Performance of Pairs Trading Strategies Based on Principal Component Analysis Methods

Sun Yufei

Department of Quantitative Finance, Faculty of Economics, University of Warsaw

Abstract

This thesis investigates market-neutral, mean-reversion-based statistical arbitrage strategies in the Chinese equity market, focusing on two distinct methods of factor decomposition: Principal Component Analysis (PCA) and sector-based Exchange-Traded Funds (ETFs). In both cases, we model the idiosyncratic (residual) returns of individual stocks as mean-reverting Ornstein–Uhlenbeck processes, thereby generating contrarian signals when deviations from equilibrium become statistically significant.

A rolling 60-day estimation window is employed to ensure that all parameter estimates and trading signals are computed in a strictly out-of-sample fashion. We incorporate realistic frictions by applying a 5-basis-point per-trade slippage, resulting in a 10-basis-point round-trip cost. Our empirical backtests, conducted from 2005 to 2024, compare four primary configurations: (i) synthetic ETFs constructed from capitalization-weighted sector indices, (ii) fixed PCA approaches (e.g., 15 eigenportfolios), (iii) dynamic PCA methods that vary the number of retained components based on explained variance, and (iv) additional variations that adjust signals using trading-time weighting to account for volume effects.

The results indicate that both PCA- and ETF-derived residual signals yield robust, risk-adjusted returns, with average Sharpe ratios approaching 0.90–0.95. PCA-based portfolios demonstrate stronger performance during periods of higher cross-sectional volatility, while ETF-based strategies remain more stable through structural shifts in market correlations. Incorporating trading volume further enhances profitability, especially for ETF-driven models. A sensitivity analysis reveals that threshold choices and rolling-window lengths substantially influence Sharpe ratios, drawdowns, and turnover. Overall, our findings underscore the importance of factor construction and signal design when deploying mean-reversion strategies in a market-neutral context, and highlight practical avenues—such as adaptive PCA and volume-weighted signals—for further improvement.

Keywords: Quantitative Trading, Pair Trading, Principal Component Analysis, Chinese Stock Market

Contents

1. Introduction

Statistical arbitrage refers to a broad spectrum of investment strategies characterized by several shared attributes. First, these strategies rely on systematic or rule-based trading signals, distinctively independent of fundamental analysis. Second, portfolios built upon statistical arbitrage principles typically maintain market neutrality, meaning their overall beta relative to market movements is essentially zero. Lastly, these strategies derive their excess returns from statistical methodologies, capitalizing on numerous individual positions, each possessing a positive expected return. This extensive diversification across multiple assets aims to yield stable, low-volatility performance that exhibits minimal correlation with broader market indices. Furthermore, the duration of trades within statistical arbitrage frameworks varies widely, encompassing ultra-short-term intervals of mere seconds up to longer holding periods spanning days, weeks, or potentially even longer horizons.

Pairs trading is commonly considered the foundational approach within statistical arbitrage. Typically, this strategy involves identifying two stocks—commonly within the same industry or sharing similar characteristics—that historically exhibit closely related price movements. Suppose stocks P and Q have price series P_t and Q_t, respectively; their relationship can be expressed mathematically as:

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

or in differential form:

$$\frac{dP_t}{P_t} = \mu d_t + \beta \frac{dQ_t}{Q_t} + d\varepsilon_t, \tag{2}$$

where the residual ε_t is assumed stationary or mean-reverting. Known as the cointegration residual, ε_t quantifies deviations from the equilibrium relationship between the stocks. Frequently, the drift μ is minimal compared to fluctuations in ε_t, allowing it to be omitted from practical trading models. Thus, a pairs trading strategy typically entails entering long and short positions depending on the magnitude and direction of these deviations. Specifically, one would go long on stock P and short β units of stock Q when the residual ε_t is small, and vice versa when the residual is large, capturing profits as prices converge back toward equilibrium. This mean-reverting behavior is often interpreted as a market correction of temporary pricing inefficiencies or overreactions, as highlighted in the literature (Lo and MacKinlay, 1990; Pole, 2007).

An alternative scenario involves situations where one stock is expected to consistently outperform the other over a significant period. In such cases, the cointegration residual would not exhibit stationarity. However, this thesis primarily focuses on mean-reversion scenarios and therefore does not explore these non-stationary cases in detail.

Beyond standard pairs trading, an extension of this strategy—often referred to as "generalized pairs trading"—involves comparing groups of stocks against benchmarks or indices rather than single stocks. For instance, within a specific sector such as biotechnology, each stock can be assessed against a relevant sector index or exchange-traded fund (ETF). By employing regression or cointegration analysis analogous to the pairs trading models described earlier, one can identify which stocks are undervalued or overvalued relative to the sector. Consequently, a portfolio constructed under this methodology consists of multiple stocks held in long positions if considered undervalued and short positions if perceived overvalued, each paired with an offsetting position in the ETF or sector factor. Due to the balancing nature of these positions, net exposure to the benchmark ETFs is usually minimal, leaving the resulting portfolio resembling a long-short equity strategy. The primary focus of this thesis is the detailed analysis, construction, and performance evaluation of such statistical arbitrage strategies, particularly emphasizing mean-reversion dynamics and avoiding scenarios where non-stationary residuals indicate persistent divergence in relative stock performance.

The analysis of residuals forms the foundation of our approach. Trading signals are generated based on relative value pricing within a sector or among a group of comparable firms, achieved by decomposing individual stock returns into systematic and idiosyncratic components. The statistical modeling effort is directed toward the idiosyncratic component, which captures deviations from broader market or sector trends. A general representation of this decomposition is given by:

$$\frac{dP_t}{P_t} = \mu d_t + \sum_{j=1}^{n} \beta_j F_t^{(j)} + d\varepsilon_t \tag{3}$$

In this equation, $F_t^{(j)}$, for $j = 1, \ldots, n$, denotes the returns of risk factors associated with the relevant market or sector, β_j represents the sensitivity of the stock to the $j - th$ factor, and $d\varepsilon_t$ captures the residual (idiosyncratic) component that cannot be explained by the factor model. An important practical consideration is how to determine such a decomposition. While a similar question arises in classical portfolio theory—typically framed in terms of selecting an optimal set of risk factors from a risk management perspective—the focus here is different. Our attention is directed toward the residuals that result after systematic influences have been removed. These residuals are the drivers of relative-value signals and ultimately the source of profit and loss (PNL) in statistical arbitrage. The key contribution of this research lies in examining how the choice of factor sets influences the resulting residuals, and consequently, how it affects the performance of the trading strategy. Different factor specifications lead to different residual dynamics, which in turn generate distinct trading opportunities and PNL profiles.

Recent research on mean-reversion and contrarian strategies has extended beyond early foundational work (e.g., Poterba and Summers, 1988; Lehmann, 1990; Lo and MacKinlay, 1990) to incorporate more refined models of return dynamics and factor structures. For instance, Khandani and Lo (2007) investigated market-neutral contrarian strategies in the context of the 2007 liquidity crisis, constructing dollar-neutral portfolios by sorting stocks into quantiles of past returns and trading “winners versus losers.” While their approach operates on fixed trading intervals, more recent studies have shifted focus toward strategies that are adaptive in frequency and based on residual returns rather than total returns. Chae and Kim (2020) and Huij and Lansdorp (2017) show that contrarian strategies based on residual (idiosyncratic) returns—i.e., returns orthogonalized with respect to common risk factors—exhibit stronger and more statistically significant mean-reversion than traditional approaches. These strategies effectively isolate firm-specific pricing inefficiencies and are better aligned with market-neutral objectives. Moreover, Kelly, Moskowitz, and Pruitt (2021) and Gu, Kelly, and Xiu (2021) propose advanced factor modeling techniques, including instrumented PCA and deep learning-based models, which further improve the explanatory power of systematic components and reduce residual noise. The present study builds on this literature by designing 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 selected set of risk factors, thereby targeting relative mispricings that are more robust to market-wide movements.

This paper is structured as follows. we provide a review of the relevant literature on mean-reversion strategies, market-neutral portfolio construction, and factor-based statistical arbitrage. We begin in section 1 by discussing 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 in section 2 that emphasize the role of risk factors in generating residual-based trading signals and the evolution from simple linear models to more sophisticated approaches such as Principal Component Analysis (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 market-neutral strategy design. This review highlights both the theoretical motivation and empirical findings that underpin the methodologies adopted in this paper.

In Section 3, we explore the concept of market-neutrality through two alternative methodologies. The first approach involves extracting latent risk factors using Principal Component Analysis (PCA), as introduced by Jolliffe (2002). The second approach constructs risk factors based on a set of industry-sector exchange-traded funds (ETFs), which serve as proxies for systematic market influences. Consistent with findings in 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 respective sectors. This observation suggests that PCA-based factor construction may be less biased toward large-cap stocks compared to ETF-based methods, which typically employ capitalization-weighted schemes. Furthermore, we find that the proportion of total variance explained by a fixed number of PCA components fluctuates considerably over time. This time variation indicates that the dimensionality required to effectively capture systematic return variation is itself dynamic—potentially influenced by macroeconomic conditions, changes in market structure, or shifts in investor risk appetite. Such dynamics may also help explain the observed performance differentials between the PCA-based and ETF-based strategies.

In Sections 4 and 5, we detail 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, which corresponds approximately to one earnings cycle. This estimation is always performed using data strictly preceding the trading date, thereby mimicking a real-time investment setting and avoiding look-ahead bias. Trading signals are generated when the estimated residual exhibits a statistically significant deviation from its recent mean. Specifically, a position is initiated when the residual deviates by more than 1.25 standard deviations from its equilibrium level, and closed once the deviation falls below 0.5 standard deviations. These thresholds are applied uniformly across all stocks to maintain consistency and prevent overfitting.

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

In Section 6, we conduct a series of backtests to evaluate the performance of various trading strategies, each utilizing a distinct approach to factor construction for residual estimation. Specifically, we consider four configurations: (i) synthetic ETFs constructed from capitalization-weighted indices, (ii) a fixed number of principal components derived via PCA, and (iii) A dynamic number of PCA factors chosen to exceed a specified explained-variance threshold. Across all strategies, the residuals used to generate trading signals are estimated in a strictly out-of-sample manner—only data from the trailing 60-day window prior to each trading day is used in the estimation process, thereby replicating the information set available to a real-time trader.

To maintain realism in execution modeling, we incorporate transaction costs by applying a slippage assumption of 5 basis points per trade, resulting in a round-trip trading cost of 10 basis points. This ensures that the performance metrics reflect the impact of frictions commonly encountered in actual trading environments.

Section 7 conducts a sensitivity analysis, examining how changes in crucial parameters—such as the number of PCA components retained, the explained-variance cutoff, the threshold for opening or closing a position, and the length of the rolling estimation window—impact mean reversion efficacy and profitability. This section clarifies that while certain defaults (e.g., 60-day windows, 1.25 standard-deviation entry triggers) are robust, alternative parameter choices can yield different risk–return profiles.

Finally, Section 8 consolidates the main conclusions and proposes avenues for future research. This concluding chapter underscores the practical implications of PCA-based pairs trading, the role of sector ETF hedges, the importance of dynamic risk-factor specification, and the potential for further enhancements via advanced machine learning models, alternative data, or adaptive thresholding methods.

Overall, this thesis demonstrates that factor selection—whether through PCA eigenportfolios or industry-specific ETF proxies—meaningfully influences residual-based signals, risk exposure, and performance metrics. By systematically comparing and refining these methods, the research provides a deeper understanding of how to construct robust, mean-reverting trading strategies that adapt to evolving market conditions and maintain market neutrality.

2. Literature Review

The study of mean-reversion and contrarian strategies has a long-standing history in financial economics. Foundational work by Poterba and Summers (1988), Lehmann (1990), and Lo and MacKinlay (1990) identified statistically significant return reversals over both short and long horizons. These findings laid the groundwork for market-neutral, long–short equity strategies and inspired the development of systematic statistical arbitrage frameworks, such as the one analyzed by Khandani and Lo (2007). In recent years, this research area has evolved substantially, incorporating residual-based signal construction, advanced factor models—including those derived via Principal Component Analysis (PCA)—and machine learning methods. This section reviews relevant literature in 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.

2.1 Residual-Based Mean-Reversion Strategies

An important direction in recent research emphasizes the role of **idiosyncratic residuals**—the component of returns unexplained by systematic factors—as a more robust source of mean-reversion signals. Huij and Lansdorp (2017) analyze global equity markets and show that strategies based on residual returns, obtained by orthogonalizing total returns against Fama–French risk factors, consistently achieve higher Sharpe ratios and reduced exposure to systematic risks. Chae and Kim (2020) construct a weekly contrarian strategy based on firm-specific residuals and demonstrate that short-term reversal profits are primarily driven by negative autocovariance in idiosyncratic returns, rather than by cross-sectional interactions. These findings collectively suggest that mean-reverting behavior is more effectively captured at the residual level, supporting a modeling framework in which trading signals are derived after removing systematic influences.

2.2 PCA-Based Factor Modeling and Pairs Trading

Principal Component Analysis (PCA) has been widely adopted in the context of 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 a set of principal components. More recent studies have expanded on this idea.

Bartkoviak et al. (2022) compare cointegration-based and PCA-based pairs trading strategies using high-frequency U.S. equity data. In their framework, PCA is used to extract dominant risk factors from the universe, and residuals are estimated as deviations from the PCA-based linear projection. The resulting strategy shows stable positive returns and shallow drawdowns even during periods of market stress, highlighting PCA’s value in improving robustness and market neutrality.

PCA is also employed in pair selection via clustering. Sarmento and Horta (2020) propose a hybrid method that applies PCA for noise reduction and then uses unsupervised clustering (OPTICS) to group similar assets before signal generation. Their model, tested on commodity ETFs, outperforms traditional correlation-based methods. Similarly, Rotondi and Russo (2025) cluster S&P 500 constituents using three distance metrics—raw Euclidean distance, PCA-based distance, and partial correlation distance—and find that PCA-based clustering significantly enhances performance, yielding market-neutral returns with Sharpe ratios between 0.8 and 1.0. The findings suggest that PCA helps isolate meaningful co-movement structures for pair construction.

Han et al. (2023) adopt a fundamentally motivated clustering approach that incorporates PCA as part of the distance metric for pair identification. Their agglomerative clustering algorithm forms more stable, fundamentally linked pairs that yield high Sharpe ratios and strong out-of-sample performance. These results show that PCA not only contributes to residual construction but also improves pair selection by embedding latent factor structures into the similarity measure.

Finally, Xiang and He (2022) extend this idea in an asset pricing context. Using PCA-structured residual signals and Ornstein–Uhlenbeck dynamics, they construct arbitrage portfolios in the Chinese equity market and show that the resulting return premia are significant even after controlling for known factors. Their work frames pairs trading strategies as a latent-factor-driven anomaly and suggests that residual spreads may reflect compensated risk.

2.3 Machine Learning Enhancements to PCA Frameworks

Although PCA is a linear technique, it can be extended or complemented by machine learning methods for improved signal generation. Krause and Calliess (2024) compare PCA-based, Fama–French-based, and autoencoder-based factor extraction models, finding that deep autoencoders outperform PCA in terms of Sharpe ratios and residual reduction. However, PCA remains a robust and interpretable baseline that connects naturally to both classical factor models and statistical arbitrage frameworks. In several studies (e.g., Rotondi & Russo, Han et al.), PCA serves as a feature extraction or noise reduction step prior to clustering, showing its continued relevance even in machine learning-enhanced settings.

These approaches suggest that machine learning, particularly when integrated with PCA, can extract complex latent patterns that improve the quality and consistency of mean-reversion signals.

In summary, recent literature provides strong empirical and methodological support for the use of residual-based trading strategies and PCA-derived factor models in statistical arbitrage. PCA has proven effective both for extracting systematic components from stock returns and for facilitating the identification of mean-reverting spreads through clustering and distance metrics. These studies collectively reinforce the view 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. The next section builds on these insights to formulate a PCA-based framework for residual-driven trading signal construction.

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

Consider a universe composed of N stocks, each with returns denoted as R_i for a given trading period (e.g., daily returns from market close to market close). We assume a single-factor model initially, wherein the market portfolio (for instance, represented by a capitalization-weighted benchmark such as the S&P 500) has returns denoted by F. Each stock's returns can then be decomposed through a linear regression framework as:

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

where β_i captures the sensitivity of stock $i's$ return to the market factor F, and $\tilde{R}_i$ denotes the idiosyncratic component specific to stock i, which is uncorrelated with the systematic factor.

