generation, producing a visibly flatter equity curve. Moreover, because the effective number of factors changes over time, portfolio weights may become less stable when correlation structures rotate abruptly, which can in turn increase turnover and weaken performance persistence.

These observations illustrate a classic bias–variance trade-off. The fixed 15-factor approach provides richer informational coverage and structural persistence, though at the cost of greater exposure to estimation error. The adaptive 55 percent method suppresses noise but sacrifices breadth, leaving it less effective in environments with high latent factor complexity. On balance, the informational advantages of the fixed-dimension approach outweigh its noise costs, yielding higher cumulative profitability and greater robustness over long horizons.

Figure 11. Cumulative profit and loss (PNL) comparison between two PCA-based strategies
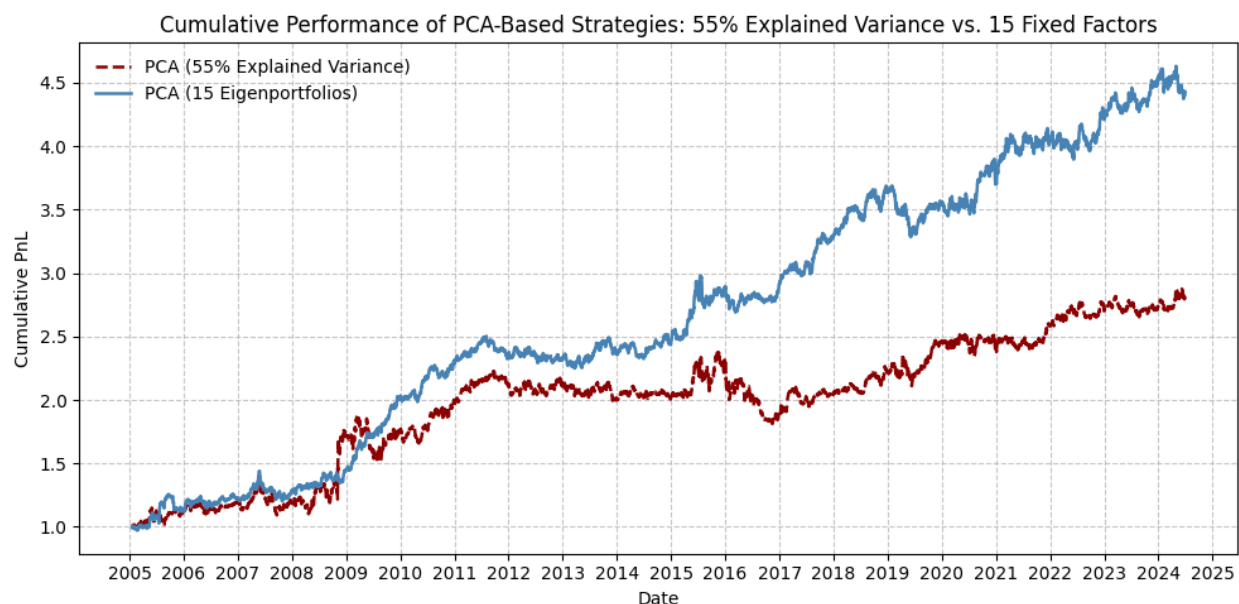
Note: Cumulative profit and loss (PNL) comparison between two PCA-based strategies from 2005 to 2024. The first strategy selects a variable number of principal components to explain 55% of total variance (dashed red line), while the second uses a fixed set of 15 eigenportfolios (solid blue line). The fixed-factor approach demonstrates superior long-term performance, particularly during trending market regimes.

[Figure 12](#) tracks, in tandem, the annualized 252-day volatility of the CSI 300 (grey bars, right axis) and the number of principal components required to explain 55 percent of cross-sectional return variance (red line, left axis). A pronounced inverse co-movement emerges: when market turbulence intensifies, the dimensionality of the return space contracts. During the 2008 global financial crisis, the 2015 margin-deleveraging episode, and the COVID-19 shock of early 2020, volatility spiked above 35 percent while the factor count occasionally dropped below three PCs, indicating that a handful of dominant latent forces overwhelmed idiosyncratic variation. Conversely, in more tranquil regimes—such as 2012–2014 and again in 2021–2023—volatility subsided toward 10–15 percent and the model required a markedly larger basis, often exceeding fifteen components, to capture the same share of dispersion. This pattern reflects a more fragmented risk structure with weaker common components.

Figure 12. Number of factors vs. CSI300 historical volatility (2005-2024)
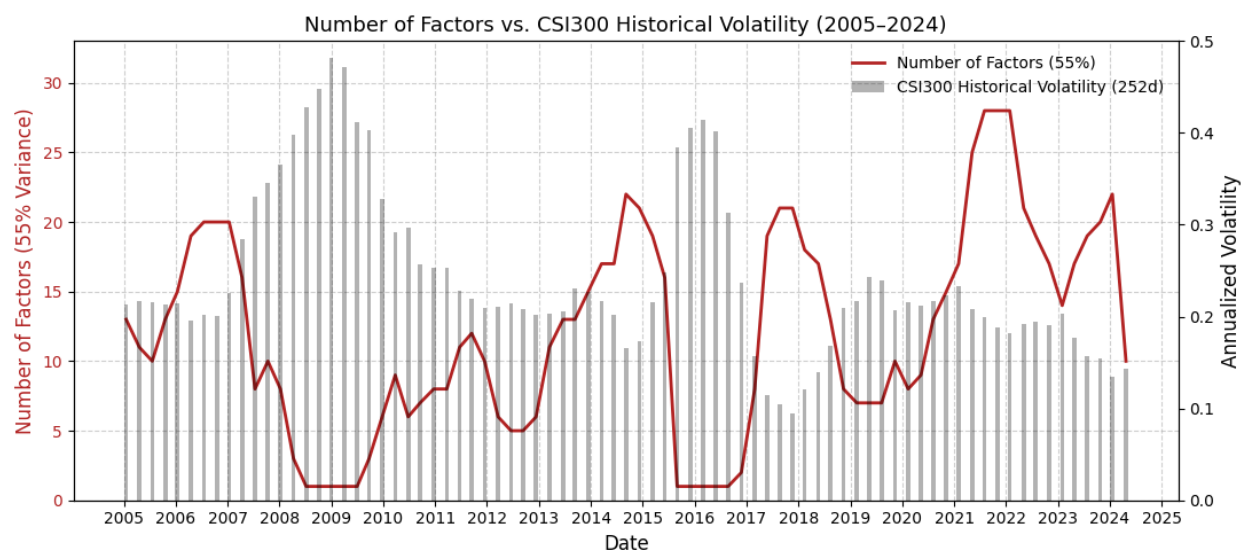
Note: Number of principal components required to explain 55% of return variance (red line) and CSI300 historical volatility (blue bars) from 2005 to 2024. The number of factors fluctuates significantly across market regimes, often increasing during volatile periods.

These dynamics corroborate established evidence in empirical asset pricing: high-volatility regimes compress cross-sectional correlations around a few systemic shocks, whereas tranquil markets disperse risk across numerous micro-drivers. From a modeling standpoint, the chart underscores a key implication of this study: factor complexity is inherently state-dependent. Using a fixed factor set risks under-fitting in highly integrated markets and over-fitting in fragmented ones. An adaptive framework—defined by a variance-explained threshold such as the

55 percent rule—better aligns effective dimensionality with prevailing conditions, offering a more economical and context-sensitive basis for residual mean-reversion trading.

6.4 Incorporating Trading Volume into Mean-Reversion Signals

This section investigates the impact of trading volume on the informational efficiency and effectiveness of mean-reversion signals. Inspired by Avellaneda and Lee (2010), we adopt a *trading-time–inspired* framework in which raw returns are rescaled by contemporaneous trading activity, effectively transforming the signal space from uniform calendar time into a liquidity-adjusted measure of time. Formally, the adjusted return is defined as:

$$R_t^{adj} = \frac{S_{t+\Delta t}-S_t}{S_t} \cdot \frac{\langle V \rangle}{V(t+\Delta t)-V(t)} = R_t \cdot \frac{\langle V \rangle}{V(t+\Delta t)-V(t)} \tag{32}$$

where R_t is the conventional return over the interval Δt, $V(t+\Delta t) - V(t)$ is the cumulative trading volume during the interval, and $\langle V \rangle$ is the trailing average daily volume over a 10-day lookback window.

This specification penalizes large returns that occur on unusually high volume and amplifies those arising on light volume. The intuition, consistent with market microstructure theory, is that price moves on heavy trading activity are more likely to reflect genuine information, while those on light activity are more likely to reflect transitory liquidity shocks or behavioral noise. Supporting evidence is provided by Chordia et al. (2005), who document that high-volume trades exhibit lower reversal probabilities and faster convergence to informational equilibrium. Similarly, Lee and Swaminathan (2000) find that reversals are statistically stronger following low-volume moves, suggesting conditional predictability tied to liquidity conditions.

While the adjustment is motivated by theory, whether it systematically improves trading performance remains an empirical question. We apply this volume-rescaled framework to both PCA-based strategies—using a fixed set of 15 eigenportfolios—and synthetic ETF-based strategies constructed via sector-level return aggregation. Performance results are summarized in Table 7 and Table 8, respectively.

For the PCA strategy, the full-sample annualized return increased to 6.57% with a Sharpe ratio of 0.64, indicating a modest improvement. However, the gains were episodic—concentrated in 2005–2006 and 2017–2019—and deteriorated sharply during stress periods such as 2020 and

2023. This pattern suggests that PCA-derived latent factors are prone to instability when subjected to volume-based normalization, particularly under regime shifts in market structure or trading behavior. This observation is consistent with Khandani and Lo (2009), who highlight the fragility of PCA-based statistical arbitrage during structural changes. Since PCA implicitly assumes a stationary covariance structure, reweighting through volume adjustments can exacerbate estimation noise or suppress economically meaningful signals.

In contrast, the synthetic ETF-based strategy delivered more robust and consistent improvements under the volume-rescaled specification. The Sharpe ratio rose to 0.75, accompanied by gains in downside risk measures such as Sortino, Calmar, and Conditional VaR. Performance remained resilient across both tranquil and turbulent regimes—including 2009–2010 and 2015–2019—indicating that the combination of sector-level economic exposure and liquidity-sensitive rescaling provides a more stable framework for signal generation.

The contrasting outcomes between PCA- and ETF-based strategies can be attributed to how volume rescaling interacts with the underlying factor structure. In PCA models, the covariance matrix of returns is estimated under the implicit assumption of stationarity. Volume-adjusted returns effectively reweight time periods, increasing the influence of episodes with unusually low trading activity while downweighting periods of heavy volume. This distortion can destabilize the estimated eigenvectors, particularly during regime shifts when correlation structures rotate abruptly, thereby amplifying estimation noise or suppressing economically meaningful factors. In other words, volume normalization in PCA can transform transitory liquidity effects into spurious common components, undermining the robustness of residual-based signals.

By contrast, synthetic ETF factors are constructed through sector-level aggregation, which inherently filters out idiosyncratic noise and yields economically interpretable exposures. Because these sector indices are tied to persistent economic structures, their covariance representation is less sensitive to short-term distortions introduced by trading-time adjustments. Volume rescaling thus acts more like a liquidity-aware filter—moderating signal intensity without destabilizing the underlying factor basis. The result is a structurally more stable improvement, as evidenced by the superior risk-adjusted performance in Table 8.

These findings echo the broader literature on liquidity and asset pricing. Pastor and Stambaugh (2003) show that liquidity risk is priced and significantly affects return dynamics, while Brunnermeier and Pedersen (2009) demonstrate that funding and market liquidity jointly determine arbitrage capacity under stress. In this context, our results suggest that ETF-based factors, by embedding sectoral economic structures, are more resilient to liquidity-sensitive transformations than purely statistical PCA factors.

Overall, the evidence supports the view that trading volume contains relevant state information that can enhance the effectiveness of mean-reversion signals. While PCA-based models show conditional and fragile improvements, synthetic ETF strategies yield structurally robust gains, underscoring the importance of factor interpretability and model design in liquidity-aware arbitrage.

