Financial Innovation

RESEARCH Open Access

Copula-based trading of cointegrated cryptocurrency Pairs

Masood Tadi[1,2*] and Jiří Witzany[2]

*Correspondence:
tadim@karlin.mff.cuni.cz

[1] Department of Probability and Mathematical Statistics Faculty of Mathematics and Physics, Charles University Prague, Prague, Czech Republic
[2] Faculty of Finance and Accounting, Prague University of Economics and Business Prague, Prague, Czech Republic

Abstract

This study introduces a novel pairs trading strategy based on copulas for cointegrated pairs of cryptocurrencies. To identify the most suitable pairs and generate trading signals formulated from a reference asset for analyzing the mispricing index, the study employs linear and nonlinear cointegration tests, a correlation coefficient measure, and fits different copula families, respectively. The strategy's performance is then evaluated by conducting back-testing for various triggers of opening positions, assessing its returns and risks. The findings indicate that the proposed method outperforms previously examined trading strategies of pairs based on cointegration or copulas in terms of profitability and risk-adjusted returns.

Keywords: Statistical arbitrage, Pairs trading, Cointegration, Copulas, Cryptocurrency market

Introduction

Pairs trading is a well-known algorithmic trading strategy that capitalizes on temporary abnormal relationships among two or multiple assets whose historical prices tend to shift together. When this relationship begins to exhibit abnormal behavior, it triggers the opening of a trading position. The position closes as soon as the assets return to their normal behavior (Vidyamurthy 2004). According to Krauss (2017), pairs trading is characterized by a formation and trading period. During the formation period, the objective is to identify pairs of assets that exhibit similar price movements.

This is commonly achieved through co-movement criteria, which can be measured using various methods. For example, the distance approach, as described by Gatev et al. (2006), Perlin (2009), Do and Faff (2010), and Do and Faff (2012), uses distance metrics such as the minimum sum of squared distances of normalized asset prices. Statistical relationships, such as cointegration rules (the cointegration approach) which was introduced by Vidyamurthy (2004), Rad et al. (2016), Clegg and Krauss (2018), Fil and Kristoufek (2020), and Tadi and Kortchemski (2021) can also be used.

Furthermore, the parameters of the trading period are estimated during the formation period. During the trading timeframe, irregularities in pairs' price movement aim to benefit from statistical arbitrage opportunities and create signals to open long or short

positions. Advanced strategies can leverage various mathematical tools to optimize the efficacy of their results. These tools encompass stochastic processes as examined in works by Elliott et al. (2005), Bertram (2010), and Bogomolov (2013); stochastic control techniques as seen in the works of Jurek and Yang (2007), Mudchanatongsuk et al. (2008), Tourin and Yan (2013), and Lintilhac and Tourin (2017); copula-based methods as explored in the studies by Liew and Yuan (2013), Rad et al. (2016), Krauss and Stübinger (2017), and Silva et al. (2023); and machine learning or deep learning methods as demonstrated in the works of Sarmento and Horta (2020), Brim (2020), and Chang (2021).

This study aims to implement a pairs trading strategy by employing a reference asset-based copula method for cointegrated cryptocurrency pairs.

Individual investors and investment companies place value on statistical arbitrage strategies, such as pairs trading in the cryptocurrency market, making this a valid area of research. We specifically focus on trading strategies that can yield consistent returns. Pairs trading can be profitable in the decentralized cryptocurrency market, offering two potential arbitrage opportunities: exchange-to-exchange and statistical arbitrage. Borri and Shakhnov (2022) offer a novel risk-based explanation for the price differences observed across cryptocurrency exchanges. To understand these discrepancies, they highlight significant factors such as transaction costs, market liquidity, and investor sentiment. Nevertheless, implementing exchange-to-exchange arbitrage can be risky and pose numerous challenges. By contrast, statistical arbitrage opportunities present similar profit potential but with lower risk (Pritchard 2018). The cryptocurrency market is noted for its volatility, which can reduce market efficiency and potentially provide arbitrage opportunities for traders, as highlighted by Al-Yahyaee (2020). Therefore, investors are driven to explore innovative trading strategies. Such algorithmic trading strategies can offer a more methodical and organized way to navigate the cryptocurrency market, leveraging market volatility to potentially enhance profitability. Moreover, in the cryptocurrency ecosystem, delving into algorithmic trading strategies is crucial for establishing a framework with best practices for traders and investors. This, in turn, contributes to the development and maturation of the market, providing the necessary stability for the sustainable growth of the cryptocurrency ecosystem.

In the subsequent sections, this study is structured as follows: Section "Literature review" reviews the literature, focusing specifically on existing pairs trading strategies and copula-based techniques to identify limitations in previously introduced methodologies. Section "Cryptocurrency exchange and data source" provides an overview of the cryptocurrency exchanges and data sources employed in this study. Section "Theoretical framework" details the theoretical framework, including discussions on unit-root tests, the concept of copulas, families of copulas, and methods for copula estimation. Section "Implementation methodology" describes the implementation methodology of our hybrid approach, which combines copula-based and cointegration-based strategies for pair trading with cryptocurrency coins. This involves selecting linear and non-linear cointegrated coins, establishing trading rules using copula conditional probabilities, and conducting thorough back-testing using historical cryptocurrency price data. Section "Key assumptions and limitations" presents the key assumptions and limitations of the proposed approach. Section "Empirical results " discusses the empirical results

and evaluates the strategy's performance using diverse metrics such as profitability, risk-adjusted return, and maximum drawdown. Section "Conclusion" concludes the findings, emphasizing the importance of selecting an appropriate entry threshold in pairs trading and demonstrating its impact on both volatility and returns. This study also contributes to the literature by comparing the performance of the proposed copula method with that of conventional methods in this domain.

Literature review

Multiple studies have used various concepts, such as the distance approach, cointegration analysis, or the concept of copulas, in pairs trading strategies. The distance approach involves calculating the historical price spread or the price difference between two related assets and monitoring this spread over time. The spread is usually calculated as the difference between the prices of two assets, either as a raw spread or a normalized spread, such as the z-score or percent difference. For example, Gatev et al. (2006) defined the normalized spread S_t^{ij} between assets i and j at time t as:

$$S_t^{ij} = P_t^{i*} - P_t^{j*} \tag{1}$$

$$P_t^{i*} = \frac{P_t^i}{P_0^i} = \prod_{\tau=1}^{t} \frac{P_\tau^i}{P_{\tau-1}^i} = \prod_{\tau=1}^{t} \left(1 + r_\tau^i\right) = cr_t^i, \tag{2}$$

where P_t^i is the price of asset i at time t, r_t^i is the t-period return on asset i at time t, and cr_t^i is the cumulative total return on asset i until time t ($cr_0^i = 1$). Spread sum of squared distance (SSD) is performed using the following equation:

$$SSD_{i,j} = \sum_{t=1}^{T} \left(S_t^{ij}\right)^2 = \sum_{t=1}^{T} (cr_{i,t} - cr_{j,t})^2. \tag{3}$$

Pairs are chosen for the trading period based on the ascending order of $SSD_{i,j}$ during the formation stage. The initial pairs at the top of the list are selected, and basic nonparametric threshold rules are employed to generate trading rules. An alternative method was used by Chen (2019), who identified the most suitable pairs for a trading period using the Pearson correlation of asset returns instead of finding the minimum sum of the squared distance. Krauss (2017) found that to maximize the profit of the distance approach strategy, the spread of each selected pair must have high volatility, indicating the potential divergence of the two assets. The pair's spread should also have a mean-reverting property. The key advantage of this approach is its simplicity and transparency, making it suitable for large-scale empirical applications.

Huck (2015) found that the cointegration approach outperformed the distance approach in selecting effective pairs. The cointegration approach aims to identify a long-term equilibrium between nonstationary time series (e.g., asset prices) that move together. This equilibrium can be linear or nonlinear. Engle and Granger (1987) introduced the first cointegration test, which is based on linear regression and the unit-root test of residuals in the equilibrium. Typically, the augmented Dickey-Fuller (ADF) test is used for the unit-root test. Other improvements of the Engle-Granger (EG) cointegration

test were introduced by Phillips and Ouliaris (1990) and Johansen (1991). Highly volatile markets such as cryptocurrencies usually exhibit nonlinear features. Therefore, we can extend the Engle-Granger cointegration test by adjusting the error correction model and applying nonlinear unit root tests to increase the reliability of the study. These extensions were studied by Enders and Siklos (2001), Hansen and Seo (2002), and Kapetanios et al. (2006).

Several studies have also explored the application of pairs trading in the cryptocurrency market, such as Lintilhac and Tourin (2017), van den Broek and Sharif (2018), Pritchard (2018), Kakushadze and Willie (2019), Leung and Nguyen (2019), Tadi and Kortchemski (2021), and Fil and Kristoufek (2020). Furthermore, Fang (2022) conducted an extensive review of cryptocurrency trading research, covering diverse aspects such as cryptocurrency trading systems, market conditions (bubbles and extreme events), volatility and return prediction, portfolio construction, and technical trading.

In addition to the commonly used methods discussed above, more advanced concepts, such as copulas, can be applied to enhance the empirical results of the strategy. Compared to correlation or linear cointegration-based methods, the copula approach provides more valuable information regarding the shape and characteristics of pairs' dependency (Ferreira 2008). Xie and Wu (2013) demonstrated that the commonly used distance and cointegration methods can be generalized as special cases of the copula method under certain conditions. This means that the dependency structure of assets in the copula approach is more robust and accurate.

Moreover, Liew and Yuan (2013) conducted a comparative study of a copula-based pairs-trading strategy with other conventional approaches such as distance and cointegration approaches. They found that the copula approach for pairs trading shows better empirical results than other approaches. It also presents more trading possibilities with higher confidence in practice and does not entail any rigid assumptions, such as the linear association between asset returns in traditional approaches. Hence, the copula approach provides closer estimates and predictions of the reality. According to Stander et al. (2013), the copula approach can significantly demonstrate the dependency of pairs because it can capture the asymmetry and heavy-tailed characteristics of asset returns to model marginal distribution functions instead of modeling them by Gaussian distribution. Furthermore, their empirical analyses reveal that there are more trading opportunities when the market is highly volatile and that the profitability of a strategy strongly depends on the liquidity of the market.

According to Xie and Liew (2016), the copula method outperforms the distance approach in describing the dependency relationship between assets and identifying more statistical arbitrage opportunities that generate greater profits. They also recommended utilizing the copula approach for high-frequency pairs trading. Rad et al. (2016) compared the performance of the copula approach with that of traditional methods using daily US stocks. They found that while the profitability of the copula method could be weaker, it is more reliable in capturing arbitrage opportunities in the US stock market. They determined that the Student-t copula is more appropriate for modeling the dependence structure of pairs in the US stock market than other copulas.