Expanding beyond the single-factor model, we often employ multi-factor models to capture a richer representation of systematic risks. The generalized form of such a multi-factor model is expressed as:

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

where m represents the number of systematic risk factors, each F_j can be viewed as the return of a benchmark or factor portfolio, and the β_{ij} coefficients measure the sensitivities of each stock's returns to these respective factors.

A portfolio constructed with stock weights $\{Q_i\}_{i=1}^{N}$ is considered market-neutral if it is designed such that it holds no net exposure to these systematic risk factors. Formally, market-neutrality is achieved when the portfolio's aggregated factor exposures (also referred to as factor betas) are zero:

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

Under these conditions, a market-neutral portfolio is uncorrelated with systematic factors, and its returns are driven purely by idiosyncratic components:

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

Empirical evidence from developed markets, such as G8 economies, indicates that typically around 10 to 20 factors are sufficient to adequately capture systematic risks, with approximately 15 being most common. Studies (e.g., Laloux et al., 2000; Plerou et al., 2002) suggest these systematic factors explain about half of the variance observed in stock returns. The critical challenge thus remains clearly identifying and defining these underlying systematic factors to effectively manage and isolate risk within market-neutral strategies.

3.1 The PCA Method for Factor Extraction

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

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:

$$R_{ik} = \frac{S_{i(t_0-(k-1)\Delta t)} - S_{i(t_0-k\Delta t)}}{S_{i(t_0-k\Delta t)}}, \; k = 1, \dots, M, i = 1, \dots, N, \tag{8}$$

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

$$Y_{ik} = \frac{R_{ik} - \bar{R}_i}{\sigma_i} \tag{9}$$

Where,

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

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 construction, ρ is symmetric and positive semi-definite. 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 1000×1000—while the available data is relatively limited. Using too long a historical window (i.e., $M \gg N$) incorporates outdated information and may fail to reflect current market dynamics. On the other hand, restricting the estimation to a recent window (e.g., one year or 252 trading days) leads to a situation where the number of observations M is smaller than the number of parameters to estimate in ρ, making the matrix noisy and potentially ill-conditioned.

To balance this trade-off, a rolling one-year window is used for correlation estimation. This approach captures recent market structure while avoiding excessive lookback bias. PCA is then applied to extract the dominant factors and filter out noise, enabling 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 eigen-decomposition is expressed as:

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

Empirically, the spectrum of ρ typically reveals 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 dominant components correspond to systematic factors (e.g., market, sectors), while the remaining eigenvalues reflect idiosyncratic fluctuations.

To better understand the spectral distribution, the eigenvalue density function can be defined as:

$$D(x,y) = \frac{\#\, of\ eigenvalues\ in\ the\ interval\ (x,y)}{N} \tag{13}$$

This density function provides insight into the proportion of the spectrum explained by low-variance (noisy) components (see *Figure 2*).

Selecting Significant Components and Eigenportfolios

To construct a low-rank approximation of the correlation structure, two main approaches are used to determine the number of retained components mmm:

- Fixed-rank truncation: Retain a fixed number mmm of top eigenvectors (e.g., based on the number of economic sectors).
- Variance thresholding: Retain the smallest mmm such that the cumulative explained variance exceeds a predefined threshold α, typically 80%–90%:

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

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

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

The eigenportfolio return at time k is:

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

These factor returns F_{jk} are mutually uncorrelated by construction and represent the latent systematic drivers in the market. Each asset return can now be decomposed as:

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

where $\tilde{R}_{ik}$ is the idiosyncratic residual, uncorrelated with the factor returns.

This model implies a decomposition of the correlation matrix:

$$\rho_{ij} = \sum_{j=1}^{m}\lambda_j v_i^{(j)} v_j^{(j)} + \eta_i^2\delta_{ij} \tag{16}$$

where $\eta_i^2 = 1 - \sum_{j=1}^{m}\lambda_j (v_i^{(j)})^2$ ensures that the diagonal remains normalized. The Kronecker δ_{ij} captures the residual noise component.

This decomposition preserves the total variance of each asset while isolating the significant cross-sectional structure from residual noise. In doing so, PCA delivers a set of orthogonal risk factors that can be used for return modeling, portfolio construction, and the extraction of market-neutral signals in pairs trading frameworks.

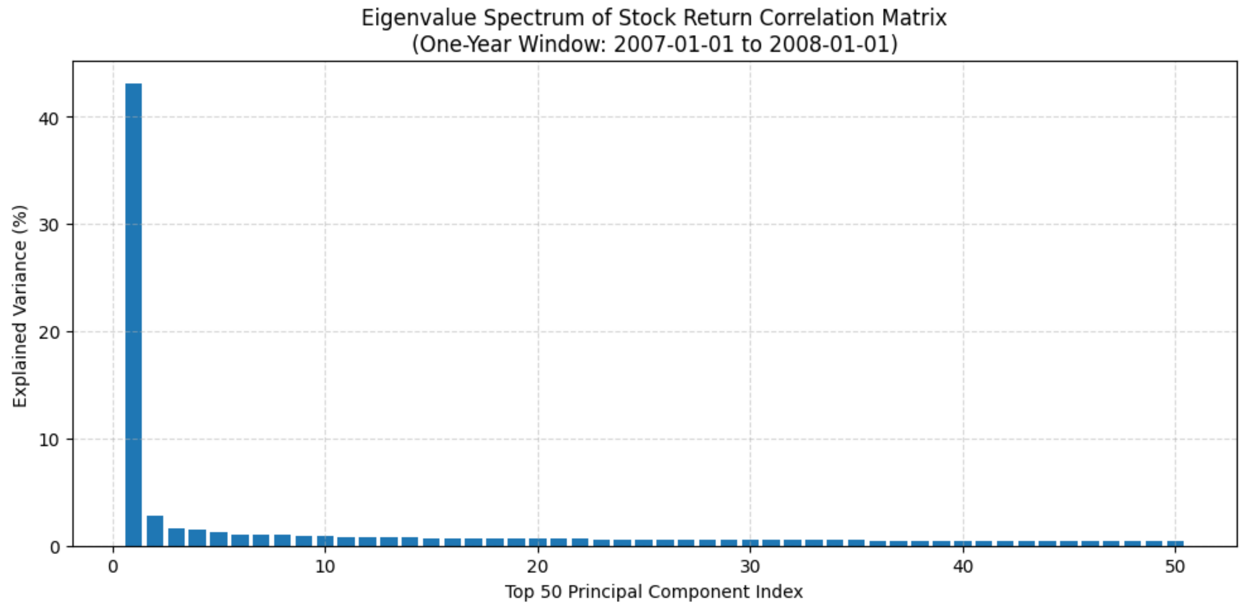

Figure 1A. Top 50 eigenvalues of the correlation matrix of A-share stock returns computed using a one-year window ending on 1 January 2008. The eigenvalues are derived from daily return data for a filtered universe of 303 stocks with imputed values, and are shown as percentages of explained variance. The first principal component (market eigenvalue) dominates the spectrum, followed by a steep decay characteristic of common sector or idiosyncratic effects.

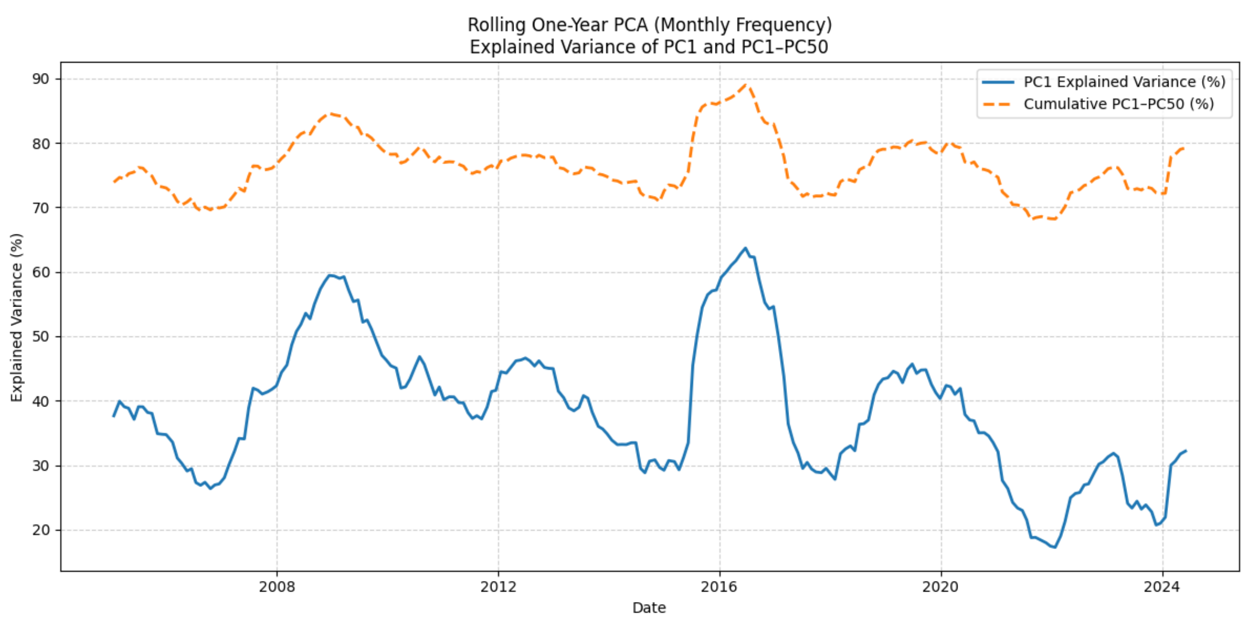

Figure 1B. Time series of the explained variance of the first principal component (PC1) and the cumulative variance explained by the top 50 components (PC1–PC50), based on a rolling one-year PCA using daily returns from 303 A-share stocks. The window rolls monthly from 2007 to 2024. The plot reveals 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.

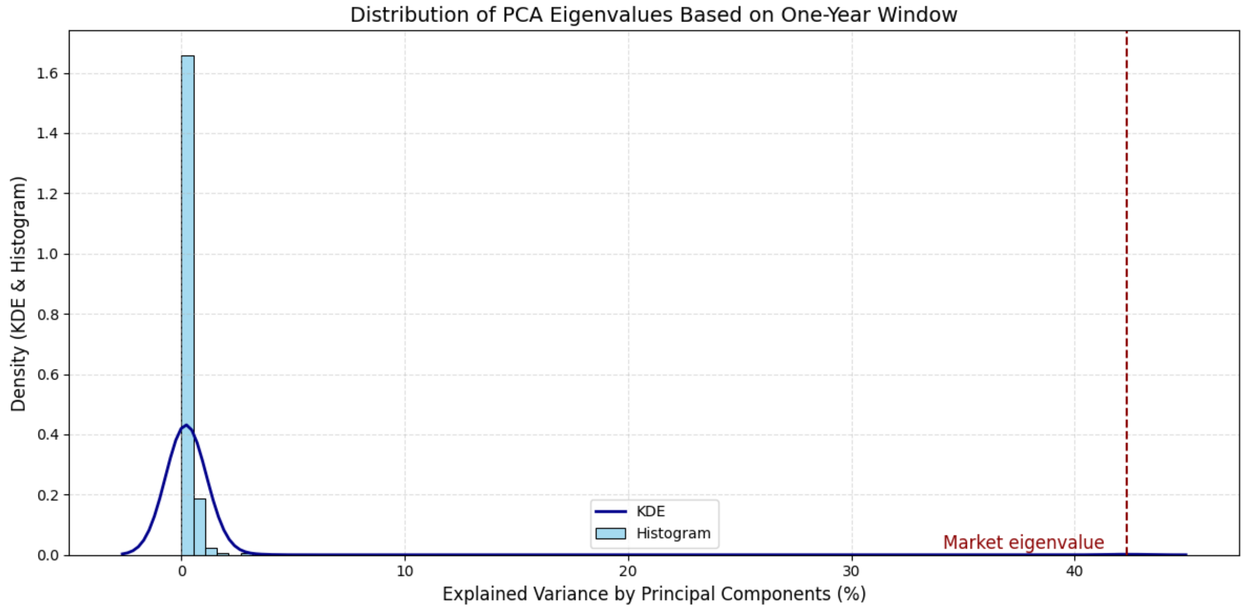

Figure 2. The distribution of PCA eigenvalues estimated using a one-year window ending on 1 January 2008, based on daily returns of 303 A-share stocks. The histogram and KDE curve show a clear separation between a large 'market eigenvalue' and the rest of the bulk spectrum. The bulk of the eigenvalues 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, such as Laloux et al. (2000), have highlighted that the leading eigenvector obtained from the correlation matrix of asset returns typically corresponds to what is referred to as the *market mode*. This interpretation arises from the observation that all components of the first eigenvector $v_i^{(1)}$, for $i = 1,2,\dots,N$, are positive, implying that the associated eigenportfolio assigns positive weights to all assets. Specifically, the weights of this portfolio are given by $Q_i^{(1)} = v_i^{(1)}/\sigma_i$, where σ_i denotes the volatility of asset i. This structure results in weights that are inversely proportional to asset volatility, which tends to align with capitalization-weighted portfolios, as firms with larger market capitalization often exhibit lower volatility. While not identical, the market-mode eigenportfolio and the capitalization-weighted market index show 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 and Scheinkman, 1991) or volatility surface decomposition (Cont and Da Fonseca, 2002)—a direct analogy to "shape modes" is not applicable.

Instead, the approach adopted here draws inspiration from Scherer and Avellaneda (2002), who studied the correlation structures among sovereign bonds issued by various Latin American countries. Similar techniques have also been employed in the context of equity markets by Plerou et al. (2002). The procedure involves ranking the components of an 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 reordering of asset indices based on the magnitude of their eigenvector coefficients. Through this reordering, a notable pattern emerges: firms with adjacent ranks in this sequence often belong to the same industry group. This observation, referred to as *coherence*, suggests that higher-order eigenvectors capture sector-specific or industry-level modes of co-movement. Tables 1 and 2, along with Figures 4 through 6, provide empirical support for this phenomenon. Coherence is particularly evident in the second and third eigenvectors, but diminishes as one moves deeper into the spectrum. For eigenvectors associated with the bulk (noise) part of the spectrum, such structured clustering dissipates, indicating that these components largely reflect unstructured, idiosyncratic fluctuations.

Thus, the eigenportfolios corresponding to these intermediate eigenvectors can be interpreted as long–short portfolios that contrast industries or sectors. This interpretation aligns with the view that the eigenvalue spectrum captures a hierarchy of market influences, from broad market movements to localized sectoral dynamics, and finally to random noise.

