Title page

Type of Study: Original paper

Title:

Research on Hierarchical Futures Pair Trading Strategy Based on Machine Learning and Kalman Filter

Authors:

Yang Shuo[1,a], Huang Ke[2,b], Chen Yao-wen[3,c*]

[1] *School of Economics, Guangxi University, Nanning, 530000, Guangxi, China*
[2] *School of Digital Economics, Nanning University, Nanning 530000, Guangxi, China*
[3] *School of ASEAN, Guangxi University for Nationalities, Nanning 530000, Guangxi, China*
[a] *E-mail:* rainseed@foxmail.com
[b] *E-mail:* kateghuang@foxmail.com
[c] *E-mail:* yaowen.chen@foxmail.com
[*] Corresponding author

keywords: Pair Trading, Kalman Filtering, Half-life, Dynamic Time Warping, Cointegration Theory

Funding Acknowledgment:
The authors gratefully acknowledge the financial support from the National Social Science Foundation of China for the Key Project titled "Research on the Deepening Cooperation Mechanism Between China and ASEAN Under the Belt and Road Initiative and Uncertainty Shocks" (Project No. 22AGJ014).

Conflict of interest disclosure: The authors declare that there is no conflict of interest regarding the publication of this paper.

Data availability statement: The authors will supply the relevant data in response to reasonable requests.

Research on Hierarchical Futures Pair Trading Strategy Based on Machine Learning and Kalman Filtering

Abstract: Pair trading, as a significant arbitrage strategy, profits from the residuals between asset pairs. This study aims to explore the application of machine learning and Kalman filtering techniques in pair trading to optimize traditional strategies, enhancing both the return and stability of the strategy. Specifically, we address two issues: (1) how to improve the pairing efficiency and stability of assets, and (2) how to reduce noise interference and generate dynamic trading signals that enhance returns. Empirical results demonstrate that the pair trading strategy based on machine learning and Kalman filtering holds a significant advantage in hedge trading. This strategy enhances the cointegration pairing efficiency using the DTW clustering algorithm and employs Kalman filtering to eliminate noise, successfully increasing the accuracy and stability of trading signals. Additionally, utilizing the half-life as a sliding window to calculate the z-value (trading signal) allows the strategy to more flexibly seize trading opportunities. The empirical results indicate that this strategy has achieved favorable returns in backtesting, with robustness and effectiveness that can provide investors with valuable strategic optimization insights.
keywords： Pair Trading, Kalman Filtering, Half-life, Dynamic Time Warping, Cointegration Theory

Fund Project: This work was supported in part by the National Natural Science Foundation of China under Grant 71763004, in part by the Guangxi University Young and Middle-Aged Teachers Research Basic Ability Improvement Project under Grant 2022KY0644, in part by the Guangxi Philosophy and Social Science Planning Research Project under Grant 22CJY003, in part by the Guangxi First-Class Discipline Applied Economics Construction, and in part by the Guangxi Big Data Analysis of Taxation Research Center of Engineering Construction Project.

1.Introduction

Quantitative investing, based on mathematical analysis, involves issuing buy and sell orders through computer programs to achieve stable returns. It thrives in the complex, volatile, and information-saturated financial markets by seeking controlled and stable investment returns. The development and application of quantitative investing have gradually evolved due to the complexity of financial markets, advances in information technology, increased market efficiency, enhanced risk control demands, and the emergence of large-scale investments. The complexity of financial markets has made traditional investment decision-making, which relies on experience and market understanding, increasingly difficult. In response, quantitative investing offers new possibilities by reducing the complexity of investment decisions through algorithms and models.

Pair trading, an essential quantitative investment strategy, has garnered attention due to the complexities of financial markets, advancements in information technology, increased market efficiency, and heightened risk control needs. In financial markets, there are stable relationships between various assets, but identifying these relationships is not intuitive and requires complex algorithms and extensive data analysis. The core strategy of pair trading is based on identifying these stable relationships, using the price deviations of cointegrated asset pairs for profit. This involves buying and selling two correlated assets that are relatively overvalued and undervalued, respectively, to achieve arbitrage and minimize market volatility's impact on portfolio returns. However, traditional pair trading strategies face several challenges in practice, such as difficulty in asset selection and unstable trading signals.

The success of trading strategies depends on two critical steps. First, it involves analyzing the price time series of assets to identify suitable asset pairs, forming a profitable asset pool. Second, it involves monitoring the price spread fluctuations between asset pairs to capture appropriate trading signals for opening and closing positions profitably.

However, increased market efficiency, growing data volumes, and the impact of market noise pose challenges in finding stable pairs and determining reasonable trading opportunities. This study aims to address two issues: (1) how to improve the pairing efficiency and stability of assets; (2) how to reduce noise interference and generate dynamic trading signals that can improve returns.

Unlike traditional pairs trading strategies, this study plans to integrate modern machine learning techniques to enhance the performance of pairs trading. We intend to use Dynamic Time Warping (DTW) combined with the k-means clustering algorithm to identify potentially correlated asset classes. DTW helps in recognizing the lead-lag effects in time series, which aids in uncovering the potential connections between futures contracts. The k-means clustering, an efficient unsupervised clustering algorithm, will cluster futures contracts based on the DTW matrix, followed by cointegration pairing to reduce false pairings and save computational power (thereby improving the efficiency and stability of asset pairing). Additionally, we plan to use Kalman filtering, based on a half-life sliding window, to eliminate noise, determine dynamic hedge ratios, and calculate dynamic trading signals. This will make our trading signals more timely and flexible (thus reducing noise interference and generating dynamic trading signals that can improve returns). The objective of this study is to design and empirically test a pairs trading strategy that performs exceptionally well in the Chinese commodity futures market. By introducing modern data mining and machine learning techniques, we hope to further enhance the performance of traditional pairs trading and provide valuable references for optimizing pairs trading strategies.

The remainder of this paper is organized as follows: Section II reviews the related research achievements of scholars both domestically and internationally, sorting through the current research status from the aspects of pair asset selection and trading timing determination, absorbing valuable research methods, and exploring directions for improvement to provide references for our study. Section III introduces the construction process of the pair trading strategy based on machine learning and Kalman filtering, and its basic theory. Section IV integrates historical data of futures contracts from the Chinese futures market into the pair trading model, constructs an optimized dynamic pair trading model, verifies the optimization effects of the model, and adjusts the model parameters adaptively. Section V tests the backtesting results through a validation set and evaluates the effectiveness of the trading model. Section VI analyzes and summarizes the empirical results, explores the research's flaws and shortcomings, and provides directions for future research improvement.

2. Reviews

The Efficient Market Hypothesis (EMH), a landmark theory introduced by Fama in 1965 through empirical analysis, rapidly became a core component of modern financial studies and spurred the development of the Mean Reversion Theory of Tradable Securities, which is a key presupposition of statistical arbitrage strategies. The theory posits that some degree of market efficiency is necessary to ensure that spreads follow mean reversion.

The success of pair trading hinges on selecting appropriate asset pairs and generating accurate trading signals, starting with the formation of a suitable asset pool. Typically, the selection of paired assets utilizes cointegration methods, but prior to this, it is essential to cluster similar target contracts. Big data techniques for classifying stocks have proven effective in identifying stocks with similar volatility. This approach aims to group similar stocks together while maximizing the differences between groups. In the financial field, clustering based on stock eigenvalues can reveal interrelations among different data characteristics.Nevertheless, these techniques often fail to fully consider the time-series nature of stock prices and their historical correlations, posing challenges in analyzing time-series data. Techniques used in stock clustering are not limited to methods based on eigenvalues but also include strategies based on raw data or models. Clustering using raw data can directly reflect the fundamental characteristics of the data, and adjusting the distance measurement methods within clustering models can achieve clustering based on data models. For instance, Dynamic Time Warping (DTW) has been proposed to measure the similarity between time series data. Berndt and Clifford (1994) and Paparrizos and Gravano (2015) both employed the DTW method in their research. Recently, Chen (2022) improved the traditional Euclidean distance calculation in the K-means algorithm by using DTW as a new metric to cluster time series of bank stock prices, identifying stocks with price correlations.

Pair trading has become a mainstream quantitative trading strategy widely applied in investment banks and hedge funds globally. This strategy's core idea is to exploit statistical relationships between two or more assets for arbitrage. Vidyamurthy (2001) outlined that constructing a pair trading strategy involves three basic steps: firstly, pairing assets based on fundamental analysis and their interrelationships; then, applying the Engle and Granger (1987) cointegration test method (EG two-step method) to assess the tradability of these assets during the strategy construction period, with a particular focus on the mean-reverting characteristics within the cointegration relationship; and finally, formulating entry and exit signals using non-parametric methods. This strategy exploits market inefficiencies by buying one asset and simultaneously selling another related asset, profiting from the variation in price differences. The strategy's success depends on precise market analysis and strict risk management measures.

Gatev, Goetzmann, and Rouwenhorst (GGR) (2006) utilized daily stock data from 1962 to 2022 to study the pair trading strategy, applying the distance method to filter potential paired stocks. This method first removed suspended stocks and standardized the remaining stock prices, selecting paired assets by minimizing differences in price time series, achieving an annualized return of 11%. However, Alexander (2001) criticized this method for potentially leading to pseudo-correlation, as the distance-based pairing did not consider the cointegration relationship, possibly causing price differences to diverge and lose the mean-reverting attribute. Additionally, Do and Faff (2001) found that 32% of stock pairs using the distance method exhibited diverging price differences during validation, failing to demonstrate strict mean reversion. In contrast, Huck and Afawubo (2015) showed that assets paired using the cointegration method had stronger mean-reverting properties, while those determined by the distance method were relatively weaker in mean reversion. These studies reveal the strengths and risks of using different methods for pairing stocks in pair trading strategies. While the distance method can achieve significant returns in some cases, its stability and reliability may not be as strong as methods based on cointegration relationships. This emphasizes the importance of selecting appropriate pairing methods in applying pair trading strategies to reduce risk and enhance investment returns.

Rad et al. (2016) built on the work of Gatev, Goetzmann, & Rouwenhorst (GGR) and Vidyamurthy, constructing a pair trading strategy using American CRSP stock data combined with the GGR method and cointegration theory, achieving an annualized excess return of 9.96%, under conditions ignoring transaction costs like fees and slippage. The study noted potential selection biases when combining the distance method with cointegration methods. Meanwhile, Huck and Afawubo (2015) found that the cointegration method had advantages over the distance method based on S&P 500 constituent stock data. Additionally, Li et al. (2014) constructed a cointegration arbitrage strategy for stocks dual-listed in the A-shares and Hong Kong markets, achieving an annual average excess return of 17.6%, demonstrating the potential and practical value of the cointegration method in cross-market arbitrage strategies.

The Kalman filter, initially applied in the aerospace sector, has increasingly found applications in economics over time, given its wide applicability and flexibility. Relying on a state-space model, which includes state and observation equations, the Kalman filter can recursively estimate and update system states without assuming the stationarity of data and noise. This characteristic makes the Kalman filter particularly suited to handling dynamically changing data series and noisy observations. Over the past decades, Kalman filtering and its improved algorithms have achieved significant success in various research fields. Chen (1993) combined the Modified Extended Kalman Filter (MEKF) with feedforward neural networks, demonstrating its superiority in convergence speed and prediction accuracy.

In summary, the Efficient Market Hypothesis and the weak-form efficiency of capital markets form the theoretical basis for statistical arbitrage strategies, aiding researchers in exploring long-term market equilibrium relationships. Research in statistical arbitrage primarily focuses on basic theory exploration and strategy optimization, forming a mature research system, including distance methods, cointegration methods, and time series analysis. These strategies have been empirically validated and continuously optimized, but they may face issues of overfitting and pseudo-pairing in practice, making data preprocessing and correct testing contract selection crucial.