Table 7. Annual performance metrics of a mean-reversion strategy that employs 15 fixed principal components as factors.

| Year	Mean	Std. Dev.	Sharpe	Sortino	Max Drawdown	Calmar	t-stat	z-stat	JB test (p-value)	CVaR (95%)	Omega
2005	0.3062	0.2458	1.2455	3.4887	-0.0874	3.5034	1.1976	1.1976	0.0000	-0.0164	1.4728
2006	0.0776	0.2311	0.3356	0.8676	-0.0983	0.7892	0.3282	0.3282	0.0000	-0.0167	1.1118
2007	-0.0378	0.1029	-0.3677	-0.5885	-0.1354	-0.2792	-0.3603	-0.3603	0.7366	-0.0133	0.9424
2008	-0.0399	0.0944	-0.4225	-0.6763	-0.1428	-0.2794	-0.4174	-0.4174	0.7124	-0.0122	0.9339
2009	0.2059	0.0802	2.5673	4.4410	-0.0506	4.0670	2.5262	2.5262	0.8472	-0.0097	1.5068
2010	0.1414	0.0745	1.8983	3.0462	-0.0620	2.2816	1.8603	1.8603	0.7155	-0.0093	1.3557
2011	0.0478	0.0635	0.7526	1.1808	-0.0638	0.7490	0.7405	0.7405	0.6916	-0.0082	1.1282
2012	-0.0190	0.0559	-0.3400	-0.5445	-0.0536	-0.3548	-0.3338	-0.3338	0.2263	-0.0076	0.9473
2013	0.0496	0.0703	0.7060	1.0454	-0.0386	1.2867	0.6861	0.6861	0.0078	-0.0094	1.1199
2014	0.1136	0.0639	1.7784	2.9777	-0.0424	2.6790	1.7535	1.7535	0.9560	-0.0079	1.3276
2015	0.0287	0.0991	0.2899	0.4159	-0.1006	0.2855	0.2853	0.2853	0.0268	-0.0144	1.0484
2016	0.0276	0.0625	0.4420	0.6549	-0.0694	0.3980	0.4350	0.4350	0.0000	-0.0086	1.0767
2017	0.1265	0.0528	2.3958	4.2086	-0.0613	2.0646	2.3575	2.3575	0.9775	-0.0061	1.4535
2018	0.0864	0.0606	1.4260	2.0898	-0.0375	2.3013	1.4003	1.4003	0.0000	-0.0082	1.2563
2019	0.0922	0.0602	1.5331	2.4750	-0.0405	2.2752	1.5085	1.5085	0.0001	-0.0076	1.2928
2020	-0.0573	0.0725	-0.7906	-1.1273	-0.1232	-0.4649	-0.7763	-0.7763	0.0000	-0.0108	0.8755
2021	0.0532	0.0670	0.7948	1.4178	-0.0452	1.1775	0.7805	0.7805	0.0595	-0.0078	1.1324
2022	0.0563	0.0760	0.7414	1.1112	-0.0774	0.7273	0.7265	0.7265	0.0181	-0.0104	1.1291
2023	0.0086	0.0544	0.1591	0.2430	-0.0666	0.1300	0.1559	0.1559	0.1222	-0.0076	1.0258
2024	0.0434	0.0615	0.7056	1.2231	-0.0339	1.2806	0.4787	0.4787	0.8085	-0.0073	1.1191
All	0.0657	0.1021	0.6433	1.2099	-0.2449	0.2681	2.7851	2.7851	0.0000	-0.0109	1.1491

Note: This table reports the annual performance metrics of a mean-reversion strategy that employs 15 fixed principal components as factors, with return signals adjusted according to the trading-time framework. The adjustment normalizes raw returns based on recent realized trading volume to account for liquidity conditions. Results are reported from 2005 through 2024, along with full-sample statistics in the last row.

Table 8. Annualized performance statistics for a statistical arbitrage strategy based on synthetic capitalization-weighted sector indices.

Year	Mean	Std. Dev.	Sharpe	Sortino	Max Drawdown	Calmar	t-stat	z-stat	JB test (p-value)	CVaR (95%)	Omega
2005	0.4588	0.2274	2.0175	3.9493	-0.1435	3.1965	1.9400	1.9400	0.0000	-0.0239	1.5854
2006	0.2669	0.2387	1.1181	2.5705	-0.1569	1.7005	1.0934	1.0934	0.0000	-0.0215	1.3249
2007	-0.0249	0.1097	-0.2268	-0.3506	-0.1478	-0.1685	-0.2223	-0.2223	0.0002	-0.0152	0.9632
2008	0.0868	0.0899	0.9656	1.4910	-0.0894	0.9711	0.9540	0.9540	0.0000	-0.0118	1.1773
2009	0.1388	0.0689	2.0148	4.1313	-0.0427	3.2486	1.9825	1.9825	0.0000	-0.0070	1.4119
2010	0.0918	0.0637	1.4424	2.4014	-0.0344	2.6659	1.4135	1.4135	0.0000	-0.0078	1.2877
2011	-0.0004	0.0479	-0.0074	-0.0105	-0.0356	-0.0100	-0.0073	-0.0073	0.0000	-0.0071	0.9987
2012	0.0168	0.0404	0.4158	0.6932	-0.0304	0.5519	0.4083	0.4083	0.0222	-0.0052	1.0706
2013	0.0436	0.0566	0.7701	1.2416	-0.0583	0.7465	0.7484	0.7484	0.7542	-0.0074	1.1318
2014	0.0271	0.0460	0.5891	1.0279	-0.0482	0.5616	0.5809	0.5809	0.8854	-0.0056	1.0966
2015	0.1289	0.0915	1.4094	2.3052	-0.0644	2.0024	1.3868	1.3868	0.0000	-0.0117	1.2674
2016	0.0422	0.0468	0.9014	1.2514	-0.0336	1.2563	0.8870	0.8870	0.0000	-0.0063	1.1611
2017	-0.0124	0.0603	-0.2057	-0.2592	-0.0660	-0.1881	-0.2024	-0.2024	0.0000	-0.0096	0.9653
2018	-0.0532	0.0526	-1.0130	-1.3033	-0.1038	-0.5131	-0.9947	-0.9947	0.0000	-0.0074	0.8407
2019	0.1215	0.0579	2.0997	3.5563	-0.0331	3.6755	2.0661	2.0661	0.0000	-0.0071	1.4358
2020	0.0059	0.0778	0.0752	0.0932	-0.1133	0.0517	0.0739	0.0739	0.0000	-0.0130	1.0137
2021	0.0077	0.0648	0.1183	0.1997	-0.1025	0.0748	0.1162	0.1162	0.0608	-0.0081	1.0194
2022	0.0600	0.0529	1.1332	1.9839	-0.0327	1.8362	1.1105	1.1105	0.6662	-0.0064	1.1973
2023	0.0048	0.0397	0.1216	0.1986	-0.0489	0.0985	0.1191	0.1191	0.0003	-0.0050	1.0199
2024	0.0502	0.0812	0.6180	0.7359	-0.0829	0.6057	0.4193	0.4193	0.0000	-0.0133	1.1216
All	0.0728	0.0969	0.7510	1.2504	-0.2598	0.2802	3.2513	3.2513	0.0000	-0.0115	1.1814

Note: This table presents annualized performance statistics for a statistical arbitrage strategy based on synthetic capitalization-weighted sector indices, incorporating volume-adjusted return signals. The strategy applies the trading-time framework to modulate signal strength based on trading intensity, with volume adjustments computed over a 10-day lookback window. The evaluation spans the period from 2005 to 2024, with aggregate results presented in the final row.

[Figure 13](#) contrasts the cumulative PnL of the 15-factor PCA strategy measured in calendar time with its trading-time counterpart, where returns are inversely scaled by contemporaneous volume. While both specifications remain profitable over 2005–2024, the volume-adjusted variant consistently underperforms, with the performance gap widening notably after 2015. This pattern indicates that, for latent-factor portfolios, inverse-volume scaling tends to attenuate rather than amplify the mean-reversion signal.

Figure 13. Comparison of cumulative PNL for PCA-based strategies
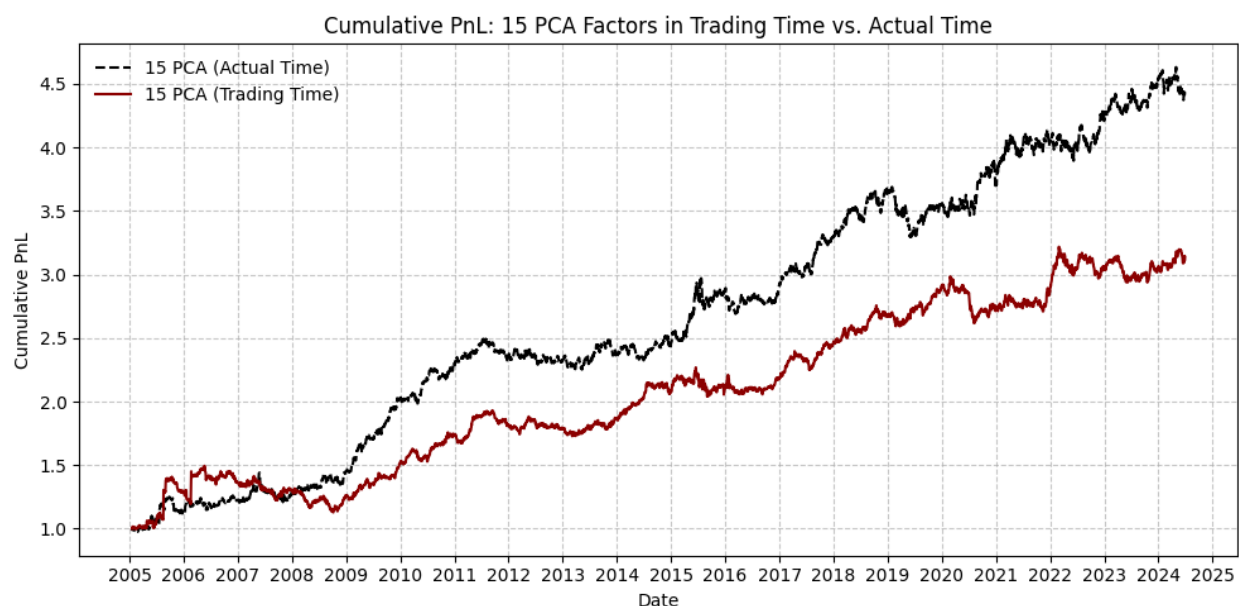
Note: Comparison of cumulative PNL for PCA-based strategies using 15 fixed principal components under trading-time and actual-time frameworks. The trading-time approach adjusts returns based on recent volume activity to account for liquidity conditions. While both approaches exhibit mean-reversion profitability over time, the actual-time strategy consistently outperforms its volume-adjusted counterpart from 2005 to 2024.

Two mechanisms likely drive this shortfall. First, attenuation of contrarian signals. PCA eigenportfolios already capture dominant covariance modes, leaving residuals that are smoothed and cross-sectionally diversified. Penalizing high-volume moves further dampens these residuals, thereby suppressing otherwise tradable reversals. Second, a timescale mismatch. Both theoretical and empirical studies in market microstructure (Lo and Wang, 2000; Hasbrouck, 2009) indicate that volume clocks are effective at isolating single-stock liquidity shocks over intraday horizons. In contrast, factor-based statistical arbitrage typically operates on multi-day to multi-week horizons; applying an intraday-motivated filter at such coarser frequencies risks injecting noise rather than enhancing informational efficiency.

Empirically, the calendar-time strategy's robustness aligns with prior evidence that PCA-derived contrarian signals profit from structural persistence (Gatev et al., 2006; Avellaneda and Lee, 2010). Volume normalization disrupts that persistence, especially in the post-2015 regime when turnover accelerates and eigenvector stability deteriorates. Hence, while trading-time formulations may benefit liquidity-sensitive, single-asset trades, traditional return measures remain superior for principal-component statistical arbitrage, preserving both signal strength and

capital-allocation efficiency. This finding echoes the broader liquidity-risk literature: Pastor and Stambaugh (2003) show that returns are predictably shaped by market-wide liquidity fluctuations, while Brunnermeier and Pedersen (2009) demonstrate that funding and market liquidity interact to destabilize strategies under stress. In this light, volume-scaling of PCA factors may inadvertently embed liquidity spirals rather than extract persistent arbitrage signals.