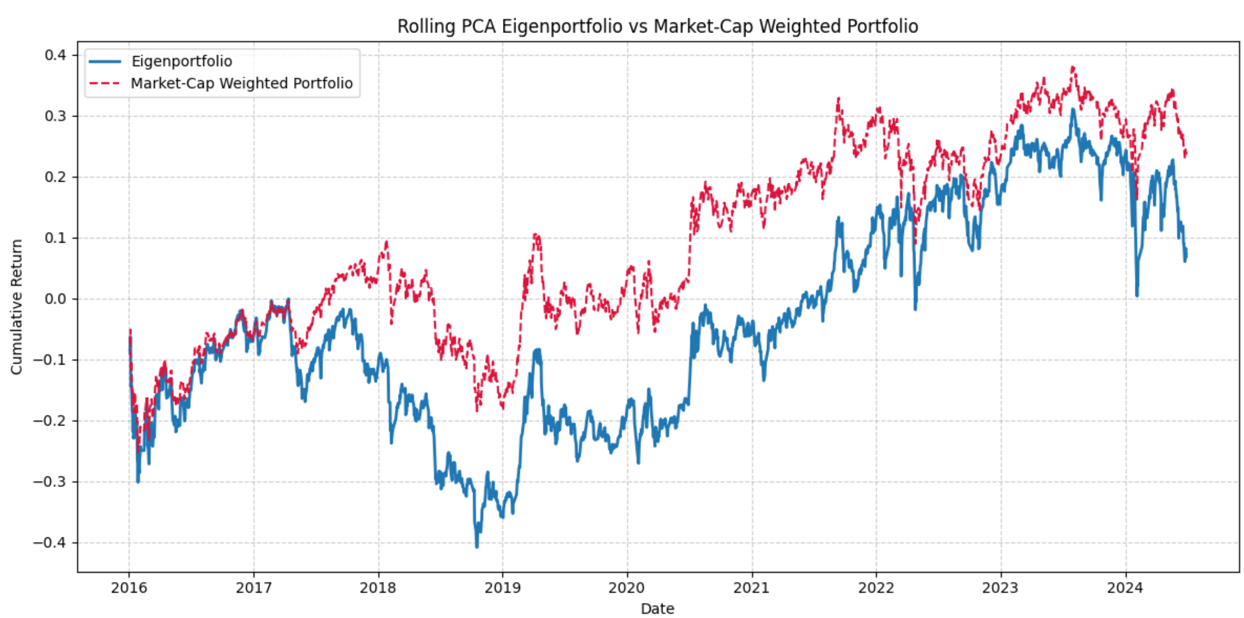

Figure 3. Comparative evolution of the principal eigenportfolio and the capitalization-weighted portfolio from January 2016 to January 2024. The eigenportfolio is constructed using rolling PCA with a one-year window and updated monthly. Both portfolios exhibit broadly similar trends over time, with the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

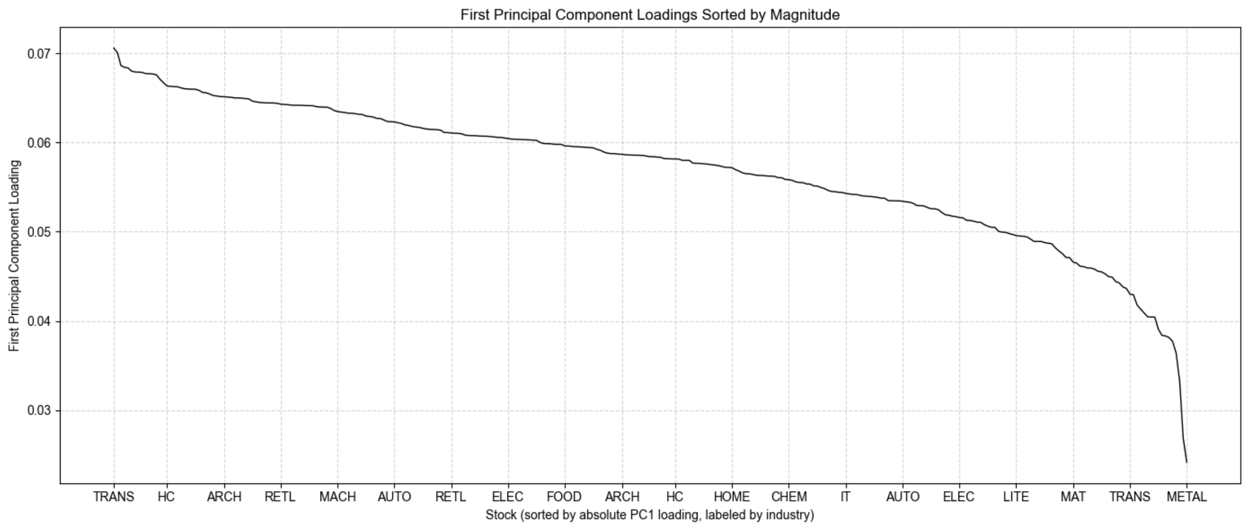

Figure 4. presents the loadings of the first principal component (i.e., the leading eigenvector) derived from the PCA of the stock return correlation matrix. The loadings are sorted by their absolute magnitude, highlighting the stocks that contribute most strongly to the market-wide common component. 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).

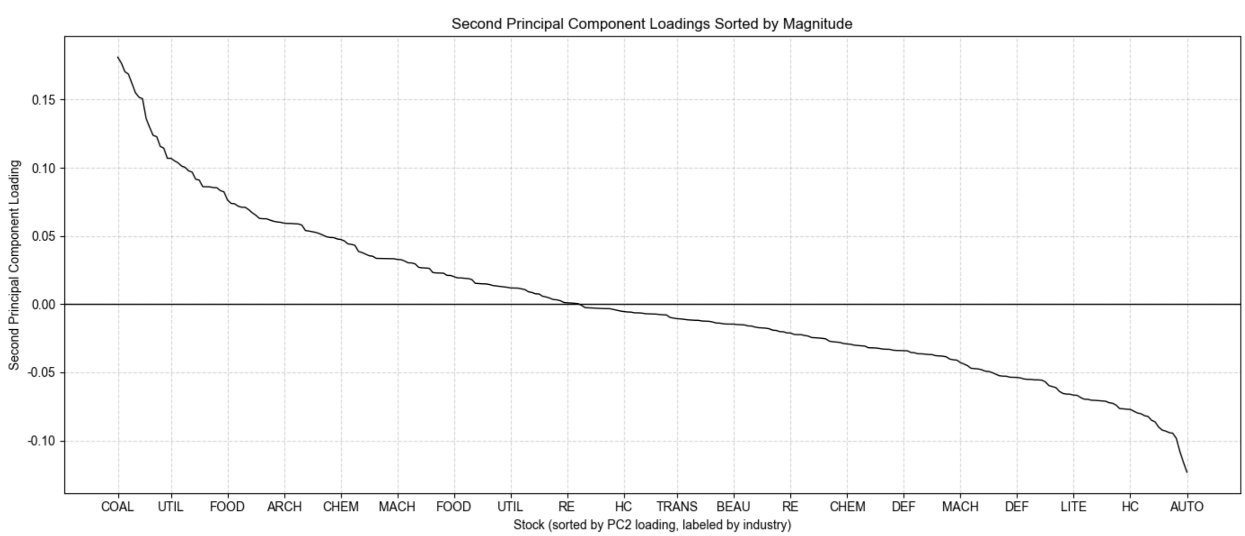

Figure 5. This figure plots the loadings of the second eigenvector obtained from the principal component analysis (PCA) of the stock return correlation matrix. The stocks are sorted in descending order of their PC2 coefficients (not by absolute value), and each is labeled with the abbreviation of its 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.

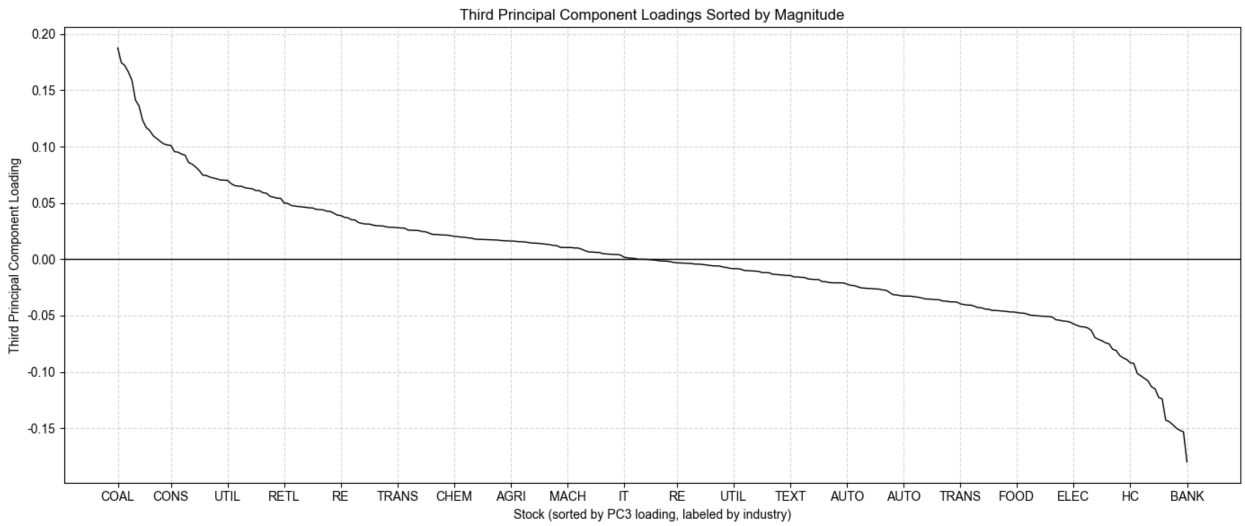

Figure 6. This figure displays the loadings of the third principal component (PC3), obtained from the PCA of the stock return correlation matrix. Stocks are ordered from highest to lowest based on their PC3 loadings, and each is labeled by the abbreviation of its corresponding industry sector. Unlike 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.

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 Feed
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

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

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

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

A practical alternative to principal component methods for isolating residuals in asset returns involves using sector-based exchange-traded funds (ETFs) as factor proxies. In this approach, the returns of ETFs that represent specific industry sectors are employed as explanatory variables in a single-factor regression framework for individual stocks.

Table 3 presents an illustrative breakdown of industry sectors, including the number of constituent stocks with a market capitalization exceeding RMB 2 billion as of January 2007. This table helps contextualize the relative size of each sector and its representation within the broader trading universe. For each industry, a corresponding sector ETF is also identified, which serves as a simplified risk factor for modeling the returns of stocks in that sector.

Unlike eigenportfolios—whose orthogonality guarantees uncorrelated factors—sector ETF returns are often significantly correlated with one another. This interdependence can introduce redundancy into the factor structure. For example, two highly correlated ETFs may both load heavily on a stock's returns, but with coefficients of opposing signs, especially when the stock is influenced by multiple sectors. This can lead to interpretability challenges and inflated noise in the residuals.

To address this issue, various regression strategies have been proposed in the literature. One promising direction involves the use of sparse regression techniques, which aim to reduce the number of explanatory variables by selecting only the most relevant ones. A notable example is the Matching Pursuit algorithm (Davis et al., 1997), which iteratively selects factors that best explain the variance in returns, encouraging parsimony. Similarly, Ridge Regression (Jolliffe, 2002) introduces a penalty term to shrink the regression coefficients, thereby mitigating the effects of multicollinearity among ETFs.

In the context of this review, a simplified and more interpretable methodology is adopted. Each stock is associated with a single sector ETF, based on its primary industry classification, as outlined in Table 3. A univariate linear regression is then performed:

$$R_i = \beta_i \cdot ETF_i + \tilde{R}_i, \tag{17}$$

where R_i is the return of stock i, ETF_i is the return of the associated industry ETF, β_i is the factor loading, and $\tilde{R}_i$ denotes the residual return component not captured by the sector factor.

This one-to-one matching approach avoids the complications arising from correlated predictors and provides a straightforward method for residual estimation, which can be directly utilized in subsequent mean-reversion or statistical arbitrage strategies.

			Market cap unit: 1M/RMB		
Index	**Industry**	**Num of Stocks**	**Average MktCap**	**Max MktCap**	**Min MktCap**
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**

Table 3. Summary statistics by industry based on market capitalization as of January 4, 2016. The table reports the number of stocks, average, maximum, and minimum market capitalization (in million RMB) for each industry sector. Total figures are shown in the last row.

4. An Equity Valuation Approach Based on Relative Value

This section introduces a quantitative, relative-value approach to equity valuation, centered on a stock's performance relative to its industry sector or factor benchmarks. The model employs either sector-based ETFs or PCA-derived synthetic factors. A further refinement, incorporating trading volume, is discussed in Section 5. Although the current formulation is solely based on price and volume data, it can be extended to include fundamental indicators such as analyst revisions, earnings momentum, and other measurable financial variables.

We adopt continuous-time modeling, 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 cases where a single sector ETF is used as the benchmark, 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 $I(t)$ being the ETF corresponding to the stock's industry classification.

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

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

where $\alpha_i dt$ represents the excess return over the market or sector, and $dX_i(t)$ captures the residual price movements driven by stock-specific factors. This component is modeled using a mean-reverting Ornstein–Uhlenbeck (OU) process:

$$dX_i(t) = \kappa_i(\mu_i - X_i(t))dt + \alpha_i dW_i(t), \kappa_i > 0, \quad (21)$$

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

The expected conditional increment of this process is given by:

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

implying that the forecasted return depends on the deviation of $X_i(t)$ from its equilibrium level 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 empirical implementation, we estimate these parameters using a 60-day rolling window and retain only those stocks with sufficiently high mean-reversion speed (i.e., large κ_i) to ensure robustness of the stationarity assumption. Details on how to estimate the model are given in the next section and in Appendix A.

Solving the OU process yields:

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

In the limit as $\Delta t \to \infty$, the process reaches its stationary distribution:

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

Based on this formulation, we construct market-neutral long-short portfolios where the investor takes a long position in the stock and a short position in the corresponding sector ETF (or factor portfolio), proportionally scaled by the factor loading. The expected one-day return from such a portfolio is:

$$\alpha_i dt + \kappa_i(\mu_i - X_i(t))dt, \quad (25)$$

where the second term captures the forecasted mean-reversion return. When $X_i(t)$ is above its long-term mean μ_i, the model predicts negative expected returns, suggesting a short signal, and vice versa when $X_i(t)$ is below μ_i.

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

$$\tau_i = 1/\kappa_i,$$

represents the characteristic timescale of mean reversion. Stocks with $\kappa_i \gg 1$ (i.e., $\tau_i \ll 1$) revert quickly, making them ideal candidates for short-term relative-value strategies. Our model selects only those stocks with sufficiently rapid mean-reversion to ensure validity of the estimation and profitability of the trading signals.

5. Construction of Trading Signals

In this section, we introduce the construction of trading signals derived from the residual processes estimated under the mean-reversion framework. Specifically, we assume that residual returns follow an Ornstein–Uhlenbeck (OU) process, and we estimate model parameters using a rolling window of 60 business days (i.e., $T1 = 60/252$). This window length is chosen to balance model responsiveness and stability, corresponding roughly to the length of one corporate earnings cycle.

As part of the signal filtering mechanism, we retain only those stocks whose estimated mean-reversion times are shorter than half a reversion cycle, i.e., less than approximately 8.4 trading days ($252/30$). This ensures that the underlying process is sufficiently mean-reverting within the estimation window, thereby validating the assumption of local stationarity and making the signals statistically meaningful for short-horizon trading.

To provide an initial empirical evaluation of the signals generated under this model, Table 5 reports the historical Sharpe ratios achieved by strategies that rely on these signals, applied to portfolios sorted by sector ETFs over the whole trading period. Although full-scale backtesting results will be presented in a later section, this table offers a preliminary performance overview.

For technical details regarding the estimation of the OU process parameters and the generation of standardized signals (e.g., s-scores).

5.1 Signal Design Based on Pure Mean-Reversion Dynamics

In the baseline signal construction approach, we consider a simplified version of the Ornstein–Uhlenbeck (OU) process by omitting the drift component. That is, we model the residual process $X_i(t)$ as purely mean-reverting without directional bias. Under this assumption, the equilibrium variance of the process can be expressed as:

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

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

To standardize deviations from equilibrium across stocks with different volatility profiles, we define a dimensionless measure called the s-score as follows:

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

where μ_i is the long-term mean of the residual process. The s-score measures the normalized distance of the residual from its equilibrium level and serves as the key decision variable in our mean-reversion trading strategy. See figure 7 for a graph showing the evolution of the s-score for residuals of CMBC against the Banking sector. A higher absolute value of $s_i(t)$ indicates a greater deviation from equilibrium and, hence, a stronger 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 results in a market-neutral position by construction. When a long signal is triggered (i.e., $s_i(t) < -s_{entry}$), the strategy involves buying one dollar of the corresponding stock and simultaneously selling β_i dollars of the associated sector ETF to hedge systematic exposure. This ensures the position is neutral with respect to the sector-level factor used in residual construction.

In the case of a multi-factor model—such as a regression on multiple ETFs or on principal components extracted via PCA—the hedging leg of the trade involves a weighted combination of 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_{ij} denotes the factor loading of stock i on factor j. These weights are obtained from the regression used in residual estimation. The reverse applies to short positions, and all trades are closed by unwinding both the stock and the hedging legs according to the same weights.

This hedging procedure preserves dollar neutrality with respect to the underlying systematic components, ensuring that the expected return of the trade depends solely on the behavior of the idiosyncratic (mean-reverting) component. This design is central to the strategy's ability to remain agnostic to broader market movements while exploiting transitory pricing inefficiencies at the stock level.

To ensure consistency and scalability across different assets, we apply uniform entry and exit thresholds. These thresholds were selected empirically through backtesting on historical data from 2010 to 2014. The optimal values found are:

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

The asymmetric closing rules—closing short positions at a lower s-score (0.75) than long positions (0.50)—were found to yield slightly better performance during the training period, and are therefore retained in our backtesting procedure.

This rule-based approach aims to capitalize on statistically significant deviations of the residual from its mean, under the assumption that such excursions are temporary and will revert over a characteristic time scale determined by $1/\theta_i$. By focusing only on substantial deviations, the strategy attempts to filter out noise and concentrate trades on high-conviction opportunities with a favorable mean-reversion profile.