The Kalman filter, due to its effectiveness in eliminating market noise and restoring true price values, is particularly important in financial market analysis. Combined with machine learning, the Kalman filter algorithm is applied in futures pair trading strategies, providing new perspectives and methods, enhancing the adaptability, profit potential, and risk management capabilities of strategies. The application of this methodology promises to improve existing trading systems, enhancing their efficiency and accuracy.

3 Research Design

3.1 Data Selection

In the quantitative research and strategy development of the futures market, selecting the correct data sources and types is crucial to ensure the accuracy of analysis and the effectiveness of strategies. This study has obtained the time series data of dominant continuous contracts through the JoinQuant quantitative platform. The reasons for choosing dominant continuous contracts include:

1.Continuity perspective: As futures contracts have a fixed expiration date, dominant continuous contracts merge contracts of different months to provide a seamless price time series, which facilitates technical analysis and model backtesting.

2.Representation of mainstream market trends: Dominant contracts, usually those with the highest trading or holding volumes, effectively reflect the primary activities of the market.

3. Simplification of data processing: Using dominant continuous contracts avoids the complexities of handling the rolling over and price adjustments associated with single expiration date contracts.

4. Support for long-term analysis: Dominant continuous contracts offer long-term data, which aids in analyzing market trends and identifying historical patterns.

Although dominant continuous contracts offer convenience for research, they also pose challenges. The dominant contract is calculated based on a position-weighted formula, which may not be suitable for price-sensitive strategies. Moreover, the dominant contract itself cannot be directly traded and only serves as an analytical tool.

When applying strategies in practice, it is necessary to adjust the strategies for specific expiration month contracts and consider the conversion logic and costs. Overall, despite the challenges, dominant continuous contracts have significant advantages in various aspects and serve as an effective data source for futures market research and strategy development.

The data for this study originates from China's five major futures exchanges: China Financial Futures Exchange (CFFEX), Shanghai Futures Exchange (SHFE), Dalian Commodity Exchange (DCE), Zhengzhou Commodity Exchange (ZCE), and China International Energy Exchange (INE). These exchanges cover multiple sectors including finance, industrial products, agricultural products, and energy. The dataset encompasses daily-level data from January 4, 2016, to November 20, 2023. This range not only captures mid-to-long term market trends but also spans multiple economic cycles, which helps to enhance the stability and universality of the strategy.

The futures contracts involved in the study include stock indices, agricultural products, energy, and metals, with a primary focus on closing prices, reflecting the market's final pricing on each trading day. Data organization and analysis form the foundation of this research, involving data cleaning, outlier handling, standardization, and normalization to accommodate advanced analytical methods such as machine learning and Kalman filtering.

The sample is divided into a clustering set, a testing set, and a validation set to support different stages of analysis and model evaluation:

1.The clustering set is used for Dynamic Time Warping (DTW) and k-means clustering to identify assets with similar price movements, providing a basis for cointegration pairing.

2.The testing set is utilized for backtesting the model and adjusting parameters to evaluate the performance of the trading strategy.

3.The validation set is employed for the final model testing to assess the stability of the strategy and its adaptability to unseen data, helping to prevent overfitting.

These steps ensure the practicality and applicability of the research findings while considering the real-world constraints of trading costs and market liquidity. This comprehensive approach to backtesting not only validates the effectiveness of the pair trading strategy under various market conditions but also highlights areas for potential optimization and adjustment.

3.2 Pair Selection

The generation of our paired asset pool is a two-step process. Initially, after data processing, futures contracts undergo Dynamic Time Warping (DTW) and clustering, which groups assets with similar time series into different families. Subsequently, within each family, futures contracts are paired using cointegration techniques, forming our final paired asset pool.

In this study, we use the K-means clustering algorithm to group futures contracts, aiming to enhance pairing precision and efficiency, thereby avoiding exhaustive cointegration calculations on the entire sample. First, the price series of each futures contract is processed with Dynamic Time Warping (DTW).

To construct the time series matrix, we build it based on two time series, with each matrix element representing the Euclidean distance between adjacent data points. Then, through backtracking, we identify the shortest path and compute the cumulative values to obtain the DTW distance.

We calculate the DTW distance for each pair of futures contracts. This process results in a distance matrix reflecting the similarity in price fluctuation patterns between the futures contracts. Subsequently, we use this distance matrix as input for the K-means clustering algorithm to perform cluster analysis. This clustering method based on K-means effectively groups together futures contracts with similar price fluctuation patterns. It provides an efficient preliminary screening tool for constructing pair trading strategies and selecting futures contract pairs with cointegration relationships.

Determining cointegration relationships is a crucial step in our research. Once we identify potential futures contract pairs through DTW and K-means clustering, we need to further verify the presence of stable cointegration relationships between these pairs. In this regard, we employ the Engle-Granger two-step method for cointegration testing.

The Engle-Granger two-step method was proposed by Nobel laureates Robert F. Engle and Clive W.J. Granger in 1987. It provides an effective approach for detecting cointegration relationships in nonstationary time series. Engle and Granger demonstrated that if two nonstationary time series exhibit a long-term equilibrium relationship, some linear combination of them can be stationary. This phenomenon is known as cointegration. In financial markets, the existence of cointegration relationships is often seen as a manifestation of market efficiency, as any deviation from equilibrium prices would trigger arbitrage activities to correct towards the equilibrium level.

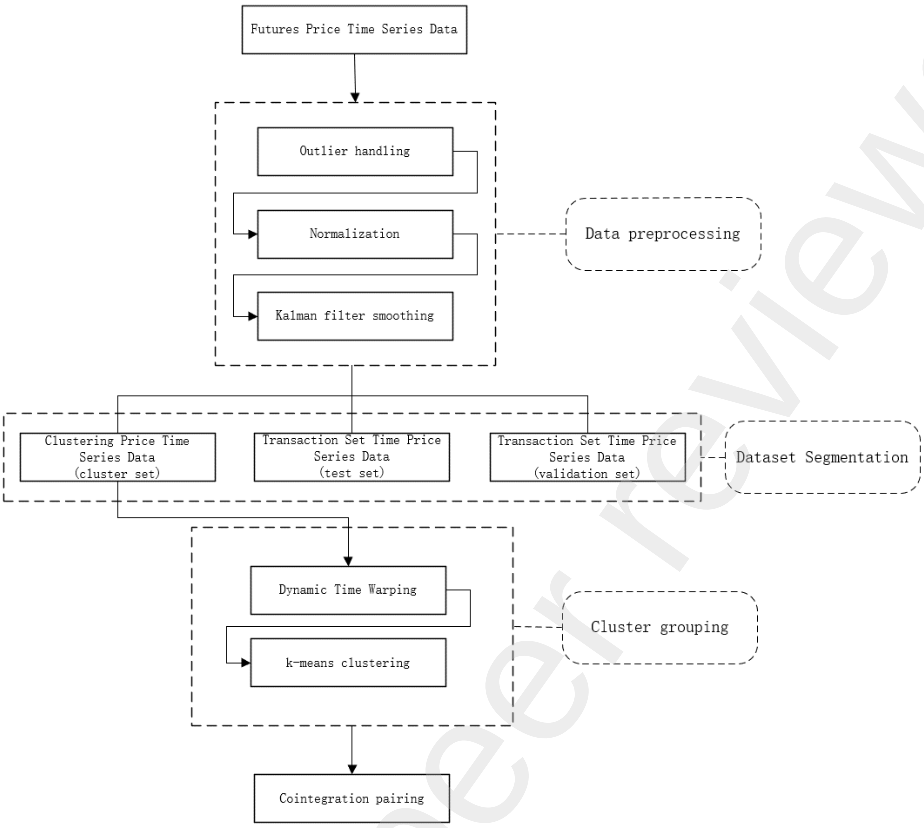

Figure1 Pairs selection flowchart

3.3 Trading Model

The determination of holding periods. This study incorporates Kalman filtering to enhance the stability and accuracy of the strategy. The filter uses a recursive algorithm to estimate system states amidst data noise, making it well-suited for dynamic and uncertain financial market environments.

Kalman filtering operates through two core steps: prediction and updating. It recursively updates state estimates and covariance matrices, accurately adjusting hedge ratios to adapt to market changes. This method not only excels in analyzing price series of financial contracts but also enhances strategy responsiveness and market adaptability through dynamic updates of model parameters. Additionally, applying a sliding window of the half-life length to Kalman filtering allows the strategy to dynamically adjust based on market volatility, improving trading efficiency and capital utilization. The theory posits that some degree of market efficiency is necessary to ensure that spreads follow mean reversion.

In our pair trading strategy, Kalman filtering is used not only to reduce noise and predict the true state of the vector space but also to estimate the hedge ratio for two asset price series. This ratio is dynamic and requires real-time estimation for making trading decisions. The Kalman filter effectively helps us achieve this goal.

For the "special linear Gaussian state space model":

$$y_t = \mu_t + \varepsilon_t, \varepsilon_t \sim N\left(0, \delta_\varepsilon^2\right) \quad (1)$$

$$\mu_{t+1} = \mu_t + \eta_t, \eta_t \sim N\left(0, \delta_\varepsilon^2\right) \quad (2)$$

where εt and ηt are both white noise sequences, and t = 1, 2, …, T, and so on.

Equation (1) represents the measurement error model with measurement error εt, and Equation (2) determines the characteristics of the state variable ηt's time variation, referred to as the state equation of ηt. Equation (1) predicts the variable based on the past's linear combination and current disturbances, while Equation (2) represents the refreshment of variable μt based on past predictions of the next period's variable μt +1 in the presence of random fluctuations. Considering that both variable yt and μt are multidimensional, the state space model can be applied in various fields such as time-varying Capital Asset Pricing Model (CAPM).

Applying the state space model to the time-varying CAPM can be written in the following form:

$$r_t = [1 r_m] \times \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + \theta_t, \theta_t \sim N\left(0, \sigma_\theta^2\right) \quad (3)$$

$$\begin{bmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \varepsilon_t \sim N\left(0, \delta_\epsilon^2\right), \eta_t \sim N\left(0, \delta_\varepsilon^2\right) \quad (4)$$

The purpose of the Kalman filter method is to solve the above state space model, essentially obtaining updated data through a series of recursive algorithms after acquiring new data. That is, obtaining new conditional distributions and data yt given μt and other historical conditions. The Kalman filter process can be summarized using the following five equations:

$$x(k \mid k-1) = AX(k-1 \mid k-1) + BU(k) \quad (5)$$

$$P(k \mid k-1) = AP(k-1 \mid k-1)A' + Q \quad (6)$$

$$X(k \mid k) = Kg(k)(Z(k) - HX(k \mid k-1)) \quad (7)$$

$$Kg(k) = P(k \mid k-1)H'/(HP(k \mid k - 1)H' + R \quad (8)$$

$$P(k \mid k) = (1 - Kg(k)H)P(k \mid k-1) \quad (9)$$

Equation (5) estimates x(k|k-1) using coefficient matrix A and control parameter B. Equation (6) estimates P(k|k-1) based on the previous period's covariance matrix

P(k-1|k-1). Kg represents the Kalman gain, showing the weight of the variance of the parameter estimation. Equation (7) corrects X(k|k) based on the introduction of information (H) for re-estimating X. Finally, re-estimating P(k|k) completes a full filtering process, and the next task is to refresh this process for t = 1, 2, …, T.

To enhance capital utilization efficiency and reduce the risk of divergence in spreads, we employ a delayed entry strategy to optimize trading signals. When analyzing the backtesting results of pairs trading, in addition to the annualized return and maximum drawdown, the average duration of individual trades is a crucial metric. Although the single trade profit in pairs trading is locked in at the moment of entry, the waiting time for exit is uncertain. If the timing strategy can minimize the duration of individual trades, the utilization efficiency of capital in the actual application of the strategy would be significantly higher than the backtesting result. Here, we adopt a delayed entry approach to reduce the time occupied by individual trades. Delayed entry refers to opening a position only when the spread breaks below the entry threshold. If the spread only breaks above the entry threshold, we maintain a short position. Furthermore, the delayed entry strategy can also reduce the risk of spread divergence, as we only establish a position when the spread crosses the entry threshold in the opposite direction. If the spread crosses the entry threshold and does not revert, we maintain a short position.