Figure 14 plots cumulative PnL in trading time for two volume-adjusted mean-reversion portfolios: a sector-based synthetic-ETF strategy (solid blue) and a 15-factor PCA strategy (dashed red). Both strategies generate positive cumulative returns over the 2005–2024 sample period. However, three salient features highlight the superior performance of the ETF-based model.

Figure 14. Cumulative profit-and-loss (PNL) of two statistical arbitrage strategies
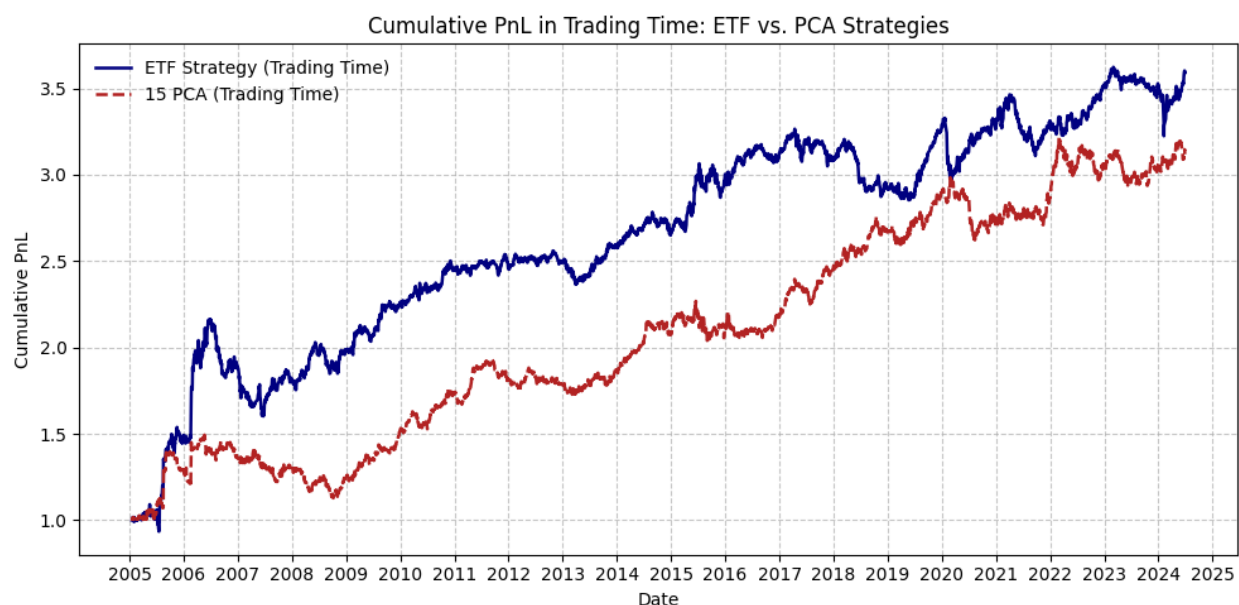
Note: This figure compares the cumulative profit-and-loss (PNL) of two statistical arbitrage strategies implemented under the trading-time framework: one using synthetic ETF-based factors and the other using 15 principal components (PCA). The ETF strategy exhibits superior performance, demonstrating a more consistent and higher cumulative return across the sample period. The results suggest that incorporating trading volume through trading-time adjustments enhances signal reliability, particularly for ETF-based strategies.

First, from the first volatility spike in 2006, the ETF curve rises more steeply and maintains a persistent lead. Even during the 2010–2014 plateau—when both strategies move largely sideways—the ETF portfolio stabilizes at a higher base level, sustaining a 0.4–0.6 PnL multiple

over its PCA counterpart. A renewed acceleration after 2017 further widens the gap, reaching nearly one unit of cumulative PnL by 2024.

Second, the ETF strategy exhibits shallower drawdowns. Major setbacks in 2008, 2011–2012, and early 2020 are visibly milder for the ETF strategy, which also recovers to new highs more quickly than the PCA portfolio. This resilience is consistent with Table 8, where the ETF model demonstrates superior risk-adjusted performance, including lower Conditional VaR and higher Sortino ratios during stress episodes.

Third, the ETF strategy demonstrates greater robustness to volume scaling. Because synthetic ETFs aggregate capitalization-weighted sector flows, their signals remain anchored to macroeconomic themes even after inverse-volume rescaling. In contrast, PCA eigenportfolios hinge on residual cross-sectional noise; scaling by volume dampens those already fragile reversals, flattening the red curve—especially evident in the post-2015 divergence.

Taken together, Figure 14 reinforces the liquidity-compatibility thesis advanced by Lo and Wang (2000) and Chordia et al. (2005): volume clocks enhance strategies whose factors correspond to stable, economically interpretable risks, but offer little benefit—and may even be detrimental—when applied to statistically extracted, correlation-sensitive portfolios such as PCA eigenvectors.

6.5 Extreme Periods Comparison and Analysis

To evaluate the robustness of statistical arbitrage across heterogeneous market environments, we examine three structurally distinct regimes: the Global Financial Crisis (GFC, 2007–2008), the Chinese Bull Market (BNB, 2014–2016), and the COVID-19 pandemic (2020–2022). These regimes capture, respectively, systemic risk, directional momentum, and heightened macroeconomic uncertainty. Table 9 and 10 report the annualized performance of PCA-based and synthetic ETF-based strategies across these regimes. Our objective is to assess not only the relative effectiveness of the two approaches but also the extent to which market conditions shape the performance of mean-reversion strategies.

During the GFC, both models delivered comparable returns with limited drawdowns, underscoring their resilience in periods of systemic distress. The PCA strategy generated an average return of 9.39% with a Sharpe ratio of 0.79, while the ETF-based strategy produced

a nearly identical return of 9.38% but slightly stronger risk-adjusted outcomes (Sharpe 0.94, Calmar 0.58). These results indicate that both residual-based and sector-anchored signals remain effective in crisis conditions, consistent with prior findings (Gatev et al., 2006; Avellaneda and Lee, 2010) that arbitrage opportunities tend to persist longer when market dislocations are severe.

By contrast, the BNB regime provided a particularly favorable environment for sector-based arbitrage. The ETF strategy achieved markedly stronger performance across all dimensions, with an average annual return of 13.36%, a Sharpe ratio of 2.05, and a Calmar ratio of 2.55. The PCA approach, while profitable (Mean 7.51%), was substantially less efficient, recording a Sharpe of 0.88 and Calmar of 0.77. This divergence highlights the ETF strategy's ability to capture macro-driven sector rotation and persistent trending behaviors, effects amplified in sustained bull markets. These findings align with Kelly et al. (2020), who document the effectiveness of sector tilts in high-dispersion environments.

The COVID-19 pandemic presents a different performance landscape. Although both strategies remained profitable, with mean returns of 6.22% (PCA) and 4.89% (ETF), the PCA strategy displayed stronger downside protection. Its Sharpe ratio (0.77) was statistically indistinguishable from that of the ETF strategy (0.81), yet the PCA portfolio experienced shallower drawdowns, more stable CVaR profiles, and a higher Omega ratio. These results suggest that residual-based methods were more adaptive to the heightened volatility and rapid regime shifts of the pandemic, as they dynamically captured evolving cross-sectional co-movements without reliance on fixed sectoral structures.

Beyond macro-level interpretations, microstructure dynamics help explain why regime conditions magnify the divergence between PCA- and ETF-based approaches. During systemic crises such as the GFC, market-wide deleveraging and liquidity freezes amplify cross-sectional co-movements, making residual-based signals more persistent and thereby sustaining PCA-driven arbitrage. In contrast, in momentum-driven expansions such as the BNB, heightened retail participation and sectoral capital inflows reinforce capitalization-weighted sector rotations, which naturally align with ETF-based factors. Under the COVID regime, characterized by abrupt volatility spikes and rapid state shifts, PCA's model-free structure adapts more readily to evolving correlation matrices, whereas sector-anchored ETFs lag in adjusting to transient shocks. This microstructural lens clarifies how liquidity provision, trading volume composition, and the

speed of correlation adjustment interact with strategy design to shape relative performance across regimes.

Taken together, the evidence demonstrates that the relative merits of PCA- and ETF-based statistical arbitrage are regime-dependent. The ETF approach is more effective in expansionary, trend-dominated markets such as the BNB period, whereas the PCA framework provides superior robustness in turbulent or uncertain environments such as the GFC and COVID-19. This divergence underscores that statistical arbitrage is not universally optimal but contingent on the interplay between model design and prevailing macroeconomic conditions. From a practical perspective, these findings highlight the potential benefits of dynamic model selection or ensemble frameworks that condition arbitrage execution on regime-specific state variables, thereby enhancing out-of-sample robustness and portfolio resilience.

Table 9. Annualized performance metrics of the PCA-based statistical arbitrage strategy across eight distinct macroeconomic regimes.

Period (Mean)	Mean	Std. Dev.	Sharpe	Sortino	Max Drawdown	Calmar	t-stat	z-stat	JB test (p-value)	CVaR (95%)	Omega
2005	0.1551	0.1518	1.0219	1.4039	-0.1095	1.4163	0.9826	0.9826	0.0000	-0.0208	1.2368
2006	0.0785	0.1068	0.7348	1.2215	-0.0750	1.0472	0.7186	0.7186	0.0000	-0.0131	1.1425
Pre F.C.	0.1164	0.1309	0.8893	1.2874	-0.1095	1.0629	1.2184	0.0000	0.0000	-0.0170	1.1930
2007	0.0573	0.1208	0.4739	0.6210	-0.1647	0.3477	0.4644	0.4644	0.0000	-0.0171	1.0812
2008	0.1298	0.1157	1.1224	1.7345	-0.0594	2.1848	1.1090	1.1090	0.0001	-0.0149	1.2047
In F.C.	0.0939	0.1182	0.7942	1.1193	-0.1647	0.5699	1.1053	0.0000	0.0000	-0.0162	1.1402
2009	0.3522	0.0953	3.6977	5.6140	-0.0430	8.1951	3.6385	3.6385	0.0002	-0.0116	1.8761
2010	0.1294	0.0756	1.7114	2.8199	-0.0452	2.8608	1.6771	1.6771	0.0013	-0.0094	1.3335
Post F.C.	0.2413	0.0862	2.7979	4.4260	-0.0452	5.3349	3.8856	0.0000	0.0000	-0.0106	1.6107
2011	0.0262	0.0658	0.3979	0.6247	-0.0715	0.3662	0.3916	0.3916	0.1852	-0.0088	1.0679
2012	-0.0085	0.0697	-0.1221	-0.1755	-0.0519	-0.1641	-0.1199	-0.1199	0.0000	-0.0101	0.9802
2013	0.0260	0.0798	0.3256	0.5406	-0.0521	0.4986	0.3164	0.3164	0.9766	-0.0101	1.0531
Pre B.N.B.	0.0145	0.0719	0.2017	0.3120	-0.1016	0.1426	0.3421	0.0000	0.0030	-0.0098	1.0334
2014	0.0387	0.0632	0.6126	0.9188	-0.0526	0.7356	0.6040	0.6040	0.1428	-0.0083	1.1056
2015	0.1701	0.1162	1.4630	2.3896	-0.0854	1.9918	1.4396	1.4396	0.0000	-0.0145	1.2921
2016	0.0167	0.0668	0.2498	0.2981	-0.0700	0.2382	0.2458	0.2458	0.0000	-0.0106	1.0441
In B.N.B.	0.0751	0.0855	0.8781	1.2734	-0.0971	0.7733	1.4976	0.0000	0.0000	-0.0119	1.1699
2017	0.1236	0.0578	2.1374	3.7191	-0.0350	3.5313	2.1032	2.1032	0.9353	-0.0069	1.4046
2018	0.1070	0.0652	1.6404	2.3339	-0.0464	2.3079	1.6109	1.6109	0.0000	-0.0092	1.3213
2019	-0.0216	0.0711	-0.3033	-0.4132	-0.1096	-0.1968	-0.2984	-0.2984	0.0000	-0.0108	0.9501
Pre Covid	0.0696	0.0650	1.0714	1.5520	-0.1096	0.6355	1.8248	0.0000	0.0000	-0.0093	1.1950
2020	0.0539	0.0842	0.6398	1.0735	-0.0508	1.0607	0.6282	0.6282	0.4600	-0.0108	1.1086