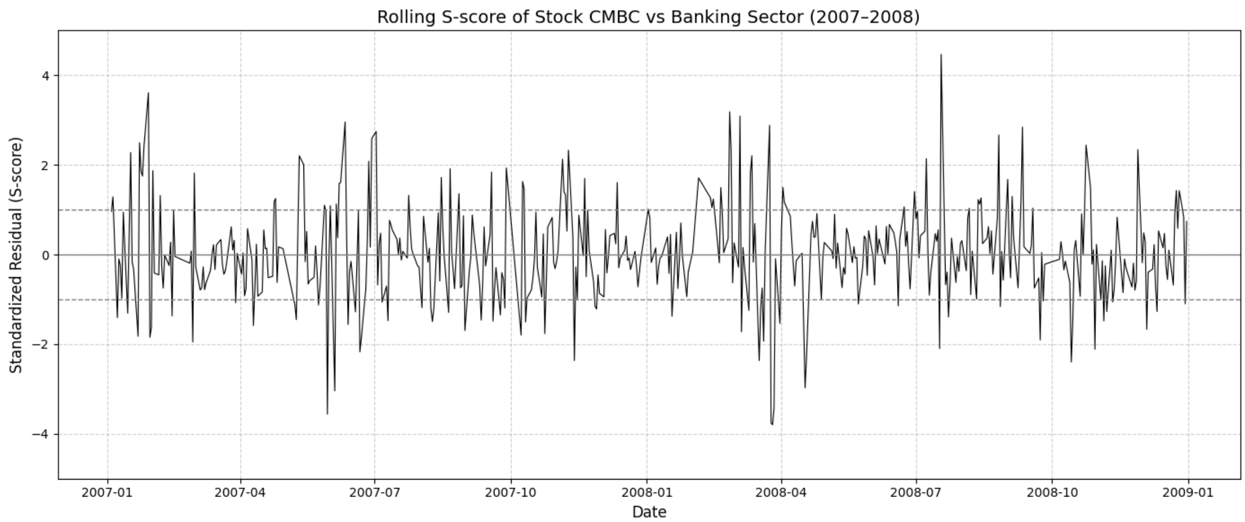

Figure 7. Rolling S-score of China Minsheng Banking Corp. (CMBC) vs. the Banking sector index during 2007–2008. The S-score represents standardized residuals from a rolling regression of stock returns on sector returns, used to detect mean-reversion signals.

5.2 Signal Design Based on Mean-Reversion with drift

In the pure mean-reversion model introduced in the previous section, the drift term μ_i of the residual process was neglected under the assumption that it is negligible compared to the amplitude of fluctuations, i.e., $\sigma_{eq,i} = \sigma_i/\sqrt{2\theta_i}$. However, it is possible to incorporate the drift explicitly into the trading signal, leading to a modified signal metric that accounts for potential directional bias in the residual.

To do so, we consider the conditional expectation of the residual return over a short time interval dt, which under the Ornstein–Uhlenbeck (OU) specification is given by:

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

This expression can be rearranged as:

$$E[dX_i(t)] = \left(\frac{\mu_i}{\frac{\sigma_i}{\sqrt{2\theta_i}}} - \theta_i s_i(t)\right) dt = \left(\frac{\mu_i}{\theta_i \sigma_{eq,i}} - s_i(t)\right)(-\theta_i)dt. \tag{29}$$

This motivates defining a modified s-score that incorporates the effect of the drift:

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

The intuition here is that if the residual has a positive drift ($\mu_i > 0$), we are less inclined to short the stock, since it tends to move upward even after adjusting for factor exposure. Conversely, a negative drift would strengthen a short signal. This formulation effectively tilts the original symmetric mean-reversion signal in the direction of drift, introducing an implicit momentum component.

To illustrate: under the original strategy, a short position would be initiated if $s_i > s_{entry}$. With the modified s-score, however, this threshold is harder to exceed if $\mu_i > 0$, thereby reducing the likelihood of falsely shorting a stock with upward drift.

From a statistical viewpoint, the drift μ_i can be interpreted as the slope of a short-term (e.g., 60-day) moving average of the residual process. As such, the incorporation of μ_i introduces a momentum filter into the otherwise purely mean-reverting strategy.

Empirical analysis of drift-adjusted signals over the 2010–2014 training period shows that the optimal threshold values determined for the pure mean-reversion strategy remain effective here as well. Moreover, we find that in practice, the estimated values of μ_i are relatively small—typically around 15 basis points—while the average reversion time is approximately 7 trading days, and the equilibrium volatility of residuals is on the order of 300 basis points. This implies that the adjustment to the s-score, given by:

$$\frac{\mu_i}{\theta_i \sigma_{eq,i}} \approx \frac{0.15 \times 7}{3.00} \approx 0.3,$$

has only a minor effect on the decision variable over short horizons.

Backtesting simulations confirm that this adjustment leads to little improvement in trading performance at the daily frequency considered. Consequently, for the sake of simplicity and robustness, we adopt the original s-score formulation in our primary implementation and omit further analysis of the drift-adjusted signal in subsequent sections.

In summary, while theoretically appealing, the inclusion of a constant drift term in the residual process does not provide material improvement to strategy performance. This suggests that, after adjusting for systematic risk factors, the idiosyncratic component of stock returns exhibits little persistent momentum at the timescales relevant for our trading framework.

6. Results

To evaluate the effectiveness of the proposed strategy, we conducted a series of back-testing experiments using historical stock data. The core idea is to simulate daily trades across the entire stock universe based on the signal rules defined earlier (Equation 27). Every trading day, we perform parameter estimation—such as betas and residuals—and signal computation, ensuring that all information used is strictly backward-looking and respects a rolling window framework.

During the estimation process, if the mean-reversion speed parameter κ_i exceeds the threshold of 8.4, we interpret this as the model being ill-suited for that stock. In such cases, we either (i) refrain from initiating new trades or (ii) unwind any existing positions associated with the affected stock.

We assume all trades are executed at that day's closing price. To reflect the presence of transaction frictions—including price slippage and other trading costs—we impose a round-trip transaction cost of 10 basis points (i.e., 0.1%).

Let E_t denote the total portfolio equity at time t. The profit-and-loss (PNL) update equation for the portfolio is as follows:

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

where:

- $Q_{it} = E_t \gamma_t$ represents the capital allocated to stock i at time t,
- R_{it} is the return of stock iii over $(t, t + \Delta t)$,
- r is the risk-free interest rate,
- $\Delta t = \frac{1}{252}$ corresponds to one trading day in a year,
- D_{it} is the dividend paid by stock iii 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 and is identical for all stocks. It is calibrated to maintain a target leverage. For instance, if we aim to hold 200 long and 200 short positions simultaneously with an overall leverage of 2+2 (i.e., two times equity on both the long and short sides), then we set $\gamma_t = \frac{2}{100}$. In other words, γ_t can also be interpreted as the maximum allowable fraction of equity to be invested in any single stock.

This leverage calibration, validated using data from 2002 to 2004, was selected to produce a portfolio with approximately 10% annualized volatility. It is worth noting that the choice of leverage does not affect risk-adjusted performance metrics such as the Sharpe ratio. Similar strategies using different leverage levels—such as the "1/2 + 1/2" contrarian portfolio from Khandani and Lo (2007)—yield comparable results once properly normalized.

Given the binary nature of the trading signals, the strategy follows a bang-bang implementation. That is, once a trading signal is activated, we take a full position in the respective stock (either long or short), and the position is entirely unwound when the exit signal is triggered. Although this all-or-nothing approach may appear suboptimal at first glance, it consistently outperforms strategies that make incremental adjustments. This outperformance is likely due to reduced model error sensitivity and minimized transaction cost accumulation associated with frequent rebalancing.

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 over a rolling 60-day window preceding each estimation date. Individual stock returns were regressed on the returns of their respective sector indices, and the residuals from these regressions were modeled as Ornstein–Uhlenbeck (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 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 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 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.

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.

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.

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

Table 4. Annual performance metrics of the synthetic ETF-based statistical arbitrage strategy with trading costs included (2005–2024). Performance includes a round-trip transaction cost of 10 basis points per trade. The final row ("All") summarizes the full-period statistics.

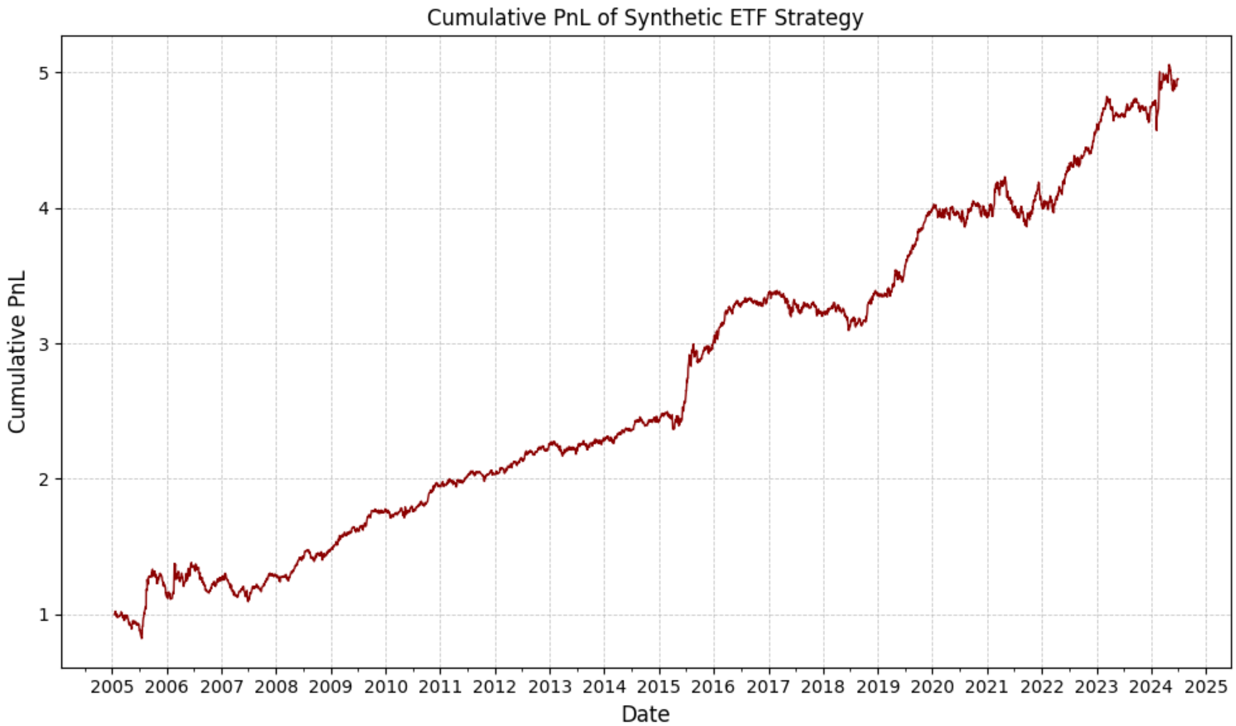

Figure 8. Historical PNL statistics for the strategy using synthetic ETFs as factors from 2005–2024.

6.2 PCA as Factors

Table 5 reports the annual performance of the statistical arbitrage strategy constructed using 15 principal components as systematic risk factors. As with the synthetic ETF-based strategy, daily regressions were conducted over a 60-day rolling window to extract residuals for each stock. Trading signals were generated based on the degree of deviation from the estimated long-term mean, and market neutrality was enforced through dynamic hedging using the SPY index.

The cumulative performance over the full sample period (2005–2024) is generally strong, with the strategy achieving an average annual return of 8.33% and a Sharpe ratio of 0.94 after accounting for trading costs. Notably, the strategy performed particularly well in 2009, 2010, 2015, 2017, and 2018—years in which the Sharpe ratio exceeded 1.6 and the Sortino ratio reached as high as 5.61. These periods were characterized by high cross-sectional dispersion and relatively low drawdowns, resulting in high Calmar and Omega ratios. In contrast, the strategy underperformed in 2012, 2019, and 2024, where negative Sharpe and Calmar ratios indicate periods of weak mean-reversion or increased residual noise.

In years with strong performance, such as 2009 and 2017, the t-statistics exceed 2.0 and the Omega ratio approaches or surpasses 1.4, confirming statistically significant excess returns and favorable gain-loss asymmetry. The Jarque–Bera test, however, frequently returns low p-values, suggesting deviations from normality in residual returns—likely due to fat tails or skewed distributions during high-volatility market episodes.

Overall, the PCA-based approach using 15 eigenportfolios offers a compelling alternative to ETF-based factor models, delivering stable and risk-adjusted returns across a variety of market conditions. The results highlight the effectiveness of PCA in capturing underlying co-movements among stocks and generating robust residual signals suitable for contrarian trading strategies.

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

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

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 from 2005 to 2024. In the earlier years (2005–2010), the PCA-based strategy shows superior performance, suggesting that principal components extracted from stock returns captured meaningful common risk factors that were not fully represented by sector-based indices. The ability of PCA to adapt to latent market structures likely contributed to the early gains. Between 2010 and 2016, both strategies performed similarly, with alternating phases of outperformance. The PCA strategy maintained higher responsiveness to price deviations, while the synthetic ETF strategy exhibited smoother and more stable returns.

From 2018 onward, the synthetic ETF strategy began to outperform more consistently. This shift may be driven by changes in market structure that enhanced the relevance of sector-based co-movements, making capitalization-weighted synthetic indices more effective for residual estimation. Meanwhile, the PCA-based strategy may have suffered from instability in the eigenportfolios due to evolving return correlations and increased noise in factor loadings. Overall, the synthetic ETF strategy appeared more robust in later years, while the PCA approach was more sensitive to regime shifts. These results highlight the complementary strengths of the two approaches across different market environments.

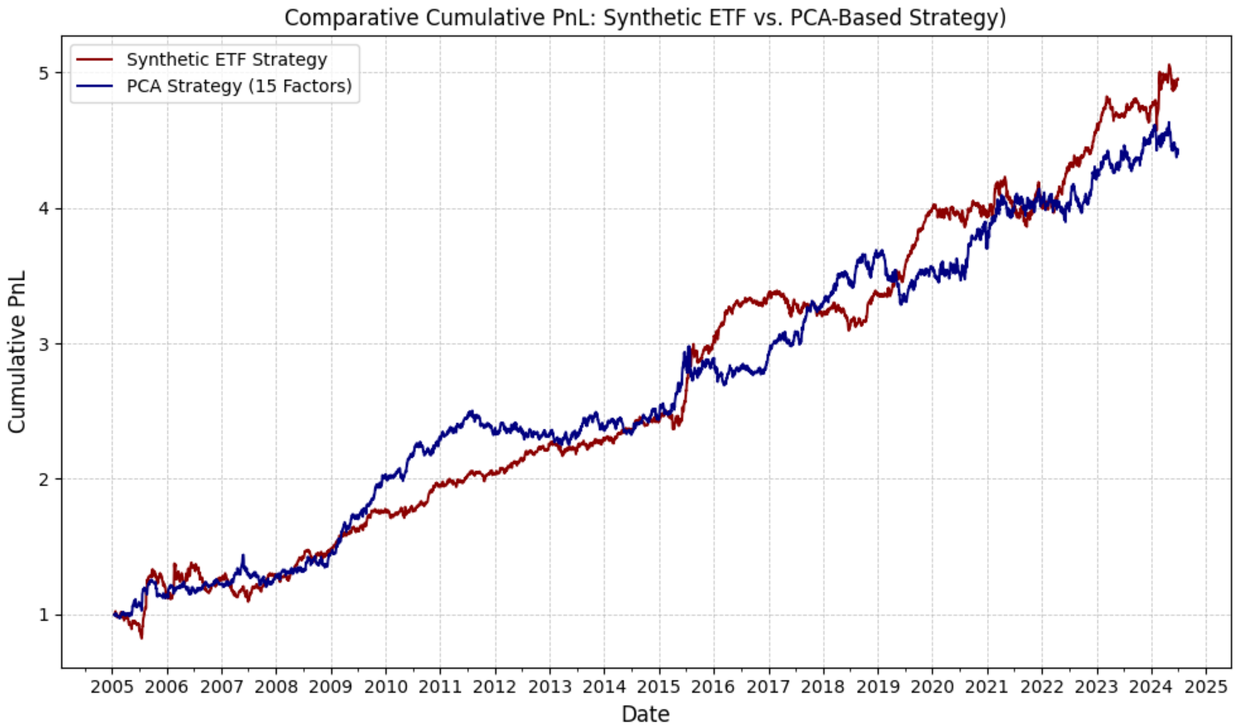

Figure 9. Cumulative PnL comparison between the synthetic ETF-based strategy and the PCA-based strategy with 15 eigenportfolios (2005–2024). The PCA strategy outperforms in earlier years, while the synthetic ETF strategy delivers more stable and superior performance in recent periods.