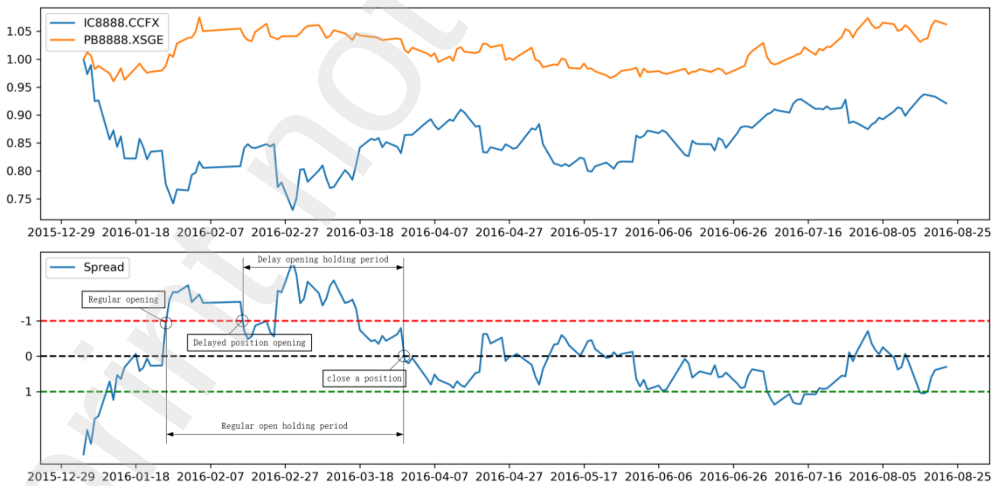

Figure 2. Schematic diagram of the principle of pair trading

Finally, according to the different opening and closing thresholds, the optimal trading signal is determined for opening and closing operations and backtesting results are obtained.

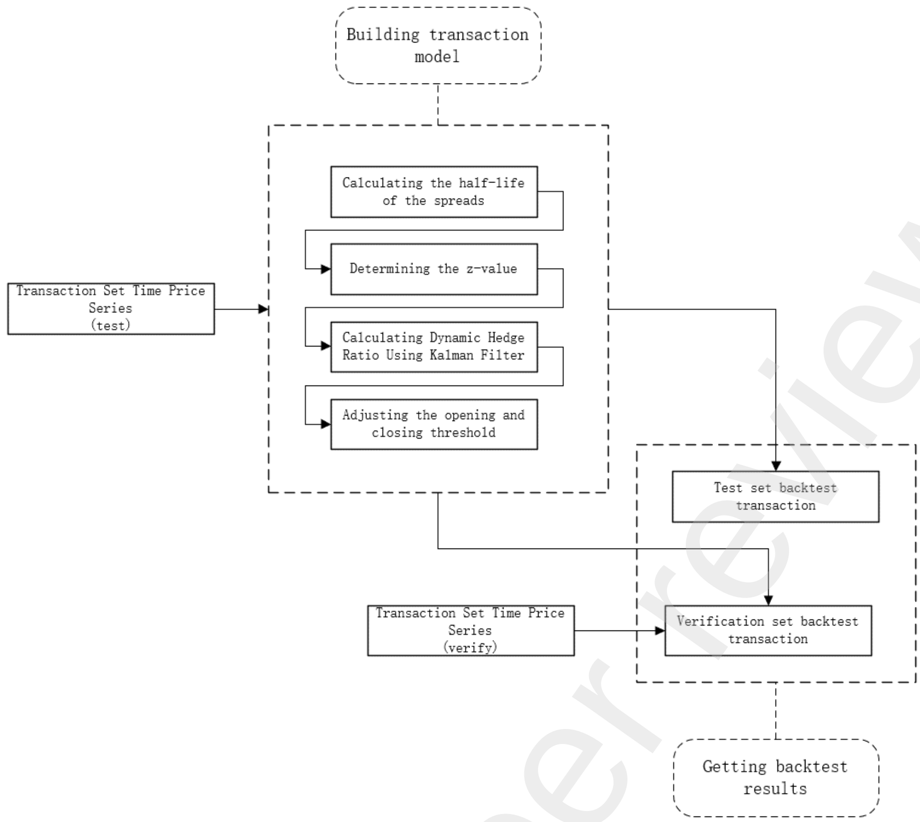

Figure 3. Simulated trading flowchart

4.Backtesting Analysis

4.1. Clustering and Cointegration

Before starting data analysis and model construction, it is necessary to clean and preprocess the raw data. This step helps eliminate erroneous or inaccurate data and ensures that our model is based on accurate and reliable information.

Firstly, data cleaning focuses on identifying and addressing errors and anomalies in the data, which includes but is not limited to checking and handling missing values, verifying data consistency, and removing duplicate records. For missing data, depending on the duration of absence, methods such as interpolation or forward filling are used to maintain data continuity; if the absence is significant, the corresponding time series may be removed.

Secondly, before conducting DTW, we need to normalize the clustering set data to ensure consistent data scales, enhance algorithm convergence speed, and improve model robustness. Different variables may have significantly different value ranges, and normalization can bring them to the same scale. This helps eliminate biases caused by different scales and makes the variables comparable during calculations. Most machine learning algorithms are sensitive to data scales, and larger-scale data can cause slower or unstable algorithm convergence. Normalization ensures that the data is within an appropriate range, thereby improving algorithm convergence speed and stability. In dynamic time warping and clustering, sensitivity to outliers or extreme values is high.

Through normalization, we can restrict the data distribution range to a smaller interval, reducing the influence of outliers or extreme values and enhancing the robustness of the model.

Lastly, before conducting backtesting for pair trading, we need to normalize the trading set data based on relative returns to eliminate scale differences. Due to the magnitude differences among different contracts, it is challenging to analyze their price trends directly. Relative return normalization is performed by dividing each data point by the value at the first time point, transforming the time series data into growth proportions relative to the initial value. This approach ensures that all data points are benchmarked against the initial value, enabling better comparison and observation of changes between different time points.

Next, we perform feature clustering on the futures contracts. The results of K-means clustering can be influenced by the selection of initial centroids. Therefore, in practical applications, we typically perform multiple random initializations to select the optimal clustering result. Additionally, the choice of K value requires careful consideration, as choosing a value that is too large can lead to excessive dispersion, while choosing a value that is too small can result in a loss of discrimination. In this study, we utilize the Silhouette Score to determine the optimal K value. The Silhouette Score provides us with a clustering evaluation method that is not dependent on the specific number of clusters and assists in selecting the best number of clusters. By calculating the Silhouette Score for different cluster numbers, we can choose the cluster number with the highest score as the optimal number of clusters.

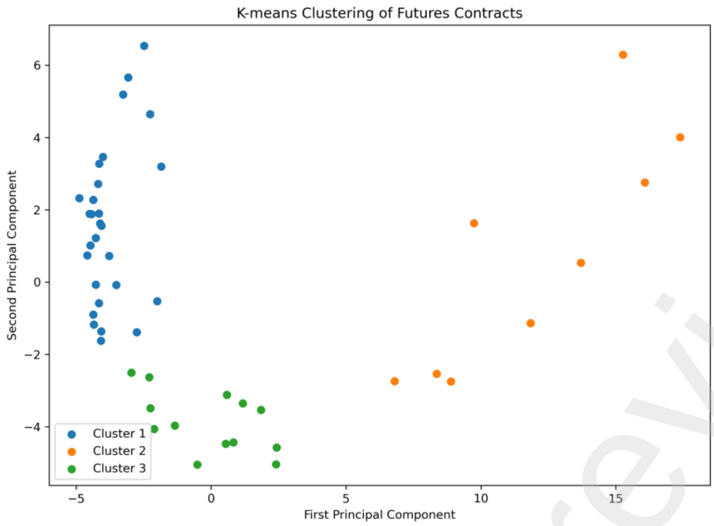

Figure 4. *K*-means clustering of future contracts

We will utilize the Silhouette Score to measure the quality of clustering. It takes into account both the compactness within clusters and the separation between clusters. The Silhouette Score ranges from -1 to 1, where a value close to 1 indicates reasonable clustering, a value close to -1 suggests that samples might have been wrongly assigned to other clusters, and a value close to 0 implies that samples are on the boundary between two clusters.

After the completion of K-means clustering, we obtained the grouping results for futures contracts, which will be utilized for subsequent pair selection and pair trading strategy development. This grouping method based on DTW and K-means clustering not only considers the price movement patterns between futures contracts but also effectively uncovers the underlying similarities among them, providing us with an effective tool for futures contract screening.We employ the Engle-Granger two-step method to determine asset pairs.

In pair trading strategies, we aim to identify such cointegration relationships. If the price sequences of two futures contracts are cointegrated, we can engage in pair trading when prices deviate from their long-term equilibrium, capitalizing on mean reversion characteristics to generate profits.

We set the significance level for the cointegration test at 5%. The results of the cointegration test are presented below.

Table 1. Cointegration test results

Index	Asset1	Asset2	co-integration	cluster
1	IF8888.CCFX	IH8888.CCFX	0.0372	1
2	IF8888.CCFX	SF8888.XZCE	0.0001	1
3	IF8888.CCFX	BB8888.XDCE	0.0131	1
4	IH8888.CCFX	SF8888.XZCE	0.0104	1
5	AL8888.XSGE	WR8888.XSGE	0.0083	1
6	AL8888.XSGE	FG8888.XZCE	0.0098	1
7	AU8888.XSGE	M8888.XDCE	0.0274	1
8	NI8888.XSGE	ZC8888.XZCE	0.0273	1
9	NI8888.XSGE	J8888.XDCE	0.0499	1
10	NI8888.XSGE	JM8888.XDCE	0.0265	1
11	SN8888.XSGE	JM8888.XDCE	0.0441	1
12	FG8888.XZCE	ZC8888.XZCE	0.0124	1
13	ZC8888.XZCE	J8888.XDCE	0.0268	1
14	A8888.XDCE	I8888.XDCE	0.0451	1
15	IC8888.CCFX	L8888.XDCE	0.008	2
16	JR8888.XZCE	SR8888.XZCE	0.020	2
17	RI8888.XZCE	SR8888.XZCE	0.015	2
18	RI8888.XZCE	WH8888.XZCE	0.018	2
19	RI8888.XZCE	L8888.XDCE	0.013	2
20	PB8888.XSGE	MA8888.XZCE	0.042	3
21	PB8888.XSGE	PP8888.XDCE	0.028	3
22	ZN8888.XSGE	RS8888.XZCE	0.046	3
23	LR8888.XZCE	RS8888.XZCE	0.015	3
24	SM8888.XZCE	TA8888.XZCE	0.028	3
25	SM8888.XZCE	PP8888.XDCE	0.045	3

We utilize heat maps to visually represent the level of cointegration among assets. In the heat map, darker colors indicate stronger cointegration.

