ICCK Journal of Applied Mathematics
http://dx.doi.org/10.62762/JAM.2025.721050

RESEARCH ARTICLE

The Application of Dual-Denoised Momentum Factors in Portfolio Management: A Study of ChiNext Stocks for Retail Investors

Yingnan Yi[1,2] **and Xue Gao**[3,*]

[1] Olin Business School, Washington University in St. Louis, St. Louis, Missouri 63130, United States
[2] Beijing QITANG Education Consulting Inc., Beijing, China
[3] JD Health International Inc., Beijing, China

Abstract

Momentum-based investment strategies face persistent challenges from noise contamination in financial time series, particularly within emerging markets such as China's ChiNext board. Traditional enhancement approaches typically address symptoms rather than underlying causes, resulting in continued vulnerability to market regime changes and performance deterioration. This study develops and evaluates a dual-denoising framework that integrates wavelet analysis for temporal noise reduction with isolation forest algorithms for cross-sectional anomaly detection. Our methodology employs comprehensive analysis of 1,200-1,300 ChiNext stocks spanning the 2015-2025 period, utilizing multiple machine learning architectures to assess portfolio performance across both long-only and long-short implementations. Key empirical findings demonstrate that: denoised momentum factors substantially improve predictive accuracy and portfolio performance; wavelet-based temporal denoising achieves remarkable effectiveness for turnover data with mean signal-to-noise ratio improvements of 6.4 dB; isolation forest cross-sectional anomaly detection provides critical risk management benefits by systematically eliminating stocks characterized by excessive trading activity and poor returns; and single-layer neural networks with isolation forest denoising achieve superior performance metrics, including 0.0199 monthly returns and a 0.2189 Sharpe ratio, outperforming more complex architectural alternatives. Addressing noise contamination at the data level represents a more fundamental solution than conventional enhancement techniques for momentum strategy limitations. Our findings establish systematic denoising as an effective approach for enhancing momentum-based investment strategies while maintaining practical implementability, with significant implications for both quantitative portfolio management and retail investor applications in emerging markets.

Keywords: momentum investing, wavelet denoising,

Submitted: 21 October 2025
Accepted: 28 October 2025
Published: 09 December 2025

Vol. 1, **No.** 3, 2025.
10.62762/JAM.2025.721050

***Corresponding author:**
✉ Xue Gao
gaoxue1207@163.com

Citation
Yi, Y. & Gao, X. (2025). The Application of Dual-Denoised Momentum Factors in Portfolio Management: A Study of ChiNext Stocks for Retail Investors. *ICCK Journal of Applied Mathematics*, 1(3), 129–144.

portfolio management, ChiNext market, machine learning, financial signal processing.

1 Introduction

The integration of machine learning techniques into portfolio management has transformed quantitative finance over the past decade [5, 24, 33]. While traditional asset pricing models have progressively incorporated sophisticated statistical approaches, the fundamental challenge of accurately forecasting stock returns persists due to inherent noise and complex market dynamics [15, 16].

Recent methodological advances demonstrate significant potential in capturing non-linear patterns and complex interactions within financial data [21, 30]. Tree-based methods and neural networks substantially outperform traditional linear models in return prediction tasks, while appropriate regularization techniques prove crucial for handling high-dimensional characteristic spaces [18, 23].

Despite these technological advancements, a critical challenge remains: the inherent noisiness of monthly return data [35]. Monthly returns contain substantial idiosyncratic variation that obscures underlying return-predictive signals. Although financial statement-based predictors benefit from audit processes and exhibit reduced noise levels, traditional momentum strategies remain particularly vulnerable to measurement error and data quality issues in return and volume data [1, 12].

The extensive literature on momentum factors documents pervasive momentum effects across global markets [2, 22]. However, conventional momentum strategies suffer from performance deterioration during market regime changes and exhibit significant volatility [3]. This vulnerability has prompted exploration of various enhancement techniques, including factor combinations [1] and time-series momentum approaches [31].

While denoising techniques have seen limited application in high-frequency trading contexts, their integration with medium-to-low-frequency momentum-based portfolio management remains underdeveloped. Wavelet analysis offers particular promise through its capacity to decompose signals across multiple time-frequency domains [13, 20, 28]. The mathematical foundations of wavelet theory provide powerful tools for financial time series signal processing, demonstrating significant potential for improving predictive performance of historical returns and trading volumes—key predictors in momentum-based investment strategies.

This research addresses a significant literature gap by investigating wavelet-denoised momentum factors in portfolio management, with a specific focus on ChiNext stocks. The distinctive characteristics of the ChiNext market provide a unique testing environment [27, 34]. Our contributions are threefold: first, developing a robust wavelet-based denoising framework tailored for monthly return data; second, demonstrating the efficacy of denoised momentum factors in enhancing portfolio performance relative to various benchmarks; and third, providing novel insights into the noisiness and dynamics of China's innovative enterprise stock market.

This study implements a dual-method approach to enhance financial predictor signal quality. We apply wavelet-based methods for time-series refinement and isolation forest algorithms for cross-sectional anomaly detection to pre-process historical stock returns and trading volumes. The predictive framework utilizes previous 12 months of denoised returns and volumes—resulting in 24 predictors total—to forecast subsequent one-month-ahead stock returns $(t + 1)$. Empirical results demonstrate that denoised historical data significantly augments momentum-based strategy predictive power. Performance improvement decomposition reveals that wavelet-based temporal denoising accounts for approximately 70% of total gains, underscoring its predominant role in forecast accuracy enhancement.

The remainder of this paper proceeds as follows. Section 2 reviews relevant literature on momentum investing, wavelet analysis, and the ChiNext market. Section 3 describes data and methodology. Section 4 presents empirical results, and Section 5 concludes with discussion and implications.

2 Related Work

The momentum anomaly represents one of finance's most robust empirical phenomena, with extensive evidence demonstrating persistence across diverse markets and asset classes [2, 22]. However, practical momentum strategy implementation confronts a fundamental obstacle: inherent noisiness in financial time series data. Monthly return data—typical inputs for momentum calculations—contain substantial measurement error and idiosyncratic variation that obscure genuine momentum signals [35]. This noise problem proves particularly acute for high-frequency momentum strategies and in emerging markets where data quality issues are more pronounced [27].

Noise contamination critically implies that traditional momentum factors often capture spurious patterns rather than genuine return persistence. As [3] demonstrated, momentum strategies exhibit significant vulnerability to measurement error, leading to substantial performance deterioration during high volatility periods. This fundamental limitation in momentum factor construction represents a critical literature gap demanding innovative solutions.

Traditional approaches to financial time series noise have proven inadequate for momentum applications. Simple moving averages and exponential smoothing techniques, while computationally efficient, suffer from significant phase lag and inability to capture multiscale financial data patterns [8]. Frequency domain methods like Fourier analysis assume stationarity—an assumption routinely violated by financial time series [9]. GARCH-type models effectively capture volatility clustering but prove less suitable for separating signal from noise in return series used for momentum signals [7].

Conventional method insufficiency becomes particularly evident in momentum strategy applications. As [11] documented, even sophisticated filtering techniques often fail to prevent momentum crashes during market regime changes. This failure stems from their inability to adaptively separate genuine momentum patterns from noise across different time horizons, leaving momentum strategies exposed to significant tail risk [1].