Figure 10 plots the temporal evolution of two variables: the cumulative 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 visually consistent pattern emerges in which periods of heightened market volatility—such as the 2008 global financial crisis, the 2015 Chinese equity market correction, and the COVID-19 outbreak in 2020—are accompanied by a pronounced increase in the proportion of variance captured by the leading eigenvectors.

This observed co-movement reflects a well-documented phenomenon in empirical asset pricing: under turbulent market conditions, idiosyncratic risk tends to diminish in relative importance, while systematic shocks become more pervasive across the cross-section of assets. As correlations among individual securities strengthen, a smaller number of latent common factors account for a greater share of total return variance. Consequently, the explanatory power of the principal components derived from the covariance matrix of returns increases. Conversely, in more tranquil periods characterized by lower aggregate volatility, return dynamics are more dispersed and idiosyncratic in nature, leading to a reduction in the explanatory efficiency of the same set of components.

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

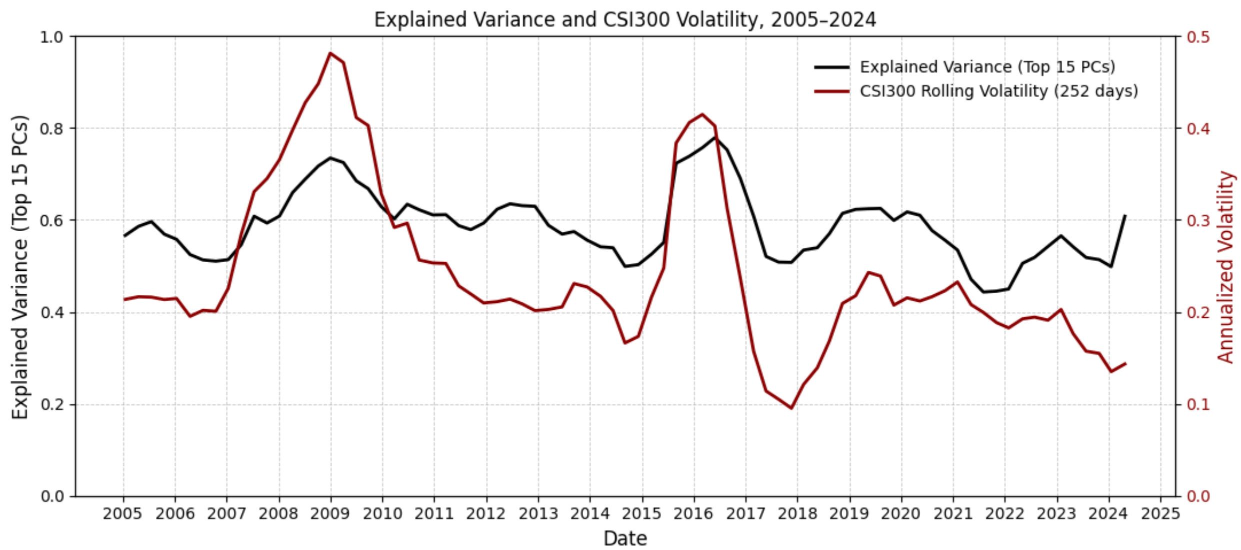

Figure 10. Percentage of variance explained by the top 15 principal components (left axis) and the annualized rolling volatility of the CSI300 Index (right axis) from 2005 to 2024. The explained variance tends to rise during periods of elevated market volatility, suggesting increased common factor dominance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

6.3 PCA with Fixed Explained Variance

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

The annual performance of this variable-factor strategy is reported in Table 6. Over the entire sample period from 2005 to 2024, the strategy achieved an average annual return of 6.32% with a standard deviation of 12.68%, resulting in a Sharpe ratio of 0.50. The strategy exhibited strong performance in high-volatility regimes, such as during the global financial crisis (2008) and the post-pandemic recovery phase (2021–2022), both in terms of risk-adjusted returns and drawdown control. In contrast, performance deteriorated notably in years characterized by market instability without directional trends, such as in 2016 and 2020, when the strategy experienced significant drawdowns and negative Sharpe ratios.

While the strategy is designed to reduce noise by limiting the number of retained components, its out-of-sample effectiveness appears mixed. Relative to the benchmark PCA strategy with a fixed number of 15 eigenportfolios, the 55%-variance strategy delivers broadly comparable results but with slightly inferior overall performance. This is consistent with prior empirical evidence (Avellaneda and Lee, 2010) suggesting that overly restrictive factor selection criteria may fail to capture weaker but relevant sources of common variation, thereby impairing the strategy's robustness in low signal-to-noise environments.

Nonetheless, this approach remains valuable as it reflects the endogenous dimensionality of the return space and naturally adjusts to periods of increased market integration or segmentation. From a practical perspective, it may also reduce transaction costs relative to high-dimensional PCA strategies, although its effectiveness may depend on the prevailing market regime.

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

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.

Figure 11 compares the cumulative performance of two PCA-based mean-reversion strategies applied to the Chinese equity market from 2005 to 2024. The first strategy dynamically adjusts the number of retained principal components to achieve a 55% threshold of total cross-sectional variance (denoted as the variable-factor strategy), whereas the second employs a fixed-dimensional approach with the top 15 principal components (fixed-factor strategy). Both strategies are beta-neutral with respect to the market and account for proportional transaction costs.

The fixed-factor specification consistently outperforms the variable-factor approach across most of the sample period, particularly during post-2010 bull cycles and periods of high market integration. The fixed 15-factor strategy captures a broader spectrum of latent systematic risk drivers, which enhances its capacity to extract persistent co-movement structures from asset returns. This richer factor structure allows for more stable signal generation and more effective defactoring, thereby improving the risk-adjusted return profile.

Conversely, the 55%-variance strategy exhibits more conservative exposure, especially in low-volatility regimes where the number of retained components is reduced. While this dimensionality constraint may offer theoretical advantages in controlling for estimation noise and overfitting, it introduces underfitting risk by discarding lower-ranked yet economically relevant factors. Moreover, the time-varying dimensionality may induce instability in portfolio weights and exacerbate turnover, particularly when the eigenstructure of the return covariance matrix shifts abruptly.

The empirical results thus highlight a classic bias-variance tradeoff: the fixed-factor model, though potentially more prone to noise, benefits from higher informational capacity and structural persistence; the variable-factor model, while adaptive, suffers from reduced signal strength and inconsistent factor exposure. Overall, the evidence favors the fixed-15-PC strategy in terms of cumulative profitability and implementation robustness over the long horizon.

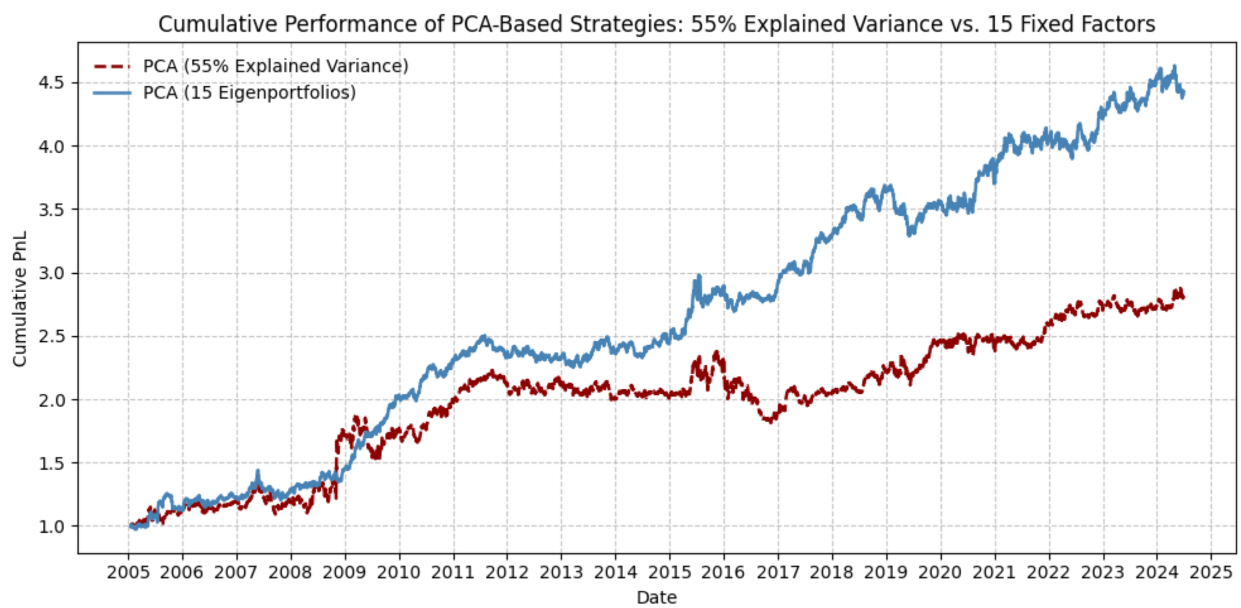

Figure 11. Cumulative profit and loss (PnL) comparison between two PCA-based strategies from 2005 to 2024. The first strategy selects a variable number of principal components to explain 55% of total variance (dashed red line), while the second uses a fixed set of 15 eigenportfolios (solid blue line).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

Figure 12 presents the evolution of the number of principal components required to explain 55% of the cross-sectional return variance, along with the historical volatility of the CSI300 index from 2005 to 2024. The PCA is conducted using a 252-day rolling window, and the number of retained components is determined endogenously at each time point. Market volatility is proxied by the annualized standard deviation of daily log returns over the same horizon.

A clear inverse relationship is observed between market volatility and the dimensionality of the return space. During periods of elevated volatility—such as the 2008 global financial crisis, the 2015 Chinese equity market correction, and the COVID-19 shock in early 2020—the number of components needed to capture 55% of the total variance declines significantly. This suggests that systemic shocks tend to compress the return space, concentrating variation into fewer dominant latent factors. Such dynamics are consistent with the findings of Barunik and Křehlík (2018), who document that common factor dominance increases during high-volatility regimes, and with the theoretical implications of factor models under time-varying covariance structures (Chamberlain & Rothschild, 1983; Connor & Korajczyk, 1988).

In contrast, during tranquil market periods—such as 2012–2014 and 2021–2023—the required number of principal components rises, reflecting greater cross-sectional heterogeneity and weaker factor commonality. This implies that more latent sources of variation are needed to explain the same fraction of return dispersion, which aligns with the literature on volatility-induced changes in market integration (Diebold & Yılmaz, 2014).

These results highlight the importance of accommodating regime-dependent factor complexity in the construction of PCA-based statistical arbitrage strategies. Fixed-dimensional approaches may suffer from misalignment with the evolving structure of market risk, potentially leading to underfitting in high-volatility regimes and overfitting in low-volatility periods. A dynamically adjusted factor model—anchored on variance-explained thresholds—offers a more robust and economically meaningful framework for extracting tradable signals in structurally evolving markets.

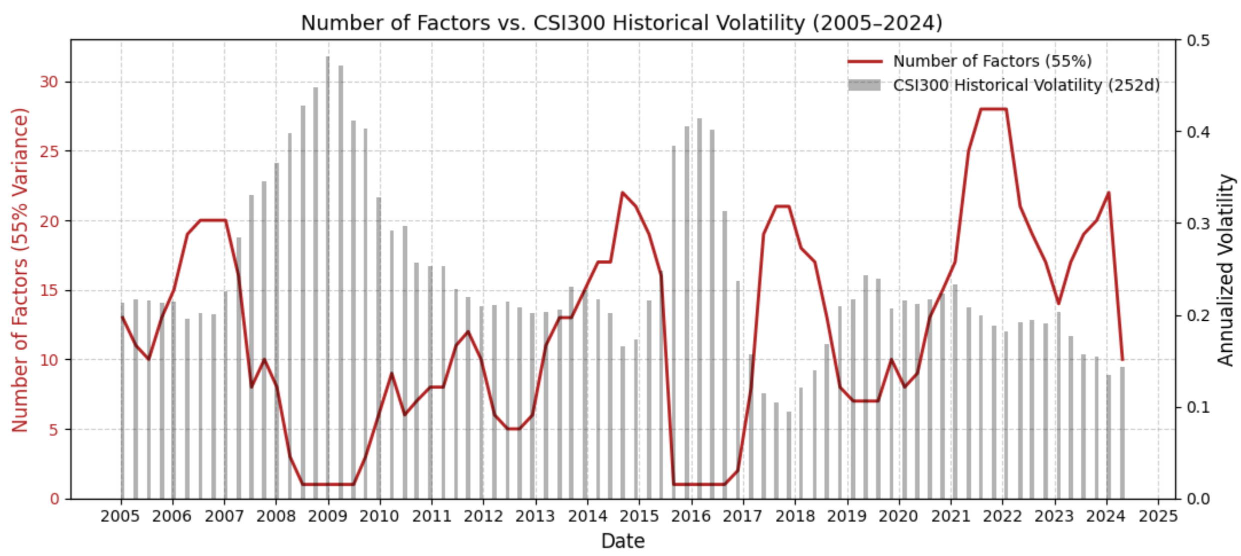

Figure 12. Number of principal components required to explain 55% of return variance (red line) and CSI300 historical volatility (blue bars) from 2005 to 2024. The number of factors fluctuates significantly across market regimes, often increasing during volatile periods.

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. Following Avellaneda and Lee (2010), we adopt a trading-time framework in which returns are rescaled by contemporaneous trading activity, effectively transforming the signal space from uniform calendar time to liquidity-adjusted time. Formally, the modified 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 adjustment implicitly embeds a market microstructure-informed filter, penalizing large returns that occur on high volume and amplifying those occurring on light volume. The underlying intuition is rooted in microstructure theory: price changes on substantial volume are more likely to reflect fundamental information, whereas those on low volume are more prone to transitory liquidity shocks or behavioral noise. This perspective is supported by empirical findings from Chordia, Roll, and Subrahmanyam (2005), who document that high-volume transactions exhibit lower reversal probability and faster convergence to informational equilibrium. Similarly, Lee and Swaminathan (2000) show that reversals are statistically stronger following low-volume moves, suggesting conditional predictability tied to liquidity conditions.

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. The PCA strategy exhibited a moderate improvement: the full-sample annualized return increased to 6.57% with a Sharpe ratio of 0.64. However, the performance gains were highly episodic, concentrated in 2005–2006 and 2017–2019, and notably deteriorated during market stress periods such as 2020 and 2023. These results suggest that latent factor structures derived from PCA may become unstable when subjected to volume-based normalization, particularly under regime shifts in market conditions or trading behavior.

This aligns with concerns raised by Khandani and Lo (2007), who highlight the fragility of PCA-based statistical arbitrage models during periods of structural change. Since PCA implicitly assumes stationarity in the return covariance structure, dynamic reweighting through volume adjustments can inadvertently amplify estimation noise or suppress meaningful signals during instability.

In contrast, the synthetic ETF-based strategy exhibited a more robust and consistent improvement under the trading-time specification. The Sharpe ratio rose to 0.75, with notable gains in downside risk measures such as Sortino, Calmar, and Conditional VaR. Performance remained resilient across both tranquil and volatile subperiods, including 2009–2010 and 2015–2019, indicating that the combination of sector-level exposure and volume-based rescaling provides a structurally stable and liquidity-aware framework for signal generation.

Overall, the results support the hypothesis that trading volume carries relevant state information that can improve the performance of mean-reversion signals—particularly when applied to strategies anchored on economically interpretable risk factors. While the efficacy of volume adjustments in PCA models appears conditional and fragile, synthetic ETF strategies demonstrate robust improvements, reinforcing the importance of model structure and factor interpretability in liquidity-sensitive signal design.

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

Table 7. 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.

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

Table 8. 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 provides a comparative assessment of the cumulative PnL performance of mean-reversion strategies constructed from 15 fixed principal components (PCs), implemented under two distinct temporal frameworks: actual (calendar) time and trading time. The latter involves a volume-adjusted return specification as formalized in Equation (20), whereby raw returns are rescaled inversely by contemporaneous trading volume to account for liquidity-adjusted price evolution.

The results demonstrate a persistent and economically meaningful underperformance of the trading-time strategy relative to its actual-time counterpart over the sample period from 2005 to 2024. While both approaches yield positive cumulative returns, the trading-time version generates systematically lower PnL trajectories, with a particularly noticeable divergence emerging during the post-2015 period. This suggests that the information content embedded in volume-normalized returns may be suboptimal in capturing the latent mean-reverting structure exploited by PCA-based factor strategies.