The performance of a pair-trading strategy based on a weighted combination of copulas was evaluated by Silva et al. (2023) against a distance methodology using a vast dataset of *S&P*500 stocks covering 25 years. This study examines the effects of financial factors on profitability. The mixed copula approach yields better results than the distance method, with higher alphas for fully invested capital and overall superior performance. The approach was notably effective under committed capital during both crisis and non-crisis periods and was fully invested during non-crisis periods.

According to Krauss and Stübinger (2017), the copula approach can be divided into two substreams:

Return-based copula method In the return-based copula method, the log-returns of two assets are calculated and their marginal distributions estimated. An appropriate copula is then chosen to represent the dependency relationship between the two assets. To generate trading signals, a mispricing index, which indicates the degree of abnormal relationships among assets, is defined. Unlike the distance and cointegration methods, which use a spread-based mispricing definition, the copula method defines mispricing based on the copula's conditional probability distribution of the corresponding assets' log-returns. The conditional distribution of the copulas can be obtained by taking the partial derivative of the copula function, as shown below[1]:

$$\begin{aligned} h^{1|2} &:= h(u_1|u_2) = P(U_1 \leq u_1|U_2 = u_2) = \frac{\partial C(u_1, u_2)}{\partial u_2} \\ h^{2|1} &:= h(u_2|u_1) = P(U_2 \leq u_2|U_1 = u_1) = \frac{\partial C(u_1, u_2)}{\partial u_1} \end{aligned} \tag{4}$$

where $C(u_1, u_2)$ is the copula distribution function; $h^{1|2}$ and $h^{2|1}$ are the conditional copula distribution functions; and U_1 and U_2 are the transformed uniform variables of the log returns. The values of $h^{1|2}$ and $h^{2|1}$ are between 0 and 1. As their values differ from 0.5, we consider this a deviation from the expected relationship between the two assets. Ferreira (2008), Liew and Yuan (2013), Stander et al. (2013), and Haddad et al. (2023) also deployed their strategy in this way.

However, a significant limitation of this return-based approach is its dependence on the returns of the previous period to generate the entry and exit signals. This method potentially overlooks long-term trends and dependencies that extend beyond immediate past returns, leading to suboptimal trading signals. Such signals may not capture the substantial market changes that affect the overall effectiveness and profitability of the strategy.

Level-based copula method To address the limitation of the return-based method, a new approach (hereafter referred to as the level-based method) was proposed by Xie and Wu (2013). They defined a new mispricing index by aggregating the surplus value of the conditional probability in Eq. (4) from 0.5 across multiple periods to determine the extent to which assets are out of balance. These cumulative mispricing indices (CMI) are defined as

[1] For more details, see Sect. "Copula concept".

$$\begin{aligned} \mathrm{CMI}_t^{1|2} &= \mathrm{CMI}_{t-1}^{1|2} + \left(h_t^{1|2} - 0.5\right) \\ \mathrm{CMI}_t^{2|1} &= \mathrm{CMI}_{t-1}^{2|1} + \left(h_t^{2|1} - 0.5\right) \end{aligned} \tag{5}$$

The studies, which investigate the level-based copula method, were conducted by Xie and Liew (2016), Rad et al. (2016), Krauss and Stübinger (2017), and Silva et al. (2023).

In practice, the application of cumulative mispricing indices (CMIs)-as detailed in Eq. (5)-within the level-based copula method presents a significant challenge. These indices are expected to exhibit mean-reverting behavior, a key assumption that underpins the profitability of this trading strategy. However, CMIs often do not consistently demonstrate such behavior. The lack of mean reversion in the CMI can adversely affect the profitability of the strategy and could lead to prolonged periods of unprofitable trades if the expected mean reversion does not materialize.

Furthermore, both return- and level-based copula methods exhibit notable limitations, particularly when applied to datasets with high granularity concerning low-liquidity assets. Given that these methods fundamentally rely on asset returns to model dependencies, they can encounter issues with finer data resolution, where frequent zero returns are common. This often results in a violation of the continuity assumption required for the random variables used in copula modeling.

To address these limitations, this study introduces a novel methodology that utilizes stationary spread processes instead of traditional log returns. This approach aims to provide a more reliable criterion for trading decisions, thereby improving the robustness and applicability of copula-based models. Furthermore, this new methodology addresses the challenge of high-frequency data, particularly in less-liquid assets, by introducing a highly liquid cryptocurrency, such as Bitcoin, as a reference asset in the spread process formula.[2]

Cryptocurrency exchange and data source

The cryptocurrency market facilitates decentralized trading of cryptocurrencies across various exchanges. Binance, established in 2017, is the largest cryptocurrency exchange worldwide in terms of daily trading volume in both spot and derivative markets. In this market, cryptocurrencies are typically quoted in pairs, where the value of one cryptocurrency is expressed in terms of another cryptocurrency or a fiat currency such as the US dollar. For example, the notation BTCUSD indicates how many US dollars one Bitcoin is worth.

Binance provides three types of derivative contracts: futures, options, and Binance leveraged tokens (BLVT). Futures contracts are classified into two main categories: COIN-margined contracts and USD(S)-margined contracts. COIN-margined contracts are inverse futures quoted in US dollars but denominated in an underlying cryptocurrency (e.g., Bitcoin). They include traditional quarterly and perpetual futures, also known as perpetual swaps, that never expire or settle. Since they lack settlement prices, their prices can deviate significantly from their spot contract prices. Binance uses funding fees to address this issue on both the long and short sides.

[2] For more details, see Sects. "Implementation methodology" and "Key assumptions and limitations".

Meanwhile, USDⓈ-margined contracts are similar to COIN-margined futures and have perpetual or quarterly expiration dates. However, they are quoted, denominated, and settled in stablecoins such as Tether (USDT) and Binance USD (BUSD), which are pegged to the US dollar value. For this study, all nominated cryptocurrency coins are USDT-margined futures (Binance Crypto Derivatives 2022). We choose USDT because its pairs generally exhibit higher liquidity than other stablecoins and many cryptocurrencies, making it ideal for implementing and evaluating our trading strategies effectively within the dynamic cryptocurrency market. Table 1 shows the Binance USDT-margined futures contracts used in this study. Notably, this subset of coins was specifically selected based on their earlier issuance and a longer history of market activity.

The minimum order price, also known as minimum price increment, is the smallest possible change in the price of a contract in exchange. This value is specific to each asset and can be adjusted over time. Assets with smaller increments have narrower bid or ask spreads. When limit orders are not entirely disclosed, traders tend to order contracts with smaller quote sizes to avoid slippage (Harris 1997).

To calculate profit and loss, all coins are valued in Tether (USDT), a stablecoin. Using the Binance application programming interface (API), we collected both historical hourly and 5-min closed prices for 20 cryptocurrency coins from 01/01/2021 to 19/01/2023.

Table 1 Binance USDT-Margined Futures Contracts Used in the Research

Symbol	Underlying crypto	Min. order price (USDT)	Max. Leverage
BTCUSDT	Bitcoin	1×10^{-2}	125x
ETHUSDT	Ethereum	1×10^{-2}	100x
BCHUSDT	Bitcoin Cash	1×10^{-2}	75x
XRPUSDT	Ripple	1×10^{-4}	75x
EOSUSDT	EOS.IO	1×10^{-3}	75x
LTCUSDT	Litecoin	1×10^{-2}	75x
TRXUSDT	TRON	1×10^{-5}	50x
ETCUSDT	Ethereum Classic	1×10^{-3}	75x
LINKUSDT	Chainlink	1×10^{-3}	75x
XLMUSDT	Stellar	1×10^{-5}	50x
ADAUSDT	Cardano	1×10^{-5}	75x
XMRUSDT	Monero	1×10^{-2}	50x
DASHUSDT	Dash	1×10^{-2}	50x
ZECUSDT	Zcash	1×10^{-2}	50x
XTZUSDT	Tezos	1×10^{-3}	50x
ATOMUSDT	Cosmos	1×10^{-3}	25x
BNBUSDT	Binance Coin	1×10^{-3}	75x
ONTUSDT	Ontology	1×10^{-4}	50x
IOTAUSDT	IOTA	1×10^{-4}	25x
BATUSDT	Basic Attention Token	1×10^{-4}	50x

Theoretical framework

Unit-root test

The cointegration property can be used to identify the most appropriate pair from multiple combinations of coins. Both linear and nonlinear cointegration tests can be utilized for evaluation. Initially, the pair spread value without an intercept was defined as follows:

$$S_t = P_t^1 - \beta P_t^2 \tag{6}$$

Suppose that P_t^1 and P_t^2 are nonstationary time series. We use unit-root tests to establish whether the spread S_t is also a non-stationary process. The augmented Dickey-Fuller (ADF) unit-root test, as described by Dickey and Fuller (1979), uses a test equation applied to the demeaned-spread process S_t in the following form:

$$\Delta S_t = \beta S_{t-1} + \sum_{i=1}^{p-1} \gamma_i \Delta S_{t-i} + \epsilon_t, \tag{7}$$

where β is the coefficient of the lagged level of the series, γ_i are the coefficient of the lagged differences, ϵ_t is the error term, and p is the number of lags in the test. The null hypothesis of the ADF test is that the series has a unit root, i.e., $\beta = 0$. If the test statistic exceeds a critical value at a given significance level, the null hypothesis is rejected, indicating that the series is stationary and does not have a unit root. Conversely, if the test statistic is below the critical value, the null hypothesis cannot be rejected, implying that the series has a unit root and is non-stationary.

Traditional unit-root tests, such as the ADF test, assume that the data-generating process is linear. However, non-linear unit-root tests are designed to account for non-linearities in time series data and provide more accurate assessments of unit roots. There are several non-linear unit-root tests available in the literature, each with its own assumptions, methodologies, and advantages. Examples include the Teräsvirta (1994) test, the Zivot and Andrews (2002) test, the Kapetanios et al. (2003) test, and the Kapetanios (2005) test. These tests often involve estimating non-linear models, such as threshold auto-regressive (TAR), smooth transition auto-regressive (STAR), or other non-linear models, and computing test statistics to compare estimated model parameters with critical values.

The general self-exciting threshold auto-regressive (SETAR) model with n regimes applied to the demeaned spread process S_t is in the following form:

$$S_t = \sum_{i=1}^{p} \left(\phi_{i1} \mathbf{1}_{\{S_{t-d} \le c_1\}} + \sum_{j=1}^{n-1} \phi_{ij} \mathbf{1}_{\{c_j < S_{t-d} \le c_{j+1}\}} + \phi_{in} \mathbf{1}_{\{S_{t-d} > c_n\}} \right) S_{t-i} + \epsilon_t, \tag{8}$$

where d denotes the transition's delay, c_j represents the j-th threshold, and ϵ_t represents the error term. Kapetanios et al. (2003) proposed a test equation where the indicator function is replaced by an exponential smooth transition function in the form:

$$S_t = S_{t-1} + \sum_{i=1}^{p} \left(\gamma_{1i} \left(1 - e^{-\theta (S_{t-1} - c)^2} \right) S_{t-i} \right) + \epsilon_t \tag{9}$$

When c is set to zero and p to one, using the Taylor approximation, Eq. (9) can be illustrated as:

$$\Delta S_t = \delta (S_{t-1})^3 + \epsilon'_t \tag{10}$$

where $\delta = \gamma_1 \theta$ and $\epsilon'_t = f(\epsilon_t)$. The null hypothesis assumes that δ is equal to zero, whereas the alternative hypothesis posits that δ is less than zero. Note that the asymptotic standard normal distribution of the t-statistic for $\delta = 0$ against $\delta < 0$ is not applicable. However, its asymptotic critical values have been determined through stochastic simulations and are documented by Kapetanios et al. (2003).