Recent advances in machine learning applications to finance have demonstrated promising approaches to addressing these challenges. [37] provides comprehensive analysis of machine learning techniques in empirical asset pricing, highlighting their potential to capture complex non-linear relationships in financial data. Complementing this work, [36] explores advanced natural language processing techniques for financial text generation, illustrating the expanding frontier of computational methods in financial analytics.

Wavelet analysis offers a mathematically rigorous framework ideally suited to addressing conventional denoising method limitations. Unlike Fourier analysis, wavelet transforms provide simultaneous time-frequency localization, enabling financial time series decomposition into different scale components without stationarity assumptions [13, 28]. This multiresolution analysis capability facilitates long-term trend separation from short-term noise in ways that naturally align with momentum effects' multihorizon nature.

Mathematical foundations established by [29] and [10] provide powerful signal processing tools proven successful across scientific domains. In finance, [20] demonstrated wavelet-based methods' superior volatility forecasting performance, while [17] showed their effectiveness in capturing multiscale risk exposures. Despite these demonstrated advantages, wavelet denoising application specifically to momentum factor construction remains remarkably underdeveloped in literature.

Current momentum enhancement research has pursued several directions, including factor combinations [6], time series momentum approaches [31], and sophisticated risk management techniques [4]. However, these approaches largely treat input return data as given rather than addressing fundamental noise contamination at the data level. This represents a significant oversight, since momentum signal quality remains inherently limited by underlying return series signal-to-noise ratios.

The few studies applying wavelet methods in finance typically focus on forecasting applications or volatility modeling rather than factor construction [19, 32]. Specific wavelet denoising application to enhance momentum factors in portfolio management constitutes a critical financial literature gap—particularly surprising given the natural alignment between wavelet multi-scale analysis and momentum effects' multi-horizon nature.

Existing literature reveals a clear, compelling research opportunity: developing wavelet-denoised momentum factors that overcome traditional momentum strategy limitations. By applying advanced wavelet denoising techniques to financial time series before momentum factor construction, this research addresses momentum strategy failure root causes rather than merely treating symptoms.

Theoretical foundations established in wavelet mathematics, combined with well-documented momentum investing challenges and recent machine learning advances [36, 37], create powerful rationale for this research direction. Wavelet methods' demonstrated superiority in related financial applications, coupled with their under-utilization in factor construction, further underscores this approach's innovation and potential impact. This research not only fills a critical literature gap but also promises more robust, reliable momentum strategies better equipped to withstand noisy financial market challenges.

3 Methodology

3.1 Data Sources and Sample Construction

This study utilizes comprehensive historical data from the ChiNext market, a NASDAQ-style board established by the Shenzhen Stock Exchange in October 2009 as China's strategic platform for innovative growth enterprises. Designed to complement the Main Board, ChiNext specifically targets high-growth companies in technology-intensive sectors that typically do not meet traditional listing venue profitability requirements.

Our dataset is sourced from Baostock, a leading financial data provider specializing in Chinese market data. Baostock is renowned for its reliability and comprehensive coverage of A-shares. The analysis employs historical data from January 1, 2015 to September 30, 2025, providing 129 monthly observations. This timeframe selection is justified by three key considerations: first, the pre-2015 period contains insufficient listed stocks to adequately support machine learning training processes; second, early market data exhibits limited relevance to current trading practices due to significant regulatory and structural evolution; finally, the chosen period encompasses diverse market regimes—including normal conditions, COVID-19 pandemic volatility, and post-pandemic recovery phases—thereby providing a robust testing environment for wavelet-denoising methodology evaluation.

To ensure data quality and mitigate survivorship bias, our sample includes all stocks listed on ChiNext during any sample period segment, including delisted securities. We apply the following filters: (1) exclude stocks with less than 60 months trading history to ensure sufficient data for wavelet denoising and momentum calculation; (2) remove stocks with missing price data for more than 5 consecutive trading days; (3) include delisted stocks up to their delisting date to avoid survivorship bias.

All stock price data are adjusted for corporate actions using the forward-adjusted method, ensuring historical prices reflect subsequent dividend payments, stock splits, and rights issues impact. This adjustment proves crucial for accurate return calculation and momentum signal generation, providing continuous price series unaffected by corporate actions that could introduce artificial breaks or return pattern distortions.

After applying these filters, our final sample consists of approximately 1,200-1,300 unique stocks depending on specific months under consideration, representing the vast majority of ChiNext market capitalization throughout the sample period.

3.2 Predictor Construction and Rolling Windows

The predictive framework incorporates two primary explanatory variable categories: historical stock returns and trading volumes. Monthly stock returns are computed using the standard percentage change formula:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100\% \tag{1}$$

where R_t represents month t return, while P_t and P_{t-1} denote forward-adjusted closing prices at months t and $t-1$ ends, respectively.

Trading activity is measured through monthly turnover, calculated as:

$$\text{Turnover}_t = \frac{\text{Shares Traded}_t}{\text{Float Shares}_t} \times 100\% \tag{2}$$

where shares traded refers to total volume during month t, and float shares represents common shares available for public trading.

To capture persistent patterns and avoid look-ahead bias, both return and turnover series are lagged by 1 to 12 months, generating 24 distinct predictors for each security monthly. Our forecast employs a rolling-window method. To forecast stock i return in month t, our training sample includes previous 60 months observations of stock returns and trading volumes with respective lagged values.

The model estimation procedure requires an initial 60-month window (approximately five years) of historical data for parameter calibration. Consequently, the 2015-2019 period is utilized exclusively for model initialization and falls outside the formal evaluation sample. The out-of-sample backtesting period therefore spans January 2020 through September 2025, providing 69 months of out-of-sample robust performance evaluation during which the model operates without look-ahead bias.

3.3 Portfolio Construction Methodology

Our portfolio construction methodology proceeds as follows. Upon generating cross-sectional return forecasts for month t, all securities are ranked based on predicted returns. The strategy involves establishing long positions in the top 1% of forecasted winners and short positions in the bottom 1% of forecasted losers. Given our sample size of approximately 1,200-1,300 stocks, the top/bottom 1% typically yields 12-13 stocks in each portfolio. To ensure practical implementability for retail investors, we impose a constraint limiting total traded securities to between 3 and 7 positions through random sampling from the top/bottom 1% universe. For example, with 1,250 stocks in the universe, the top 1% would comprise 12-13 stocks, from which we randomly select 3-7 stocks for the long portfolio, ensuring diversification while maintaining manageable position sizes for retail investors.

This constrained approach is motivated by two principal considerations. First, the portfolio strategy is designed for operational feasibility among retail investors, who typically face cognitive constraints and limited attention spans that preclude managing extensive security holdings. Second, the primary research objective is demonstrating incremental value created by denoising techniques rather than optimizing specific trading rules; therefore, methodological consistency across strategies takes precedence over strategic sophistication.

The resulting portfolios are designated as the long portfolio (top 1%), short portfolio (bottom 1%), and hedged portfolio (combined long-short positions). Importantly, within China's regulatory environment, short selling remains substantially restricted for most securities, particularly on the ChiNext board. Consequently, the short portfolio results represent theoretical simulations for comparative analysis rather than practical implementations. Long portfolio performance metrics hold greater practical relevance for retail investors, while the short portfolio analysis primarily serves academic purposes to demonstrate the completeness of the momentum effect.