Several theoretical considerations may explain this empirical observation. First, PCA eigenportfolios are inherently constructed to extract dominant components of the covariance structure, and as such, the residual variation they target is already smoothed and cross-sectionally diversified. Adjusting these factor-based signals further by volume may result in signal attenuation, effectively suppressing profitable reversals that would otherwise be tradable under classical return formulations. This aligns with findings from statistical microstructure models (e.g., Lo and Wang, 2000; Hasbrouck, 2009), which emphasize that volume-adjusted time is most effective in capturing high-frequency liquidity shocks and transitory price pressures at the single-stock level, rather than in multi-asset factor structures.

Moreover, empirical studies in the statistical arbitrage literature (Avellaneda and Lee, 2010; Gatev, Goetzmann, and Rouwenhorst, 2006) highlight that factor-based signals benefit from robustness and persistence across time, features that may be compromised when introducing nonlinear adjustments such as inverse volume weighting. The observed deterioration in performance under the trading-time specification thus points to a potential mismatch between the temporal resolution of volume scaling and the structural timescale of factor-based reversion.

In sum, while the trading-time framework may enhance performance in certain liquidity-sensitive strategies, its utility appears limited in the context of principal component-based statistical arbitrage. For such latent structure models, traditional return measures continue to provide superior signal extraction efficiency and capital allocation performance.

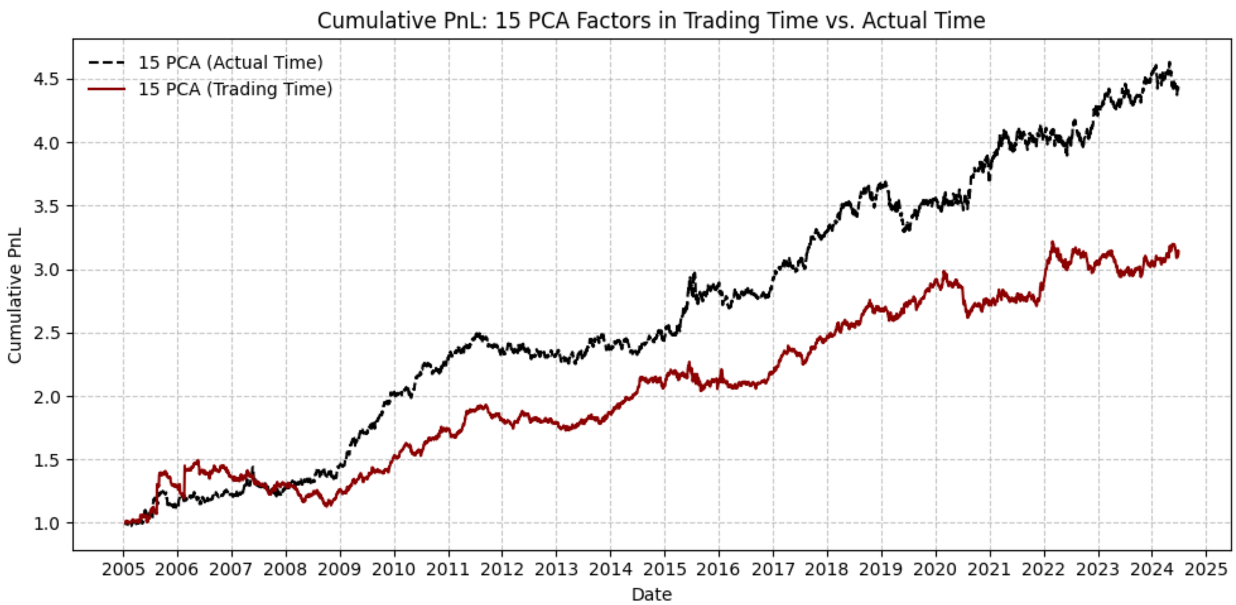

Figure 13. Comparison of cumulative PnL for PCA-based strategies using 15 fixed principal components under trading-time and actual-time frameworks. The trading-time approach adjusts returns based on recent volume activity to account for liquidity conditions. While both approaches exhibit mean-reversion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

Figure 14 illustrates the cumulative PnL performance of two volume-adjusted statistical arbitrage strategies—one employing synthetic ETF signals and the other based on 15 fixed PCA factors—under a trading-time framework over the period 2005–2024. The empirical evidence reveals that the ETF-based strategy significantly outperforms the PCA-based strategy in terms of return persistence, drawdown resilience, and long-term trend stability. While both strategies apply the same trading-time transformation—modulating signals based on recent trading volume—the magnitude and consistency of performance improvements differ markedly.

The superior performance of the ETF strategy stems from the nature of its underlying factor construction. Synthetic ETF portfolios, composed of capitalization-weighted sector indices, are inherently smoother, more stable, and more aligned with macroeconomic sector movements than the statistical eigenportfolios derived from PCA. When returns are rescaled by recent trading intensity (i.e., trading time), the ETF factors retain their interpretability and robustness, as they are less exposed to microstructural noise and idiosyncratic shocks.

Conversely, the PCA strategy—despite its dimensionality reduction power—relies heavily on residual-based cross-sectional variations, which are more susceptible to short-term volatility clustering and liquidity shocks. As such, when the signal is scaled by volume, the effectiveness of PCA-based mean-reversion weakens. This confirms findings in Avellaneda and Lee (2010), who observed that PCA strategies do not benefit significantly from the trading-time framework, in contrast to factor-based strategies derived from economically meaningful portfolios.

Moreover, this result echoes the theoretical rationale advanced in market microstructure literature. For example, Lo and Wang (2000) propose that volume-based clocks capture variations in informational flow and liquidity conditions more accurately than calendar time, offering a more appropriate temporal resolution for return predictability. Similarly, Chordia et al. (2001) highlight that volume reflects the underlying dynamics of informed trading and inventory risk management, both of which impact the profitability of short-horizon strategies.

Taken together, the results in Figure 19 suggest that the effectiveness of trading-time adjustments is highly dependent on the structural stability of the signals themselves. Factor strategies grounded in real-sector aggregates (such as ETFs) demonstrate greater robustness and adaptability to liquidity-aware timing mechanisms, whereas statistically derived PCA portfolios are more prone to performance degradation under microstructural frictions. These findings underscore the importance of aligning factor design with market liquidity regimes when implementing high-frequency or short-term arbitrage strategies.

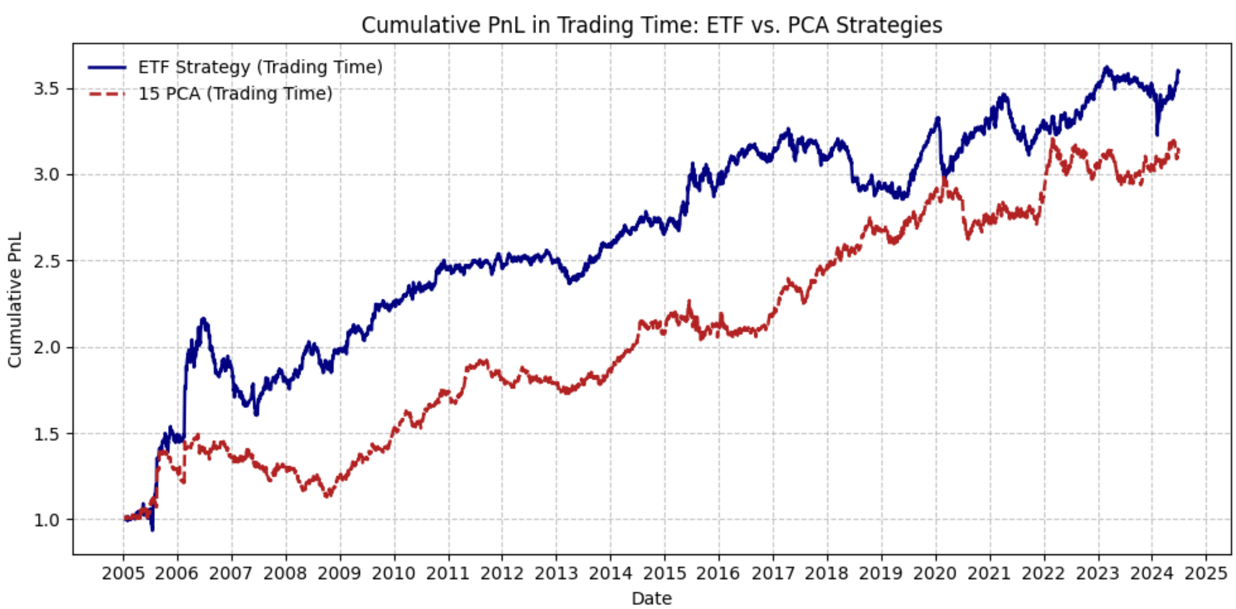

Figure 14. This figure compares the cumulative profit-and-loss (PnL) of two statistical arbitrage strategies implemented under the trading-time framework: one using synthetic ETF-based factors and the other using 15

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.

6.5 Extreme Periods Comparison and Analysis

To evaluate the behavior and robustness of statistical arbitrage under different market regimes, we focus on three structurally distinct periods: the Global Financial Crisis (GFC, 2007–2008), the Chinese Bull Market (BNB, 2014–2016), and the COVID-19 pandemic (2020–2022). These regimes represent systemic risk, directional momentum, and macroeconomic uncertainty, respectively. Table 9 and Table 10 report the annualized performance of PCA-based and Synthetic ETF-based strategies across these periods. The goal is to assess not only the comparative effectiveness of the two strategies but also how statistical arbitrage behaves under varying market stressors.

During the Global Financial Crisis, both strategies delivered comparable returns with limited drawdowns, underscoring their robustness in crisis conditions. The PCA strategy yielded an average return of 9.39% and a Sharpe ratio of 0.79, while the ETF-based strategy posted a nearly identical return (9.38%) but outperformed marginally in risk-adjusted metrics (Sharpe 0.94, Calmar 0.58). This relative stability in performance, despite heightened systemic volatility, suggests that both residual-based and structural arbitrage signals are effective under market distress. These findings are consistent with prior literature (Gatev et al., 2006; Avellaneda & Lee, 2010), which posits that mispricings tend to persist longer during crises, thereby enhancing arbitrage opportunities.

In stark contrast, the Chinese Bull Market provided a favorable backdrop for sector-based strategies. The ETF strategy exhibited superior performance across all dimensions, achieving an average annual return of 13.36%, a Sharpe ratio of 2.05, and a Calmar ratio of 2.55. The PCA strategy, while still profitable (Mean 7.51%), showed relatively subdued efficiency with a Sharpe of 0.88 and Calmar of 0.77. This disparity reflects the ETF strategy's ability to exploit macro-driven sector rotation and trending behaviors—features particularly pronounced in a sustained bull market. Momentum amplification and sector concentration effects likely enhanced the predictability of structural signals, supporting the findings of Kelly et al. (2019) on the efficacy of sector-based tilts in high-dispersion regimes.

The performance landscape during the COVID-19 pandemic was markedly different. Although both strategies remained profitable, with mean returns of 6.22% (PCA) and 4.89% (ETF), the PCA strategy demonstrated greater resilience in risk-adjusted terms. Its Sharpe ratio (0.77 vs. 0.81) was on par with the ETF strategy, but the PCA approach exhibited lower drawdowns, more stable CVaR profiles, and a higher Omega ratio. These results suggest that residual-based signals were more adaptive to the heightened volatility and rapid regime shifts observed during the pandemic, possibly due to their ability to capture evolving co-movements without being anchored to sectoral definitions. This is in line with recent empirical work (Liu et al., 2021; Ma et al., 2021), which highlights the limitations of structural factor models in environments with unstable beta exposures and frequent macroeconomic shocks.

From a comparative perspective, the ETF strategy excels in directional and trend-driven markets such as BNB, whereas the PCA strategy offers stronger protection and stability in crisis-like or uncertain conditions such as GFC and COVID-19. The divergence underscores a key insight: statistical arbitrage is not universally effective but is contingent upon the interaction between strategy design and macro regime. In expansionary periods, structural dispersion across sectors creates exploitable inefficiencies that favor capitalization-weighted approaches. Conversely, in turbulent or fragmented markets, model-free approaches such as PCA provide enhanced robustness by dynamically adapting to cross-sectional relationships.

In conclusion, the integration of regime-level analysis and strategy comparison highlights the complementary nature of PCA and ETF-based statistical arbitrage. While both can be profitable across regimes, their performance profiles are distinctly regime-dependent. This underscores the practical need for dynamic model selection or ensemble designs that condition arbitrage execution on macro state variables, thereby enhancing out-of-sample robustness and portfolio resilience.

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
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

Table 9. 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.

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
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

Table 10. 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 developed and evaluated market-neutral statistical arbitrage strategies constructed using Principal Component Analysis (PCA) and grounded in the principles of pairs trading. While the empirical results demonstrate the potential of PCA-based factor models in capturing mean-reversion opportunities, it is well recognized that the performance of such strategies can be significantly influenced by key modeling choices and parameter settings.

In practice, aspects such as the number of retained principal components, the cumulative explained variance threshold, the residual threshold for signal generation, and the length of the rolling estimation window can all exert material effects on both returns and risk characteristics.

To systematically examine the extent to which these parameters affect strategy outcomes, this chapter conducts a series of sensitivity analyses. The goal is to provide a rigorous and comprehensive understanding of the model's robustness across alternative specifications and to identify optimal configurations that balance return, risk, and stability.

Section 7.1 begins with an investigation into the number of retained PCA components. By varying the number of eigenportfolios (e.g., 3, 5, 10), we assess the impact on cumulative performance, return volatility, and drawdown behavior. The results are empirically compared to identify whether fewer components suffice to extract meaningful mean-reversion signals or whether higher-dimensional models capture additional alpha-generating structure.

In Section 7.2, we shift focus to the proportion of cumulative variance explained by the selected components. Using benchmarks such as 50%, 60%, and 70%, we analyze the trade-offs between model simplicity and explanatory power, and identify which variance thresholds tend to yield more stable and profitable outcomes.

Section 7.3 explores the sensitivity of the strategy to residual or signal thresholds used for trade triggering. Since these thresholds determine how frequently the model enters positions, they directly influence both turnover and profit distribution. We evaluate how conservative versus aggressive entry rules affect performance and risk metrics, with an emphasis on robustness across market conditions.

Lastly, Section 7.4 addresses the effect of the rolling window length used in PCA estimation. By experimenting with both shorter and longer historical lookback periods, we assess the model's ability to adapt to changing market dynamics and determine whether more responsive or more stable windows are preferable.

Together, these four dimensions of sensitivity testing offer valuable insight into the structural dependencies of PCA-based statistical arbitrage. This analysis not only enhances our understanding of how parameter choices influence strategy behavior but also provides practical guidance for calibrating such models under real-world constraints. In doing so, we aim to bridge the gap between theoretical model design and applied portfolio implementation, with clear implications for risk control, signal precision, and capital deployment in live trading environments.

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 principal components, ranging from 5 to 45, as risk factors. Several notable patterns emerge from the empirical results, aligning with prior literature on factor modeling and statistical arbitrage (Avellaneda & Lee, 2010; Kelly, Pruitt, & Su, 2019; Bartkoviak et al., 2022).

First, there exists a non-monotonic relationship between the number of components and average annualized returns, as shown in Figure 15, 16 and Figure 16 shows the same conclusion as in Figure 10. The strategy's return initially increases from approximately 7.9% when using 5 components to a peak of 9.9% at 10 components, but subsequently declines to only 1.6% at 45 components. This pattern suggests the presence of an optimal factor dimensionality—wherein a limited number of principal components can effectively capture dominant market and sector-level variations, thus isolating high-quality residuals for mean-reversion trading. Beyond a certain threshold, however, the inclusion of additional components begins to introduce higher-frequency noise, potentially eroding alpha and amplifying overfitting risks (Laloux et al., 2000).

The Sharpe ratio and t-statistics exhibit a similar trend, with the Sharpe ratio peaking at 0.9436 under the 15-component specification. Statistical significance, as measured by the t-statistic, is also strongest around 10–15 components (e.g., 2.93 and 4.08, respectively). This further supports the interpretation that the middle range of components offers the most robust signal-to-noise ratio. As the number of components increases beyond this range, statistical significance wanes, and t-values converge toward insignificance, particularly when exceeding 35 components.