Copula concept

Consider a continuous random variable X with a probability distribution function defined by $F_X(x) := \mathbb{P}(X \leq x)$. If F_X is strictly increasing, then F_X^{-1} is defined by $F_X^{-1}(u) = x \Leftrightarrow F_X(x) = u$. However, if F_X is constant on some interval, then the inverse function is not well defined by $F_X^{-1}(u) = x$. To avoid this problem, we can define $F_X^{-1}(u)$ for $0 < u < 1$ by the generalized inverse function such that

$$F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}, \tag{11}$$

Now, in the same way, we define another continuous random variable Y with distribution function F_Y and generalized inverse function F_Y^{-1} similar to Eq. (11). Given two continuous random variables X and Y, with distribution functions F_X and F_Y, respectively, the joint distribution function $F_{X,Y}$ can be written as:

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(F_X(X) \leq F_X(x), F_Y(Y) \leq F_Y(y)) \tag{12}$$

where the last equality follows from the fact that F_X and F_Y are both increasing. Then, we define random variables U and V such that $U := F_X(X)$ and $V := F_Y(Y)$. According to the probability integral transformation theorem, the probability distribution functions F_U and F_V of the random variables U and V, respectively, are uniformly distributed on [0, 1]. (See Casella and Berger (2021, p. 54–55) for the proof).

Definition A two-dimensional copula C is a function that maps the unit square $[0, 1]^2$ into the unit interval [0, 1], satisfying the following requirements:

1. $C(0, v) = C(u, 0) = 0$, for $0 \leq u, v \leq 1$.
2. $C(u, 1) = u$, and $C(1, v) = v$, for $0 \leq u, v \leq 1$.
3. $C(u_1, v_1) - C(u_1, v_2) - C(u_2, v_1) + C(u_2, v_2) \geq 0$, For $1 \geq u_1 > u_2 \geq 0$, and $1 \geq v_1 > v_2 \geq 0$

We can define several copula functions; however, the three requirements above should be satisfied by $C(u, v)$ to have a well-defined joint distribution function (Cherubini et al. 2011).

According to Sklar's theorem, there is a copula function C that could connect the uniform random variables U and V to the joint distribution function $F_{X,Y}$ as follows

$$F_{X,Y}(X,Y) = F_{X,Y}(F_X^{-1}(U), F_Y^{-1}(V)) := C(U,V), \tag{13}$$

Hence, we can rewrite the joint distribution function of X and Y in terms of standard uniform random variables U and V such that

$$F_{X,Y}(x,y) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)) := C(u,v) = C(F_X(x), F_Y(y)), \tag{14}$$

where $u = F_X(x)$ and $v = F_Y(y)$. Considering that $F_{X,Y}(x,y) = C(u,v)$, we can determine the copula density function $c(u,v)$ by

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u\, \partial v} = \frac{\partial^2 F_{X,Y}(x,y)}{\partial F_X(x)\, \partial F_Y(y)} = \frac{\frac{\partial^2 F_{X,Y}(x,y)}{\partial x\, \partial y}}{\frac{\partial F_X(x)}{\partial x}\frac{\partial F_Y(y)}{\partial y}} = \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} \tag{15}$$

Sklar's theorem enables us to separate the modeling of the marginal distributions $F_X(x)$ and $F_Y(y)$ from the dependence relation represented in C. We can characterize the conditional distribution functions by utilizing copula functions. The conditional distribution of Y given $X = x$ can be determined by computing the first partial derivative of the copula function, expressed as follows: $F_{Y|X}(y) = \frac{\partial}{\partial u} C(u,v)$ (Cherubini et al. 2011).

Copulas are invariant concerning strictly increasing transformations of the marginal distributions. For more details about the characterization of invariant copulas, see Klement et al. (2002). Additionally, we can increase the range of dependence captured by copulas by rotating them. The following equations show how to rotate a copula by 90, 180, and 270 degrees:

$$\begin{aligned} C_{90}(u_1,u_2) &:= C(1-u_2, u_1) \\ C_{180}(u_1,u_2) &:= C(1-u_1, 1-u_2) \\ C_{270}(u_1,u_2) &:= C(u_2, 1-u_1) \end{aligned} \tag{16}$$

Families of copulas

There are several types of copulas. This study focuses on three popular families of copulas: elliptical, Archimedean, and extreme value. Each copula type is characterized by its properties, such as its dependence structure and tail behavior, and is often used in different areas of statistics and finance. We chose these families of copulas based on their widespread use and established utility in financial modeling, as evidenced by the existing literature. These copula families offer distinct advantages for capturing different dependence structures in financial data.

An elliptical copula is constructed from a multivariate elliptical distribution. The density function of any elliptical distribution f_X is in the form

$$f_X(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = k_n |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g\left(\left(\boldsymbol{x} - \boldsymbol{\mu} \right)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right), \tag{17}$$

where $k_n \in \mathbb{R}$ is the normalizing constant and depends on the dimension n, $\boldsymbol{x}$ is an n-dimensional random vector with mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and a symmetric positive

Table 2 Pickands dependence function of some extreme-value copulas

Name	Pickands function $A(t)$	Parameters
Gumbel	$\left[t^{\theta}+(1-t)^{\theta}\right]^{1/\theta}$	$\theta \geq 1$
Tawn Type 1	$(1-\alpha)t+\left[(\alpha(1-t))^{\theta}+t^{\theta}\right]^{1/\theta}$	$\theta \geq 1,\ 0 \leq \alpha \leq 1$
Tawn Type 2	$(1-\beta)(1-t)\left[(1-t)^{\theta}+(\beta t)\theta\right]^{1/\theta}$	$\theta \geq 1,\ 0 \leq \beta \leq 1$

Table 3 Trading rules in terms S^1 and S^2

Trading rule	Signals
If $h^{1\|2} < \alpha_1$ and $h^{2\|1} > 1-\alpha_1$	open long S^1 and short S^2
If $h^{1\|2} > 1-\alpha_1$ and $h^{2\|1} < \alpha_1$	open short S^1 and long S^2
If $\|h^{1\|2}-0.5\| < \alpha_2$ and $\|h^{2\|1}-0.5\| < \alpha_2$	close both positions

Table 4 Trading rules in terms P^1 and P^2

Trading rule	Signals
If $h^{1\|2} < \alpha_1$ and $h^{2\|1} > 1-\alpha_1$	open long $\beta^2 \times P^2$ and short $\beta^1 \times P^1$
If $h^{1\|2} > 1-\alpha_1$ and $h^{2\|1} < \alpha_1$	open short $\beta^2 \times P^2$ and long $\beta^1 \times P^1$
If $\|h^{1\|2}-0.5\| < \alpha_2$ and $\|h^{2\|1}-0.5\| < \alpha_2$	close both positions

definite matrix $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$, and some function $g(.)$ which is independent of the dimension n (Czado 2019). This study used the multivariate Gaussian and Student-t, the most popular elliptical distributions. For more details about elliptical copulas, see the appendix.

Archimedean copulas are based on a generator function and can model dependence with tail dependence that decreases logarithmically or exponentially. The generator function determines the shape of the copula and influences the degree of tail dependence. According to Nelsen (2007), Archimedean copulas are defined as follows:

$$C(u_1,\ldots,u_n) = \phi^{[-1]}(\phi(u_1)+\cdots+\phi(u_n)) \tag{18}$$

where $\phi : [0,1] \to [0,\infty]$ is a continuous generator, strictly decreasing, and convex function such that $\phi(1) = 0$. In addition, $\phi^{[-1]} : [0,\infty] \to [0,1]$ is called the pseudo-inverse of ϕ function and is defined by

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & , 0 \leq t \leq \phi(0) \\ 0 & , \phi(0) \leq t \leq \infty \end{cases} \tag{19}$$

If $\phi(0)$ equals infinity, then the pseudo-inverse function $\phi^{[-1]}$ is equivalent to the inverse function ϕ^{-1}. In this research, both one-parameter (such as the Gumbel, Clayton, Frank, and Joe copulas) and two-parameter Archimedean copulas (such as BB1, BB6, BB7, and

BB8) are utilized. Table 10 in the appendix provides a list of the bivariate Archimedean copulas utilized in this study.

Extreme value copulas are used to model dependence with strong tail dependence, making them suitable for modeling extreme events. Suppose that $X_i = (X_{i1}, X_{i2})^T$, and $i \in \{1, 2, \cdots, n\}$ are independent and identically distributed random vectors with joint distribution function F and marginal distributions F_1 and F_2. According to Gudendorf and Segers (2010), we can define the bivariate vector of component-wise maxima $M_n = (M_{n1}, M_{n2})^T$ such that

$$M_{nj} := \max_{i \in \{1,2,\cdots,n\}} \left(X_{ij}\right), \quad j = 1, 2. \tag{20}$$

Then, the bivariate copula C_n of M_n is obtained by

$$C_n(u_1, u_2) = C_F\left(u_1^{1/n}, u_2^{1/n}\right)^n, \quad (u_1, u_2) \in [0, 1]^2. \tag{21}$$

In Eq. (21), if C_F exists such that

$$\lim_{n\to\infty} C_F(u_1^{1/n}, u_2^{1/n})^n = C(u_1, u_2), \quad (u_1, u_2) \in [0, 1], \tag{22}$$

then the bivariate copula C in (22) is called an extreme-value copula. Bivariate extreme-value copulas can be demonstrated in terms of a function $A(t)$ in this form:

$$C(u_1, u_2) = (u_1 u_2)^{A(\ln(u_2)/\ln(u_1 u_2))}, \quad (u_1, u_2) \in (0, 1]^2 \setminus \{(1, 1)\}, \tag{23}$$

where the function $A : [0, 1] \to [1/2, 1]$, which is called the Pickands dependence function, is convex and satisfies $\max(1 - t, t) \le A(t) \le 1$ for all $t \in [0, 1]$ (Gudendorf and Segers 2010). Some Archimedean copulas, such as the Gumbel copula, can be expressed by the extreme-value family. These special copulas create a hybrid category, including both the Archimedean and the extreme-value copulas, and are called Archimax copulas. Table 2 shows extreme-value copulas used in this research. Note that Tawn copula has three parameters and its Pickands' dependence function is in the form

$$A(t) = (1 - \beta) + (\beta - \alpha)t + \left[(\alpha(1 - t))^\theta + (\beta t)\theta\right]^{1/\theta}, \tag{24}$$

where $\theta \ge 1$ and $\alpha, \beta \in [0, 1]$. The simplified Tawn copula cases with $\beta = 1$ and $\alpha = 1$ are respectively called Tawn Type 1 and Tawn Type 2 copula and have two parameters.