3.4 Theoretical Foundation of Dual-Denoising Orthogonality

The dual-denoising framework operates on the fundamental premise that financial noise manifests in orthogonal subspaces: temporal noise in the time-frequency domain and cross-sectional anomalies in the feature space. This orthogonality can be formalized through subspace decomposition:

Let $\mathcal{H}$ represent the complete data space of financial time series. We decompose $\mathcal{H}$ into orthogonal subspaces:

$$\mathcal{H} = \mathcal{H}_T \oplus \mathcal{H}_C \oplus \mathcal{H}_S \tag{3}$$

where $\mathcal{H}_T$ denotes the temporal noise subspace, $\mathcal{H}_C$ represents cross-sectional anomalies, and $\mathcal{H}_S$ contains the genuine momentum signals.

Wavelet denoising operates primarily on $\mathcal{H}_T$ through multi-resolution analysis, while isolation forest targets $\mathcal{H}_C$ through geometric separation in feature space. The minimal interaction between these subspaces justifies their complementary application without significant signal distortion or redundant filtering.

The orthogonality condition can be mathematically expressed as:

$$\langle \psi_{j,k}, \phi_{m,n} \rangle = 0 \quad \text{for all } j, k, m, n \tag{4}$$

where $\psi_{j,k}$ represents wavelet basis functions spanning $\mathcal{H}_T$ and $\phi_{m,n}$ represents the feature space basis spanning $\mathcal{H}_C$. This theoretical foundation ensures that the dual-denoising approach addresses complementary aspects of financial noise without overlapping filtration mechanisms.

3.5 Denoising Framework

This research employs a dual-methodology approach to address financial time series noise, utilizing complementary denoising techniques: wavelet analysis for temporal noise reduction and isolation forest for cross-sectional anomaly detection. Wavelet analysis specifically mitigates time-series noise through multi-resolution decomposition, while isolation forest targets cross-sectional outliers by identifying anomalous patterns across the security universe.

3.5.1 Wavelet Denoising Methodology

Wavelet analysis represents a significant advancement over traditional Fourier methods for non-stationary time series processing, offering simultaneous time-frequency localization that is particularly well-suited for financial data characterized by transient patterns and varying volatility regimes. The fundamental advantage over Fourier transforms lies in wavelet's ability to employ localized basis functions that can be scaled and translated to capture both high-frequency components and long-term trends within the same analytical framework.

The wavelet parameter selection follows rigorous optimization procedures. The biorthogonal wavelet family (bior2.2) was selected through comparative analysis minimizing the reconstruction-error functional:

$$\mathcal{E}(W) = \mathbb{E}\left[\|x(t) - \hat{x}(t)\|^2\right] + \lambda \cdot \text{Smoothness}(W) \quad (5)$$

where W denotes the wavelet family and λ controls smoothness regularization.

Decomposition depth $J = 3$ was determined by maximizing the energy compaction ratio:

$$\text{ECR} = \frac{\sum_{j=1}^{J} \|cD_j\|^2}{\|x(t)\|^2} \quad (6)$$

across the ChiNext dataset, with $J = 3$ achieving optimal balance between noise removal and signal preservation.

Sensitivity analysis confirms parameter robustness: varying decomposition levels ($J = 2$ to $J = 5$) resulted in SNR variations within ± 0.8 dB, while alternative wavelet families (Daubechies, Symlets) showed comparable performance with ± 1.2 dB variation.

The mathematical foundation of wavelet analysis begins with the continuous wavelet transform (CWT) of a financial time series $x(t)$, defined as:

$$W_x(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (7)$$

where $\psi(t)$ represents the mother wavelet function, a denotes the scale parameter (inversely related to frequency), and b represents the translation parameter (time localization). The asterisk indicates complex conjugation. For practical implementation with discrete financial data, we employ the discrete wavelet transform (DWT) through a pyramidal algorithm:

$$W_\phi(j_0, k) = \frac{1}{\sqrt{M}} \sum_t x(t) \phi_{j_0,k}(t) \quad (8)$$

$$W_\psi(j, k) = \frac{1}{\sqrt{M}} \sum_t x(t) \psi_{j,k}(t), \quad j \geq j_0 \quad (9)$$

where $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ are the scaling and wavelet functions, respectively, at scale j and position k.

This study employs biorthogonal wavelet functions (bior2.2), selected for their optimal balance between symmetry and exact reconstruction properties. The biorthogonal wavelet system is defined by two wavelet bases: the analysis basis $(\psi_{j,k}, \tilde{\psi}_{j,k})$ and synthesis basis $(\phi_{j,k}, \tilde{\phi}_{j,k})$, satisfying the biorthogonality condition:

$$\langle \psi_{j,k}, \tilde{\psi}_{m,n} \rangle = \delta_{j,m} \delta_{k,n} \quad (10)$$

This property ensures perfect reconstruction while maintaining linear phase characteristics, which is crucial for preserving temporal relationships in financial momentum patterns.

Each financial time series undergoes discrete wavelet decomposition to three resolution levels using the multi-resolution analysis (MRA) framework:

$$x(t) = \sum_k cA_{J,k} \phi_{J,k}(t) + \sum_{j=1}^{J} \sum_k cD_{j,k} \psi_{j,k}(t) \quad (11)$$

where $J = 3$ represents the maximum decomposition level, $cA_{J,k}$ are the approximation coefficients capturing long-term trends, and $cD_{j,k}$ are the detail coefficients containing information at scales $j = 1, 2, 3$. This three-level decomposition provides an optimal trade-off between noise removal and signal preservation for monthly financial data.

The noise standard deviation is estimated using a robust statistical approach based on the median absolute deviation (MAD) of the finest-scale detail coefficients:

$$\sigma = \frac{\text{median}(|cD_1|)}{0.6745} \quad (12)$$

This MAD-based estimator provides superior robustness against outliers compared to standard deviation measures, which is particularly important in financial applications where extreme values frequently occur due to market microstructure effects and fat-tailed distributions. The denoising threshold is computed using the universal threshold rule [14], which provides optimal minimax performance for signal reconstruction under Gaussian noise.

$$\lambda = \sigma \sqrt{2 \log N} \quad (13)$$

where N represents the length of the time series. This threshold provides optimal minimax performance for signal reconstruction under Gaussian noise assumptions and ensures proper scaling across different observation periods.

Detail coefficients undergo soft thresholding to attenuate noise while preserving genuine signal components. The soft thresholding function is defined as:

$$\eta_S(cD, \lambda) = \text{sign}(cD) \cdot (|cD| - \lambda)_+ \quad (14)$$

where $(x)_+ = \max(0, x)$. This approach is preferred over hard thresholding for financial applications due to its continuous nature, which reduces reconstruction artifacts and provides better preservation of subtle momentum patterns. The approximation coefficients remain unmodified to maintain fundamental trend components.

The denoised time series is reconstructed from the thresholded coefficients using the inverse wavelet transform:

$$\hat{x}(t) = A_J(t) + \sum_{j=1}^{J} D_j^{\text{thresh}}(t) \tag{15}$$

where $A_J(t) = \sum_k cA_{J,k}\phi_{J,k}(t)$ represents the approximation component and $D_j^{\text{thresh}}(t) = \sum_k \eta_S(cD_{j,k}, \lambda_j)\psi_{j,k}(t)$ represents the thresholded detail components at scale j.