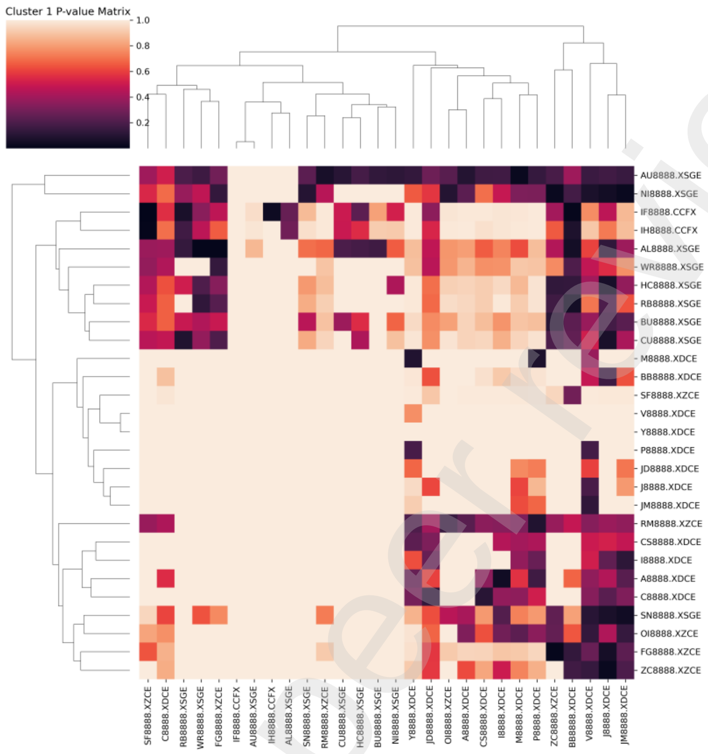

Figure 5. The heat map of Cluster I cointegration

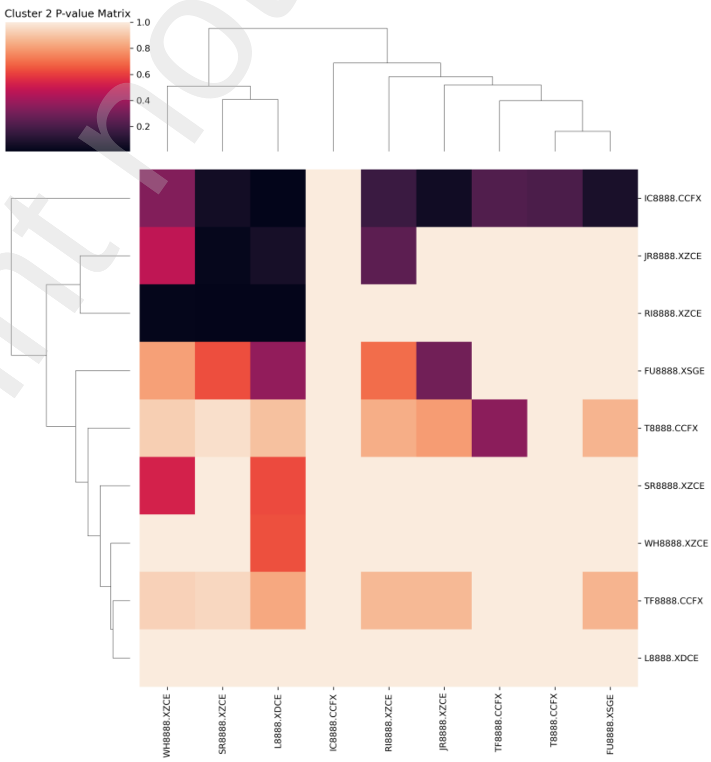

Figure 6. The heat map of Cluster II cointegration

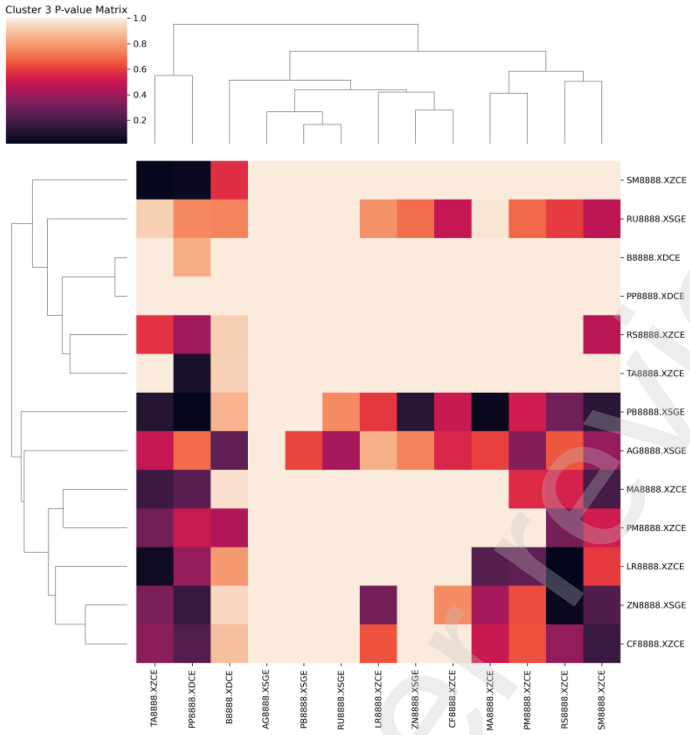

Figure 7. The heat map of Cluster III cointegration

4.2. Simulated Trading

4.2.1. Determining Trading Signals

(1) Calculating the half-life: The half-life can be used as a method to determine an appropriate moving window size. The half-life measures the speed at which autocorrelations decay in a time series. It is commonly employed to assess whether a sequence exhibit mean reversion. If mean reversion is present, the half-life represents the number of periods required for the sequence to regress halfway back to its long-term mean after deviating from it.

By setting the moving window size equal to the half-life, the data within the window can be regarded as the most recent, most correlated, and most reflective of the current state. This choice of window size ensures that when calculating z-values, we consider the most relevant and recent data, thus better reflecting the current market conditions.

(2) Establishing a Kalman filter model: Based on the price data of paired assets, we can construct a Kalman filter model. This model allows us to track the price

relationship between the paired assets, which can be expressed as a linear regression model with time-varying parameters such as slope and intercept.

(3) Updating the Kalman filter model: As new price data arrives, it is necessary to continually update the Kalman filter model. This updating process can be achieved through the prediction-update cycle of the Kalman filter. Specifically, we first predict the state at the next time step and then update our prediction based on the actual observed data.

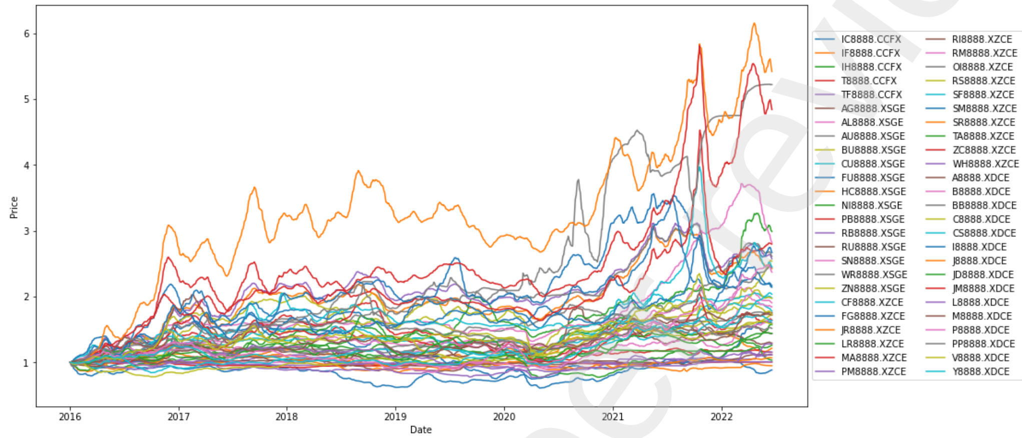

Figure 8. Filter data processing

(4) Computing the price spread and generating trading signals: With the assistance of the Kalman filter model, we can calculate the price spread of the paired assets and normalize it into z-values. Subsequently, we can establish certain thresholds to generate trading signals. When the z-value crosses above $-\hat{z_0}$ (i.e., when z-value transitions from being below -$\hat{z_0}$ at time t-1 to being above $-\hat{z_0}$ at time t), we initiate a long position by buying contract pair1 and selling contract pair2. We close the position when the z-value reaches $-\hat{z_c}$. Conversely, when the z-value crosses below $\hat{z_0}$. (i.e., when z-value transitions from being above $\hat{z_0}$. at time t-1 to being below $\hat{z_0}$. at time t), we initiate a short position by selling contract pair1 and buying contract pair2. We close the position when the z-value reaches about $\hat{z_c}$.

Average spread:

$$\bar{s} = (\sum_{i=1}^{hl} s_i)/hl \tag{10}$$

Standard deviation of spread:

$$\sigma = \sqrt{1/n \sum_{i=j}^{j+hl} (s_i - \bar{s})} \tag{11}$$

z-value:

$$z = \frac{s - \bar{s}}{\sigma} \tag{12}$$

Position opening threshold: $\hat{z}_0 = 1.25$

Closing Threshold: $\hat{z}_c = -0.08$

We identified the threshold values that yielded relatively high returns by experimenting with different opening and closing thresholds, and subsequently applied this set of thresholds in the out-of-sample testing.

4.3.2. Calculating the hedge ratio:

In pair trading, an appropriate hedge ratio is crucial for avoiding risk exposure in asset positions. Since different futures contracts have varying prices and margin ratios, it is necessary to dynamically adjust the position sizes of paired asset contracts to cover risk exposure.

The measurement equation for computing the dynamic hedge ratio using the Kalman filter is as follows:

$$Y_t = X_t \beta_t + n_t \tag{13}$$

Here, Yt and Xt represent the prices of the paired assets (observable variables), nt denotes the noise term following a normal distribution, and β represents the hedge ratio (unobservable variable). The state transition equation for the unobservable variable βt can be expressed as:

$$\beta_t = \beta_{t-1} + m_{t-1} \tag{14}$$

Here, βt-1 represents the hedge ratio at time t-1, and mt-1 is the noise factor following a normal distribution. Considering the conditional distribution of the unobservable variable when the observable variables are known at time t-1, we define:

State Prediction equation:

$$E(\beta_{t|t-1}) = E(\beta_{t-1|t-1}) \tag{15}$$

State Covariance Prediction equation:

$$\alpha_{t|t-1} \equiv \alpha_{t-1|t-1} + V_w \tag{16}$$

Here, $E(\beta_{t|t-1})$ represents the expected value of β at time t conditioned on t-1. $\alpha_{t|t-1}$= $cov[\beta_t - E(\beta_{t|t-1})\,]$ represents the corresponding prediction of the residual covariance of the unobservable variable, which is a 2x2 matrix. The equation defining the parameter Vm as δ/(δ-1)R is presented, where Vm denotes the posterior parameter and R represents a 2x2 identity matrix. The closer δ is to 1, the greater the impact of the latest observed values on the β. With known $E(\beta_{t-1|t-1})$ and $\alpha_{t-1|t-1}$, we proceed to the prediction phase:

Measurement Prediction equation:

$$E(\beta_{t|t-1}) = X_t E(\beta_{t-1|t-1}) \quad (17)$$

Measurement Variance Prediction equation:

$$F_t \equiv X'_t \alpha_{t|t-1} X_t + \tau \quad (18)$$

Here, F_t represents the variance of the expected residual terms, and εt is defined as:

$$\varepsilon_t \equiv T_t - X_e E(\beta_{t|t-1}) \quad (19)$$

Here, $\varepsilon_t$represents the deviation of Yt from the expected value at time t-1. After the prediction phase, we move to the update phase:

State Update equation:

$$E(\beta_{t|t}) = E(\beta_{t|t-1}) + K_t \varepsilon_t \quad (20)$$

State Covariance Update equation:

$$\alpha_{t|t-1} = \alpha_{t-1|t-1} - K_t X_t \, \alpha_{t-1|t-1} \quad (21)$$

Here, Kt = $\alpha_{t|t-1}$Xt/Ft represents the Kalman gain. When we set E(βt|t-1) = 0, as the observed variables are functions of unobservable variables that are correlated with noise, the above model iteratively updates the expected values of unobservable variables based on the most recent observed values.