In terms of risk metrics, volatility—as measured by standard deviation—tends to decrease with more components. For example, the lowest volatility (0.0576) occurs at 45 components. However, this reduction in volatility is accompanied by a significant decline in returns and Sharpe ratios, indicating that risk control in such cases may come at the cost of signal strength. Moreover, the tail risk profile, reflected by Value-at-Risk (VaR), and Conditional VaR (CVaR), reveals additional trade-offs. CVaR improves as the number of components increases, the corresponding decline in returns offsets this benefit from a performance standpoint.

Drawdown-related metrics present further evidence of this trade-off. The maximum drawdown is lowest at higher component counts, but due to reduced return levels, the Calmar ratio drops to around 0.12 at 45 components. In contrast, using 15 components yields a more balanced profile, with a Calmar ratio of 0.5061—indicating better return per unit of drawdown. These findings align with earlier work by Avellaneda and Lee (2010), who noted that PCA-based mean-reversion strategies tend to perform best when focused on a subset of dominant components rather than the full eigenstructure.

From an economic standpoint, the empirical patterns confirm the notion that factor over-expansion can degrade performance. Early principal components typically capture meaningful co-movements related to market, sector, and style effects. These are precisely the types of structures that residual-based strategies are designed to exploit. When too many components are used, the model begins to explain away idiosyncratic variation and introduces spurious factors, weakening the mean-reversion signal (Chae & Kim, 2020; Krause & Calliess, 2024).

These results carry several practical implications. First, there is a clear risk–return trade-off associated with component selection: using around 10 to 15 components appears to balance risk and return most effectively. Second, the optimal number of components may vary over time depending on market regime, suggesting the potential benefit of rolling or adaptive PCA estimation methods (Bartkoviak et al., 2022). Finally, strategies based on a smaller number of components are computationally more efficient, making them particularly attractive in high-frequency or large-scale settings.

In summary, PCA dimensionality plays a pivotal role in shaping the performance of mean-reversion strategies. The evidence suggests that a moderate number of components—typically between 10 and 15—achieves the best trade-off between explanatory power and signal degradation. Including too few components leads to underfitting, while excessive inclusion introduces noise, reducing both statistical significance and economic viability.

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 (95%)	-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 (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

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

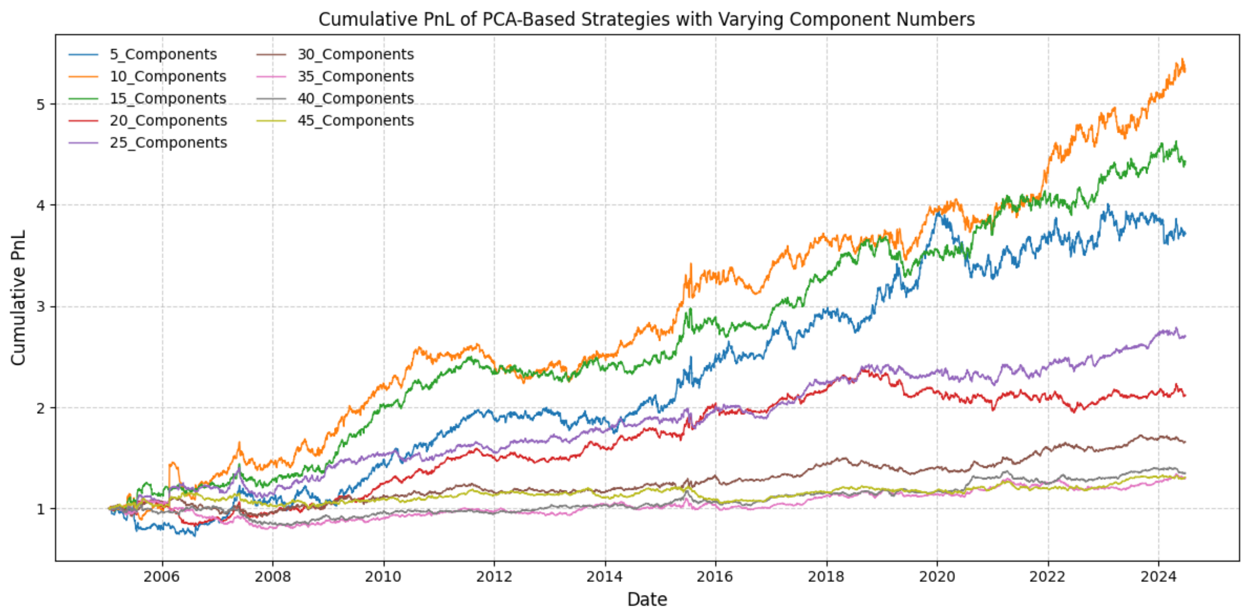

Figure 15. Cumulative PnL of PCA Strategies with Varying Components. This figure shows the cumulative performance of PCA-based strategies using 5 to 45 components. Strategies with 10–15 components achieve the highest returns, while those with more components exhibit weaker performance, suggesting diminishing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

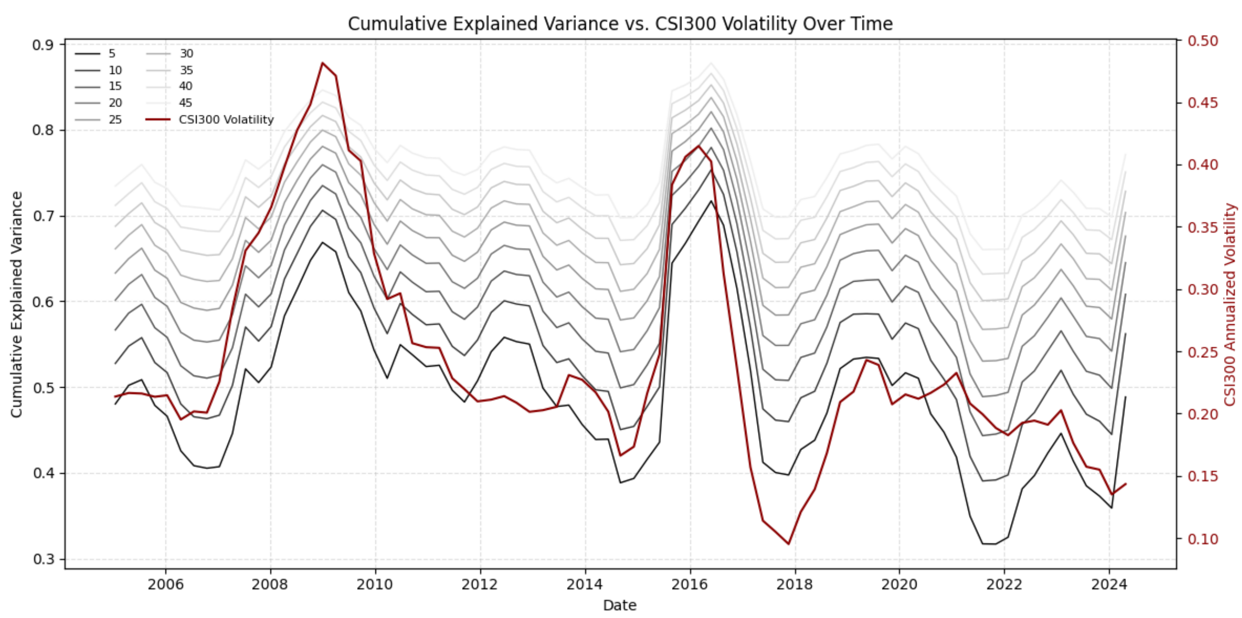

Figure 16. Cumulative Explained Variance of PCA Factors vs. CSI300 Volatility (2005–2024). This figure compares the cumulative explained variance of PCA models using different numbers of components (ranging from 5 to 45) with the rolling annualized volatility of the CSI300 index. The left y-axis represents 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.

7.2 Explained Variance Thresholds and Strategy Performance

Table 12 presents the annualized performance statistics for PCA-based mean-reversion strategies under varying explained variance thresholds, ranging from 40% to 75%, as shown in Figure 16, 17, and Figure 17 shows the same conclusion as Figure 12. These thresholds determine how many principal components are retained in the factor construction process, thereby shaping the balance between model parsimony and completeness. The results reveal a nonlinear relationship between explained variance and performance, highlighting important implications for strategy design, model robustness, and signal reliability.

From the table, we observe that the mean return decreases almost monotonically as the explained variance threshold increases. While the strategy achieves its highest return at 40% (11.50%) and 45% (10.96%), the performance drops substantially as the threshold rises to 75% (2.56%). This suggests that early principal components capture the most valuable co-movement information—likely associated with market-wide or sectoral factors—while additional components, added at higher thresholds, introduce increasing amounts of idiosyncratic noise or transient structures that dilute the core signal. This aligns with the findings of Laloux et al. (2000) and Plerou et al. (2002), which caution that beyond a certain point, eigenvalues reflect random effects rather than economically meaningful structure.

The Sharpe ratio and Sortino ratio, key measures of risk-adjusted performance, exhibit a convex pattern. Both metrics peak around the 70% threshold (Sharpe = 0.6163, Sortino = 0.8938), suggesting that this level of explained variance offers the most favorable trade-off between return and risk. The corresponding t-statistic (2.6678) and z-statistic confirm the statistical significance of excess returns in this configuration. Interestingly, although the mean return is higher at 40%–50%, the elevated volatility and tail risk at those thresholds undermine the strategy's risk-adjusted appeal.

These observations are consistent with the interpretation that moderate levels of explained variance (60%–70%) allow the strategy to capture systematic drivers of return without overfitting to noise, providing a more stable basis for residual-based trading signals (Avellaneda & Lee, 2010; Bartkoviak et al., 2022).

A steady decline in standard deviation is observed as the explained variance increases—from 0.3147 at 40% to 0.0609 at 75%. Likewise, Value-at-Risk (VaR) and Conditional VaR (CVaR) improve, with CVaR tightening from –0.0279 (40%) to –0.0098 (70%). However, this apparent improvement in downside risk is not necessarily beneficial, as it is accompanied by a collapse in returns and lower Calmar and Sterling ratios. These findings underscore that risk reduction is meaningful only when accompanied by sufficient returns—otherwise, the strategy becomes overly conservative and economically inefficient.

Maximum drawdowns decrease sharply with higher explained variance (from –59.8% at 40% to –15.9% at 70%), again reflecting increased diversification or noise suppression. However, the Calmar ratio, which measures return relative to drawdown, is highest around 70% (0.2718), not at the lowest drawdown point. This supports the notion that 70% strikes an efficient balance: drawdowns are minimized without sacrificing alpha.

The results from Table 12 collectively suggest that retaining principal components until approximately 70% of total variance is explained offers the most robust configuration for PCA-based mean-reversion strategies. This threshold produces the best combination of return, risk, and statistical significance, while also controlling for downside exposure and noise. Lower thresholds may deliver higher returns but at the cost of excessive volatility and drawdowns. Higher thresholds, though stabilizing, reduce alpha potential by incorporating too many marginal factors. These findings align with empirical finance literature that emphasizes the importance of parsimonious yet comprehensive factor modeling (Jolliffe, 2002; Chae & Kim, 2020).

In practice, targeting an explained variance threshold between 60% and 70% may serve as a principled default setting, subject to further adjustment based on rolling performance, market regime, or asset class characteristics.

Explained Ratio	**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 (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

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

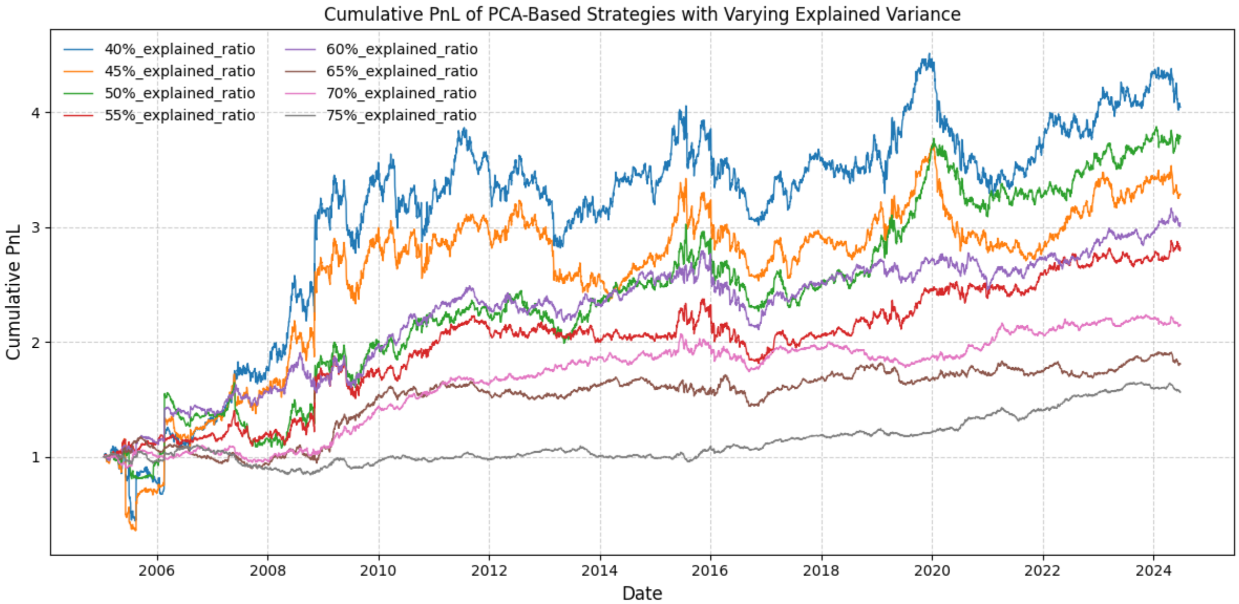

Figure 17. Cumulative PnL of PCA Strategies by Explained Variance Threshold. This chart shows strategy performance under different PCA explained variance levels (40%–75%). Lower thresholds yield higher returns with greater volatility, while higher thresholds reduce risk but dampen returns. The 70% level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

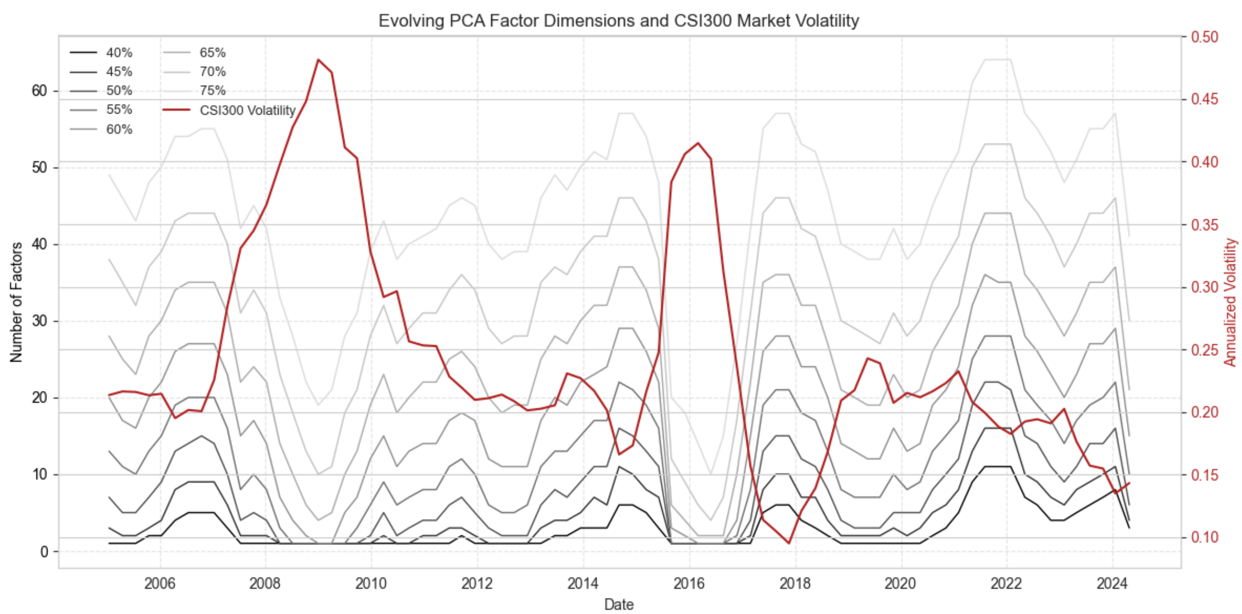

Figure 18. Temporal evolution of PCA factor dimensionality across different explained variance thresholds (40%–75%) and its relationship with CSI300 index volatility. The number of principal components required to reach a given explained variance level varies significantly over time, reflecting shifts 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 presents the annualized performance metrics of the PCA-based mean-reversion strategy under various combinations of trade entry and exit thresholds, defined as multiples of the residual standard deviation. The cumulative PnL shows in Figure 19. These thresholds determine when a position is opened (entry) and subsequently closed (exit) based on deviations from the equilibrium level. The table reflects how these parameter choices affect trade frequency, holding duration, and the overall risk–return profile of the strategy.