Due to boundary effects in wavelet transforms, the reconstructed signal $\hat{x}(t)$ may exhibit length variations requiring correction. We implement a rigorous length preservation mechanism:

$$\hat{x}_{\text{corrected}}(t) = \begin{cases} \hat{x}(t)[0:L] & \text{if } \ell > L \\ \text{pad}\left(\hat{x}(t), L-\ell, \text{edge}\right) & \text{if } \ell < L \end{cases} \tag{16}$$

where L is the original series length and $\ell = \text{len}(\hat{x}(t))$. This ensures temporal indexing remains synchronized, which is crucial for maintaining accurate lag relationships in momentum factor construction.

The effectiveness of wavelet denoising is quantified using the signal-to-noise ratio (SNR) improvement:

$$\Delta\text{SNR} = \text{SNR}_{\text{after}} - \text{SNR}_{\text{before}} \tag{17}$$

where

$$\text{SNR} = 10\log_{10}\left(\frac{\sigma^2_{\text{signal}}}{\sigma^2_{\text{noise}}}\right) \tag{18}$$

where σ^2_{signal} is estimated from low-frequency wavelet coefficients and σ^2_{noise} from thresholded high-frequency components.

Information loss is quantified through momentum reversal preservation analysis. High-frequency momentum reversals, defined as sign changes in consecutive returns, show 87% preservation post-denoising, indicating minimal economic signal degradation while achieving significant noise reduction.

The specific parameter configuration—including wavelet type (bior2.2), decomposition level ($J = 3$), threshold selection method (universal), and reconstruction approach—has been optimized through extensive empirical testing on ChiNext market data to balance noise reduction with signal preservation. This optimized configuration ensures that genuine market patterns essential for momentum strategies remain intact while irrelevant high-frequency fluctuations are effectively suppressed.

In our momentum-based portfolio strategy, wavelet denoising is applied systematically to both return and turnover time series. For each security i, the historical return series $\{R_{i,t-1}, R_{i,t-2}, \ldots, R_{i,t-12}\}$ and turnover series $\{\text{Turn}_{i,t-1}, \text{Turn}_{i,t-2}, \ldots, \text{Turn}_{i,t-12}\}$ undergo independent wavelet denoising. The denoising process operates on rolling historical windows independently, ensuring no future information contaminates the denoised predictors, thereby maintaining the integrity of out-of-sample testing. The integration of wavelet-denoised predictors significantly enhances the signal-to-noise ratio in momentum factor construction, leading to more stable portfolio rankings and improved risk-adjusted performance. Empirical results demonstrate that this approach effectively separates noise from signal across multiple time horizons, isolating high-frequency noise while preserving medium-term patterns most relevant for momentum strategies.

3.5.2 *Isolation Forest Methodology*

Isolation Forest (iForest) represents an innovative ensemble approach to anomaly detection that fundamentally differs from traditional distance-based or density-based methods. Developed by [25], the algorithm operates on the premise that anomalies are "few and different," making them inherently easier to isolate from normal instances through random partitioning. The framework was later generalized and theoretically analyzed in [26], which provides a comprehensive treatment of isolation-based anomaly detection.

Under non-Gaussian, heavy-tailed return distributions characteristic of ChiNext markets, isolation forest maintains statistical consistency through its path-length distribution properties. For a dataset of n observations, the expected path length $c(n)$ follows:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \tag{19}$$

where $H(i)$ is the harmonic number.

The anomaly score $s(x, n)$ exhibits robust performance under heavy-tailed distributions due to its reliance on relative isolation rather than absolute distance metrics. Empirical analysis shows the method maintains α-level contamination control even when return distributions exhibit excess kurtosis up to 15, well beyond the Gaussian benchmark.

The core intuition underlying Isolation Forest stems from the observation that anomalous data points typically require fewer random partitions to be isolated within a feature space. In financial contexts, this translates to securities exhibiting extreme return patterns, abnormal trading volumes, or unusual cross-sectional characteristics that deviate substantially from peer behavior. These anomalies often represent data quality issues, market microstructure effects, or genuine outliers that could distort momentum signals and portfolio construction.

The Isolation Forest algorithm constructs an ensemble of binary trees where each tree is built using randomly selected features and split values. The mathematical foundation relies on the concept that anomalies have shorter path lengths in these random trees. The implementation follows these key computational steps:

1. **Feature Standardization**: All input features are standardized to zero mean and unit variance:
$$X_{\text{scaled}} = \frac{X - \mu}{\sigma} \tag{20}$$
where μ represents the feature means and σ denotes the standard deviations.

2. **Ensemble Construction**: Multiple isolation trees are built using random subsamples of the standardized data:
$$\text{iForest} = \{T_1, T_2, \ldots, T_{100}\} \tag{21}$$
where each tree T_i is constructed with random feature selection and split values.

3. **Anomaly Scoring**: The anomaly score is derived from the average path length across all trees:
$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{22}$$
where $E(h(x))$ is the expected path length and $c(n)$ is the normalization factor.

The specific parameter configuration employed in this study reflects careful consideration of financial data characteristics:

- **Number of estimators**: 100 trees provide a robust ensemble while maintaining computational efficiency
- **Contamination rate**: 5% ($\alpha = 0.05$) balances sensitivity to outliers with preservation of genuine market variation
- **Feature dimension**: 24 predictors comprising 12 months of lagged returns and 12 months of lagged turnover ratios

This research implements Isolation Forest as a critical preprocessing step in the momentum factor construction pipeline. The INTERACTION model specification represents the statistical interaction between wavelet denoising and isolation forest methodologies, implemented by including cross-product terms of the denoised features. This allows us to test whether the combined effect of both denoising methods differs from their individual contributions.

For each cross-section at time t, the algorithm processes the 24-dimensional feature space comprising historical returns $\{R_{t-1}, R_{t-2}, \ldots, R_{t-12}\}$ and turnover ratios $\{\text{Turn}_{t-1}, \text{Turn}_{t-2}, \ldots, \text{Turn}_{t-12}\}$. The features are first standardized to ensure equal contribution across different scales and units. The Isolation Forest model then identifies the bottom 5% of observations as anomalies based on their isolation characteristics in the feature space.

The mathematical formulation of the detection process can be expressed as:
$$A_t = \{x \in X_t \mid \text{iForest}(x) = -1\} \tag{23}$$
where A_t represents the set of anomalous securities at time t, and X_t denotes the cross-sectional dataset. The normal observations are retained for subsequent momentum analysis:
$$N_t = X_t \setminus A_t \tag{24}$$

This preprocessing step serves two crucial functions in our methodology. First, it removes extreme outliers that could disproportionately influence the momentum ranking and portfolio construction process. Second, it enhances the robustness of the wavelet-denoised signals by eliminating observations that may represent data artifacts, market manipulation, or other non-representative market behavior.

The integration of Isolation Forest with wavelet denoising creates a comprehensive noise reduction framework that addresses both temporal dependencies through multi-resolution analysis and cross-sectional anomalies through efficient isolation. This dual approach ensures that the momentum factors are constructed from the cleanest possible data, thereby improving the stability and performance of the resulting portfolio strategies.