Period (Mean)	Mean	Std. Dev.	Sharpe	Sortino	Max Drawdown	Calmar	t-stat	z-stat	JB test (p-value)	CVaR (95%)	Omega
2021	0.0916	0.0779	1.1769	2.0863	-0.0413	2.2179	1.1557	1.1557	0.5679	-0.0093	1.2024
2022	0.0410	0.0792	0.5182	0.8431	-0.0530	0.7746	0.5078	0.5078	0.0010	-0.0100	1.0887
In Covid	0.0622	0.0804	0.7741	1.3077	-0.0592	1.0503	1.3157	0.0000	0.0432	-0.0102	1.1322
2023	0.0777	0.0580	1.3407	2.1632	-0.0425	1.8275	1.3138	1.3138	0.6199	-0.0075	1.2371
2024	-0.0441	0.0719	-0.6134	-0.9274	-0.0558	-0.7896	-0.4162	-0.4162	0.8901	-0.0097	0.9050
Post Covid	0.0382	0.0628	0.6089	0.9404	-0.0558	0.6850	0.7258	0.0000	0.2863	-0.0085	1.1028
All	0.0833	0.0883	0.9434	1.3736	-0.1647	0.5060	4.0840	4.0840	0.0000	-0.0120	1.1806

Note: This table summarizes the annualized performance metrics of the PCA-based statistical arbitrage strategy across eight distinct macroeconomic regimes, including the Global Financial Crisis (2007–2008), the Chinese Bull Market (2014–2016), and the COVID-19 pandemic (2020–2022). Regime averages are computed across the relevant subperiods to facilitate macro-level performance attribution.

Table 10. Annual and regime-level performance of the Synthetic ETF-based strategy using capitalization-weighted sector indices as factors.

Year	Mean	Std. Dev.	Sharpe	Sortino	Max Drawdown	Calmar	t-stat	z-stat	JB test (p-value)	CVaR (95%)	Omega
2005	0.1694	0.2317	0.7312	1.5110	-0.1964	0.8627	0.7031	0.7031	0.0000	-0.0241	1.1714
2006	0.1229	0.2499	0.4920	0.7823	-0.1631	0.7539	0.4811	0.4811	0.0000	-0.0288	1.1321
Pre F.C.	0.1461	0.2411	0.6059	1.0749	-0.1964	0.7439	0.8301	0.0000	0.0000	-0.0270	1.1520
2007	0.0442	0.1110	0.3984	0.5505	-0.1608	0.2750	0.3904	0.3904	0.0000	-0.0158	1.0711
2008	0.1426	0.0887	1.6078	2.1284	-0.0577	2.4718	1.5885	1.5885	0.0000	-0.0124	1.3167
In F.C.	0.0938	0.1003	0.9351	1.2584	-0.1608	0.5834	1.3013	0.0000	0.0000	-0.0145	1.1752
2009	0.1885	0.0677	2.7821	4.7027	-0.0239	7.8738	2.7376	2.7376	0.3413	-0.0078	1.5838
2010	0.0978	0.0675	1.4498	2.3811	-0.0405	2.4126	1.4207	1.4207	0.0000	-0.0082	1.2861
Post F.C.	0.1433	0.0676	2.1204	3.5393	-0.0405	3.5356	2.9446	0.0000	0.0000	-0.0081	1.4313
2011	0.0467	0.0505	0.9243	1.3723	-0.0372	1.2554	0.9095	0.9095	0.0000	-0.0070	1.1718
2012	0.1048	0.0446	2.3513	3.7359	-0.0190	5.5260	2.3089	2.3089	0.4876	-0.0057	1.4645
2013	0.0233	0.0595	0.3927	0.6748	-0.0476	0.4901	0.3816	0.3816	0.0031	-0.0077	1.0658
Pre B.N.B.	0.0585	0.0518	1.1296	1.7854	-0.0476	1.2283	1.9160	0.0000	0.0000	-0.0069	1.2063
2014	0.0612	0.0465	1.3149	2.0348	-0.0272	2.2461	1.2965	1.2965	0.1470	-0.0060	1.2463
2015	0.2273	0.0922	2.4643	4.2807	-0.0523	4.3459	2.4249	2.4249	0.0000	-0.0105	1.5266
2016	0.1126	0.0454	2.4782	4.0652	-0.0201	5.6042	2.4386	2.4386	0.4631	-0.0055	1.4874
In B.N.B.	0.1336	0.0652	2.0492	3.2296	-0.0523	2.5544	3.4949	0.0000	0.0000	-0.0082	1.4400
2017	-0.0466	0.0529	-0.8804	-1.1969	-0.0570	-0.8163	-0.8664	-0.8664	0.0000	-0.0083	0.8628

Year	Mean	Std. Dev.	Sharpe	Sortino	Max Drawdown	Calmar	t-stat	z-stat	JB test (p-value)	CVaR (95%)	Omega
2018	0.0499	0.0527	0.9460	1.2858	-0.0624	0.7994	0.9290	0.9290	0.0000	-0.0073	1.1704
2019	0.1785	0.0506	3.5270	6.2941	-0.0252	7.0835	3.4705	3.4705	0.0000	-0.0056	1.8424
Pre Covid	0.0606	0.0523	1.1585	1.6738	-0.0867	0.6990	1.9731	0.0000	0.0000	-0.0073	1.2156
2020	-0.0124	0.0541	-0.2295	-0.3609	-0.0413	-0.3010	-0.2254	-0.2254	0.5567	-0.0071	0.9642
2021	0.0252	0.0719	0.3505	0.6629	-0.0871	0.2893	0.3442	0.3442	0.0005	-0.0083	1.0580
2022	0.1344	0.0531	2.5299	4.1298	-0.0304	4.4142	2.4792	2.4792	0.7352	-0.0066	1.5060
In Covid	0.0489	0.0604	0.8104	1.3781	-0.0871	0.5619	1.3775	0.0000	0.0000	-0.0075	1.1401
2023	0.0374	0.0414	0.9044	1.5572	-0.0398	0.9408	0.8863	0.8863	0.1487	-0.0050	1.1559
2024	0.0978	0.0833	1.1740	1.6487	-0.0464	2.1100	0.7965	0.7965	0.0000	-0.0119	1.2523
Post Covid	0.0570	0.0583	0.9781	1.4396	-0.0519	1.0982	1.1658	0.0000	0.0000	-0.0077	1.1980
All	0.0900	0.0982	0.9172	1.3689	-0.2088	0.4312	3.9706	3.9706	0.0000	-0.0119	1.2276

Note: This table reports the annual and regime-level performance of the Synthetic ETF-based strategy using capitalization-weighted sector indices as factors. The strategy is evaluated over the same macroeconomic regimes as in Table 9, covering the pre-, during-, and post-periods of major market dislocations such as the GFC, BNB rally, and COVID-19 shock. Performance metrics mirror those in Table 9, allowing for direct comparison between structurally-defined and statistically-derived mean-reversion signals under varying market conditions.

7. Sensitivity Analysis

In the preceding chapters, we constructed and evaluated market-neutral statistical arbitrage strategies based on PCA and grounded in the principles of pairs trading. While the empirical evidence confirms the effectiveness of PCA-based factor models in identifying mean-reversion opportunities, it is well recognized that the performance of such strategies is highly sensitive to modeling choices and parameter specifications. In practice, factors such as the number of retained principal components, the variance threshold for dimensionality reduction, the entry and exit thresholds for signal generation, and the length of the rolling estimation window can materially affect both return and risk characteristics.

To systematically assess how these design choices influence strategy outcomes, this chapter conducts a series of sensitivity analyses. The objective is to evaluate robustness across a range of plausible parameter configurations and to identify settings that deliver a balanced trade-off between profitability, risk management, and stability.

Section 7.1 examines the impact of varying the number of principal components. By comparing strategies that retain 3, 5, and 10 components, we evaluate their effects on cumulative returns, volatility, and drawdowns. This analysis addresses whether a parsimonious factor structure is sufficient to capture mean-reversion opportunities, or whether higher-dimensional representations contribute additional alpha.

Section 7.2 shifts the focus to cumulative variance explained. Using thresholds of 50%, 60%, and 70%, we investigate the trade-off between parsimony and explanatory power, and assess which thresholds yield more robust and profitable strategies.

Section 7.3 explores the role of the signal threshold in trade initiation. As this parameter governs the model's sensitivity to deviations in residuals, it directly influences trading frequency, portfolio turnover, and the distribution of returns. By comparing conservative versus aggressive thresholds, we evaluate their implications for strategy dynamics and risk-adjusted performance across different market environments.

Finally, Section 7.4 considers the length of the rolling estimation window used in PCA. We contrast shorter versus longer historical windows to assess how responsiveness and stability interact in shaping the model's adaptability to changing market conditions. This provides insights into the trade-off between timely signal detection and resilience against noise.

Collectively, these sensitivity tests deepen our understanding of the structural dependencies embedded within PCA-based statistical arbitrage models. They provide practical guidance for parameter tuning in live trading environments and demonstrate how seemingly minor modeling choices can materially alter returns, risk exposures, and capital efficiency. Importantly, this analysis highlights that sensitivity to design choices directly conditions the out-of-sample robustness of mean-reversion portfolios, thereby bridging the gap between theoretical model design and real-world implementation.

7.1 Impact of PCA Factor Dimensionality on Strategy Performance

Table 11 reports the annualized performance metrics for PCA-based strategies using different numbers of retained components, ranging from 5 to 45, as risk factors. Several clear patterns emerge from the empirical results, consistent with prior findings on factor

dimensionality in statistical arbitrage (Laloux et al., 2000; Avellaneda and Lee, 2010). We highlight the most salient observations below.

First, the relationship between the number of components and average annualized returns is distinctly non-monotonic (Figure 15 and 16, see also Figure 10). Average returns rise from 7.9% with 5 components to a peak of 9.9% at 10 components, before steadily declining to just 1.6% with 45 components. This suggests the existence of an optimal factor dimensionality: a limited number of components can effectively capture dominant market- and sector-level co-movements, yielding statistically robust residual signals for mean-reversion trading. Beyond this point, however, the inclusion of additional components introduces high-frequency noise, erodes alpha, and increases the risk of overfitting (Laloux et al., 2000), thereby undermining the robustness of out-of-sample performance.