Copula estimation

When the marginal probability density of X_1 and X_2 and their corresponding copula density $c(.)$ are given in their parametric with unknown parameters, we can estimate the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})^T$ with the maximum likelihood estimation (MLE) method, where $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\right)^T$ represent the marginal parameters and $\boldsymbol{\alpha}$ represents the copula parameters. The log-likelihood function for (X_1, X_2), where $X_i = (x_{i1}, \ldots, x_{in})$, can be expressed as

$$l(\boldsymbol{\theta}) = \sum_{j=1}^{n} \left[\log c\left(F_1(x_{1j};\boldsymbol{\beta}_1), F_2(x_{2j};\boldsymbol{\beta}_2);\boldsymbol{\alpha}\right) + \log f_1(x_{1j};\boldsymbol{\beta}_1) + \log f_2(x_{2j};\boldsymbol{\beta}_2)\right], \quad (25)$$

and the maximum likelihood estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, l(\boldsymbol{\theta}). \quad (26)$$

However, given that this approach is computationally expensive, an alternative method called the inference for the margins (IFM) two-step method can be used instead. This method is computationally easier to obtain compared to the full maximum likelihood estimation approach. First, we estimate the margins' parameters by performing the estimation of the univariate marginal distributions using the log-likelihood function, where the maximum likelihood estimator of $\boldsymbol{\beta}_i$ is

$$\hat{\boldsymbol{\beta}}_i = \underset{\boldsymbol{\beta}_i}{\operatorname{argmax}} \sum_{j=1}^{n} \left[\log f_i(x_{ij};\boldsymbol{\beta}_i)\right]. \quad (27)$$

Then, given the estimated marginal parameters, we transform data to the copula scale, develop the copula model, and estimate the copula parameters $\boldsymbol{\alpha}$ as follows:

$$\hat{\boldsymbol{\alpha}}_{ML} = \underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \sum_{j=1}^{n} \left[\log c\left(F_1(x_{1j};\hat{\boldsymbol{\beta}}_1), F_2(x_{2j};\hat{\boldsymbol{\beta}}_2);\boldsymbol{\alpha}\right)\right]. \quad (28)$$

We can also employ a semiparametric approach known as canonical maximum likelihood (CML) to estimate copula parameters without specifying the marginals. The empirical cumulative distribution function of $X_i = (x_{i1}, \ldots, x_{in})$ is

$$\widehat{F}_n^i(t) = \frac{\#\{X_i \le x\}}{n+1} = \frac{1}{n+1} \sum_{j=1}^{n} \mathbf{1}_{\{x_{ij} \le t\}}. \quad (29)$$

Then, the copula parameters are estimated using maximum likelihood estimation as follows:

$$\hat{\boldsymbol{\alpha}}_{ML} = \underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \sum_{j=1}^{n} \left[\log c\left(\widehat{F_n^1}(x_{1j}), \widehat{F_n^2}(x_{2j});\boldsymbol{\alpha}\right)\right]. \quad (30)$$

Implementation methodology

Two approaches have been used so far in Chapter 2 to calculate the mispricing index: one utilizes Eq. (4), whereas the other employs Eq. (5). Both approaches have certain limitations: the return-based method's entry and exit signals are solely connected to the previous period's return, while the level-based method does not inherently demonstrate mean-reverting behavior, potentially adversely affecting the overall profitability of the strategy. This research proposes a new method that addresses the limitations inherent in both of the previously discussed methods, as outlined below.

Reference-asset-based copula method The proposed methodology deviates from conventional approaches by incorporating stationary spread processes instead of log-returns. For each asset i, the spread process S_t^i is defined as:

$$S_t^i = P_t^{Reference} - \hat{\beta}^i P_t^i \tag{31}$$

where $P_t^{Reference}$ represents the price of a specified reference asset at time t, while $\hat{\beta}^i$ is the estimated linear regression coefficient between the reference asset and the asset i. S^i is a stationary process wherein the reference asset and asset i are cointegrated. Identifying assets cointegrated with the reference asset facilitates the derivation of stationary spread processes. By utilizing these spread processes as our tradable assets, we can build copula models for each pair of stationary spreads. This approach provides signals for the mispricing index without the need to explicitly accumulate the index itself. Importantly, this signifies that the decision-making process for entering or exiting a trade is not exclusively dependent on the most recent market movement, as we have employed a stationary price process instead of relying on log-returns.

Within the pairs-trading strategy framework employed in this study, the spread process is defined as a linear combination of BTCUSDT (as a reference coin) and other cryptocurrency coins. Bitcoin is chosen for its substantial influence on the overall cryptocurrency market, often accounting for a significant portion of the total cryptocurrency market cap. Price movements play an essential role in shaping trends and sentiments in the cryptocurrency market. Furthermore, compared to newer or less established cryptocurrencies, Bitcoin is often perceived as relatively more stable. These factors make BTCUSDT an ideal reference asset for our analysis. The spread process for each coin pair is then computed as follows:

$$S_t^i = \text{BTCUSDT}_t - \hat{\beta}^i P_t^i \quad i = 1, 2, \ldots, 19 \tag{32}$$

This allows us to identify 19 pairs and select the optimal pairs during the formation period for trading in the subsequent trading period. To determine the optimal pairs, we use the linear Engle-Granger (EG) two-step method and the non-linear Kapetanios-Shin-Snell (KSS) cointegration tests to identify cointegrated coins. However, because there may be multiple cointegrated pairs, we must add another criterion to rank them. To do this, we calculate Kendall's Tau (τ), which is a measure of correlation for ranked data and defined as:

$$\tau(S^i, S^j) = \tau_{ij} = \frac{\text{Number of concordant pairs} - \text{Number of discordant pairs}}{\text{Total number of pairs}} \tag{33}$$

where

$$\text{Number of concordant pairs} = \sum_{n=1}^{N-1} \sum_{m=n+1}^{N} \text{sgn}(S_{[n]}^i - S_{[m]}^i)\text{sgn}(S_{[n]}^j - S_{[m]}^j),$$

$$\text{Number of discordant pairs} = \sum_{n=1}^{N-1} \sum_{m=n+1}^{N} \text{sgn}(S_{[n]}^i - S_{[m]}^i)\text{sgn}(S_{[m]}^j - S_{[n]}^j),$$

and total number of pairs $= N(N-1)/2$. Here, N is the number of data points; $S^i_{[n]}$ and $S^j_{[n]}$ are the rankings of the n-th data point in two different variables; and $\text{sgn}(x)$ is the sign function, which is 1 if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$.

After calculating τ of BTCUSDT with each of the 19 altcoins, we select the two altcoins that have the highest τ with BTCUSDT and create their corresponding pairs. During the course of one week, the chosen pairs are traded and can be substituted with different pairs at the start of each trading period. Instead of trading a pair of coins, we employ a pair of spreads where each of them contains BTCUSDT. In this case, a long position in one spread and a short position in the other implies that BTCUSDT will not be traded at all and only plays an intermediary role between two other coins.

Now, we estimate the probability distribution function of spread processes (marginals). We fit various distributions such as Gaussian, Student-t, and Cauchy to the data to identify the best-fitting distribution. By employing statistical methods or maximum likelihood estimation, we estimate the specific parameters associated with each distribution. To evaluate the goodness of fit, we calculate the Akaike information criterion (AIC) values for the candidate distributions, ultimately selecting the distribution with the lowest AIC value as the best-fitting option. Suppose that F^i is the fitted cumulative distribution function of the spread process S^i.

The Probability Integral Transform is employed to convert the spreads to random variables $\mathbf{U_1} := F^1(\mathbf{S^1})$ and $\mathbf{U_2} := F^2(\mathbf{S^2})$ with a standard uniform distribution. The next step is to determine a fitting copula model for $\mathbf{U_1}$ and $\mathbf{U_2}$. We select some potential copulas and estimate the corresponding parameters by the maximum likelihood method.

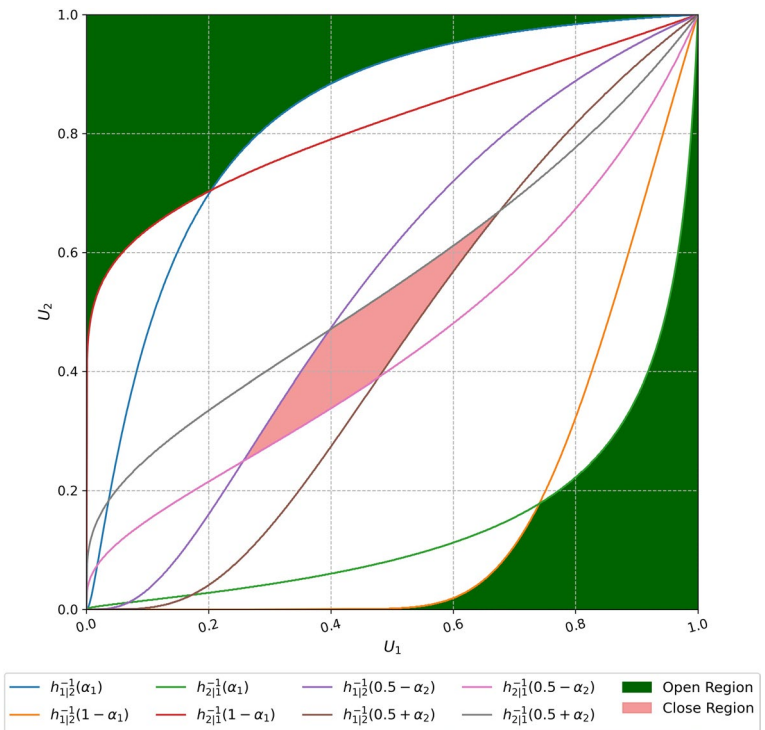

Fig. 1 Confidence bands of Gumbel copula ($\theta = 2$) at $\alpha_1 = 5\%$ and $\alpha_2 = 10\%$

Finally, using the Akaike information criterion (AIC), we distinguish the most fitted copula model.

During the trading period, using the hourly realizations of random variables U_1 and U_2, which are the transformed values of spread processes, we calculate the copula conditional probabilities $h^{1|2}$ and $h^{2|1}$ defined in Eq. (4) for the selected pairs of the week. When $h^{1|2}$ is higher (lower) than 0.5, the first coin can be considered to be overvalued (undervalued) relative to the second one. Similarly, when $h^{2|1}$ is higher (lower) than 0.5, the second coin can be considered to be overvalued (undervalued) relative to the first one. Therefore, we can interpret mispricing as conditional probabilities in Eq. (4) minus 0.5. We denote the trading thresholds as α_1 and α_2. We study the optimal triggers via back-testing. The opening and closing signals are generated by the rules outlined in Tables 3 and 4.