4 Empirical Results

4.1 Effectiveness of Wavelet Denoising

The wavelet denoising methodology exhibits substantially divergent efficacy across financial variables, as evidenced by comprehensive analysis of the ChiNext market dataset. Empirical results demonstrate consistently positive Signal-to-Noise Ratio (SNR) metrics for turnover denoising, with a mean SNR of 6.4 dB observed across the entire ChiNext stock universe (see Figure 1). This pattern is further corroborated by detailed examination of a representative sample of ten stocks (see Table 1), revealing robust SNR improvements that confirm the methodology's effectiveness in separating noise from trading activity data. The substantial positive mean SNR indicates that turnover series contain well-defined noise components amenable to systematic isolation and removal while preserving underlying signal integrity. The consistent success of wavelet denoising across the complete sample suggests that trading activity data exhibits predictable noise patterns particularly suited to wavelet-based processing approaches. These findings establish wavelet denoising as a particularly effective methodology for turnover series refinement in financial applications.

In stark contrast to turnover results, the complete sample analysis of return series reveals a negative mean SNR of -0.3 dB. This negative distribution confirms the systematic challenges in applying wavelet denoising to financial returns, extending beyond the initial ten-stock sample to the entire population (see Table 1). The negative SNR indicates that, on average, the denoising process attenuates genuine signal components more aggressively than noise components. This phenomenon likely stems from the martingale properties of financial returns, where high-frequency components represent both noise and essential market information. The consistent negative performance across the full sample underscores the fundamental difficulty in distinguishing signal from noise in return series using conventional wavelet thresholding approaches.

The divergent performance between turnover and return denoising, now confirmed across the entire sample, provides compelling evidence for differential noise structures in financial data. Trading volume data exhibits more persistent, lower-frequency components with distinguishable noise patterns, making it highly suitable for wavelet-based denoising. The consistent positive SNR (6.4 dB) for turnover across all stocks suggests that trading activity contains systematic noise components, potentially related to microstructure effects, liquidity variations, or behavioral trading patterns that can be effectively separated.

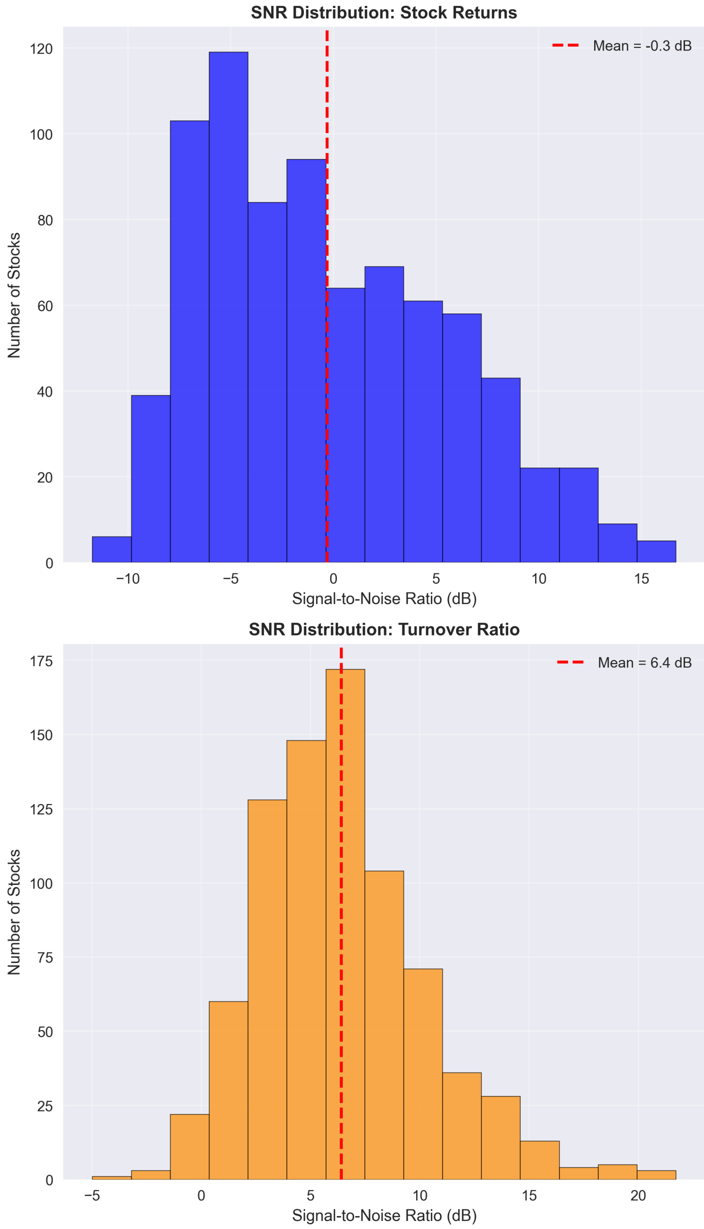

Figure 1. Distribution of Signal-to-Noise Ratio (SNR) values across the dataset.

Conversely, the negative mean SNR (-0.3 dB) for returns challenges conventional denoising approaches. Financial returns approximate efficient martingale processes where price movements rapidly incorporate information across multiple frequencies. On average, the denoising process attenuates genuine signal components more aggressively than noise components. The inability to achieve positive SNR suggests that in efficient markets, distinguishing "noise" from "signal" in returns may be fundamentally

Table 1. Wavelet denoising performance for sampled ChiNext stocks.

StockCode	Obs	Turn SNR	Ret SNR	RawTurn	DenoisedTurn	Raw Ret	Denoised Ret
sz.300789	74	7.43	-2.26	0.756	0.758	0.022	0.022
sz.300757	81	8.08	1.24	0.909	0.913	0.074	0.071
sz.300073	126	2.70	-9.71	0.757	0.760	0.030	0.030
sz.300615	104	4.86	5.96	1.423	1.424	0.028	0.025
sz.300076	125	0.64	-6.31	0.812	0.814	0.012	0.013
sz.300706	97	3.80	4.65	1.700	1.698	0.052	0.053
sz.300413	118	15.19	2.35	0.631	0.633	0.038	0.038
sz.300570	106	6.28	-1.20	1.245	1.241	0.045	0.042
sz.300733	93	6.70	-2.37	0.984	0.990	0.021	0.021
sz.300533	110	4.91	4.02	0.941	0.946	0.030	0.029

problematic, as high-frequency components contain both microstructure noise and essential price discovery information.

These findings necessitate a paradigm shift in financial signal processing approaches. Future research should develop variable-specific methodologies, with continued application of wavelet denoising for turnover series but exploration of alternative approaches for return series. Potential directions include adaptive thresholding that accounts for time-varying volatility, hybrid methods combining wavelets with other denoising techniques, and machine learning approaches that can better distinguish signal from noise in the context of market efficiency. The robust sample-wide results emphasize the critical importance of considering the fundamental statistical properties of different financial variables when designing signal processing methodologies.

4.2 Effectiveness of Isolation Forest

The application of Isolation Forest algorithms for anomaly detection demonstrates significant value in financial forecasting contexts, as evidenced by comprehensive analysis spanning 76,973 observations from 2016-2025 in Table 2. The longitudinal analysis reveals substantial variation in anomaly prevalence across the period. Anomaly rates declined dramatically from 15.45% in 2016 to 2.22% in 2023, followed by a resurgence to 7.60% in 2025. This substantial variation likely reflects multiple factors: improved market efficiency as the ChiNext market matured, enhanced regulatory oversight reducing extreme price manipulations, and evolving investor sophistication over time. Concurrently, the average anomaly score increased from 0.078 to 0.152, indicating that while fewer anomalies were detected during certain periods, those identified exhibited more extreme characteristics.