**Figure 15.** Cumulative PNL of PCA Strategies with Varying Components
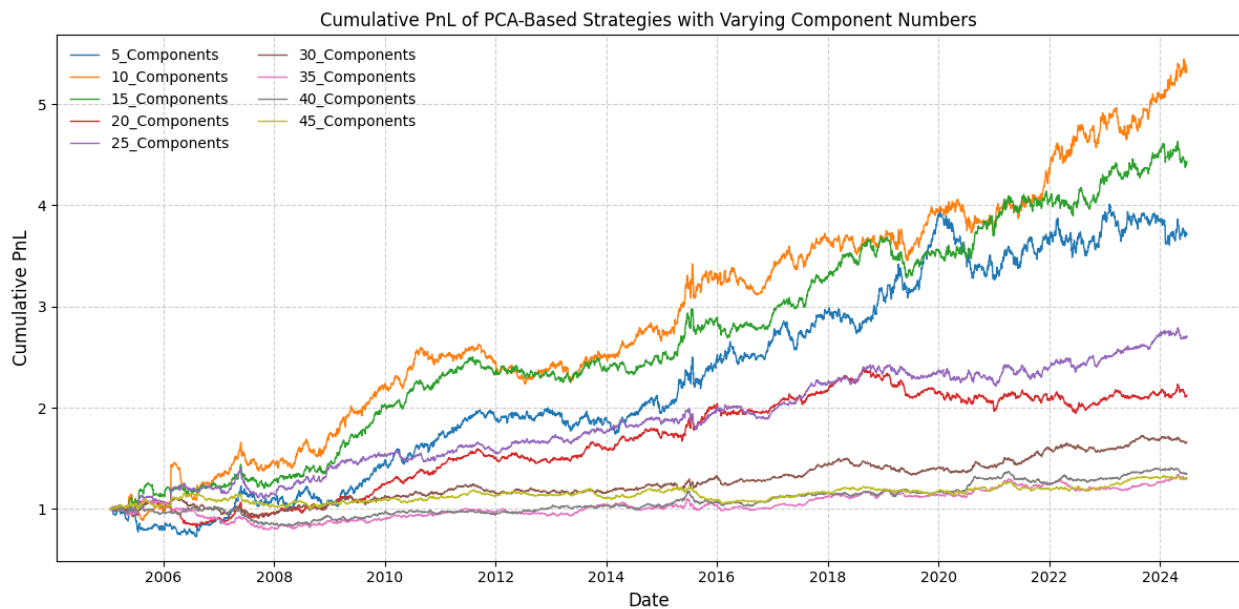
Note: This figure shows the cumulative performance of PCA-based strategies using 5 to 45 components. Strategies with 10–15 components achieve the highest returns, while those with more components exhibit weaker performance, suggesting diminishing marginal benefits from adding factors.

The Sharpe ratios and t-statistics display similar non-monotonic patterns. The Sharpe ratio peaks at 0.94 when using 15 components, while the corresponding t-statistics are highest around 10 to 15 components (2.93 and 4.08, respectively). These results indicate that mid-range dimensionality delivers the most favorable balance between return and statistical confidence.

Beyond this range, the statistical significance of returns weakens, with t-values approaching insignificance once more than 35 components are included.

From a risk perspective, volatility—as measured by standard deviation—declines as additional components are added, reaching its lowest value (0.0576) with 45 components. Tail risk, reflected in VaR and CVaR, also improves under higher-dimensional specifications. However, these risk reductions come at the cost of markedly lower returns and Sharpe ratios, implying that the decline in profitability outweighs the benefits of reduced volatility. Consequently, the overall risk-adjusted efficiency deteriorates, limiting the practical usefulness of excessively high-dimensional PCA models.

Figure 16. Cumulative Explained Variance of PCA Factors vs. CSI300 Volatility (2005–2024)
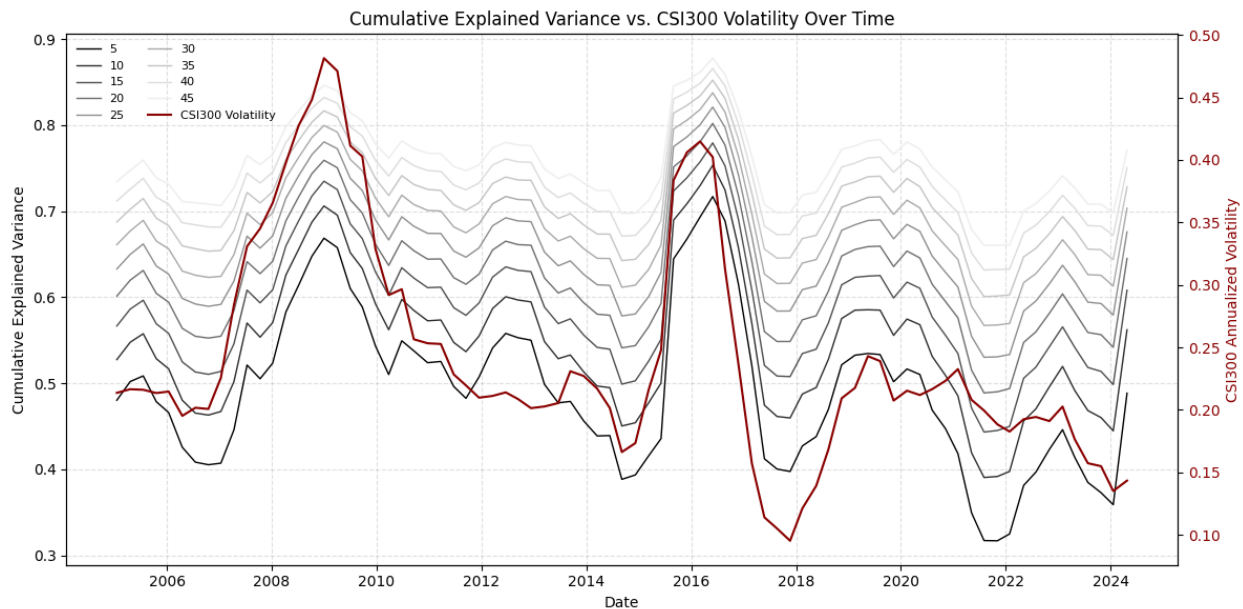
Note: This figure compares the cumulative explained variance of PCA models using different numbers of components (ranging from 5 to 45) with the rolling annualized volatility of the CSI300 index. The left y-axis represents the proportion of total variance explained by the first n principal components, where darker lines correspond to smaller component counts. The right y-axis shows the 252-day rolling volatility of the CSI300 as a proxy for market stress. The visualization reveals that during periods of heightened volatility—such as the 2008 financial crisis, the 2015 market crash, and the COVID-19 outbreak—more components are required to achieve the same level of explanatory power, indicating increased noise and dispersion in the underlying return structure.

Drawdown-related metrics provide additional evidence of the return–risk trade-off associated with PCA dimensionality. While the maximum drawdown decreases as more components are included, the accompanying decline in returns causes the Calmar ratio to

deteriorate, falling to just 0.12 at 45 components. By contrast, the 15-component strategy achieves a much more favorable profile, with a Calmar ratio of 0.51, reflecting a superior balance between return and drawdown risk. These findings are consistent with Avellaneda and Lee (2010), who argue that PCA-based mean-reversion strategies perform best when anchored to a small set of dominant components rather than the full eigenspectrum.

From an economic perspective, this pattern highlights the dangers of over-expanding the factor set. The earliest principal components capture systematic co-movements linked to market, sector, and style effects—precisely the structures that residual-based trading strategies are designed to exploit. Adding too many components, however, leads the model to explain idiosyncratic variation or even spurious noise, thereby weakening the quality of residuals and diminishing mean-reversion opportunities. Similar findings of diminishing benefits from higher-order principal components are reported in Connor and Korajczyk (1988) for asset pricing factors, further supporting the view that only the dominant components contain economically meaningful information.

The results carry several practical implications. First, they demonstrate that retaining approximately 10 to 15 components provides the most effective trade-off between explanatory power and signal degradation. Second, the optimal dimensionality is likely regime-dependent, implying that adaptive or dynamic PCA approaches could enhance robustness across heterogeneous market environments. Third, strategies with fewer retained components are computationally more efficient, which is especially advantageous in high-frequency or large-scale portfolio applications.

In sum, PCA dimensionality emerges as a key determinant of statistical arbitrage performance. Too few components risk underfitting by failing to capture dominant structures, while too many introduce noise that erodes statistical and economic significance. A moderate number of components—typically between 10 and 15—achieves the most favorable balance between risk and return, offering both theoretical justification and practical guidance for model design.

Table 11. Performance Metrics under Different Numbers of PCA Components.

Components	5	10	15	20	25	30	35	40	45
Mean	0.0791	0.0992	0.0833	0.0432	0.0562	0.0291	0.0160	0.0176	0.0160
t-stat	2.5586	2.9289	4.0844	2.3583	3.0457	1.9289	1.0782	1.2977	1.1992
Std. Dev.	0.1338	0.1466	0.0883	0.0793	0.0799	0.0654	0.0640	0.0588	0.0576
Sharpe	0.5911	0.6766	0.9436	0.5448	0.7036	0.4456	0.2491	0.2998	0.2770
z-stat	2.5586	2.9289	4.0844	2.3583	3.0457	1.9289	1.0782	1.2977	1.1992
VaR (95%)	-0.0125	-0.0095	-0.0083	-0.0074	-0.0068	-0.0063	-0.0062	-0.0058	-0.0055
CVaR	-0.0190	-0.0146	-0.0120	-0.0114	-0.0099	-0.0093	-0.0089	-0.0083	-0.0078
Omega	1.1146	1.1769	1.1806	1.0997	1.1446	1.0793	1.0446	1.0525	1.0493
Sortino	0.7606	1.2497	1.3742	0.7779	1.1176	0.6412	0.3375	0.4241	0.4303
Max D.	-0.3199	-0.2561	-0.1647	-0.2379	-0.2131	-0.1696	-0.2716	-0.2074	-0.1377
Calmar	0.2472	0.3873	0.5061	0.1815	0.2638	0.1717	0.0587	0.0849	0.1159
Sterling	0.1883	0.2785	0.3149	0.1278	0.1795	0.1080	0.0429	0.0573	0.0671
Burke	0.8407	1.4814	1.7935	0.5058	1.1305	0.5088	0.1412	0.2292	0.2746
JB test	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: This table reports the full-sample annualized performance statistics for PCA-based strategies using different numbers of components (5 to 45). All metrics are annualized.

7.2 Explained Variance Thresholds and Strategy Performance

Table 12 presents the annualized performance metrics for PCA-based mean-reversion strategies under varying explained variance thresholds, ranging from 40% to 75%. These thresholds determine how many principal components are retained in constructing the factor model, shaping the trade-off between parsimony and completeness. As illustrated in Figure 16, 17 and 18—reinforcing the conclusion observed in Figure 12—the relationship between explained variance and strategy performance follows a broadly declining trend with local reversals, indicating a nonlinear pattern. This carries important implications for model robustness, the quality of residual-based signals, and overall portfolio profitability.

The results indicate that mean returns exhibit an overall declining trend as the explained variance threshold increases, though with minor reversals at intermediate levels (e.g., 60% and 70%). The strategy achieves its highest returns at 40% (11.50%) and 45% (10.96%), but performance drops sharply at higher thresholds, reaching just 2.56% at 75%. This suggests that the early principal components capture the most economically meaningful co-movements—typically driven by broad market and sectoral dynamics—whereas higher-order components may reflect transient, idiosyncratic noise that weakens the residual-based trading signal. This pattern

is consistent with findings from Laloux et al. (2000) and Plerou et al. (2000), who warn that smaller eigenvalues often represent noise rather than informative structure.

Risk-adjusted performance, as measured by the Sharpe and Sortino ratios, exhibits a convex shape. Both metrics peak around the 70% threshold (Sharpe = 0.6163, Sortino = 0.8938), suggesting that this level of retained variance provides the most favorable balance between return and risk. The corresponding t-statistic of 2.6678 further confirms the statistical significance of returns at this level.

Although mean returns are higher at lower thresholds (40–50%), these configurations are associated with greater volatility and elevated tail risk, which ultimately undermines their overall efficiency. These results support the interpretation that moderate thresholds—typically between 60% and 70%—offer the optimal balance, allowing the model to capture systematic co-movements without overfitting to noise. This configuration generates more stable and reliable residuals that serve as effective signals for mean-reversion trading.

As the explained variance threshold increases, standard deviation declines sharply—from 0.3147 at 40% to 0.0609 at 75%—and tail risk metrics such as VaR and CVaR also improve. For instance, CVaR narrows from –0.0279 to –0.0098 across this range. However, these improvements in downside risk are offset by a collapse in return potential and significantly lower Calmar and Sterling ratios. This pattern suggests that strategies emphasizing risk minimization without sufficient return generation become economically inefficient, limiting their practical applicability.

Maximum drawdowns also decline as more components are retained—falling from –59.8% at the 40% threshold to –15.9% at 70%; at 75% they rise again to –23.8%. Importantly, the highest Calmar ratio (0.2718) coincides with the 70% threshold, which is also the point of smallest drawdown, reinforcing that 70% strikes an effective balance between return generation and drawdown control.