By following the steps of the Kalman filter algorithm, we can estimate the unobservable variable βt, i.e., the dynamic hedge ratio, on a period-by-period basis. Furthermore, we use a posterior approach to fine-tune the parameters δ and τ and validate them using out-of-sample data. Through this process, we can implement a pair trading strategy based on the Kalman filter, which offers better adaptability and flexibility to market changes. We iterate through the pairings in the asset pool and calculate the returns for each pair.

4.2.3. Backtesting Results

For this experiment, 25 rounds of pair trading tests were conducted simultaneously, with two futures contracts paired in each round. A total of 30 futures contracts entered the pair trading asset pool. The distribution of these contracts across different exchanges was as follows: 8 from the Dalian Commodity Exchange (XDCE), 7 from the Shanghai Futures Exchange (XSGE), 12 from the Zhengzhou Commodity Exchange (XZCE) , 3 from the China Financial Futures Exchange (CCFX) , and 3 from the China Financial Futures Exchange (CCFX).No contracts from the Shanghai International Energy Exchange (XINE) or Guangzhou Futures Exchange (GFEX) were selected for inclusion.

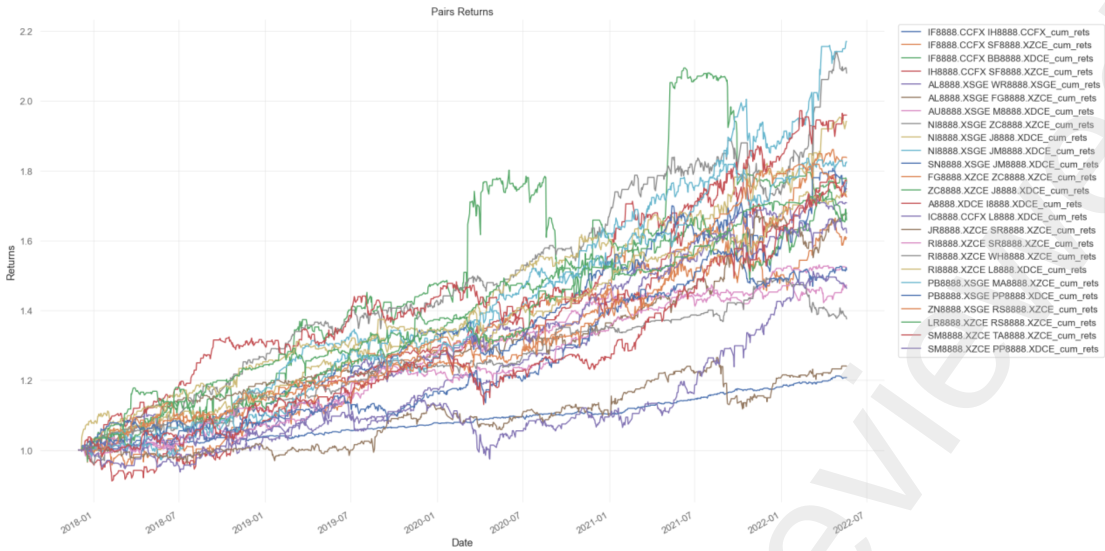

Figure 9. Cumulative yield on paired trades

Table 2. Pair trading backtesting results

Index	pairs	Sharpe ratio	CAGR
1	IF8888. CCFX——IH8888. CCFX	1. 93	0. 0295
2	IF8888. CCFX——SF8888. XZCE	1. 00	0. 0761
3	IF8888. CCFX——BB8888. XDCE	0. 63	0. 0920
4	IH8888. CCFX——SF8888. XZCE	1. 22	0. 0900
5	AL8888. XSGE——WR8888. XSGE	1. 50	0. 0862
6	AL8888. XSGE——FG8888. XZCE	1. 54	0. 0833
7	AU8888. XSGE——M8888. XDCE	1. 36	0. 0613
8	NI8888. XSGE——ZC8888. XZCE	1. 57	0. 1196
9	NI8888. XSGE——J8888. XDCE	1. 39	0. 1078
10	NI8888. XSGE——JM8888. XDCE	1. 70	0. 1270
11	SN8888. XSGE——JM8888. XDCE	1. 25	0. 0917
12	FG8888. XZCE——ZC8888. XZCE	1. 50	0. 0985
13	ZC8888. XZCE——J8888. XDCE	1. 21	0. 0812
14	A8888. XDCE——I8888. XDCE	1. 48	0. 1094
15	IC8888. CCFX——L8888. XDCE	1. 37	0. 0775
16	JR8888. XZCE——SR8888. XZCE	0. 61	0. 0338
17	RI8888. XZCE——SR8888. XZCE	1. 74	0. 0670
18	RI8888. XZCE——WH8888. XZCE	1. 15	0. 0504
19	RI8888. XZCE——L8888. XDCE	1. 85	0. 0893
20	PB8888. XSGE——MA8888. XZCE	1. 67	0. 0973
21	PB8888. XSGE——PP8888. XDCE	1. 27	0. 0663
22	ZN8888. XSGE——RS8888. XZCE	1. 46	0. 0878

23	LR8888.XZCE——RS8888.XZCE	1.50	0.0831
24	SM8888.XZCE——TA8888.XZCE	1.38	0.0924
25	SM8888.XZCE——PP8888.XDCE	0.95	0.0605

The trading results show that when the opening threshold is set to 1.25 times the standard deviation of the centralized price spread and the closing threshold is set to -0.08 times the standard deviation, the overall weighted annualized return is 12.964%, with a cumulative return of 67.068%. The alpha value is 0.13, the Sharpe ratio is 4.35, the maximum drawdown is -2.396%, the annualized volatility is 2.812%, and the beta value is 0.01. Compared to the benchmark CSI 300 Index, the strategy performed very stably over the 5-year backtesting period, resembling a nearly straight line.

Table 3. Evaluation index of backtesting results

Backtest _index	Annual return	Cumulative return	Alpha	Sharpe ratio	Max drawdown	Annual volatility	Beta
Value	12.964%	67.068%	0.13	4.35	-2.396%	2.812%	0.01

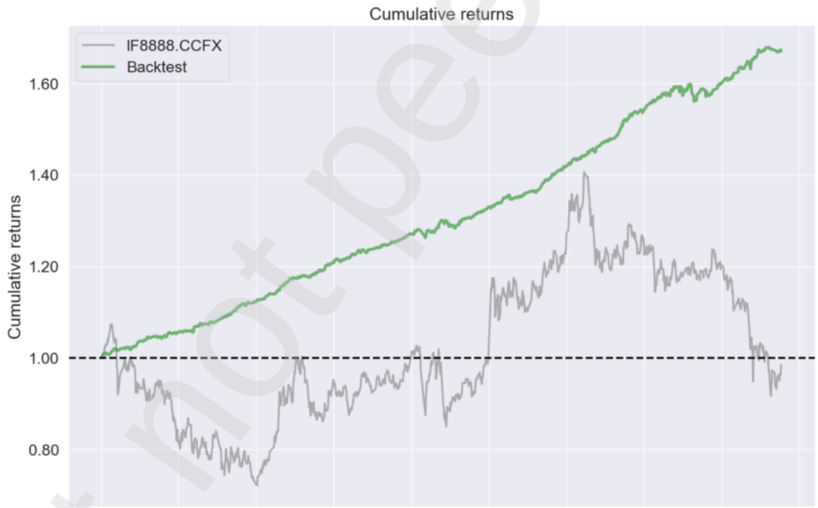

Figure 10. Cumulative returns

During the backtesting period, there were a total of five significant drawdowns. The most severe occurred from October 9, 2021, to December 27, 2020, lasting 50 days.

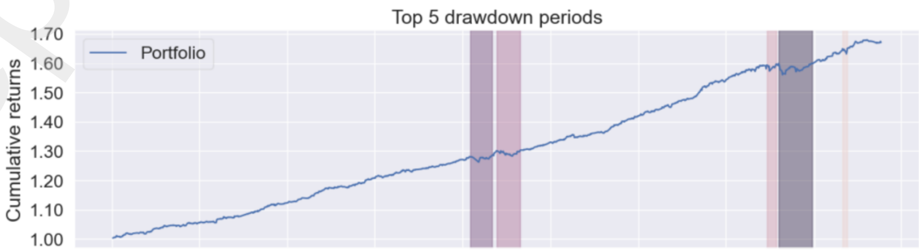

Figure 11. Top 5 drawdown periods

Table 4. Top 5 drawdown periods

Worst drawdown periods	Net drawdown in %	Peak date	Valley date	Recovery date	Duration
1	2.40	2021-10-19	2021-10-27	2021-12-27	50
2	1.42	2020-01-16	2020-02-03	2020-03-02	33
3	1.40	2020-03-11	2020-04-13	2020-04-29	36
4	1.28	2021-09-24	2021-09-30	2021-10-14	15
5	1.04	2022-03-01	2022-03-09	2022-03-11	9

Using a six-month window to calculate the rolling volatility, it can be observed that compared to the return rate of the CSI 300 Index, the volatility of the pair trading strategy has consistently remained at a lower level, indicating greater stability.

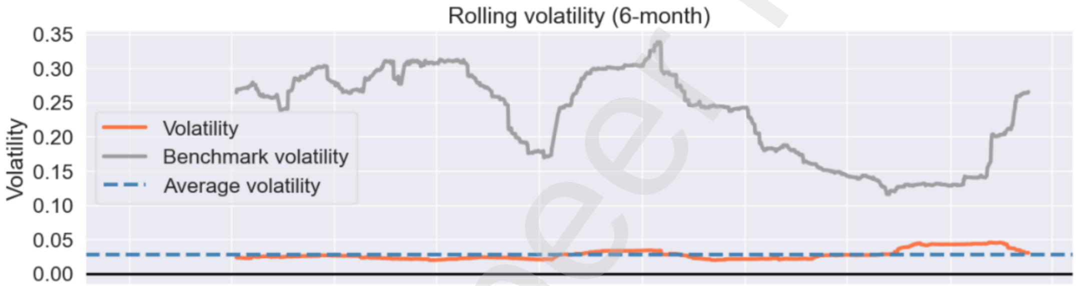

Figure 12 Rolling volatility(6-month)

The box plots illustrate that the volatility levels of various evaluation indicators fall within a reasonable range. This indicates that our experimental results possess a high level of credibility and provide valuable references.

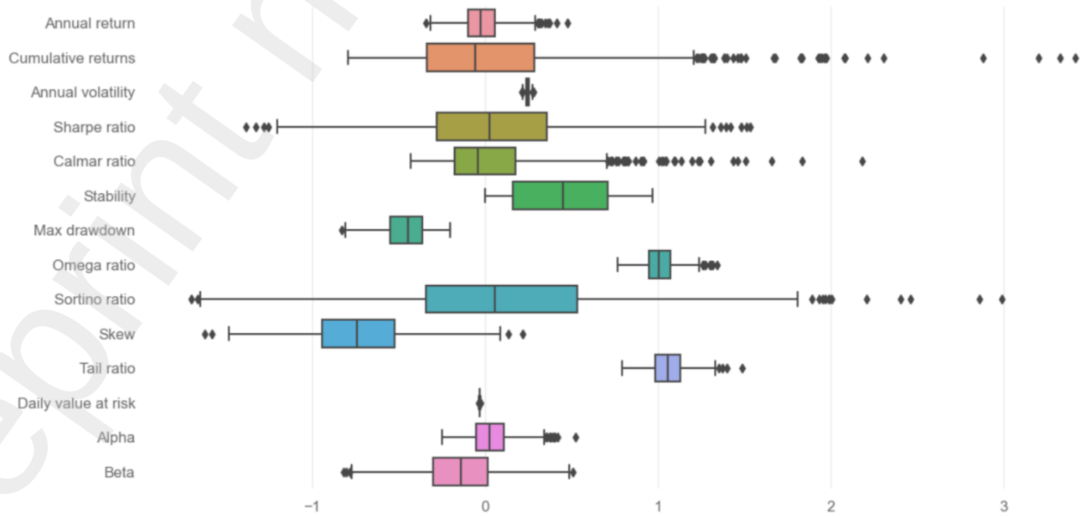

Figure 13. Evaluation indicator box plot

5.Validity Test

5.1. Out-of-Sample Backtesting

We extended the sample size from the end of the experimental backtesting period (May 20, 2022) to November 20, 2023, and obtained extended backtesting results while maintaining the same backtesting framework and parameters. The overall weighted annualized return rate was 13.318%, with a cumulative return of 19.556%. The alpha value was 0.14, the Sharpe ratio was 4.19, the maximum drawdown was -1.359%, the annualized volatility was 2.998%, and the beta value was 0.02. During the same period, the benchmark CSI 300 Index experienced negative returns, indicating an overall bear market. However, this strategy continued to perform robustly.