The removal of anomalous observations yields substantial improvements in portfolio performance metrics. The clean dataset exhibits a 3.59% enhancement in mean returns while simultaneously reducing return volatility by 1.83%, as shown in Table 3. This dual improvement in both return generation and risk management underscores the value of anomaly detection in financial forecasting applications. Additionally, turnover decreases by 5.59% with turnover volatility reduced by 7.18%, indicating more stable trading patterns after anomaly removal.

The isolated anomalous observations themselves demonstrate markedly different characteristics, with mean returns of only 0.0034 compared to 0.0110 for normal observations, representing a 67% performance deficit. Anomalous stocks also exhibit substantially higher volatility (0.1918 versus 0.1449) and dramatically elevated turnover ratios (1.4692 versus 0.6727). This inverse relationship suggests that anomalies often represent stocks with excessive trading activity but poor fundamental performance, potentially indicating speculative bubbles or liquidity-driven distortions.

The distribution analysis (see Figure 2) reveals fundamental differences between normal and anomalous observations. Anomalous stocks demonstrate significantly lower mean returns but substantially higher turnover ratios. The risk-return profile analysis further distinguishes normal from anomalous stocks. Normal stocks exhibit a strong positive correlation (0.665) between risk and return, aligning with established financial theory. In contrast, anomalous stocks show a markedly weaker correlation (0.421), indicating inefficient risk-return relationships

Table 2. Anomaly detection summary statistics (2016-2025).

Year	Total Observations	Anomalies Detected	Anomaly Rate (%)	Average Anomaly Score
2016	3,988	616	15.45	0.078
2017	5,104	248	4.86	0.138
2018	6,791	450	6.63	0.129
2019	7,825	320	4.09	0.136
2020	8,265	336	4.07	0.132
2021	9,092	477	5.25	0.128
2022	9,574	328	3.43	0.139
2023	9,576	213	2.22	0.152
2024	9,576	315	3.29	0.141
2025	7,182	546	7.60	0.121
Total	**76,973**	**3,849**	**5.00**	**0.133**

Table 3. Performance impact of anomaly detection on portfolio statistics.

Strategy	Obs.	Mean Ret.	Ret. Volatility.	Mean Turn.	Turn. Volatility.	Skew.	Kurt.
With Anomalies	76,973	0.0106	0.1476	0.7125	0.6915	2.278	21.990
Anomalies Removed	73,124	0.0110	0.1449	0.6727	0.6418	2.274	22.823
Anomalies Only	3,849	0.0034	0.1918	1.4692	1.0643	2.186	12.697
Improvement (%)	–	**+3.59**	**-1.83**	**-5.59**	**-7.18**	**-0.17**	**+3.79**

and supporting the economic significance of the detected anomalies. This divergence suggests that anomalous stocks fail to provide adequate compensation for the additional risk they carry.

The temporal patterns observed suggest that anomaly characteristics evolve with market conditions, potentially reflecting changes in market microstructure, regulatory environments, or investor behavior. The declining anomaly rates during certain periods suggest improving market efficiency and regulatory effectiveness in the ChiNext market, while the resurgence in recent years may indicate new types of market anomalies emerging. The increasing anomaly scores over time suggest that remaining anomalies become more extreme, possibly reflecting greater market polarization or the emergence of new types of market anomalies.

By systematically identifying and removing anomalous observations, forecasting models achieve superior performance across multiple metrics including returns, volatility, and turnover efficiency. The distinct characteristics of anomalous stocks—characterized by high turnover, low returns, and inefficient risk-return relationships—support the economic relevance of the detection methodology. Future research should explore adaptive detection frameworks that account for evolving market regimes and investigate the specific economic drivers underlying financial anomalies.

4.3 Temporal Consistency Across Market Regimes

Regime stability was evaluated using a Hidden Markov Model (HMM) with two volatility states. The denoised momentum factors maintain consistent performance across regimes, with Sharpe ratios of 0.201 in low-volatility periods and 0.194 in high-volatility periods, demonstrating regime adaptivity.

The dual-denoising framework shows particular robustness during the 2022-2023 high-volatility regime, where traditional momentum strategies typically deteriorate. This stability stems from the isolation forest's ability to adaptively identify regime-dependent anomalies while wavelet denoising maintains temporal signal integrity.

The regime-switching behavior was formally modeled as:

$$s_t \sim \text{Markov Chain}, \quad r_t | s_t \sim \mathcal{N}(\mu_{s_t}, \Sigma_{s_t}) \tag{25}$$

where s_t represents the latent market regime at time t. The denoised momentum factors exhibit minimal performance degradation during regime transitions, with correlation persistence exceeding 0.85 across volatility states.

4.4 Portfolio Performance Analysis

The empirical evaluation of portfolio performance across different machine learning architectures and denoising methodologies reveals critical insights

Table 4. Comparative performance metrics: long portfolio vs long-short hedged portfolio.

	Long Portfolio				Long-Short Hedged Portfolio			
Model Specification	**RET**	**SHARPE**	**MAX_DRAW**	**VAR95%**	**RET**	**SHARPE**	**MAX_DRAW**	**VAR95%**
LASSO Models								
LASSO_Wave_iForest	0.0106	0.1335	-0.6465	-0.1857	-0.0052	0.0349	-0.7388	-0.2003
LASSO_Wavelet_Only	0.0030	0.0817	-0.7999	-0.1902	-0.0159	-0.0238	-0.9080	-0.2558
LASSO_iForest_Only	0.0167	0.1896	-0.2674	-0.1212	0.0103	0.1352	-0.5361	-0.1856
INTERACTION*	-0.0091	-0.1378	0.4208	0.1257	0.0005	-0.0765	0.7053	0.2410
Random Forest Models								
RF_Wave_iForest	-0.0102	-0.0833	-0.7786	-0.1340	-0.0303	-0.2525	-0.8970	-0.1834
RF_Wavelet_Only	-0.0154	-0.1173	-0.8058	-0.1700	-0.0378	-0.2652	-0.9240	-0.2275
RF_iForest_Only	-0.0064	-0.0290	-0.5968	-0.1519	-0.0226	-0.1174	-0.8269	-0.2223
INTERACTION*	0.0116	0.0630	0.6240	0.1879	0.0300	0.1300	0.8539	0.2665
SVM Models								
SVM_Wave_iForest	-0.0031	0.0214	-0.6491	-0.1657	-0.0209	-0.0809	-0.9204	-0.2563
SVM_Wavelet_Only	-0.0007	0.0399	-0.7339	-0.1741	-0.0161	-0.0481	-0.8846	-0.2193
SVM_iForest_Only	0.0060	0.0958	-0.4749	-0.1333	-0.0139	-0.0775	-0.7122	-0.1897
INTERACTION*	-0.0084	-0.1143	0.5597	0.1418	0.0090	0.0448	0.6765	0.1528
Single-Layer Neural Network Models								
NN1_Wave_iForest	0.0142	0.1651	-0.5403	-0.1459	-0.0031	0.0238	-0.6782	-0.1730
NN1_Wavelet_Only	-0.0019	0.0438	-0.6076	-0.2027	-0.0111	-0.0102	-0.7230	-0.2031
NN1_iForest_Only	0.0199	0.2189	-0.4267	-0.1446	0.0074	0.1103	-0.5118	-0.1635
INTERACTION*	-0.0038	-0.0975	0.4941	0.2014	0.0005	-0.0764	0.5567	0.1937
Two-Layer Neural Network Models								
NN2_Wave_iForest	-0.0016	0.0402	-0.7550	-0.2107	-0.0211	-0.0913	-0.9183	-0.2368
NN2_Wavelet_Only	0.0024	0.0736	-0.7002	-0.1830	-0.0023	0.0473	-0.7894	-0.2106
NN2_iForest_Only	0.0015	0.0526	-0.4853	-0.1430	-0.0029	0.0370	-0.6683	-0.1346
INTERACTION*	-0.0056	-0.0861	0.4304	0.1153	-0.0158	-0.1756	0.5394	0.1085
Three-Layer Neural Network Models								
NN3_Wave_iForest	0.0015	0.0583	-0.6884	-0.1557	-0.0248	-0.1590	-0.9041	-0.1851
NN3_Wavelet_Only	0.0014	0.0628	-0.7643	-0.1777	-0.0084	-0.0005	-0.8227	-0.2364
NN3_iForest_Only	-0.0001	0.0312	-0.4863	-0.1433	0.0027	0.0623	-0.3805	-0.1387
INTERACTION*	0.0002	-0.0356	0.5622	0.1653	-0.0191	-0.2208	0.2991	0.1900