Figure 1 illustrates the confidence bands in the trading rules for the Gumbel copula with a parameter value of $\theta = 2$ under the conditions $\alpha_1 = 5\%$ and $\alpha_2 = 10\%$. If the data point (u_{1t}, u_{2t}) falls within the top green (down green) area, this suggests that S_1 is undervalued (overvalued) and S_2 is overvalued (undervalued), which may indicate an opportunity to open a position. By contrast, if (u_{1t}, u_{2t}) falls within the red area, it can signal the need to close the positions.

Key assumptions and limitations

The effective execution of our strategy relies on certain assumptions and acknowledges specific limitations. The key factors shaping our approach include:

- As we mentioned earlier, a pairs trading cycle comprises formation and trading periods. In this research, cycles of pairs trading are conducted within a month, with three weeks allocated for the formation step and the remaining week designated for the trading step.
- We carry out 104 cycles that move dynamically over time, with each cycle sharing three-quarters of its data with its previous or subsequent cycle. For more clarification, see Fig. 2.
- If we fail to identify a minimum of two stationary spread processes within a specified formation period, we abstain from trading during its corresponding trading period.

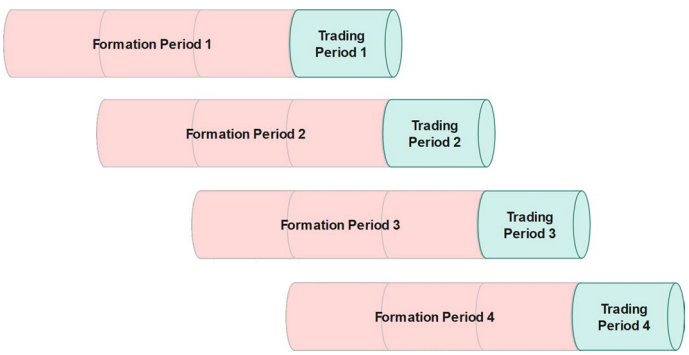

Fig. 2 The scheme of formation and trading Periods

- When a position is opened during the trading week, it will be closed at the end of that week, regardless of whether the pair of spreads passes or fails the cointegration tests in the following week.
- The ADF unit-root test is conducted with a significance level of 10%. For the KSS unit-root test, the asymptotic critical value at a 10% significance level is −1.92.
- To assess the performance of our proposed trading strategy (reference-asset-based approach), we examine various entry thresholds ($\alpha_1 = 10\%, 15\%,$ and 20%). Although we analyzed different values for the exit threshold (α_2), we have chosen not to include these results in this study as they do not significantly impact the profitability of the strategy. Consequently, we keep α_2 fixed at 10%.
- To assess the performance of our proposed trading strategy (reference-asset-based approach), we conduct a comprehensive performance comparison with established methodologies employed in previous research studies. Our evaluation includes backtesting the cointegration approach, the return-based copula approach, and the value-based copula approach. The process for selecting the optimal pairs is the same across all three approaches. The distinction lies in the manner in which trading signals are generated. In the cointegration approach, the z-score provides a measure of how far the current spread deviates from its historical average in terms of standard deviations. In the return-based approach, the mispricing index in Eq. (4) generates long or short signals. In the level-based approach, the trading signals are determined by the cumulative mispricing index defined in Eq. (5).
- In the cointegration approach, the entry threshold (t_o) is set at ± 2, whereas the exit threshold (t_c) is established at ± 1. The look-back window (N) employed for calculating the z-score spans 1 day (24 h).
- For the return-based copula approach, the entry threshold (α_1) is defined as 10%, and the exit threshold (α_2) is consistently maintained at 10%.
- For the level-based copula approach, the entry threshold (t_o) is specified at ± 1, whereas the exit threshold (t_c) is fixed at 0.
- When utilizing 5-min level data, the traditional return-based and level-based copula approaches often encounter issues, particularly if the analyzed pairs include cryptocurrencies with low liquidity. In such cases, the 5-min log returns for the less liquid assets may frequently equal zero, which contradicts the continuous random variable assumption fundamental to copula models. In contrast, our new reference asset-based approach does not rely on log-returns and instead uses BTCUSDT-a highly liquid cryptocurrency-as the reference asset. This method ensures that the spread processes defined in Eq. (32) maintain continuity, thereby aligning more closely with the underlying assumptions of the model and providing a more robust framework for analyzing cryptocurrency pairs.
- Realized profit and loss are determined by considering the commission fees and the difference between the opening and closing prices of a position.
- We initially invest 20,000 USDT, with the coins' weights set to ensure that each side has a maximum initial capital of around 20,000 USDT. It is important to note that the size of an order relative to the trading volume of the market significantly influences market impact. As highlighted by Bağcı et al. (2004), minimizing potential trading losses due to price impacts requires the capital size to ideally be less than 1% of the

Table 5 Occurrence rate of copulas in the study

Copulas and their rotations	Hourly data		5-Min data	
	EG Test (%)	KSS Test (%)	EG Test (%)	KSS Test (%)
Gaussian	4.0	6.7	4.0	7.7
Student-t	6.1	3.8	5.9	5.8
Clayton	6.1	8.7	4.0	1.9
Frank	5.1	3.8	3.0	1.9
Gumbel	0.0	5.8	0.0	2.9
Joe	8.1	5.8	8.9	6.7
BB1	6.1	3.8	4.0	2.9
BB6	1.0	1.9	3.0	2.9
BB7	16.2	13.5	10.9	14.4
BB8	10.1	8.7	15.8	13.5
Tawn type 1	23.2	24.0	20.8	25.0
Tawn type 2	14.1	13.5	19.8	14.4

daily market volume. To align with this recommendation, we analyze the daily trading volume of each cryptocurrency coin in our study over the analysis period. From this analysis, we determine that setting the initial capital size at 20,000 USDT aligns with the suggested threshold.

- In practice, at the start of each trading week, we determine a fixed quantity for each selected cryptocurrency pair, ensuring the product of the opening price and the quantity does not exceed 20,000 USDT for each coin. Specifically, for coin1, we set the quantity Q_1 such that the product of Q_1 and its opening price P_1 is less than or equal to 20,000 USDT. We apply the same principle to coin2 with Q_2 and its price P_2. Once the quantities Q_1 and Q_2 are set, they remain constant throughout the week. Whenever we receive a market signal, we either go long or short on coin1 using the predetermined quantity Q_1 and take the opposite position on coin2 with quantity Q_2. Thus, the total investment amount may vary from the 20,000 USDT limit because of price fluctuations throughout the week. For instance, if market conditions push the total value of our positions to 22,000 USDT, we may use leverage to cover the additional 2,000 USDT, or conversely, the total investment might drop to 18,000 USDT if prices fall. The quantities Q_1 and Q_2 are updated at the beginning of each subsequent week.
- It is assumed that all trades are executed using market orders, which typically incur higher fees (known as taker fees) compared to the lower fees charged for limit orders (known as maker fees). For instance, on Binance, the maker fee of USDT-margined futures is set at 0.02%, while the taker fee is set at 0.04%.

Empirical results

Table 5 presents the occurrence rate of selected copulas and their rotations over 104 trading weeks. The results indicate that copulas of extreme value, such as Tawn type 1 and 2, and certain two-parameter Archimedean copulas, particularly BB7 and BB8, play a significant role in the process of selecting the appropriate model.

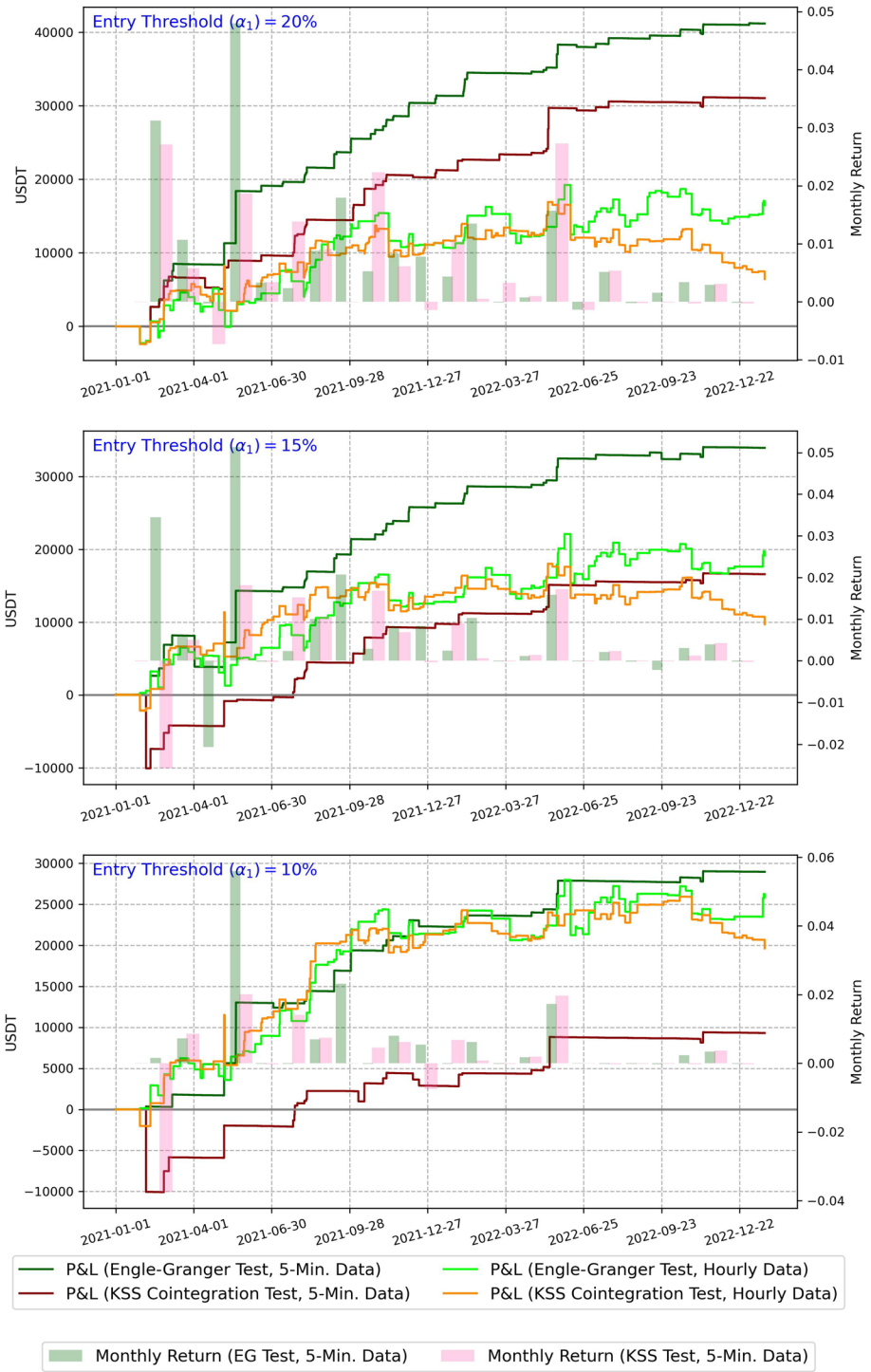

Fig. 3 Copula-based pairs trading strategy P&L and monthly returns

Tables 8 and 9 in the Appendix display the cryptocurrency pairs chosen for each trading week. The results of our trading strategy's profit and loss calculations with varying entry thresholds (α_1) are illustrated in Fig. 3. We assess the risk-adjusted performance of a strategy using the Sharpe ratio. Table 6 reports the annual returns, volatility, and