Taken together, these results suggest that retaining components to explain about 70% of total variance yields the most robust configuration: it maximizes the trade-off among alpha generation, statistical significance, and downside protection. Lower thresholds can boost average returns but at the cost of excessive volatility and instability, whereas higher thresholds stabilize performance but dilute the signal by incorporating weaker, noise-driven factors.

In practice, an explained-variance threshold in the 60%–70% range provides a reasonable default for PCA-based arbitrage strategies, subject to refinement based on regime shifts, asset-class specifics, or out-of-sample validation.

Table 12. Performance of PCA Strategies with Different Explained Variance Levels.

Explained	40%	45%	50%	55%	60%	65%	70%	75%
Mean	0.1150	0.1096	0.0868	0.0632	0.0652	0.0349	0.0433	0.0256
t-stat	1.5810	1.5131	2.0257	2.1575	2.5446	1.8770	2.6678	1.8217
Std. Dev.	0.3147	0.3136	0.1854	0.1268	0.1109	0.0805	0.0702	0.0609
Sharpe	0.3652	0.3495	0.4680	0.4984	0.5878	0.4336	0.6163	0.4208
z-stat	1.5810	1.5131	2.0257	2.1575	2.5446	1.8770	2.6678	1.8217
VaR (95%)	-0.0159	-0.0149	-0.0128	-0.0109	-0.0087	-0.0075	-0.0065	-0.0058
CVaR (95%)	-0.0279	-0.0260	-0.0201	-0.0173	-0.0141	-0.0114	-0.0098	-0.0085
Omega	1.1242	1.1298	1.1228	1.1048	1.1304	1.0805	1.1146	1.0771
Sortino	0.6600	0.4851	0.7936	0.7216	0.9239	0.6229	0.8938	0.6160
Max D.	-0.5977	-0.6917	-0.3470	-0.2368	-0.2461	-0.2334	-0.1591	-0.2379
Calmar	0.1923	0.1585	0.2500	0.2667	0.2649	0.1495	0.2718	0.1078
Sterling	0.1648	0.1385	0.1941	0.1875	0.1884	0.1047	0.1669	0.0759
Burke	0.8058	0.6579	0.6858	0.7625	0.9126	0.4594	0.7552	0.2997
JB test	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: This table reports the performance of PCA-based strategies using varying explained variance thresholds (40%–75%). While higher thresholds generally reduce volatility and tail risk, the best risk-adjusted returns (e.g., Sharpe and Sortino ratios) are observed around the 70% level, suggesting an optimal balance between factor completeness and noise control.

Figure 17. Cumulative PNL of PCA Strategies by Explained Variance Threshold
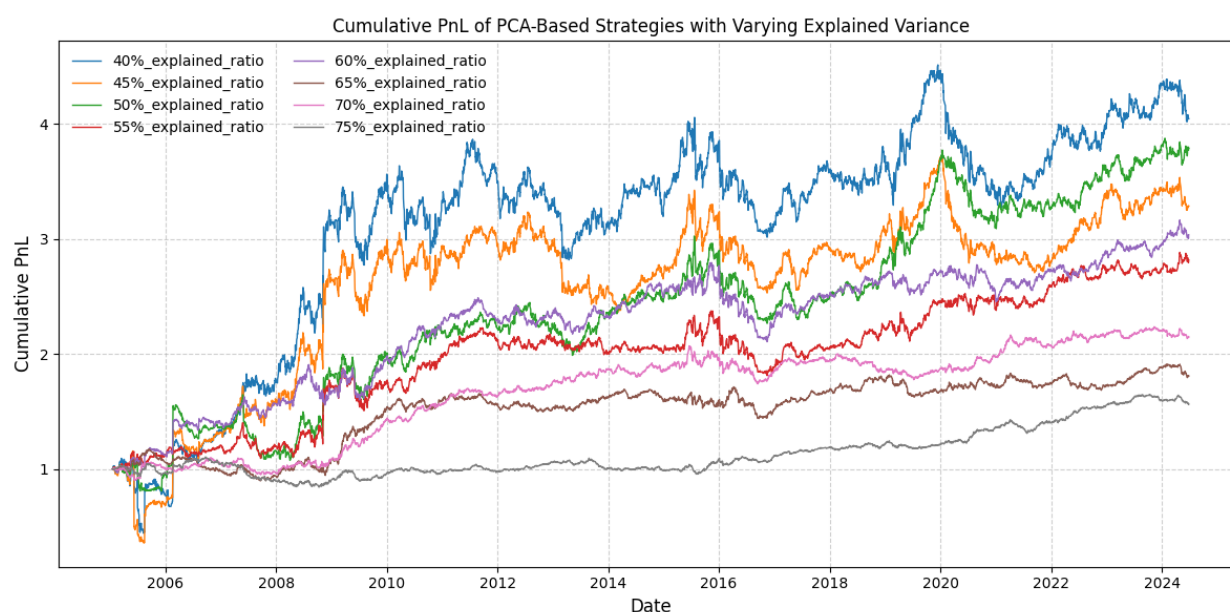
Note: This chart shows strategy performance under different PCA explained variance levels (40%–75%). Lower thresholds yield higher returns with greater volatility, while higher thresholds reduce risk but dampen returns. The 70% level offers the best overall balance.
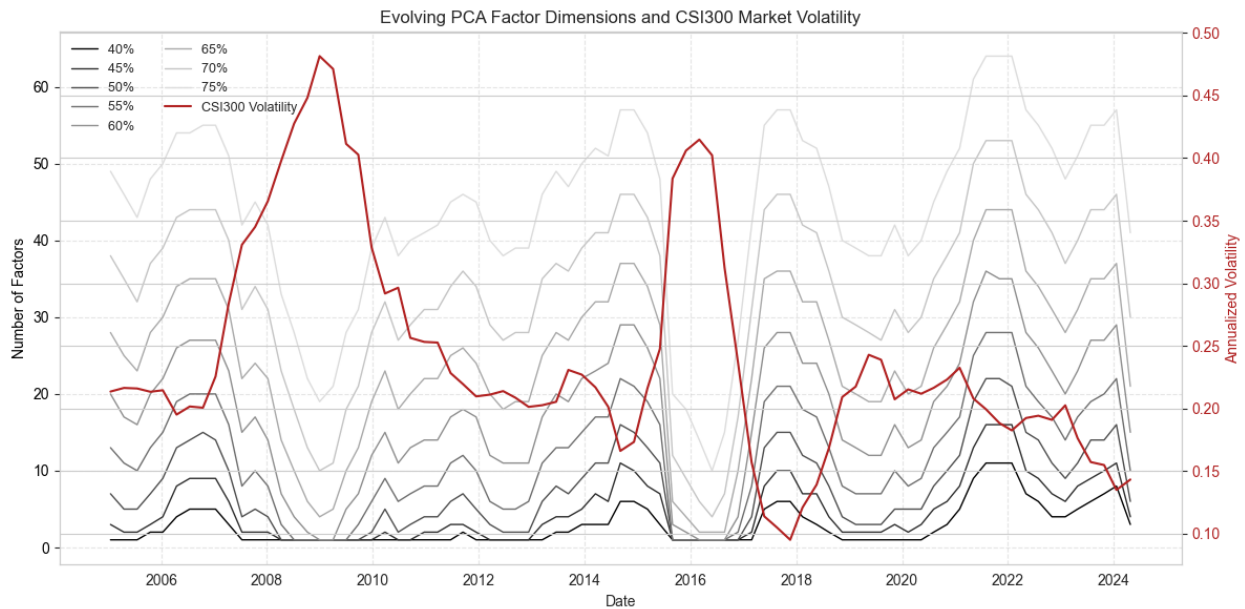
Figure 18. Temporal evolution of PCA factor dimensionality and CSI300 market volatility

Figure 18. Temporal evolution of PCA factor dimensionality across different explained variance thresholds (40%–75%) and its relationship with CSI300 index volatility. The number of principal components required to reach a given explained variance level varies significantly over time, reflecting shifts in market structure and cross-sectional correlation. The CSI300 annualized volatility (right axis, red line) tends to increase during market stress periods and coincides with a rise in the number of required PCA factors, indicating increased noise and reduced factor dominance during turbulent regimes.

7.3 Impact of Entry and Exit Thresholds on Strategy Performance

Table 13 reports the annualized performance metrics of the PCA-based mean-reversion strategy under different combinations of entry and exit thresholds, expressed as multiples of the residual standard deviation. Figure 19 plots the corresponding cumulative PnL. These thresholds govern the conditions for initiating and closing positions based on residual deviations from the estimated equilibrium level, thereby directly influencing trade frequency, average holding period, and the overall risk–return characteristics of the strategy.

Lower entry thresholds (e.g., 1.0) generate more frequent trading signals by reacting to relatively small deviations. While this increases exposure to potential arbitrage opportunities, it also raises susceptibility to noise and false signals, particularly during volatile market conditions. In contrast, higher entry thresholds (e.g., 1.5) act as stricter filters, initiating trades only when deviations are substantial. Although this reduces sensitivity to noise, the excessive stringency

results in fewer executed trades and missed opportunities, ultimately weakening overall performance.

Exit thresholds exert an equally important influence. Tighter exits (e.g., 0.3 or 0.5) trigger earlier closures, securing profits once partial mean-reversion occurs. While this approach helps reduce drawdowns and downside exposure, it also limits the ability to capture the full extent of price convergence. Looser exits (e.g., 0.75) allow positions to remain open longer, capturing a larger portion of the reversion at the expense of higher exposure to residual fluctuations.

Among all configurations, the (1.0, 0.75) setting delivers the strongest performance, achieving the highest Sharpe ratio (1.0835) and the largest t-statistic (4.6901). This indicates that moderate entry thresholds combined with relatively loose exits strike the most favorable balance between trade frequency, signal reliability, and return capture. The configuration denoted as (Main)—(1.25, 0.5, 0.75)—also performs robustly, with a Sharpe ratio of 0.9436 and relatively low drawdown (–0.1107), producing a high Calmar ratio of 0.8021. Notably, this specification incorporates a two-stage exit mechanism, in which partial profit-taking occurs at 0.5 and final closure at 0.75, thereby mitigating downside risk while preserving the ability to benefit from further convergence.

By contrast, configurations such as (1.5, 0.3) or (1.5, 0.5) produce markedly lower Sharpe ratios (0.6650 and 0.6395, respectively) and deeper drawdowns. These outcomes reflect a trade initiation barrier that is too high relative to a narrow exit window, leading to infrequent and low-quality trades while missing meaningful reversion opportunities.

Overall, these results highlight the critical importance of threshold calibration. Entry thresholds that are too low can result in overtrading and excess sensitivity to noise, whereas thresholds that are too high filter out profitable trades and weaken performance. Likewise, exit thresholds that are asymmetrically lower than entry thresholds—as in the main configuration—help lock in profits before full mean-reversion, improving both downside protection and capital efficiency.

Future extensions could explore adaptive threshold schemes that adjust dynamically to market conditions, such as volatility or liquidity regimes. Taken together, the evidence suggests that combining moderate entry thresholds (1.0–1.25) with relatively loose exits (around 0.75), or employing multi-stage exit rules, yields the most robust and economically viable strategy design

Table 13. Performance Metrics under Different Entry and Exit Thresholds.