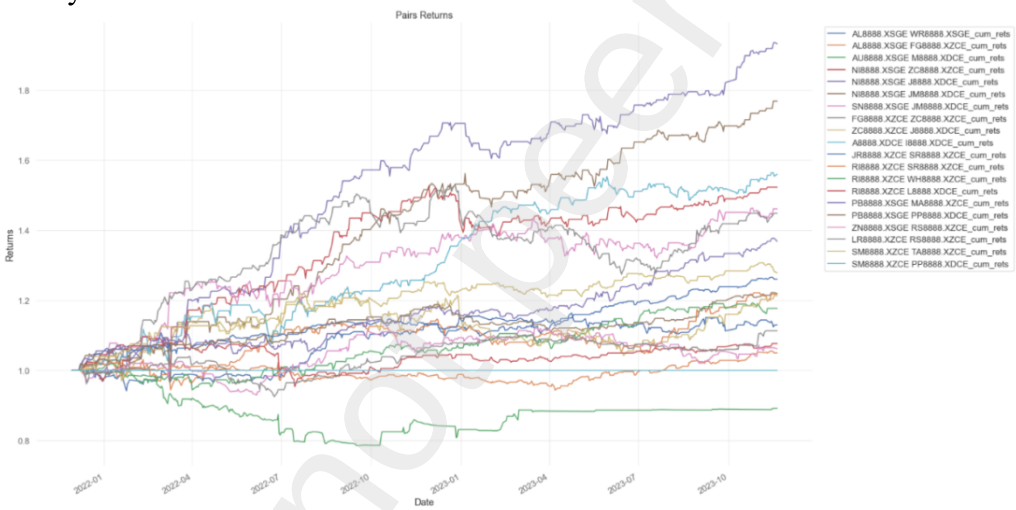

Figure 14. Cumulative yield on paired contracts

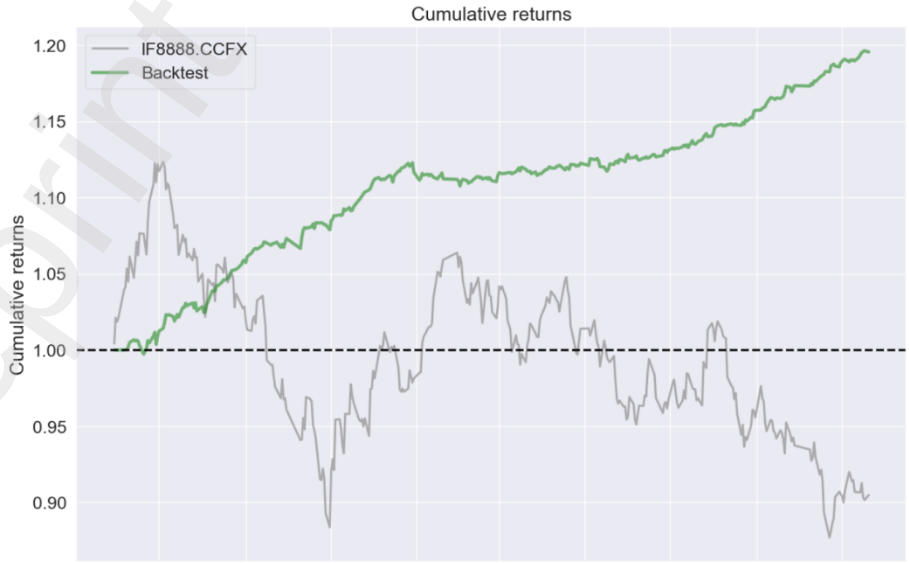

Figure 15. Weighted average cumulative rate of return

5.2 Comparison with Classical Pair Trading Strategy

Traditional cross-commodity futures pair trading strategies primarily rely on statistical methods to identify and execute trading opportunities. They scan all target contracts for cointegration pairing or perform correlation analysis to select asset pairs, utilizing fixed window lengths for parameter calculations. In contrast, the strategy discussed in this study extensively applies machine learning techniques, dynamic parameter adjustments (such as half-life), or advanced signal processing methods (such as Kalman filtering). However, to ensure comparability between the classical pair trading strategy's backtesting results and the out-of-sample backtesting results, we still use the same asset pool and backtesting period.

Using the same asset pool as the experimental strategy, we load the validation set data into the classical pair trading strategy for backtesting. The comparative results are as follows:

Table 5. Simulated trading results comparison table

Backtest_index	Annual return	Cumulative return	Alpha	Sharpe ratio	Max drawdown	Annual volatility	Beta
Optimized	13.318%	19.556%	0.14	4.19	−1.359%	2.998%	0.02
Classical	6.326%	9.237%	0.07	1.40	−3.234%	4.466%	0.03

Table 5 provides a comparison of the experimental strategy and the classical pair trading strategy across several key financial metrics. These metrics offer a deeper understanding of how each strategy performed in historical simulations. The experimental strategy outperforms the classical strategy in terms of return on investment, risk resistance, maximum potential loss, and market correlation. During the out-of-sample backtesting period, the experimental strategy exhibits superior risk/reward characteristics, demonstrating the effectiveness of the optimized techniques.

6.Conclusion

During the pairing phase, this study combined Dynamic Time Warping (DTW) with the K-means clustering algorithm to perform clustering analysis on contracts before conducting cointegration pairing. The empirical results indicate that the contracts grouped by this method exhibit significant internal connections. Compared to methods that rely solely on correlation analysis or cointegration pairing, this approach significantly improves the efficiency and stability of asset pairing.

In the trading phase, the study integrated Kalman filtering and half-life techniques to optimize the traditional pairs trading strategy. Using the half-life as a moving window significantly enhanced the model's flexibility and adaptability, while the introduction of Kalman filtering allowed for precise calculation of dynamic hedge ratios and the generation of dynamic trading signals. This not only reduced human intervention but also substantially lowered risk exposure, maximizing the advantages of the pairs trading strategy. Additionally, the application of Kalman filtering in smoothing data effectively eliminated noise interference, preventing the costs associated with frequent trading. As a result, this method successfully reduced noise interference and generated dynamic trading signals that enhance profitability.

In-sample backtesting demonstrated an annualized return of 12.964%, a Sharpe ratio of 4.35, and an annualized volatility of 2.812%. By setting different thresholds for opening and closing positions, the strategy consistently achieved stable returns. In out-of-sample backtesting, the strategy averaged an annualized return of 13.314%, a Sharpe ratio of 4.19, and an annualized volatility of 2.998%, effectively proving the strategy's stability and effectiveness. Compared to traditional strategies, the trading strategy combining machine learning and Kalman filtering showed clear advantages in terms of returns and stability.

This research delved deeply into the field of statistical arbitrage, employing machine learning technologies and state-space models, successfully validating the practicality of the pair trading model that integrates machine learning and Kalman filtering in the Chinese futures market. This provides valuable methodologies and insights for future investors and researchers. However, the trading strategy constructed in the study has some limitations. Firstly, a uniform commission rate of one-thousandth was applied to all contracts, ignoring the different actual trading costs associated with various contracts, which may lead to deviations from real market conditions. Secondly, the fixed nature of the asset pool might lead to decreased efficiency in capital utilization and profit levels. Lastly, although k-means was used as the main clustering method, considering the availability of more efficient clustering technologies in the field of machine learning, future work could seek better classification effects by comparing different clustering algorithms.

Appendix A

1. Dynamic Time Warping (DTW)

In this study, we use the K-means clustering algorithm to group futures contracts, aiming to enhance pairing precision and efficiency, thereby avoiding exhaustive cointegration calculations on the entire sample. First, the price series of each futures contract is processed with Dynamic Time Warping (DTW).

To construct the time series matrix, we build it based on two time series, with each matrix element representing the Euclidean distance between adjacent data points. Then, through backtracking, we identify the shortest path and compute the cumulative values to obtain the DTW distance.

Initial condition: Lmin = M(1,1) (A.1)

Recurrence rule:

Lmin(i,j) = min{Lmin(i,j-1), Lmin(i-1,j), Lmin(i-1,j-1)}+M(i, j) (A.2)

Euclidean distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{A.3}$$

DWT distance:

$$DWT(x,y) = \sqrt{\sum_{i=1}^{n} L_{min}^2} \tag{A.4}$$

Sequence1 = [1,4,5,7,6,6,9,7]

Sequence2 = [0,2,4,4,5,4,5,6]

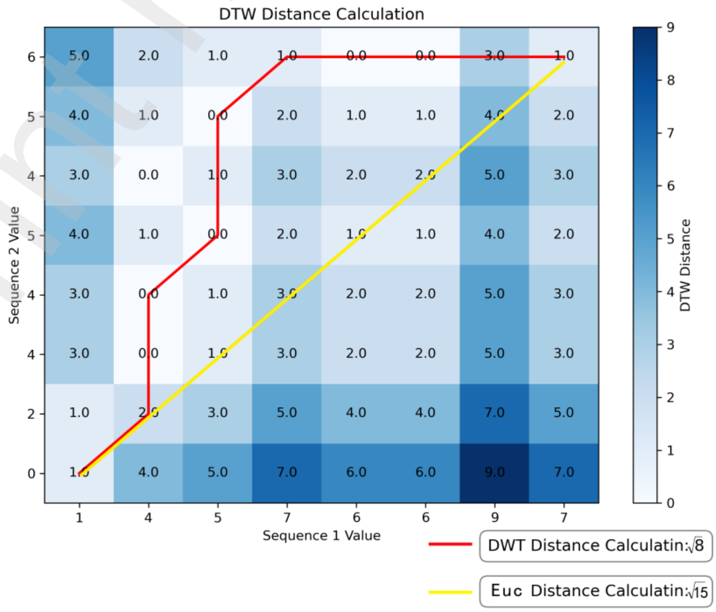

Figure A.1 DWT distance matrix

2.Cointegration test principle

Cointegration testing is an important statistical tool widely used in economics and finance to examine whether there exists a long-term equilibrium relationship between two or more time series. In our research, we utilize the Engle-Granger two-step method for cointegration testing to determine whether the selected futures contract pairs exhibit such long-term and stable cointegration relationships. If an m-dimensional vector Xt, with each component being an integrated series of order d, denoted as Xt ~ I(d), and there exists a nonzero vector Yt such that XtYt ~ I(d-b), then we say the components of the m-dimensional vector Xt are cointegrated at order (d-b), denoted as Xt ~ CI(d-b), where Xt is integrated of order d-b.