Table 6 Results of the proposed pairs trading strategy in this study utilizing closed prices of twenty cryptocurrencies from 22/01/2021 to 19/01/2023

	Hourly data			**5-min Data**		
	$\alpha_1 = 0.10$	$\alpha_1 = 0.15$	$\alpha_1 = 0.20$	$\alpha_1 = 0.10$	$\alpha_1 = 0.15$	$\alpha_1 = 0.20$
The Proposed Pairs Trading Strategy (Reference-Asset-Based Copula Approach with EG Test)						
Total Gross Return	144.1%	113.0%	101.9%	155.1%	183.9%	222.3%
Transaction Cost Percentage	−14.9%	−17.2%	−19.5%	−10.3%	−14.1%	−16.4%
Total Net Return	129.2%	95.8%	82.3%	144.8%	169.8%	205.9%
Annualized Net Return	51.6%	40.0%	35.1%	56.7%	64.5%	75.2%
Annualized Standard Deviation	35.5%	37.7%	41.4%	21.3%	25.2%	19.9%
Annualized Sharpe Ratio	1.45	1.06	0.85	2.66	2.56	3.77
Maximum Drawdown	−36.6%	−39.6%	−41.6%	−37.1%	−39.2%	−30.5%
Total Return over Maximum Drawdown	3.53	2.42	1.98	3.90	4.33	6.76
Number of Transactions	176	204	228	188	242	266
The Proposed Pairs Trading Strategy (Reference-Asset-Based Copula Approach with KSS Test)						
Total Gross Return	113.6%	67.3%	53.6%	55.7%	94.4%	170.3%
Transaction Cost Percentage	−15.5%	−18.8%	−21.5%	−9.1%	−11.4%	−15.2%
Total Net Return	98.1%	48.5%	32.1%	46.6%	83.0%	155.1%
Annualized Net Return	40.9%	21.9%	15.0%	21.1%	35.4%	60.0%
Annualized Standard Deviation	31.9%	33.0%	34.8%	48.5%	46.9%	17.7%
Annualized Sharpe Ratio	1.25	0.66	0.43	0.44	0.75	3.39
Maximum Drawdown	−34.6%	−34.7%	−38.5%	−160.0%	−160.0%	−163.5%
Total Return over Maximum Drawdown	2.84	1.40	0.84	0.29	0.52	0.95
Number of Transactions	184	222	252	164	200	256

Sharpe ratios of the proposed strategy, together with the maximum drawdown and the return over maximum drawdown (RoMaD), which serve as alternatives to the Sharpe ratio, over two data sampling frequencies (hourly and 5-min).[3]

As shown in Table 6, our proposed pairs trading strategy using the EG cointegration test yields satisfactory total net returns for both hourly and 5-min data. Specifically, 5-min data significantly outperform hourly data, with the best total net return at $\alpha_1 = 0.20$ achieving 205.9%. Using 5-min data (with the EG test), the annualized net return increases with higher entry thresholds, ranging from 56.7% at $\alpha_1 = 0.10$ to 75.2% at $\alpha_1 = 0.20$. This is because the increased trading frequency allows for capitalizing on more frequent short-term price discrepancies or trends, which are more prevalent in high-frequency data. In contrast, using hourly data (with the EG test), the annualized return decreases with higher entry thresholds, ranging from 51.6% in $\alpha_1 = 0.10$ to 35.1% at $\alpha_1 = 0.20$. This suggests that the lower frequency of data may result in missing out on optimal trading times, making higher thresholds less effective in capturing profitable opportunities in less volatile or slower-moving market conditions.

Transaction costs increase with higher α_1, which could indicate more frequent trading. Despite the higher costs, the total net return remains strong, especially for finer (5-min) data. Although the 5-min strategy shows higher returns, its risk (measured by

[3] Transaction fees are taken into account in all calculations.

the annualized standard deviation) is lower than that of the hourly data for higher levels α_1. This suggests a more efficient market capture on a finer scale.

For the EG test, maximum drawdowns in the hourly data range between −36.6% and −41.6% and increase with parameter α_1. This finding suggests a direct correlation between the risk threshold α_1 and exposure to potential losses. RoMaD values decrease from 3.53 to 1.98 as α_1 increases, indicating less favorable risk-return ratios at higher thresholds. In contrast, the 5-min data displays slightly better maximum drawdown values and significantly higher RoMaD values (3.90, 4.33, and 6.76), suggesting that shorter sampling intervals may provide more efficient risk-return trade-offs under the EG test, especially at $\alpha_1 = 0.20$, which combines a relatively lower maximum drawdown with higher returns.

A comparison of the EG and KSS tests for the proposed cryptocurrency pairs trading strategy reveals distinct performance profiles in both hourly and 5-min data sets. Under the EG test, the hourly and 5-min data strategies demonstrate higher overall net returns and more favorable risk-adjusted returns, as indicated by higher Sharpe ratios and RoMaD values. Specifically, the 5-min data in the EG test achieve an impressive annualized net return of up to 75.2% and a Sharpe ratio of 3.77, suggesting an efficient balance between risk and return. In contrast, the results of the KSS test are significantly inferior, with the 5-min data suffering from extraordinarily high maximum drawdowns, exceeding −160%, and significantly lower RoMaD values. This indicates a less favorable balance between risk and return. Furthermore, the net returns and Sharpe ratios under the KSS test are consistently lower than those observed in the EG test, underscoring the potential volatility and risk associated with the KSS testing approach, particularly with high-frequency data.

In addition, we assess our strategy by comparing it to the conventional approaches employed in prior research studies, including the cointegration approach, the return-based copula approach, and the level-based copula approach. As shown in Table 7, the Pairs Trading Strategy, using a cointegration approach, exhibited a distinct disparity between performance on hourly and 5-min data. The results for the hourly data were relatively moderate, with total gross returns at 33.8% and 19.3% for EG and KSS tests, respectively. This strategy was heavily impacted by transaction costs, which were particularly high and resulted in negative net returns across both tests. The 5-min data showed a more pronounced effect, yielding higher gross returns; however, it also faced astronomical transaction costs, leading to drastically negative net returns. This suggests that while the approach can capture gains, the cost structure makes it unsustainable, especially at higher trading frequencies. Note that in cases where the calculation of the geometric annualized net return was not feasible due to the highly negative values of total net returns, the arithmetic annualized net return was computed instead.

The return-based copula approach to pairs trading delivered higher gross returns compared to the cointegration method, particularly in the 5-min data where returns reached up to 249.6% (EG Test). However, this approach also suffered from extreme transaction costs, which drastically eroded profits and resulted in significant net losses (−40.1 to −1138.9% across various tests and data frequencies). The approach appears to aggressively capitalize on market movements but at a cost inefficiency that undermines its viability. This is reflected in very negative annualized Sharpe ratios and substantial maximum drawdowns, indicating high risk without proportional returns.

Table 7 Results from pairs trading strategies employed in prior research studies utilizing closed prices of twenty cryptocurrencies from 22/01/2021 to 19/01/2023

Pairs trading strategy (cointegration approach)				
$t_o = \pm 2,\ t_c = \pm 1,\ N = 24$	**Hourly data**		**5-min Data**	
	EG Test	**KSS Test**	**EG Test**	**KSS Test**
Total Gross Return	33.8%	19.3%	121.4%	63.6%
Transaction Cost Percentage	−57.5%	−28.5%	−681.2%	−339.9%
Total Net Return	−23.7%	−9.2%	−559.8%	−276.3%
Annualized Net Return	−12.8%	−4.6%	−283.4%	−140.0%
Annualized Standard Deviation	14.7%	7.6%	1828.4%	675.4%
Annualized Sharpe Ratio	−0.87	−0.61	−0.15	−0.21
Maximum Drawdown	−37.6%	−25.9%	−558.5%	−558.5%
Total Return over Maximum Drawdown	−0.63	−0.24	−1.00	−0.49
Number of Transactions	1324	1296	15676	15692
Pairs Trading Strategy (Return-Based Copula Approach)				
$\alpha_1 = 10\%,\ \alpha_2 = 10\%$	**Hourly data**		**5-min Data**	
	EG Test	**KSS Test**	**EG Test**	**KSS Test**
Total Gross Return	61.9%	39.1%	249.6%	127.1%
Transaction Cost Percentage	−102.0%	−52.0%	−1388.4%	−689.2%
Total Net Return	−40.1%	−12.9%	−1138.9%	−562.1%
Annualized Net Return	−22.7%	−6.5%	−577.0%	−284.8%
Annualized Standard Deviation	22.6%	10.4%	11880.2%	684.5%
Annualized Sharpe Ratio	−1.00	−0.62	−0.05	−0.42
Maximum Drawdown	−53.5%	−53.3%	−1102.9%	−1102.9%
Total Return over Maximum Drawdown	−0.75	−0.24	−1.03	−0.51
Number of Transactions	2204	2172	31354	31420
Pairs Trading Strategy (Level-Based Copula Approach)				
$t_o = \pm 1,\ t_c = 0$	**Hourly Data**		**5-min Data**	
	EG Test	**KSS Test**	**EG Test**	**KSS Test**
Total Gross Return	22.5%	5.5%	22.2%	8.9%
Transaction Cost Percentage	−3.1%	−1.7%	−4.3%	−2.0%
Total Net Return	19.4%	3.8%	17.9%	6.9%
Annualized Net Return	9.3%	1.8%	8.6%	3.3%
Annualized Standard Deviation	9.8%	5.4%	9.0%	4.5%
Annualized Sharpe Ratio	0.95	0.34	0.96	0.74
Maximum Drawdown	−9.5%	−9.6%	−15.9%	−15.9%
Total Return over Maximum Drawdown	2.03	0.40	1.13	0.44
Number of Transactions	66	70	136	122
Buy & hold strategy				
	Bitcoin buy & hold		**Portfolio buy & hold**	
Total Net Return	−31.2%		28.3%	
Annualized Net Return	−17.1%		13.3%	
Annualized Standard Deviation	78.1%		103.4%	
Annualized Sharpe Ratio	−0.22		0.13	

Table 7 (continued)

Buy & hold strategy		
	Bitcoin buy & hold	**Portfolio buy & hold**
Maximum Drawdown	−77.4%	−82.7%
Total Return over Maximum Drawdown	−0.40	0.34

In the level-based copula approach, we initiated a long-short trade when one of the CMIs in Eq. (5) surpasses the opening threshold (t_o), while simultaneously, the other CMI drops below $-t_o$ for the chosen pair. The positions are terminated when the CMI of the short position falls below the closing threshold (t_c) and the CMI of the long position rises above $-t_c$. According to the data in Table 7, the level-based copula approach presents a more conservative and stable outcome than other conventional strategies. It generates positive net returns across all test scenarios with lower transaction costs and maximum drawdowns. This strategy achieved annualized net returns up to 9.3% and maintained Sharpe ratios close to or above 0.95, which still implies that our proposed pairs trading method with $\alpha_1 = 0.20$ exhibits superior risk-adjusted performance compared to the level-based approach. Furthermore, it's noteworthy that our proposed strategy entails a notably higher frequency of transactions in contrast to the level-based approach.