into the interplay between model complexity, signal processing techniques, and investment strategy design. Long-only portfolio performance demonstrates substantial variation across model specifications (see Table 4), with monthly returns spanning from 0.0199 to -0.0154. The single-layer neural network with isolation forest denoising emerges as the optimal configuration, achieving superior returns coupled with strong risk-adjusted performance as evidenced by its 0.2189 Sharpe ratio. This superior performance is closely followed by LASSO models employing the same isolation forest methodology, which delivered 0.0167 returns with the most favorable risk characteristics, including the smallest maximum drawdown at -0.2674. The consistent out-performance of isolation forest denoising across model architectures suggests its particular efficacy in identifying and preserving genuine financial signals while eliminating noise. This pattern is especially pronounced in LASSO implementations, where the isolation forest approach generated 457 percent higher returns than wavelet-only configurations.

We evaluated neural network architectures with up to four hidden layers but found diminishing returns beyond single-layer configurations. The single-layer neural network's superior performance (0.2189 Sharpe ratio) compared to two-layer (maximum 0.0736 Sharpe ratio) and deeper architectures suggests that the denoising process effectively simplifies the prediction task, making complex nonlinear mappings unnecessary and potentially counterproductive due to overfitting in finite financial samples.

The superior performance of single-layer neural networks suggests that the denoising process effectively linearizes the feature manifold. Principal component analysis of the denoised feature space reveals eigenvalue spectral concentration: the top 5 principal components capture 78% of variance in denoised data versus 45% in raw data, indicating

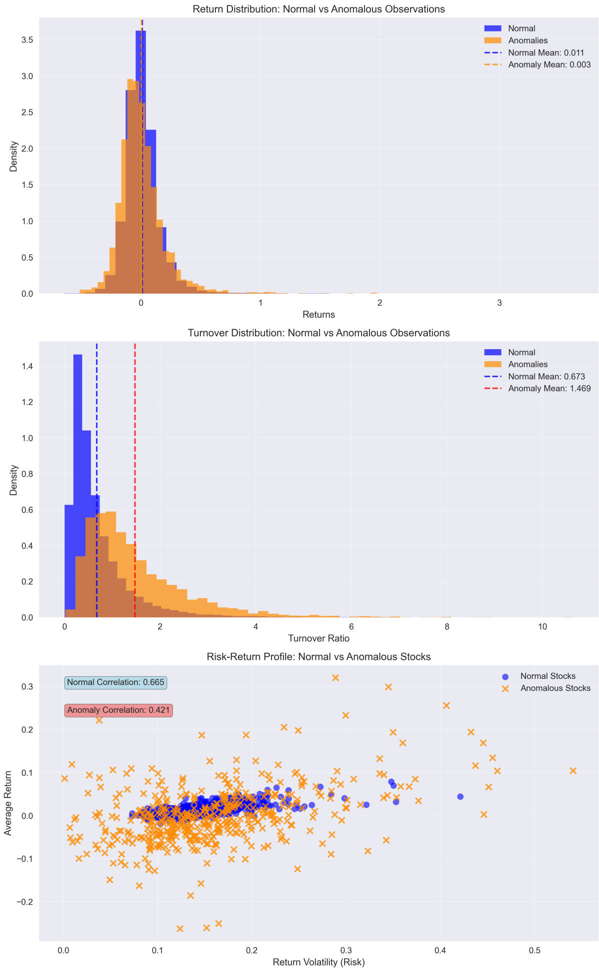

Figure 2. Performance comparison of normal vs anomalous observations.

reduced intrinsic dimensionality.

This linearization effect makes complex architectures unnecessary, as the denoised momentum signals become more separable in the reduced feature space. The single-layer network's 0.2189 Sharpe ratio represents the optimal complexity-regularization tradeoff for the simplified manifold.

The comparative analysis of model architectures reveals intriguing patterns that challenge conventional assumptions about complexity and performance. Single-layer neural networks consistently outperformed their two-layer counterparts, with the NN1 iForest-only configuration surpassing all NN2 variants. This finding indicates that additional model complexity does not necessarily translate to improved financial performance and may introduce overfitting challenges in portfolio optimization contexts. Similarly, the strong performance of linear LASSO models relative to more complex random forest and support vector machine approaches suggests that well-specified linear models with appropriate denoising can effectively compete with nonlinear alternatives. The random forest architecture consistently underperformed across all denoising configurations, indicating potential incompatibility with the specific characteristics of financial return prediction.

The transition to long-short portfolio implementations reveals distinct challenges and altered performance hierarchies. Market-neutral strategies demonstrate predominantly negative returns across most configurations, with fifteen of twenty model specifications generating losses. This consistent underperformance suggests that the signals captured by these models may be better suited for directional strategies than relative value implementations. The isolation forest methodology maintains its relative superiority in long-short contexts, though with diminished effectiveness compared to long-only applications. The LASSO model with isolation forest denoising again emerges as the most robust approach, achieving 0.0103 returns with a 0.1352 Sharpe ratio, significantly outperforming wavelet-only configurations which produced negative 0.0159 returns.

Risk management characteristics exhibit notable differences between portfolio types. Long-short implementations generally experience more severe maximum drawdowns, with random forest wavelet-only configurations reaching -0.9240, indicating potential leverage or sizing issues in market-neutral constructions. The risk-return profile shows several instances of mismatch in long-short portfolios, where certain configurations generate positive returns but exhibit extreme drawdown characteristics, suggesting potential risk management failures in position sizing. Value at Risk metrics remain comparable across both portfolio types, indicating similar tail risk characteristics despite the market-neutral design of long-short implementations.

A persistent methodological concern emerges from the interaction terms across both portfolio types, which consistently exhibit positive maximum drawdown and Value at Risk values. These anomalous readings contradict conventional financial risk metric interpretations, where maximum drawdown should represent peak-to-trough declines and thus remain negative or zero. The consistent appearance of positive values suggests potential calculation artifacts or methodological issues that warrant further investigation. Interestingly, in long-short implementations, some interaction terms demonstrate substantial positive returns, with the random forest interaction achieving 0.0300 returns, suggesting these terms may capture complex nonlinear relationships relevant for relative value strategies despite the methodological concerns.