Thresholds	1.0_0.3	1.0_0.5	1.0_0.75	1.25_0.3	1.25_0.5	1.25_0.5_0.75 (Main)	1.25_0.75	1.5_0.3	1.5_0.5	1.5_0.75
Mean	0.1112	0.1109	0.1154	0.0786	0.0809	0.0833	0.0888	0.0787	0.0763	0.0859
t-stat	4.4973	4.6213	4.6901	3.8827	3.9714	4.0844	4.3435	2.8785	2.7683	3.1139
Std. Dev.	0.1070	0.1039	0.1065	0.0876	0.0881	0.0883	0.0885	0.1184	0.1193	0.1194
Sharpe	1.0389	1.0676	1.0835	0.8970	0.9175	0.9436	1.0034	0.6650	0.6395	0.7194
z-stat	4.4973	4.6213	4.6901	3.8827	3.9714	4.0844	4.3435	2.8785	2.7683	3.1139
VaR (95%)	-0.0078	-0.0077	-0.0077	-0.0081	-0.0083	-0.0083	-0.0084	-0.0090	-0.0090	-0.0089
CVaR (95%)	-0.0110	-0.0111	-0.0110	-0.0119	-0.0120	-0.0120	-0.0120	-0.0130	-0.0131	-0.0130
Omega	1.2572	1.2549	1.2661	1.1713	1.1743	1.1806	1.1921	1.1529	1.1467	1.1663
Sortino	2.0520	2.0421	2.1528	1.3074	1.3373	1.3742	1.4882	1.1943	1.1530	1.3201
Max D.	-0.1756	-0.1761	-0.1739	-0.1651	-0.1642	-0.1647	-0.1107	-0.2774	-0.2209	-0.1590
Calmar	0.6332	0.6300	0.6636	0.4760	0.4926	0.5061	0.8021	0.2837	0.3455	0.5401
Sterling	0.4034	0.4018	0.4214	0.2965	0.3061	0.3149	0.4214	0.2085	0.2378	0.3316
Burke	2.7983	2.7924	3.1143	1.6700	1.6991	1.7935	2.2051	1.0351	1.1931	1.6964
JB test	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: This table reports the annualized performance statistics of the mean-reversion strategy under various combinations of entry and exit thresholds. Each column represents a threshold pair in the format (entry, exit), where entry refers to the standard deviation trigger for opening a position, and exit denotes the closing threshold. The configuration labeled (Main) corresponds to the primary model employed in this study.

Figure 19. Cumulative PNL under different entry/exit thresholds
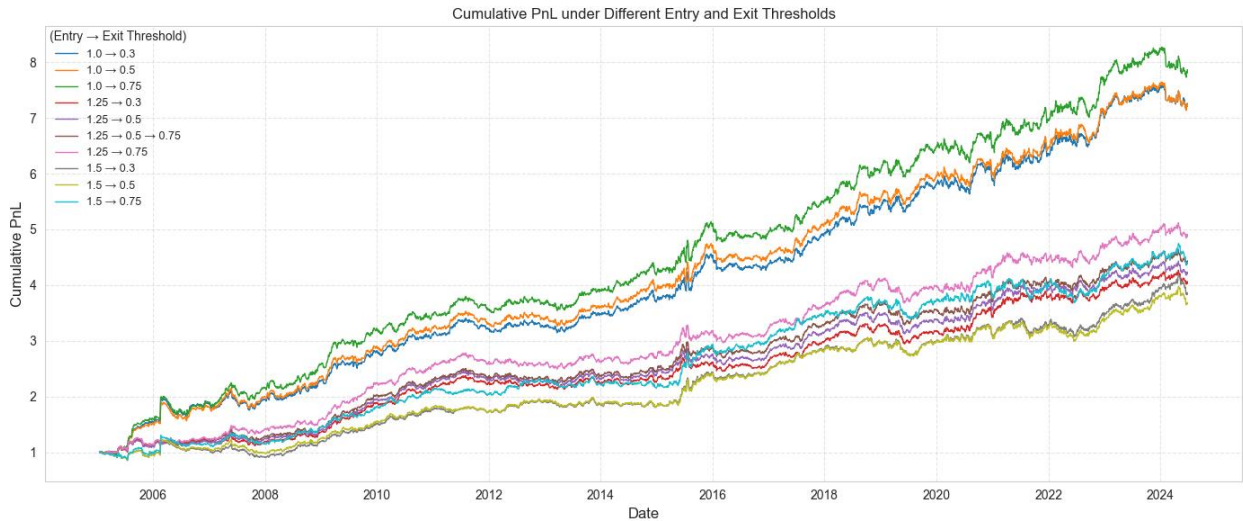
Note: This figure compares the cumulative PNL of PCA-based mean-reversion strategies under different combinations of entry and exit thresholds, defined as standard deviation multiples of residual signals. Lower entry thresholds (e.g., 1.0) generate more frequent trades, while looser exit thresholds (e.g., 0.75) allow positions to capture deeper mean reversion. The configuration (1.0→0.75) achieves the best overall performance, while the benchmark strategy (1.25→0.5→0.75) strikes a balance between return and drawdown.

7.4 Impact of Rolling Window Length on Strategy Performance

Table 14 summarizes the performance of the PCA-based mean-reversion strategy across a range of rolling estimation window lengths, from 30 to 120 trading days. Figure 20 displays the corresponding cumulative PnL. The reported annualized metrics—including mean return, t-statistic, and Sharpe ratio—reveal a clear pattern: shorter windows (30–50 days) tend to exhibit higher volatility and weaker statistical significance, while longer windows (100–120 days) consistently deliver improved risk-adjusted returns and more statistically significant alpha.

Short estimation windows (e.g., 30–50 days) respond more quickly to recent market changes and shifting factor structures. This sensitivity allows the model to adapt to transient dynamics but also introduces substantial estimation noise. For example, the 30-day window produces a very low mean return (0.0083) and an insignificant t-statistic (0.43), suggesting that such short samples fail to extract reliable principal components or generate stable residuals. Statistically, short windows are equivalent to small-sample estimation, which is more susceptible to the influence of extreme returns or temporary structural breaks. From a financial perspective, this translates into unstable factor structures and residual signals that lack robustness.

In contrast, medium to longer windows (60–120 days) improve statistical reliability by smoothing out short-term fluctuations and reducing overfitting. A longer historical lookback increases the effective sample size, yielding more stable PCA loadings and residual estimates. This stability enhances the precision of the signals and strengthens alpha consistency. For instance, the 120-day configuration achieves the highest Sharpe ratio (1.44) and t-statistic (6.22), underscoring that longer windows provide both statistical and economic benefits.

Longer windows yield more stable PCA loadings by averaging across a larger information set, thereby reducing the influence of outliers and transient noise. This stability in factor structure translates into more reliable residual estimates, which are essential for generating effective mean-reversion signals.

From a statistical perspective, extending the window is equivalent to enlarging the sample size, lowering the variance of eigenvalue and eigenvector estimates and mitigating the emergence of spurious components driven by random fluctuations. From a financial perspective, shorter windows resemble small-sample estimation and are therefore more vulnerable to extreme returns or temporary shocks, producing unstable factor structures. By contrast, longer windows

filter out such idiosyncratic events and emphasize persistent co-movements that better reflect underlying market and sector dynamics. Consistent with prior studies, reducing estimation noise in factor models has been shown to enhance portfolio performance.

However, excessively long windows may render the model unresponsive during regime shifts—such as those triggered by macroeconomic shocks, policy changes, or sector rotations—delaying signal adjustment and weakening alpha capture. Hence, an inherent trade-off exists between statistical robustness and adaptability.

The 60-day window, adopted as the baseline in this study, represents a practical compromise between responsiveness and estimation stability. Its prevalence in both academic research and industry practice is well documented, with Avellaneda and Lee (2010) and Gu et al. (2021) among the studies supporting its use. Nevertheless, dynamic window adjustment—conditional on market volatility or signal stability—may further enhance performance. For instance, longer windows can be deployed in tranquil markets to exploit stable factor structures, while shorter windows may help capture shifts during turbulent regimes.

Table 14 underscores that rolling window length is a critical design choice in PCA-based statistical arbitrage. Shorter windows adapt quickly to evolving dynamics but suffer from instability and weak statistical reliability, whereas longer windows (80–120 days) generate stronger Sharpe ratios, lower drawdowns, and more consistent signal extraction. In practice, a window length of 60–100 trading days appears to strike the most effective balance between adaptability and robustness, particularly in environments with moderate structural variation. This trade-off aligns with prior findings in the factor modeling literature, which emphasize that reducing estimation noise enhances portfolio robustness without entirely sacrificing responsiveness.

Table 14. The performance of the PCA-based mean-reversion strategy under varying rolling window lengths (30–120 trading days).

Windows	30_days	40_days	50_days	60_days	70_days	80_days	90_days	100_days	110_days	120_days
Mean	0.0083	0.0495	0.0967	0.0833	0.0911	0.1063	0.0680	0.1041	0.0972	0.1435
t-stat	0.4316	2.1621	3.7179	4.0844	3.7864	4.9797	2.9260	4.4991	4.3176	6.2196
Std. Dev.	0.0828	0.0990	0.1126	0.0883	0.1042	0.0924	0.1006	0.1001	0.0975	0.0999
Sharpe	0.0997	0.4995	0.8589	0.9436	0.8747	1.1504	0.6759	1.0394	0.9974	1.4368
z-stat	0.4316	2.1621	3.7179	4.0844	3.7864	4.9797	2.9260	4.4991	4.3176	6.2196
VaR (95%)	-0.0075	-0.0075	-0.0080	-0.0083	-0.0085	-0.0085	-0.0091	-0.0088	-0.0090	-0.0089

Windows	30_days	40_days	50_days	60_days	70_days	80_days	90_days	100_days	110_days	120_days
CVaR (95%)	-0.0106	-0.0110	-0.0115	-0.0120	-0.0124	-0.0125	-0.0141	-0.0127	-0.0131	-0.0126
Omega	1.0190	1.1104	1.2167	1.1806	1.1894	1.2222	1.1312	1.2080	1.1893	1.2850
Sortino	0.1616	0.8888	1.7136	1.3742	1.4388	1.6166	0.8196	1.6641	1.4634	2.3224
Max D.	-0.3830	-0.1969	-0.1306	-0.1647	-0.1795	-0.2331	-0.3606	-0.1658	-0.1687	-0.1226
Calmar	0.0215	0.2512	0.7406	0.5061	0.5076	0.4562	0.1885	0.6279	0.5761	1.1707
Sterling	0.0171	0.1666	0.4195	0.3149	0.3260	0.3192	0.1476	0.3916	0.3617	0.6447
Burke	0.0388	0.6028	2.6725	1.7935	1.5578	1.9587	0.6093	2.4399	1.7169	3.8388
JB test	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: This table reports the performance of the PCA-based mean-reversion strategy under varying rolling window lengths (30–120 trading days). As the estimation window expands, both the Sharpe ratio and statistical significance (t-stat) generally improve, with the best results observed at 120 days (Sharpe = 1.44, t-stat = 6.22). Longer windows provide more stable factor estimates, enhancing signal reliability and reducing overfitting. However, windows that are too short (e.g., 30 days) yield weak performance and high drawdowns, reflecting unstable factor structures and noisy residuals. Overall, a window length between 80 and 120 days appears to offer the best risk-adjusted trade-off.

Figure 20. Cumulative PNL under various rolling estimation windows
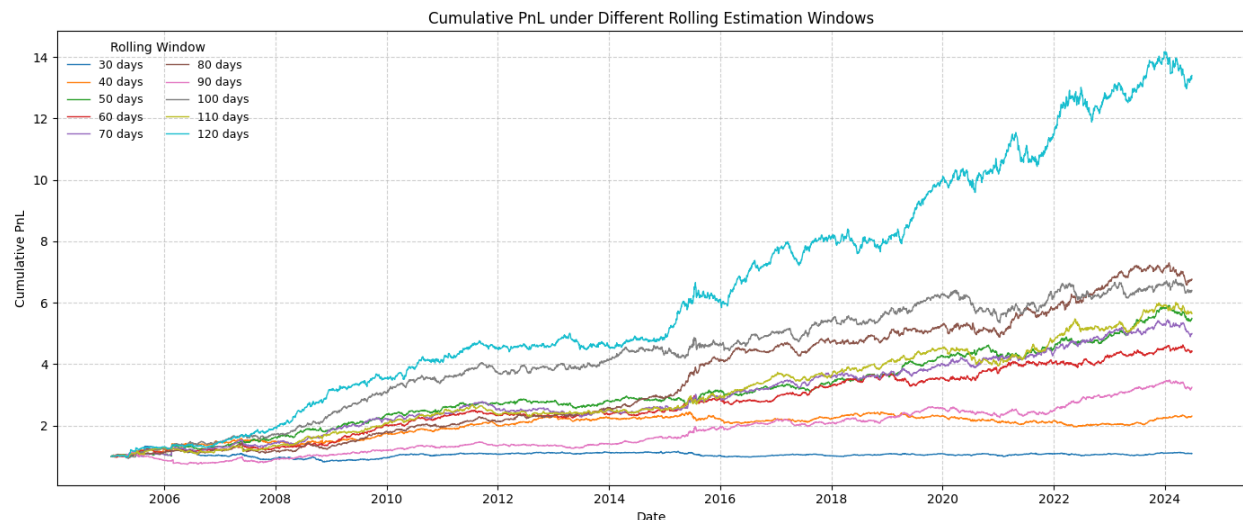

Figure 20. This figure illustrates the cumulative PNL of a PCA-based mean-reversion strategy under various rolling estimation windows, ranging from 30 to 120 trading days. Longer windows (e.g., 100–120 days) tend to yield smoother and more profitable trajectories, suggesting that more stable factor estimates enhance signal quality. In contrast, very short windows (e.g., 30–40 days) appear to introduce more noise, leading to lower overall performance and noisier cumulative return paths. These findings highlight the trade-off between adaptability and statistical robustness when selecting the window length for PCA-based signal generation.