Finally, to compare our strategy's results with those of the buy-and-hold strategy, we use a passive investment approach that involves holding a relatively steady portfolio for an extended period, despite short-term market fluctuations. On the one hand, the Bitcoin buy-and-hold strategy shows a negative annualized Sharpe ratio (−0.22), indicating poor risk-adjusted performance. On the other hand, the portfolio buy-and-hold strategy[4] shows a positive annualized Sharpe ratio (0.13), indicating better risk-adjusted performance than the Bitcoin buy-and-hold strategy. The proposed pairs-trading strategy demonstrates impressive performance, surpassing the portfolio buy-and-hold strategy by up to 29 times.

In summary, the proposed pairs-trading strategy employing a reference-asset-based copula approach with EG and KSS tests significantly outperforms other approaches examined in terms of both risk-adjusted returns and overall profitability. This strategy demonstrates a robust capability to achieve high net returns of up to 205.9% and superior Sharpe ratios as high as 3.77, indicating an excellent balance between return and risk. In contrast, other strategies, such as the cointegration approach, return-based, and level-based copula methods, though varying in their risk and return profiles, consistently underperformed with either negative net returns or significantly lower Sharpe ratios. Additionally, conventional buy & hold strategies lag, suffering from substantial drawdowns and lower overall returns. The reference asset-based approach not only minimizes drawdowns compared to these methods but also maximizes efficiency and profitability, making it the most advantageous strategy among those evaluated within the volatile cryptocurrency market.

[4] The buy-and-hold strategy in the portfolio involves investing in all twenty cryptocurrency coins with equal weights at the start of the study, retaining them throughout the trading periods, and ultimately selling them at the end of the study period.

Conclusion

This study develops a novel pairs trading framework for twenty Binance USDT-margined futures coins, combining copula-based and cointegration-based approaches. The methodology sets BTCUSDT as the reference asset and identifies other cryptocurrency coins that are cointegrated with it. We employ both the EG two-step method and the KSS cointegration test to detect cointegration, ranking the cointegrated coins based on Kendall's Tau correlation coefficients. The top two correlated assets are selected for trading over a one-week period, with weekly updates. Trading rules are based on the copula conditional probabilities of the spread processes corresponding to the selected assets. Various trading triggers are established and the strategy is rigorously back-tested.

Our strategy significantly outperforms traditional pairs trading methods including: the cointegration approach, return-based copula approach, level-based copula approach, and various buy-and-hold strategies across key performance metrics, such as Sharpe ratios and net returns. It demonstrates notable improvements in risk-adjusted returns and efficiency, particularly in high-frequency trading environments like the 5-min data sampling. This performance underscores the advantages of our approach in capturing short-term price discrepancies more effectively than traditional methods. Moreover, our findings emphasize the importance of choosing the appropriate entry thresholds and trading frequencies, as these factors critically influence the profitability and volatility of the trading strategy.

This study contributes to the academic literature by providing a refined methodological approach to pairs trading. Further, it also offers practical insights that could be beneficial for practitioners in the cryptocurrency trading space. By effectively integrating different concepts and adapting them to the unique characteristics of cryptocurrencies, our proposed strategy promises enhanced profitability, serving as a valuable tool for traders and investors seeking to exploit inefficiencies in rapidly evolving markets.

Appendix

Elliptical Copulas: The density function of any elliptical distribution f_X is shown in Eq. (17). In the case of bivariate Gaussian distribution $g(x) := e^{-x/2}$, $k_n := 1/(2\pi)$, and the probability density function of $\boldsymbol{X} = (X_1, X_2)$ is

Table 8 Selected pairs using unit-root tests and Kendall's τ coefficient for weeks 1–52

Week	ADF unit-root test (5-min data)			KSS unit-root test (5-min data)			ADF unit-root test (hourly data)			KSS unit-root test (hourly data)		
	Pair	*P*-Value (S_1)	*P*-Value (S_2)	Pair	t-stat (S_1)	t-stat (S_2)	Pair	*P*-Value (S_1)	*P*-Value (S_2)	Pair	t-stat (S_1)	t-stat (S_2)
1	LTC-BCH	0.087	0.059	LTC-BCH	−2.83	−2.41	ETH-LTC	0.075	0.070	BCH-ETC	−2.72	−1.96
2	LTC-BCH	0.026	0.016	LTC-XRP	−3.90	−4.94	LTC-BCH	0.023	0.010	LTC-BCH	−3.07	−2.09
3	ETH-LTC	0.048	0.032	ETH-LTC	−2.01	−4.33	ETH-LTC	0.075	0.037	ETH-LTC	−2.10	−2.90
4	ETH-LTC	0.058	0.022	ETH-LTC	−2.52	−2.15	ETH-LTC	0.095	0.022	ETH-ETC	−2.35	−2.64
5	LTC-EOS	0.097	0.059	LTC-BCH	−2.74	−2.80	LTC-EOS	0.096	0.082	LTC-BCH	−2.48	−2.66
6	TRX-XRP	0.027	0.082	LTC-BCH	−2.80	−2.66	LINK-TRX	0.090	0.069	LTC-BCH	−2.44	−2.46
7	TRX-LINK	0.036	0.097	LTC-BCH	−2.57	−2.40	LINK-TRX	0.097	0.051	LTC-BCH	−2.44	−2.32
8	ETH-LTC	0.010	0.010	ETH-BCH	−2.73	−2.25	ETH-LTC	0.010	0.010	ETH-BCH	−2.66	−2.47
9	ETH-LTC	0.076	0.010	LTC-BCH	−2.90	−2.03	ETH-LTC	0.042	0.022	LTC-BNB	−3.43	−4.80
10	BCH-BNB	0.036	0.016	LTC-BCH	−2.70	−3.88	ETH-BCH	0.097	0.073	LTC-BCH	−3.03	−2.02
11	LTC-BCH	0.056	0.010	LTC-BCH	−1.99	−3.55	LTC-BCH	0.079	0.010	LTC-BCH	−2.58	−2.63
12	ADA-ATOM	0.010	0.010	LTC-LINK	−2.24	−1.94	ATOM-ADA	0.010	0.010	BCH-ATOM	−1.96	−3.02
13	–	–	–	ADA-XTZ	−3.28	−2.33	BAT-ONT	0.095	0.021	ADA-XTZ	−2.61	−2.07
14	ADA-EOS	0.025	0.010	LTC-BCH	−2.35	−2.90	ADA-EOS	0.064	0.010	LTC-ADA	−2.15	−3.23
15	LTC-EOS	0.027	0.041	LTC-EOS	−2.26	−2.72	EOS-LTC	0.044	0.027	EOS-LTC	−2.33	−2.14
16	BNB-TRX	0.062	0.081	TRX-XRP	−4.00	−3.52	BAT-IOTA	0.022	0.010	TRX-BNB	−2.56	−2.06
17	TRX-BNB	0.043	0.035	TRX-LTC	−2.30	−2.20	BNB-TRX	0.067	0.036	BCH-BNB	−1.96	−3.08
18	LTC-TRX	0.090	0.023	BNB-TRX	−2.61	−4.03	BCH-LTC	0.082	0.086	TRX-IOTA	−3.08	−2.64
19	ETH-TRX	0.065	0.010	ETH-LTC	−2.83	−2.12	ETH-TRX	0.059	0.023	DASH-BCH	−2.77	−3.42
20	ETH-LINK	0.089	0.082	ETH-ONT	−4.79	−1.93	ETH-LTC	0.078	0.088	BCH-EOS	−2.07	−2.46
21	TRX-LTC	0.019	0.012	TRX-LTC	−2.18	−1.97	TRX-LTC	0.010	0.010	TRX-LTC	−2.17	−2.05
22	–	–	–	BNB-LTC	−2.05	−2.15	–	–	–	LTC-ETC	−1.99	−2.39
23	LINK-XMR	0.080	0.052	BNB-LINK	−2.17	−3.46	–	–	–	BNB-LINK	−2.10	−2.06
24	ETH-BNB	0.057	0.076	ETH-BNB	−2.07	−2.64	ETH-BNB	0.072	0.092	ETH-BNB	−2.05	−2.93
25	LTC-BCH	0.031	0.036	BNB-LTC	−2.04	−2.12	LTC-EOS	0.018	0.018	LTC-BNB	−1.97	−2.32
26	LTC-BCH	0.010	0.010	LTC-BCH	−2.81	−2.96	ETH-LTC	0.085	0.010	LTC-XRP	−2.50	−3.91
27	BNB-XRP	0.049	0.079	BCH-ONT	−2.08	−1.95	BNB-DASH	0.041	0.043	XRP-DASH	−2.02	−2.48

Table 8 (continued)