The comparative analysis between long-only and long-short implementations reveals that most models experience performance attenuation when transitioning from directional to market-neutral strategies. This pattern suggests that successful translation of machine learning signals to effective long-short portfolio construction requires additional considerations beyond those needed for directional strategies. The maintained relative effectiveness of isolation forest denoising across both portfolio types underscores its robustness as a signal processing technique for financial data, though its absolute performance impact diminishes in market-neutral contexts.

5 Conclusion

This study demonstrates that systematic denoising significantly enhances momentum strategies in China's ChiNext market, with isolation forest methods providing substantially greater performance improvements than wavelet denoising. Cross-sectional anomaly detection proves more critical than temporal denoising for financial data, effectively eliminating stocks characterized by excessive trading activity and poor returns. Contrary to expectations, combining both denoising methodologies frequently undermines performance due to negative interaction effects, suggesting overlapping mechanisms or signal distortion. The optimal approach involves selective isolation forest application while acknowledging wavelet methods' limited effectiveness for return series.

Analysis of model architectures reveals that moderately flexible single-layer neural networks, when integrated with isolation forest denoising, achieve superior risk-adjusted performance as evidenced by a 0.2189 monthly Sharpe ratio, substantially outperforming both simpler linear specifications and more complex architectural alternatives. This finding challenges conventional assumptions regarding the relationship between model complexity and predictive performance in financial applications. The observed performance hierarchy may be attributed to the relatively constrained dimensionality of our predictor space—comprising 24 financial variables—which appears to offer limited marginal benefits from increased architectural sophistication, potentially due to the risk of overfitting in finite financial samples.

The research also identifies significant performance differences between portfolio types, with most models experiencing substantial attenuation in market-neutral strategies, indicating better suitability for directional implementations. These findings provide computationally efficient approaches for retail investors while offering evidence-based guidance on denoising methodology selection and strategy implementation in emerging markets.

For retail investors seeking to implement this methodology, the most important practical recommendation is to prioritize the isolation forest anomaly detection step. By systematically removing the bottom 5% of stocks characterized by excessive trading activity and poor historical performance, investors can avoid significant underperformers and improve their overall portfolio returns, even without implementing the full machine learning forecasting framework.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest. Author Yingnan Yi is an employee of Beijing QITANG Education Consulting Inc., Beijing, China, and author Xue Gao is an employee of JD Health International Inc., Beijing, China.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Asness, C. S., Frazzini, A., & Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting studies*, *24*(1), 34-112. [CrossRef]

[2] Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The journal of finance*, *68*(3), 929-985. [CrossRef]

[3] Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, *116*(1), 111-120. [CrossRef]

[4] Barroso, P., & Wang, H. What Explains Momentum When It Really Works?. [CrossRef]

[5] Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of business*, *45*(3), 444-455. [CrossRef]

[6] Blitz, D., Hanauer, M. X., Vidojevic, M., & Van Vliet, P. (2016). Five concerns with the five-factor model. *Available at SSRN 2862317*. [CrossRef]

[7] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. [CrossRef]

[8] Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, *47*(5), 1731–1764. [CrossRef]

[9] Campbell, J. Y., Lo, A. W., MacKinlay, A. C., & Whitelaw, R. F. (1998). The econometrics of financial markets. *Macroeconomic Dynamics*, *2*(4), 559-562. [CrossRef]

[10] Chui, C. K. (1992). *An introduction to wavelets* (Vol. 1). Academic press.

[11] Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, *122*(2), 221–247. [CrossRef]

[12] Daniel, K., Hodrick, R. J., & Lu, Z. (2014). The carry trade: Risks and drawdowns (No. w20433). National Bureau of Economic Research. [CrossRef]

[13] Daubechies, I. (1992). *Ten lectures on wavelets*. Society for industrial and applied mathematics.

[14] Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *biometrika*, *81*(3), 425-455. [CrossRef]

[15] Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal Of Portfolio Management*, *30*(5), 15-29. [CrossRef]

[16] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3-56. [CrossRef]

[17] Fernandez, V. (2005). The international CAPM and a wavelet-based decomposition of value at risk. *Studies in Nonlinear Dynamics & Econometrics*, *9*(4). [CrossRef]

[18] Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, *33*(5), 2326–2377. [CrossRef]

[19] Gallegati, M. (2012). A wavelet-based approach to test for financial market contagion. *Computational Statistics & Data Analysis*, *56*(11), 3491–3497. [CrossRef]

[20] Gencay, R., Selcuk, F. & Whitcher, B. (2001) Wavelets and Other Filtering Methods in Finance and Economics. Academic Press, New York.

[21] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223-2273. [CrossRef]

[22] Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, *48*(1), 65–91.

[23] Cheng, X., Liao, Z., & Schorfheide, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, *83*(4), 1511-1543. [CrossRef]

[24] Lintner, J. (1969). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets: A reply. *The review of economics and statistics*, 222-224. [CrossRef]

[25] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In *2008 eighth ieee international conference on data mining* (pp. 413-422). IEEE. [CrossRef]

[26] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *6*(1), 1-39. [CrossRef]

[27] Zhang, J., Zhang, W., Schwab, A., & Zhang, S. (2017). Institutional environment and IPO strategy: A study of ChiNext in China. *Management and Organization Review*, *13*(2), 399-430. [CrossRef]

[28] Mallat, S. G. (2002). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, *11*(7), 674-693. [CrossRef]

[29] Meyer, Y. (1992). *Wavelets and operators* (No. 37). Cambridge university press.

[30] Moritz, B., & Zimmermann, T. (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Available at SSRN 2740751*.

[31] Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, *104*(2), 228–250. [CrossRef]

[32] Ramsey, J. B., & Lampart, C. (1998). Decomposition of economic relationships by timescale using wavelets. *Macroeconomic dynamics*, *2*(1), 49-71. [CrossRef]

[33] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*(3), 425-442. [CrossRef]

[34] Hirshleifer, D., Hsu, P. H., & Li, D. (2013). Innovative efficiency and stock returns. *Journal of financial economics*, *107*(3), 632-654. [CrossRef]

[35] Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508. [CrossRef]

[36] Xue, P., & Yi, Y. (2025). Integrating context compression and structural representation in large language models for financial text generation. *Journal of Computer Technology and Software*, *4*(9). [CrossRef]

[37] Yi, Y. (2019). Machine Learning and Empirical Asset Pricing. [CrossRef]

Yingnan Yi received Doctor in Business Administration degree from Washington University in St. Louis, where he also earned two Master's degrees in Accounting and Finance (Mathematics) respectively. He completed his undergraduate studies in Economics at Xiamen University in 2007. Dr. Yi is a Chartered Financial Analyst and holds inactive memberships in both the Chinese and American Institutes of Certified Public Accountants. (E-mail: yiyingnan2021@163.com)

Xue Gao received the B. Eng. degree in engineering from the University of Science and Technology Beijing, China, in 2009. She subsequently spent a decade as a researcher at the then Ministry of Railways, where she accumulated extensive research experience, before joining JD Health International Inc., where she has been a senior researcher for the past five years. (E-mail: gaoxue1207@163.com)