8. Conclusions

This thesis has undertaken a comprehensive investigation of pairs trading and broader statistical arbitrage strategies grounded in mean-reversion principles. A central element of the

analysis is the decomposition of stock returns into systematic and idiosyncratic components, with the latter forming the basis of residual-driven, market-neutral trades. Two factor-construction methodologies were examined: (i) PCA, which extracts latent "eigenportfolios" from the return correlation matrix; and (ii) ETF-based factor modeling, which regresses stock returns on sector ETF benchmarks.

The empirical evidence shows that the choice of factor specification materially influences the behavior of residuals and, in turn, the profitability of mean-reversion signals. PCA-derived residuals better isolate latent co-movements and are less biased toward large-cap exposures, whereas ETF-based models provide intuitive sectoral interpretations and facilitate direct hedging with tradable instruments. Both approaches generate robust signals, though the optimal number of retained PCA components or choice of ETFs is time-varying. Importantly, the adoption of a dynamic PCA rule—retaining components sufficient to explain a fixed proportion of variance—enhances adaptability to evolving cross-sectional structures and improves risk-adjusted returns.

Trading signals are systematically generated when residuals deviate from a rolling 60-day equilibrium band, defined in terms of standardized residuals (z-scores). Predefined entry and exit thresholds capture temporary mispricings while reducing overfitting risk. Transaction costs are incorporated through a realistic slippage model, highlighting the role of trading frictions. Further, weighting signals by trading volume improves robustness in low-liquidity or high-volatility markets.

Sensitivity analyses confirm that strategy outcomes are shaped by methodological choices. While the 60-day window and $\pm 1.25\sigma$ thresholds are effective baseline settings, alternative calibrations may be warranted in markets with different volatility or liquidity regimes. Stress tests during extreme episodes demonstrate that mean-reversion strategies may suffer drawdowns but recover once residual mispricings re-emerge, particularly when the factor model parsimoniously captures systematic variation.

Overall, this research highlights the value of rigorous factor decomposition—whether PCA- or ETF-based—in constructing market-neutral strategies. It shows that residual-based signals can consistently generate alpha across regimes, provided factor structures are robust, dynamically updated, and risk controls enforced. Future research could integrate machine

learning for dynamic factor selection, incorporate macroeconomic or fundamental variables to enrich valuation signals, or explore adaptive thresholding linked to volatility and liquidity states. Ultimately, the central insight is that robust identification of systematic structure is essential to isolating idiosyncratic mispricings and sustaining profitability in mean-reversion–based statistical arbitrage.

References

Avellaneda, M., Lee, J.H., 2010. Statistical arbitrage in the US equities market. Quantitative Finance, 10(7), pp. 761-782.

Brunnermeier, M.K., Pedersen, L.H., 2009. Market liquidity and funding liquidity. The review of financial studies, 22(6), pp. 2201-2238.

Caneo, F., Kristjanpoller, W., 2021. Improving statistical arbitrage investment strategy: Evidence from Latin American stock markets. International Journal of Finance & Economics, 26(3), pp. 4424-4440.

Chordia, T., Roll, R., Subrahmanyam, A., 2005. Evidence on the speed of convergence to market efficiency. Journal of financial economics, 76(2), pp. 271-292.

Connor, G., Korajczyk, R.A., 1988. Risk and return in an equilibrium APT: Application of a new test methodology. Journal of financial economics, 21(2), pp. 255-289.

Cont, R., Da Fonseca, J., 2002. Dynamics ofimplied volatility surfaces. Quantitative finance, 2(1), p. 45.

Davis, G., Mallat, S., Avellaneda, M., 1997. Adaptive greedy approximations. Constructive approximation, 13(1), pp. 57-98.

Gatev, E., Goetzmann, W.N., Rouwenhorst, K.G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. The review of financial studies, 19(3), pp. 797-827.

Gatta, F., Iorio, C., Chiaro, D., Giampaolo, F., Cuomo, S., 2023. Statistical arbitrage in the stock markets by the means of multiple time horizons clustering. Neural Computing and Applications, 35(16), pp. 11713-11731.

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics, 222(1), pp. 429-450.

Han, C., He, Z., Toh, A.J.W., 2023. Pairs trading via unsupervised learning. European Journal of Operational Research, 307(2), pp. 929-947.

Hasbrouck, J., 2009. Trading costs and returns for US equities: Estimating effective costs from daily data. The Journal of Finance, 64(3), pp. 1445-1477.

Jolliffe, I., 2011. Principal component analysis, in: International encyclopedia of statistical science. Springer, Berlin, Heidelberg, pp. 1094-1096.

Kelly, B.T., Pruitt, S., Su, Y., 2020. Instrumented principal component analysis. Available at SSRN 2983919.

Khandani, A., Lo, A.W., 2009. What Happened To The Quants In August 2007? SSRN.

Krause, F., Calliess, J.P., 2024. End-to-End Policy Learning of a Statistical Arbitrage Autoencoder Architecture. arXiv preprint arXiv:2402.08233.

Laloux, L., Cizeau, P., Potters, M., Bouchaud, J.P., 2000. Random matrix theory and financial correlations. International Journal of Theoretical and Applied Finance, 3(03), pp. 391-397.

Lee, C.M., Swaminathan, B., 2000. Price momentum and trading volume. the Journal of Finance, 55(5), pp. 2017-2069.

Lehmann, B.N., 1990. Fads, martingales, and market efficiency. The Quarterly Journal of Economics, 105(1), pp. 1-28.

Litterman, R., 1991. Common factors affecting bond returns. Journal of fixed income, pp. 54-61.

Lo, A.W., MacKinlay, A.C., 1990. When are contrarian profits due to stock market overreaction? The review of financial studies, 3(2), pp. 175-205.

Lo, A.W., Wang, J., 2000. Trading volume: definitions, data analysis, and implications of portfolio theory. The Review of Financial Studies, 13(2), pp. 257-300.

Pástor, Ľ., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political economy, 111(3), pp. 642-685.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.N., Stanley, H.E., 2000. A random matrix theory approach to financial cross-correlations. Physica A: Statistical Mechanics and its Applications, 287(3-4), pp. 374-382.

Pole, A., 2011. Statistical arbitrage: algorithmic trading insights and techniques. John Wiley & Sons.

Poterba, J.M., Summers, L.H., 1988. Mean reversion in stock prices: Evidence and implications. Journal of financial economics, 22(1), pp. 27-59.

Rotondi, F., Russo, F., 2024. Machine Learning for Pairs Trading: a Clustering-based Approach. Available at SSRN 5080998.

Sarmento, S.M., Horta, N., 2020. Enhancing a pairs trading strategy with the application of machine learning. Expert Systems with Applications, 158, p. 113490.

Scherer, K.P., Avellaneda, M., 2002. All for one… one for all? A principal component analysis of Latin American brady bond debt from 1994 to 2000. International Journal of Theoretical and Applied Finance, 5(01), pp. 79-106.

Xiang, Y., He, J., 2022. Pairs trading and asset pricing. Pacific-Basin Finance Journal, 72, p. 101713.

Annexes

Appendix A: Estimation of the residual process

This appendix outlines the procedure for modeling regression residuals as OU processes and for deriving standardized s-scores. While the method described here is not the most advanced or efficient, it is transparent, straightforward to implement, and provides a solid foundation that can be improved in practice.

A.1 OU Parameter Estimation for ETF Regressions

For clarity, we illustrate the procedure in the context of ETF-based factor regressions; the PCA-based case follows an analogous approach. For each stock S, we regress its daily returns on a corresponding sector ETF:

$$R_n^S = \gamma_0 + \gamma\, R_n^I + \epsilon_n, n = 1, 2, \dots, 60$$

where R_n^S and R_n^I are, respectively, the stock and ETF returns over 60 consecutive daily observations (chronologically ordered), and ϵ_n is the regression residual. Noting the continuous-time model in Equation (18), we define

$$\mu = \gamma_0 \times 252$$

interpreting γ_0 as the daily drift, then annualizing it with the factor 252 (the approximate number of trading days in a year).

Next, we construct the cumulative sum of residuals:

$$X_k = \sum_{j=1}^{k} \epsilon_j, k = 1, 2, \dots, 60$$

This sequence $\{X_k\}$ can be viewed as a discrete analog of the continuous OU process $X(t)$. Because γ and γ_0 are estimated over the same 60 data points, we observe $X_{60} = 0$, which is an artifact of the regression procedure forcing the mean of the in-sample residuals to zero.

We then fit the discrete 1-lag regression model:

$$X_{n+1} = a + b\, X_n + v_{n+1}, n = 1, 2, \dots, 59$$

where v_{n+1} is noise. According to the OU model (Equation (23) in the main text), the parameters (a, b) relate to the continuous-time parameters (m, θ, σ) as follows:

$$a = m(1 - e^{-\theta \Delta t}), b = e^{-\theta \Delta t}, Var(v) = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta \Delta t})$$

Here, $\Delta t = 1/252$ is the daily time step. From these relationships, we solve for the OU parameters:

$$\theta = -\frac{\ln(b)}{\Delta t} \approx -\ln(b) \times 252, m = \frac{a}{1 - b}, \sigma = \sqrt{\frac{Var(v)(2\theta)}{1 - b^2}}, \sigma_{eq} = \frac{\sigma}{\sqrt{2\theta}} = \sqrt{\frac{Var(v)}{1 - b^2}}$$

For practical purposes, we require relatively fast mean reversion over the 60-day window (i.e., θ large). A common cutoff is $\theta > 252/30$, meaning the characteristic reversion time is below 1.5 months. In that scenario, $b = e^{-\theta \Delta t}$ is safely below one. Conversely, if b is close to unity, the implied mean-reversion speed is too slow, and the model is discarded for that stock.

A.2 s-Score Computation

Given the OU process $X(t)$ with long-term mean mmm and equilibrium standard deviation σ_{eq}, the theoretical s-score is:

$$s = \frac{X(t) - m}{\sigma_{eq}}$$

However, due to the regression constraint $X_{60} = 0$ in the 60-day sample, we have $X(t) = 0$ at the end of the estimation window, implying:

$$s = \frac{-m}{\sigma_{eq}} = \frac{-a}{(1 - b)\sigma_{eq}}$$

To reduce systematic biases, it is often beneficial to re-center the mean by subtracting the cross-sectional average across all stocks in the universe:

$$m \leftarrow m - \langle m \rangle$$

where $\langle m \rangle$ is the mean across all stocks in the universe. Hence, we replace $a/(1 - b)$ with $(a/(1 - b)) - \langle a/(1 - b) \rangle$. The resulting s-score thus becomes:

$$s = \frac{-1}{\sigma_{eq}}(\frac{a}{1 - b} - \langle \frac{a}{1 - b} \rangle)$$

This empirical adjustment mitigates potential cross-sectional biases and typically yields more robust trading signals in backtests, even though it is not theoretically required by the OU framework.