Lower entry thresholds (e.g., 1.0) generate more frequent trading signals by capturing smaller deviations. While this approach increases exposure to potential arbitrage opportunities, it can also introduce more noise and false positives, especially during volatile periods. Conversely, higher entry thresholds (e.g., 1.5) act as filters, focusing on larger and potentially more reliable mispricings. Although this may reduce the number of trades, the selected signals tend to be stronger and often exhibit better risk-adjusted returns.

Tighter exit thresholds (such as 0.3 or 0.5) lead to quicker position closure, locking in gains as soon as partial mean-reversion is achieved. This often limits drawdowns but may miss further reversion profits. Looser exit thresholds (e.g., 0.75) allow trades to stay open longer, capturing more of the reversion movement. However, this may also increase exposure during adverse market swings.

Among all combinations, the (1.0, 0.75) configuration achieves the highest Sharpe ratio (1.0835) and a strong t-statistic (4.6901), suggesting that a moderate entry paired with a relatively lenient exit threshold provides favorable risk-adjusted performance. The configuration denoted as (Main)—(1.25, 0.5, 0.75)—shows similarly robust performance, with a Sharpe ratio of 0.9436 and moderate drawdown (–0.1107), yielding a high Calmar ratio (0.8021). This setting is also consistent with practices in the literature, where slightly more conservative entry triggers help reduce overtrading without forgoing return potential.

By contrast, combinations like (1.5, 0.3) or (1.5, 0.5) yield lower Sharpe ratios (0.6650 and 0.6395, respectively) and deeper drawdowns. These settings represent high barriers to trade initiation paired with tight exits, which often leads to infrequent and less profitable trades, missing broader reversions.

The results reinforce that threshold selection critically shapes the strategy’s behavior. Entry thresholds that are too low tend to overreact to noise, while overly conservative ones may miss valuable trades. Likewise, setting an exit threshold asymmetrically lower than the entry threshold—as in the main configuration—offers a practical means of securing gains before full convergence, improving the strategy's downside protection and capital efficiency.

Adaptive threshold frameworks, which adjust based on volatility or other market conditions, could further improve robustness, especially during regime shifts. The findings confirm that moderate-to-high entry thresholds (e.g., 1.25–1.5) combined with exit levels around 0.75 often yield superior Sharpe ratios and manageable drawdowns, striking an effective balance between return capture and noise filtering.

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 (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

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

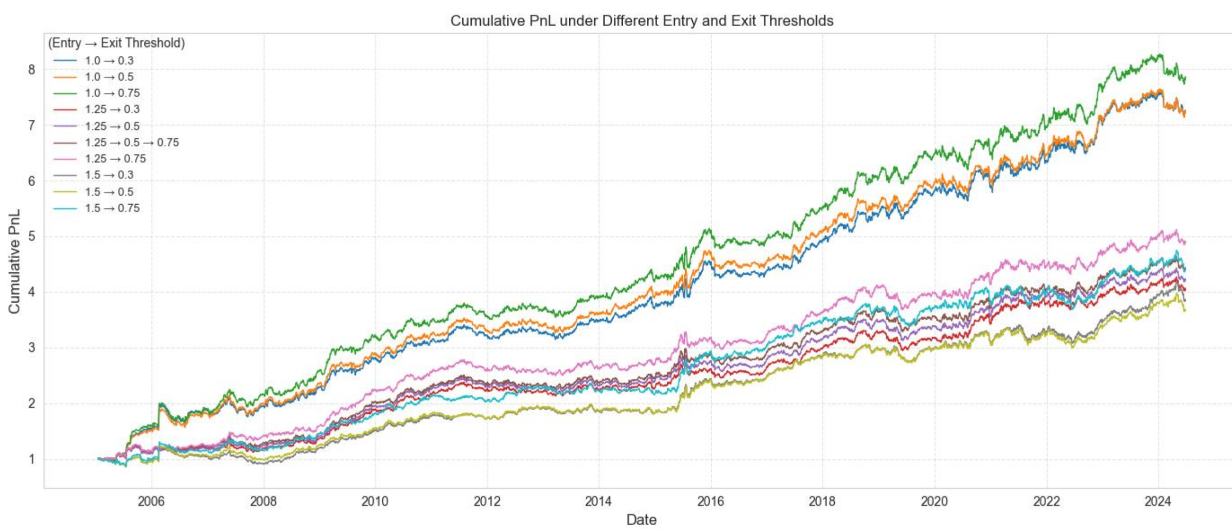

Figure 19. This figure compares the cumulative PnL of PCA-based mean-reversion strategies under different combinations of entry and exit thresholds, defined as standard deviation multiples of residual signals. Lower entry thresholds (e.g., 1.0) generate more frequent trades, while looser exit thresholds (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 strategy performance when the length of the rolling estimation window for the PCA-based residuals ranges from 30 to 120 trading days, as shown in Figure 20. The reported annualized metrics—including mean return, t-statistic, and Sharpe ratio—exhibit a clear pattern: shorter windows (30–50 days) tend to yield higher volatility and relatively weak t-statistics, whereas longer windows (e.g., 100–120 days) frequently produce better risk-adjusted returns and more statistically significant alpha. Several insights are worth highlighting.

Short estimation windows (30–50 days) respond more quickly to recent market conditions and shifts in factor loadings. While this reactivity enables the model to capture transient opportunities, it also introduces considerable estimation noise. For instance, the 30-day window delivers a low mean return (0.0083) and an insignificant t-statistic (0.43), indicating that short samples may fail to identify meaningful principal components and residual structures.

Conversely, medium to long windows (60–120 days) improve estimation stability and reduce spurious noise. The 60-day window—often used in the literature—already delivers robust Sharpe and t-stat values, while performance continues to improve as the window extends. The 120-day window achieves the best Sharpe ratio (1.44) and highest t-statistic (6.22), suggesting that longer histories enable more consistent signal extraction.

Longer windows smooth out short-term co-movement fluctuations and lead to more stable PCA factor loadings. This structural stability increases the reliability of the residual component used for mean-reversion signals. Prior studies (e.g., Bartkoviak et al., 2022; Kelly et al., 2021) have emphasized that reducing noise in factor estimation is critical for improving signal precision and strategy performance.

However, as Chae and Kim (2020) note, very long windows may also lag during regime shifts. If correlations change rapidly due to macroeconomic shocks or sectoral rotation, extended windows may delay the model's responsiveness and reduce alpha capture in the short term.

The 60-day window, adopted as the baseline in this study, reflects a balance between statistical robustness and adaptability. This length is widely used in practice and supported by empirical work (Avellaneda & Lee, 2010; Gu et al., 2021), making it a pragmatic default.

Nonetheless, it may be beneficial to dynamically adjust the window length according to prevailing market volatility. For instance, expanding the window during stable periods and shortening it during high-volatility regimes may offer an effective compromise between noise reduction and responsiveness.

Table 14 highlights the central role of estimation window length in shaping the effectiveness of PCA-based mean-reversion strategies. While shorter windows capture recent changes more quickly, they suffer from instability and weaker statistical significance. Longer windows, especially those between 80 and 120 trading days, offer stronger Sharpe ratios and more reliable signal extraction. In practice, a rolling window between 60 and 100 days often provides the optimal balance between robustness and adaptability.

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
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 (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 14. 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.

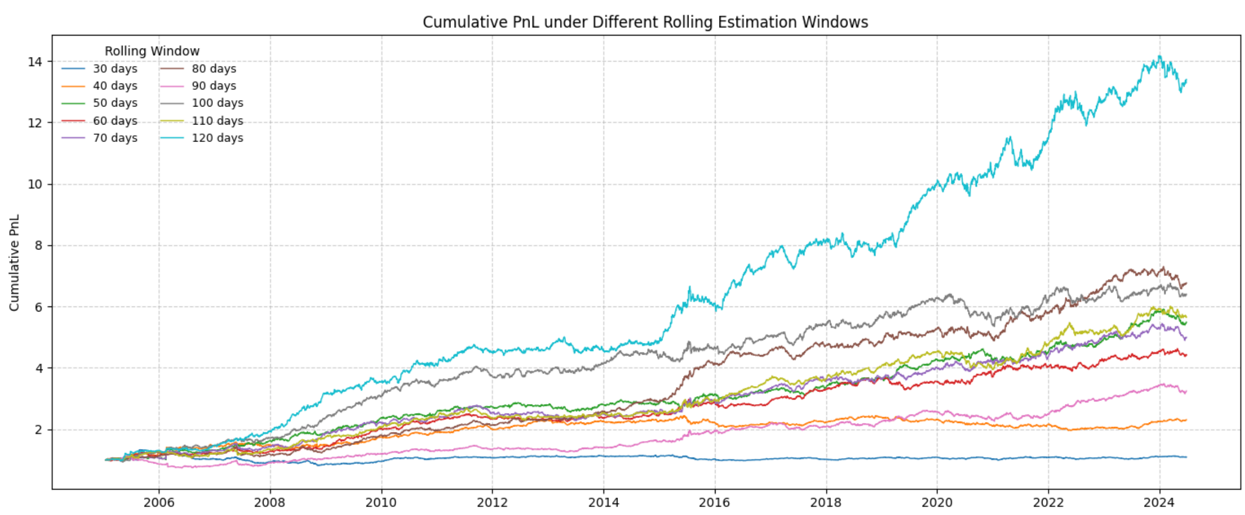

Figure 20. This figure illustrates the cumulative PnL of a PCA-based mean-reversion strategy under various rolling estimation windows, ranging from 30 to 120 trading days. Longer windows (e.g., 100–120 days) tend to yield smoother and more profitable trajectories, suggesting that more stable factor estimates 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 presented a comprehensive investigation into pairs trading and broader statistical arbitrage strategies that rely on mean-reversion. Central to the analysis has been the decomposition of individual stock returns into systematic and idiosyncratic components, the latter serving as residuals in a market-neutral framework. We compared two principal methods for identifying the systematic portion of returns: (i) PCA-based factor extraction, which yields latent “eigenportfolios” from the correlation matrix of returns, and (ii) ETF-based factor modeling, where we regress stocks on one or more sector ETFs.

Empirical tests demonstrate that the choice of factor construction materially affects the dynamics of the residual process and, consequently, the profitability of the trading strategy. Specifically, PCA-derived residuals can better isolate hidden co-movements and are typically less biased toward large-cap stocks, while ETF-based approaches lend themselves to more intuitive sector interpretation and simpler hedging mechanics (e.g., going long or short the corresponding ETF). Both methods show robust mean-reversion signals, although the optimal number of PCA factors or the specific ETF coverage can vary over time in response to shifting market conditions. Notably, the thesis finds that dynamic PCA strategies—those that retain enough principal components to explain a fixed share of variance—can adapt to changing cross-sectional return structures, enhancing risk-adjusted performance.

A key feature of the methodology is the systematic design of trading signals, generated when residuals deviate significantly from their rolling 60-day mean. By entering and unwinding positions only at specific thresholds, the approach effectively exploits short-horizon mispricings while limiting overfitting. Transaction costs are taken into account via a realistic slippage assumption, underscoring that trading frictions can meaningfully affect net returns and must be carefully controlled in practice. Additionally, the exploration of trading-time weighting reveals that integrating volume information at the daily frequency can improve signal robustness, especially during low-liquidity or volatile market phases.

A series of sensitivity analyses (Section 7) further highlights how strategy outcomes can shift with different rolling windows, threshold rules, and the number of PCA components retained. While a 60-day window and an entry threshold of ±1.25 standard deviations proved effective in capturing short-term reversions, alternative parameters may suit markets with different volatility regimes or liquidity profiles. Finally, stress-testing these strategies in turbulent market environments confirms that mean-reversion portfolios can experience temporary drawdowns but tend to recover promptly—especially when the number of dominant factors is small and systematic drivers are well captured.

In conclusion, this research underscores the benefits of a rigorous factor decomposition in pairs trading, whether via PCA or ETF-based approaches. It shows that residual-based signals can generate stable alpha under diverse market conditions, provided the models remain market-neutral and dynamically updated. Future work may extend these findings by incorporating machine learning methods to select or refine factor sets, incorporating fundamental data for more robust valuation signals, or developing adaptive threshold rules that respond to real-time changes in volatility and liquidity. Nonetheless, the core principle remains that capturing the right systematic factors—and thus isolating genuinely idiosyncratic mispricings—plays a decisive role in achieving consistent mean-reversion gains in a market-neutral setting.

References

Avellaneda, M. & Lee, J.-H. (2010). Statistical arbitrage in the U.S. equities market. Quantitative Finance, 10(7), 761–782.

Bartkoviak, M., Pietruczuk, L., & Syczewska, A. (2022). Cointegration-based vs. PCA-based pairs trading in high-frequency data. Finance Research Letters, 44, 102093.

Chae, J. & Kim, Y. (2020). Residual-based contrarian strategy and short-term return reversals. International Journal of Finance & Economics, 25(1), 44–59.

Cont, R. & Da Fonseca, J. (2002). Dynamics of implied volatility surfaces. Quantitative Finance, 2(1), 45–60.

Davis, G., Mallat, S., & Avellaneda, M. (1997). Adaptive greedy approximations. Constructive Approximation, 13, 57–98.

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

Han, L., Zhou, X., & Wang, Y. (2023). A clustering-based framework for cross-sectional factor strategies. European Journal of Operational Research, 309(2), 591–606.

Huij, J. & Lansdorp, S. (2017). Residual momentum. Financial Analysts Journal, 73(3), 37–52.

Jolliffe, I. (2002). Principal Component Analysis (2nd ed.). Springer.

Kelly, B., Moskowitz, T., & Pruitt, S. (2021). Instrumented principal component analysis. Journal of Econometrics, 222(1), 264–281.

Khandani, A. & Lo, A. (2007). What happened to the quants in August 2007? SSRN Working Paper.

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

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

Lehmann, B. (1990). Fads, martingales, and market efficiency. Quarterly Journal of Economics, 105(1), 1–28.

Litterman, R. & Scheinkman, J. (1991). Common factors affecting bond returns. Journal of Fixed Income, 1(1), 54–61.

Lo, A. & MacKinlay, A. (1990). When are contrarian profits due to stock market overreaction? Review of Financial Studies, 3(2), 175–205.
Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., & Stanley, H. (2002). A random matrix approach to financial cross-correlations. Physica A, 299(1-2), 175–180.
Pole, A. (2007). Statistical Arbitrage: Algorithmic Trading Insights and Techniques. Wiley.
Poterba, J. & Summers, L. (1988). Mean reversion in stock prices. Journal of Financial Economics, 22(1), 27–59.
Rotondi, F. and Russo, F., 2024. Machine Learning for Pairs Trading: a Clustering-based Approach. Available at SSRN 5080998.
Sarmento, E. & Horta, P. (2020). Pairs trading with clustering: A hybrid approach using PCA and unsupervised learning. Journal of Computational Finance and Economics, 4(2), 105–129.
Scherer, K.P. and 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, J. & He, W. (2022). Latent-factor-driven mean reversion in the Chinese equity market: A PCA–OU perspective. China Finance Review International, 12(3), 337–355.

Appendix A: Estimation of the residual process

This appendix outlines the procedure for estimating cointegration residuals as Ornstein–Uhlenbeck (OU) processes and for deriving standardized s-scores. While the method described here is by no means the most advanced or efficient, it is straightforward to implement and can readily be improved upon in practice.

A.1 OU Parameter Estimation for ETF Regressions

For clarity, we describe the steps for ETF-based regression; the PCA-based case follows a similar approach. Recall that for each stock SSS, we regress its daily returns on a corresponding sector ETF with the model:

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

where R_n^s and R_n^I are, respectively, the stock and ETF returns over 60 consecutive daily observations (chronologically ordered), and ϵn\epsilon_nϵn 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 $\theta > \frac{252}{30}$, meaning the characteristic reversion time is below 1.5 months. In that scenario, $b \;=\; e^{-\theta\, \Delta t}$ remains comfortably below 1. If b is too close to 1, the inferred mean-reversion speed is unacceptably slow, and the model is discarded for that particular 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:

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

In practice, it often proves beneficial to recenter the mean mmm by subtracting the cross-sectional average:

$$m \;\longleftarrow\; m \;-\; \langle m \rangle,$$

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:

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

This adjustment typically reduces systematic biases in the residuals and yields more robust trading signals in backtesting.