The Engle-Granger two-step method can be summarized as follows:

(1) Estimation of Coefficients and Residuals using Ordinary Least Squares (OLS):

Assuming that variables Xt and Yt both follow integrated order one (I(1)), we establish the regression equation:

$$Yt = \beta 0 + \beta 1\ Xt + \varepsilon t \qquad (A.5)$$

By performing OLS regression, we obtain the estimates of the cointegration coefficients β0 and β1 as:

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t \qquad (A.6)$$

The regression residual sequence $\hat{e}_t$ is obtained as:

$$\hat{e}_t = Y_t - \hat{Y}_t \qquad (A.7)$$

(2) Augmented Dickey-Fuller (ADF) Test:

The ADF test is applied to the residual sequence $\hat{e}_t$. If the test result indicates stationarity, it implies that Xt and Yt are cointegrated at the (1,1) order. However, if the residual sequence et is stationary at order d, it suggests that Xt and Yt are cointegrated at the (d,1) order.

3.Normalize the clustering set data

Through normalization, we can restrict the data distribution range to a smaller interval, reducing the influence of outliers or extreme values and enhancing the robustness of the model.

$$O_{\_score} = \frac{X - \bar{x}}{\delta} \tag{A.8}$$

Here, X represents the original time series data, $\bar{x}$ represents the mean, and δ represents the standard deviation. O_{score} represents the normalized price time series data.

Lastly, before conducting backtesting for pair trading, we need to normalize the trading set data based on relative returns to eliminate scale differences. Due to the magnitude differences among different contracts, it is challenging to analyze their price trends directly. Relative return normalization is performed by dividing each data point by the value at the first time point, transforming the time series data into growth proportions relative to the initial value. This approach ensures that all data points are benchmarked against the initial value, enabling better comparison and observation of changes between different time points.

$$Relative_{\_return} = \frac{X - x_0}{x_0} \tag{A.9}$$

Here, X represents the original time series data and x0 represents the value at the first time point.

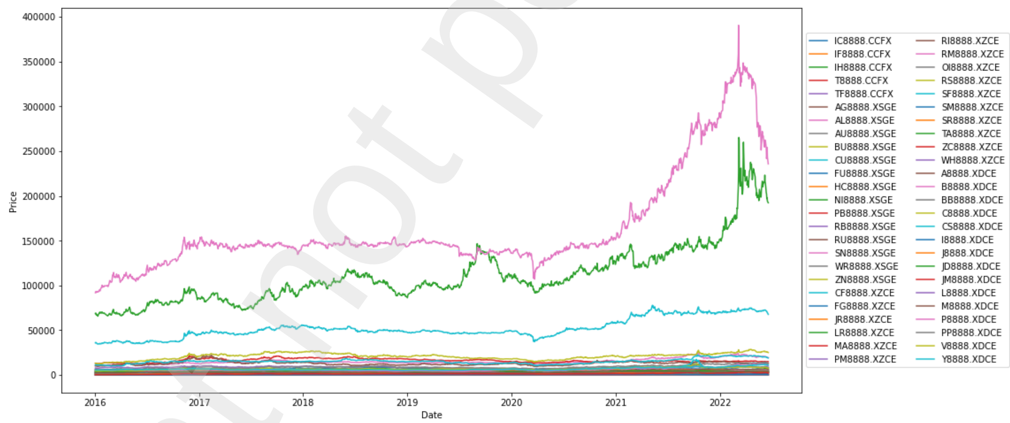

Figure A.2 Futures time series chart

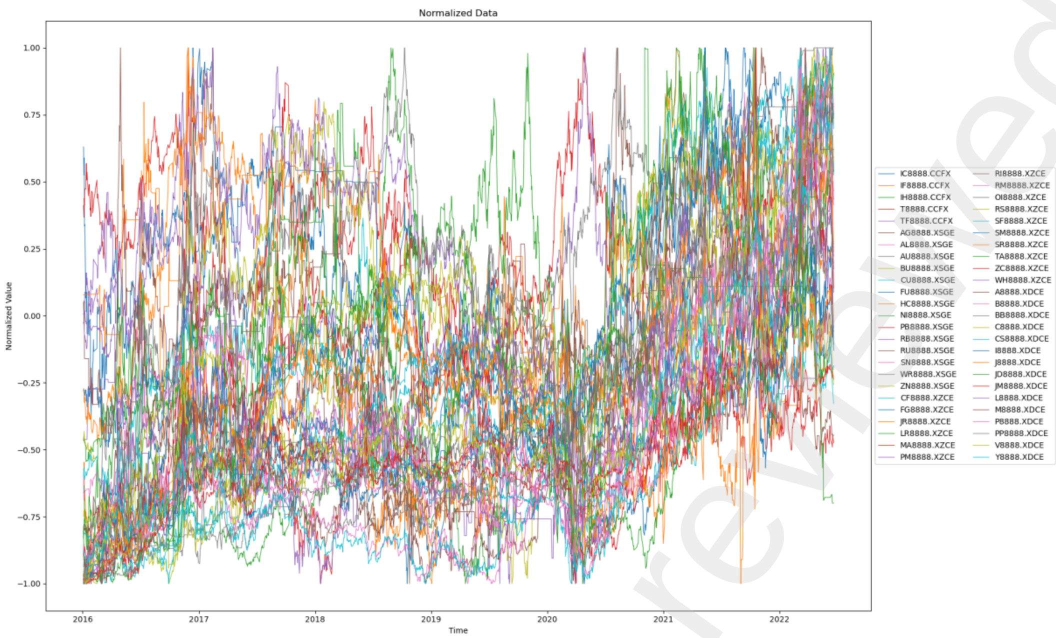

Figure A.3 Normalize time series plots

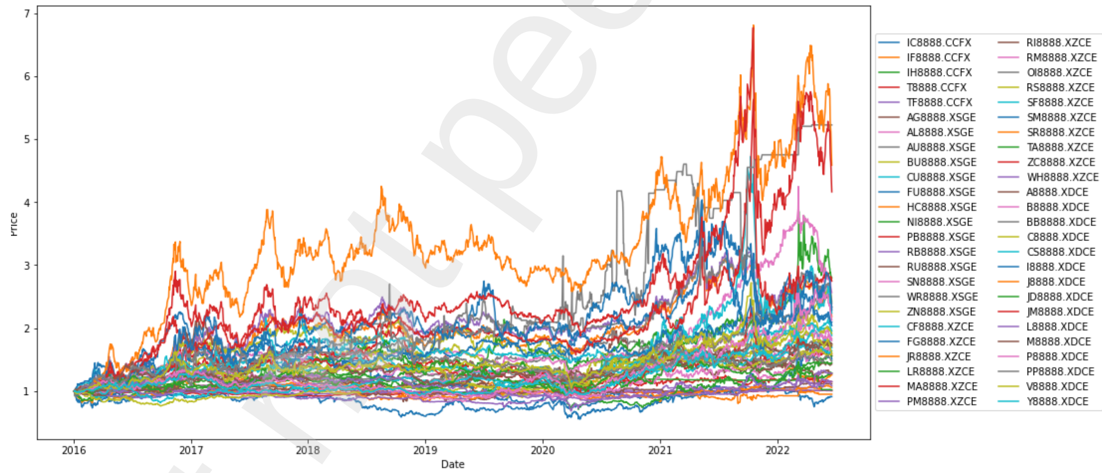

Figure A.4 Relative return normalized time price series chart

4. Silhouette Score

We will utilize the Silhouette Score to measure the quality of clustering. It takes into account both the compactness within clusters and the separation between clusters. The Silhouette Score ranges from -1 to 1, where a value close to 1 indicates reasonable clustering, a value close to -1 suggests that samples might have been wrongly assigned to other clusters, and a value close to 0 implies that samples are on the boundary between two clusters.

The Silhouette Score is calculated as:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{A.10}$$

where: a(i) represents the average distance between sample xi and all other points in the same cluster, indicating the cohesion within the cluster; b(i) is the average distance between sample xi and all points in the nearest neighboring cluster, obtained by iterating over all other clusters and it represents the separation between clusters as well.

By computing the Silhouette Score for all data points and calculating the average, we obtain the overall Silhouette Score for the current clustering. The Silhouette Score ranges from -1 to 1.

In our clustering results, we achieved a relatively optimal Silhouette Score of 0.54 when K=3.

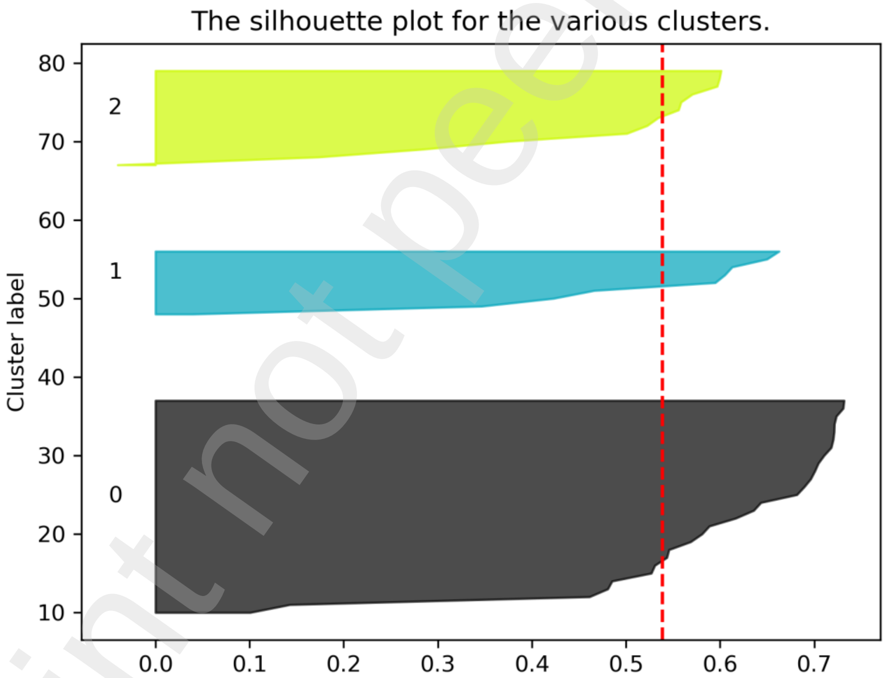

Figure A.5 The Silhouette Score

5. Out-of-sample backtest result analysis

From the distribution of returns shown in Histogram A.6 and Kernel Density Estimate (KDE), the backtesting results on the validation set demonstrate a strong central tendency, with most of the returns clustering around the mean. This suggests that the strategy's returns mostly fluctuate around their average value. The KDE curve forms a bell shape, indicating that the distribution of returns is approximately close to a normal distribution. Additionally, the average return line (red dashed line) is positioned above zero, allowing us to infer that, on average, the pair trading strategy achieved positive returns during this out-of-sample period.

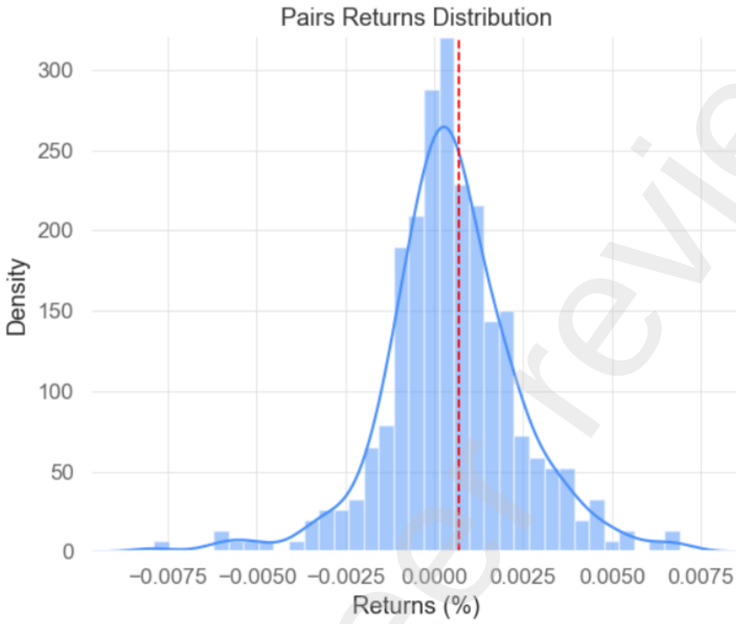

Figure A.6 Distribution of Returns on Validation Set

A Probability Plot, often known as a Q-Q (Quantile-Quantile) Plot, is a graphical tool used to assess whether a data set follows a certain theoretical distribution, in this case, a normal distribution. The Q-Q plot displays the quantiles of the sample data against the quantiles of the theoretical distribution, allowing for a visual comparison.

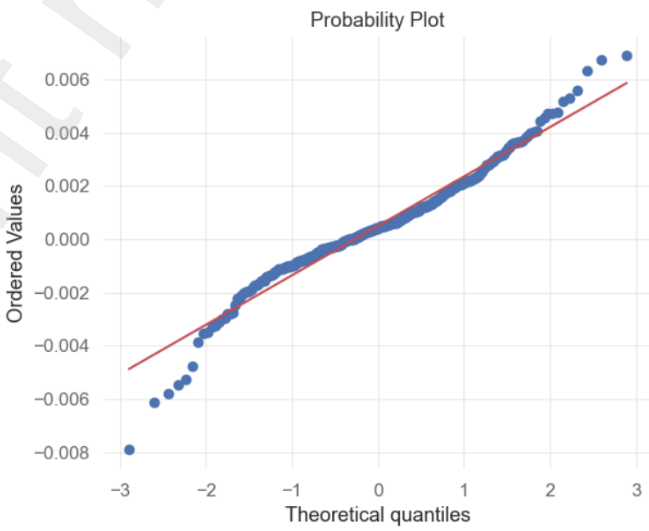

Figure A.7 Quantile-Quantile Plot

Figure A.8 shows that the trading strategy or portfolio exhibited a stable positive growth trend during the backtesting period. Although there were some months with losses, the overall monthly return distribution tends to positive gains, indicating that the strategy is both stable and profitable. Figure A.9 reveals that during the out-of-sample backtesting period, the strategy achieved positive returns most of the time. The peak in the histogram indicates that in most months, the strategy's returns are concentrated in a lower positive return range, but there are also some months where higher returns are achieved.

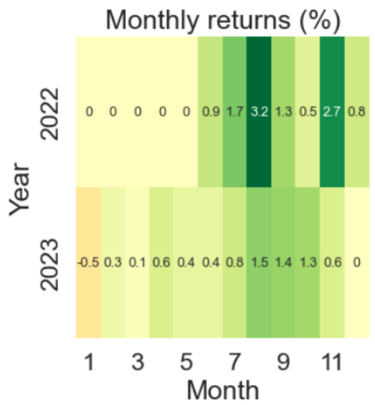

Figure A.8 Monthly Returns Heatmap

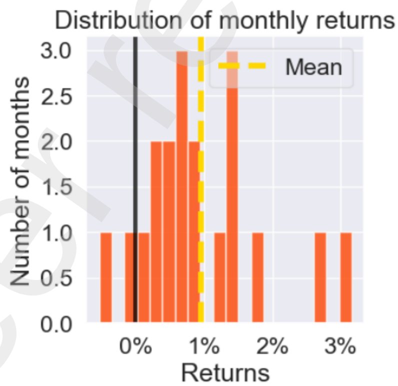

Figure A.9 Histogram of Monthly Returns Distribution

From the box plot indicators, the strategy generally achieves positive returns most of the time and performs well on risk-adjusted metrics such as the Sharpe ratio, Omega ratio, and Sortino ratio. Although the strategy exhibits some volatility, it shows good performance in controlling maximum drawdowns. The distribution of the Alpha value indicates that the strategy often generates returns exceeding the benchmark, while the low distribution of the Beta value suggests that the strategy has a relatively low sensitivity to market fluctuations.

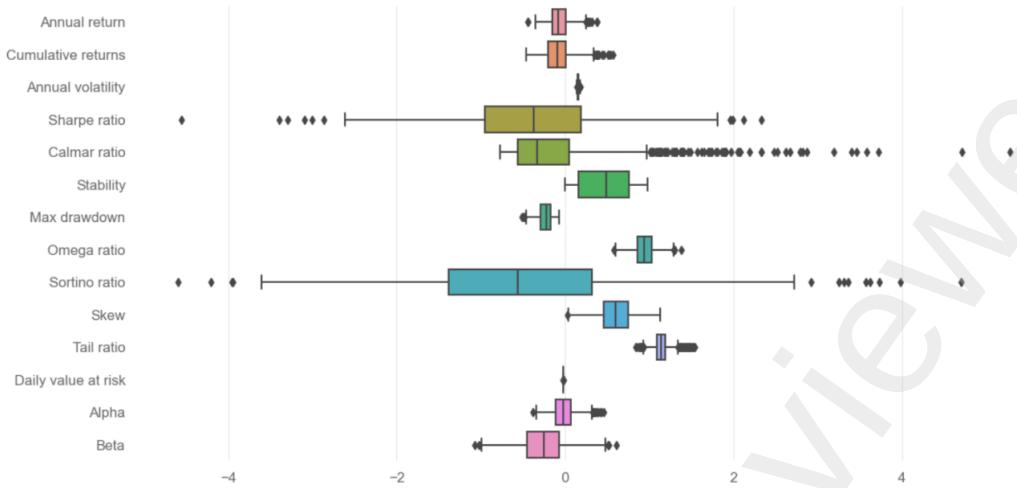

Figure A.10 Box Plot of Performance Metrics

6. Evaluation Metrics

The evaluation metrics used include return indicators and risk indicators. The return indicators comprise annualized rate of return, compound annual growth rate (CAGR), cumulative return, and alpha. The risk indicators consist of the Sharpe ratio, maximum drawdown, strategy volatility, beta, and information ratio.

(1) Annualized Rate of Return: The annualized rate of return measures the investment return and represents the average return rate obtained from the investment over one year. It considers the time span of the investment and standardizes it into a yearly growth rate for comparison between different trades.

$$R_P = \left((1+P)^{\frac{250}{n}} - 1\right) \times 100\% \tag{A.11}$$

Where P denotes the strategy's return, n represents the number of backtesting days.

(2) CAGR: CAGR is used to assess the average annual growth rate of an investment or asset and takes compounding into account. CAGR helps investors evaluate the overall performance of an asset over a certain period, allowing comparisons between different investments or assets at different time intervals. It considers the compounding growth of the asset and reflects its sustained growth capability. A higher CAGR signifies faster asset growth, while a lower CAGR indicates slower growth.

$$CAGR = \left(\frac{EV}{BV}\right)^{\left(\frac{1}{n}\right)} - 1 \quad (A.12)$$

Where EV denotes the end value, BV represents the beginning value, and n indicates the years of investment.

(3) Cumulative Return: Cumulative return refers to the total accumulated profit generated by an investment over a specific period. It measures the overall return on investment considering the changes between the initial and final values.

$$CR = \frac{(P_n - P_0)}{P_0} \times 100\% \quad (A.13)$$

Where Pn represents the ending value and P0 denotes the initial value.

(4) Alpha: Alpha is an indicator used to evaluate the relative performance of an investment strategy or portfolio compared to the market. It measures the portfolio's excess return relative to a market benchmark.

$$Alpha = R_p - [R_f + \beta_p(R_m - R_f)] \quad (A.14)$$

Where Rp represents the strategy's annualized rate of return, Rm is the benchmark's annualized rate of return (in this context, the Shanghai and Shenzhen 300 Index is selected as the benchmark), Rf represents the risk-free rate, and βp denotes the strategy's beta value.

(5) Sharpe Ratio: The Sharpe ratio measures the relationship between the excess return of an investment portfolio and its risk. It evaluates the excess return earned per unit of risk taken by the portfolio. A higher Sharpe ratio indicates a higher risk-adjusted return for each unit of risk. Generally, a higher Sharpe ratio signifies better portfolio performance.

$$SR = \frac{R_p - R_f}{\delta_p} \quad (A.15)$$

Where Rp is the portfolio's expected return (average return), Rf represents the risk-free rate (usually represented by short-term government bond rate), and δp denotes the standard deviation of the portfolio, used to measure its volatility or risk.

(6) Maximum Drawdown: Maximum Drawdown is a metric that measures the maximum loss suffered by a portfolio or asset during a specific period. It is used to assess investment risk and resilience. Maximum Drawdown reflects the largest decline

experienced by a portfolio or asset after reaching its historical peak.

$$Maximum\ Drawdown = \frac{P_t - P_d}{P_t} \tag{A.16}$$

Where Pt represents the peak value of the asset or portfolio during a certain period, and Pd represents the trough value, which is the lowest point of the asset or portfolio after the peak.

(7) Strategy Volatility: Strategy Volatility is an indicator that measures the level of risk in an investment strategy or portfolio. It reflects the fluctuation of returns, i.e., the range of changes in returns, for the strategy or portfolio.

$$\sigma_p = \sqrt{\frac{250}{n-1} \sum_{i=1}^{n} \left(r_p - \overline{r_p}\right)^2} \tag{A.17}$$

Where rp denotes the daily returns, $\overline{r_p}$ represents the mean daily return, and n is the number of days the strategy is executed.

(8) Beta (β): Beta is a measure used to gauge the relative volatility of an asset or investment portfolio compared to the overall market, i.e., systematic risk. It quantifies the correlation between portfolio returns and market returns.

$$\beta = \frac{Cov(D_p, D_m)}{Var(D_m)} \tag{A.18}$$

Where D_p represents the daily returns of the strategy, and D_m denotes the daily benchmark returns.

Reference:

[1] Alexander, Carol. Market models: A guide to financial data analysis. University of Sussex, 2001.

[2] Berndt, Donald J., and James Clifford. "Using dynamic time war** to find patterns in time series." Proceedings of the 3rd international conference on knowledge discovery and data mining. 1994.

[3] Chen, Guanrong, and Haluk ÖGˇmen. "Modified extended Kalman filtering for supervised learning." International journal of systems science 24.6 (1993): 1207-1214.

[4] Chen, Yufeng, **wang Wu, and Zhongrui Wu. "China's commercial bank stock price prediction using a novel K-means-LSTM hybrid approach." Expert Systems with Applications 202 (2022): 117370.

[5] Do, Binh, and Robert Faff. "Does simple pairs trading still work?." Financial Analysts Journal 66.4 (2010): 83-95.

[6] Engle, Robert F. "Cointegration and error correction." Econometrica 55.2 (1987): 143-159.

[7] Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst. "Pairs trading: Performance of a relative-value arbitrage rule." *The Review of Financial Studies* 19.3 (2006): 797-827.

[8] Huck, Nicolas, and Komivi Afawubo. "Pairs trading and selection methods: is cointegration superior?." Applied Economics 47.6 (2015): 599-613.

[9] Huck, Nicolas, and Komivi Afawubo. "Pairs trading and selection methods: is cointegration superior?." Applied Economics 47.6 (2015): 599-613.

[10] Li, Mao Liang, Chin Man Chui, and Chang Qing Li. "Is pairs trading profitable on China AH-share markets?." Applied Economics Letters 21.16 (2014): 1116-1121.

[11] Paparrizos, John, and Luis Gravano. "k-shape: Efficient and accurate clustering of time series." Proceedings of the 2015 ACM SIGMOD international conference on management of data. 2015.

[12] Rad, Hossein, Rand Kwong Yew Low, and Robert Faff. "The profitability of pairs trading strategies: distance, cointegration and copula methods." Quantitative Finance 16.10 (2016): 1541-1558.

[13] Vidyamurthy, Ganapathy. Pairs trading: Quantitative methods and analysis. Vol. 217. John Wiley & Sons, 2004.

[14] Fama, Eugene F. "The behavior of stock-market prices." The journal of Business 38.1 (1965): 34-105.