Week	ADF unit-root test (5-min data)			KSS unit-root test (5-min data)			ADF unit-root test (hourly data)			KSS unit-root test (hourly data)		
	Pair	*P*-Value (S_1)	*P*-Value (S_2)	Pair	t-stat (S_1)	t-stat (S_2)	Pair	*P*-Value (S_1)	*P*-Value (S_2)	Pair	t-stat (S_1)	t-stat (S_2)
28	ETH-LINK	0.058	0.057	ETH-LINK	−1.98	−2.38	ETH-BCH	0.060	0.096	ETH-BCH	−1.95	−2.35
29	BNB-LTC	0.060	0.012	BNB-LTC	−2.24	−2.56	BNB-LTC	0.068	0.015	BNB-LTC	−2.42	−2.99
30	BNB-LTC	0.065	0.015	BNB-LTC	−2.11	−2.41	BNB-EOS	0.044	0.030	BNB-EOS	−2.22	−3.77
31	ETH-LTC	0.023	0.077	LTC-LINK	−2.67	−2.84	ETH-LINK	0.030	0.015	LTC-LINK	−2.52	−2.72
32	ETH-LTC	0.049	0.051	ETH-LTC	−2.75	−2.17	ETH-LINK	0.058	0.096	ETH-LTC	−2.46	−1.95
33	LTC-EOS	0.043	0.074	BNB-LTC	−2.19	−2.99	LTC-XRP	0.063	0.032	LTC-BNB	−2.73	−1.93
34	XRP-EOS	0.031	0.066	BNB-XRP	−4.68	−3.72	EOS-XRP	0.049	0.041	EOS-LTC	−2.36	−1.97
35	ETH-XRP	0.071	0.043	BNB-XRP	−4.47	−2.54	ETH-EOS	0.061	0.041	EOS-BNB	−2.15	−3.96
36	BNB-XRP	0.095	0.096	LTC-XMR	−6.78	−2.72	BNB-XRP	0.095	0.096	LTC-XMR	−6.78	−2.72
37	ETH-BNB	0.080	0.032	ETH-ETC	−2.54	−2.04	ETH-BNB	0.080	0.032	ETH-ETC	−2.54	−2.04
38	ETH-BNB	0.017	0.091	BNB-LTC	−1.98	−2.62	ETH-LINK	0.013	0.063	BCH-ONT	−2.49	−2.10
39	DASH-ONT	0.086	0.020	XRP-DASH	−2.44	−2.00	DASH-ONT	0.079	0.010	LTC-DASH	−2.08	−2.20
40	LTC-BNB	0.093	0.098	DASH-TRX	−2.22	−1.97	LTC-BNB	0.091	0.090	DASH-BNB	−2.18	−2.07
41	ETC-BCH	0.061	0.059	ETC-BCH	−2.83	−2.30	ETC-DASH	0.091	0.035	ETC-DASH	−1.95	−2.52
42	BCH-ETC	0.021	0.010	BCH-ETC	−3.15	−7.94	BCH-ETC	0.021	0.010	BCH-ETC	−3.15	−2.30
43	ETC-EOS	0.010	0.010	ETC-BCH	−8.47	−2.12	ETH-ETC	0.075	0.010	ETC-EOS	−3.81	−2.44
44	ETH-ETC	0.053	0.010	ETH-LINK	−3.32	−2.64	ETH-ETC	0.045	0.010	ETH-ETC	−2.66	−3.13
45	ETC-EOS	0.010	0.027	ETC-EOS	−4.93	−3.90	EOS-ETC	0.021	0.010	EOS-ETC	−4.45	−3.32
46	ETC-EOS	0.010	0.062	ETC-LINK	−6.09	−2.05	ETC-EOS	0.027	0.010	ETC-EOS	−2.46	−3.56
47	ETC-EOS	0.055	0.046	ETC-EOS	−8.45	−4.26	ETC-EOS	0.055	0.046	ETC-EOS	−8.45	−4.26
48	LTC-BNB	0.034	0.052	LTC-ETC	−3.01	−3.96	ETH-LTC	0.089	0.088	LTC-DASH	−3.29	−4.01
49	ETH-LTC	0.020	0.010	ETH-LTC	−2.92	−1.92	ETH-ETC	0.026	0.010	ETH-ETC	−2.43	−2.18
50	ETH-LTC	0.011	0.010	ETH-LTC	−3.72	−4.14	ETH-LTC	0.070	0.010	ETH-LTC	−2.96	−2.75
51	ETH-ETC	0.079	0.010	ETH-ETC	−2.60	−3.94	ETC-LTC	0.013	0.020	ETH-ETC	−2.43	−3.19
52	ETC-LTC	0.098	0.010	LTC-EOS	−4.40	−3.06	LTC-EOS	0.010	0.050	LTC-EOS	−3.39	−2.61

Table 9 Selected pairs using unit-root tests and Kendall's τ coefficient for weeks 53–104

Week	ADF unit-root test (5-min Data)			KSS unit-root test (5-min Data)			ADF unit-root test (hourly data)			KSS unit-root test (hourly data)		
	Pair	*P*-Value (S_1)	*P*-Value (S_2)	Pair	t-stat (S_1)	t-stat (S_2)	Pair	P-Value (S_1)	P-Value (S_2)	Pair	t-stat (S_1)	t-stat (S_2)
53	EOS-ETC	0.021	0.024	LTC-EOS	−1.95	−2.86	EOS-XRP	0.032	0.010	EOS-XRP	−2.43	−3.51
54	BNB-BAT	0.091	0.092	BNB-ETC	−1.97	−2.50	–	–	–	LTC-BNB	−2.03	−2.24
55	XLM-DASH	0.043	0.096	EOS-XLM	−2.04	−2.15	DASH-XLM	0.041	0.010	EOS-LTC	−2.20	−2.01
56	ETH-BNB	0.061	0.021	ETH-BNB	−2.50	−2.01	ETH-BNB	0.049	0.016	ETH-LTC	−2.69	−2.77
57	ETH-LTC	0.010	0.051	ETH-LTC	−2.83	−2.54	ETH-BNB	0.010	0.018	ETH-BNB	−2.83	−3.20
58	ETH-BNB	0.050	0.010	ETH-BNB	−2.59	−2.46	ETH-BNB	0.028	0.023	ETH-BNB	−2.78	−4.00
59	ETH-BNB	0.024	0.036	ETH-BNB	−3.08	−3.15	ETH-BAT	0.024	0.068	ETH-BAT	−3.41	−1.96
60	ETH-BNB	0.027	0.022	ETH-BNB	−2.43	−3.19	ETH-BNB	0.020	0.020	ETH-EOS	−2.41	−1.95
61	ETH-BNB	0.077	0.010	BNB-ADA	−3.55	−2.51	ETH-BNB	0.085	0.010	BNB-ADA	−3.02	−2.04
62	BNB-ONT	0.010	0.048	ETH-BNB	−1.99	−3.10	BNB-ONT	0.010	0.083	ETH-BNB	−1.95	−2.79
63	BNB-LTC	0.040	0.010	ETH-BNB	−2.00	−2.45	BNB-LTC	0.060	0.022	LTC-BNB	−3.47	−2.32
64	LINK-LTC	0.010	0.077	LINK-LTC	−2.57	−2.16	LINK-ONT	0.029	0.095	XRP-LINK	−2.14	−2.34
65	ADA-LTC	0.085	0.018	ETH-ADA	−1.92	−2.31	XRP-XLM	0.010	0.095	XRP-ADA	−3.35	−2.16
66	ETH-BNB	0.010	0.030	ETH-BNB	−3.81	−2.93	ETH-BNB	0.010	0.047	ETH-BNB	−3.15	−3.10
67	ETH-BNB	0.013	0.011	ETH-BNB	−3.73	−2.61	ETH-BNB	0.041	0.010	ETH-BNB	−3.40	−3.23
68	ETH-BNB	0.010	0.010	ETH-BNB	−2.96	−3.35	ETH-BNB	0.017	0.010	ETH-BNB	−3.56	−3.42
69	–	–	–	BNB-BCH	−2.47	−2.60	–	–	–	BCH-XLM	−2.27	−2.03
70	BNB-ETC	0.017	0.034	BNB-LINK	−2.79	−2.05	BNB-ETC	0.031	0.033	LINK-BNB	−2.07	−2.39
71	BNB-LINK	0.092	0.073	BNB-ETC	−2.71	−3.16	LINK-ETC	0.070	0.062	ETC-BNB	−2.29	−2.42
72	BNB-LINK	0.052	0.024	BNB-LINK	−3.43	−2.51	LINK-EOS	0.046	0.025	BNB-LINK	−2.28	−2.37
73	ETC-XRP	0.063	0.024	ETC-XRP	−3.39	−2.02	DASH-ETC	0.037	0.044	EOS-DASH	−1.96	−3.22
74	BNB-DASH	0.088	0.086	ETH-BNB	−2.25	−2.97	ONT-ZEC	0.026	0.010	BNB-DASH	−2.56	−1.95
75	ZEC-XTZ	0.037	0.018	ETH-BNB	−2.67	−2.36	ZEC-BCH	0.047	0.041	ETH-BNB	−2.36	−2.16
76	ETH-ZEC	0.010	0.096	ETH-ZEC	−2.69	−3.14	ETH-XTZ	0.010	0.029	ETH-ZEC	−2.83	−2.52
77	ETH-BNB	0.044	0.047	ETH-BNB	−3.03	−2.88	ETH-BNB	0.057	0.064	ETH-BNB	−2.77	−2.42
78	ETH-BNB	0.026	0.010	ETH-BNB	−3.76	−3.04	ETH-BNB	0.022	0.010	ETH-BNB	−3.17	−2.19
79	BNB-ADA	0.010	0.079	BNB-ADA	−2.82	−2.34	BNB-LTC	0.021	0.037	BNB-LTC	−2.93	−3.50

Table 9 (continued)

Week	ADF unit-root test (5-min Data)			KSS unit-root test (5-min Data)			ADF unit-root test (hourly data)			KSS unit-root test (hourly data)		
	Pair	*P*-Value (S_1)	*P*-Value (S_2)	Pair	t-stat (S_1)	t-stat (S_2)	Pair	P-Value (S_1)	P-Value (S_2)	Pair	t-stat (S_1)	t-stat (S_2)
80	BNB-ADA	0.010	0.010	BNB-ADA	−3.11	−2.52	BNB-LTC	0.017	0.061	BNB-LTC	−3.08	−2.09
81	ETH-LTC	0.012	0.012	LTC-ADA	−3.50	−2.72	ETH-LTC	0.010	0.010	LTC-LINK	−3.25	−2.52
82	ADA-LTC	0.010	0.017	ADA-LTC	−2.79	−3.17	LTC-XRP	0.014	0.017	LTC-XRP	−2.68	−4.02
83	ADA-DASH	0.010	0.052	ADA-DASH	−2.28	−3.07	LTC-DASH	0.010	0.047	LTC-DASH	−4.27	−2.64
84	ETH-ADA	0.029	0.082	ETH-ADA	−2.65	−2.40	ETH-LTC	0.044	0.010	ETH-LTC	−2.40	−3.95
85	ETH-LINK	0.087	0.047	ETH-LINK	−1.95	−1.95	ETH-IOTA	0.090	0.010	IOTA-DASH	−2.63	−2.87
86	BAT-BNB	0.010	0.010	DASH-BAT	−2.09	−4.25	BAT-IOTA	0.010	0.010	BAT-DASH	−3.54	−1.96
87	BNB-DASH	0.055	0.085	ETC-BAT	−2.81	−2.61	BNB-ETC	0.023	0.068	BAT-ETC	−2.49	−2.85
88	BNB-BAT	0.052	0.080	DASH-ETC	−2.18	−2.37	BNB-TRX	0.020	0.084	DASH-BAT	−2.35	−2.43
89	ETH-BNB	0.045	0.089	ETH-DASH	−2.58	−3.13	ETH-DASH	0.090	0.010	ETH-DASH	−2.71	−4.19
90	ETH-LTC	0.010	0.020	ETH-LTC	−2.58	−2.87	ETH-DASH	0.010	0.014	ETH-DASH	−2.95	−2.78
91	ETH-BAT	0.036	0.057	ETH-LTC	−3.28	−2.36	ETH-BAT	0.095	0.032	ETH-DASH	−2.54	−2.26
92	ETH-LTC	0.010	0.040	ETH-LTC	−5.78	−3.76	ETH-BAT	0.010	0.084	ETH-BAT	−3.08	−2.15
93	LTC-BNB	0.034	0.010	LTC-BNB	−3.17	−3.15	BNB-LTC	0.014	0.023	BNB-LTC	−2.22	−3.47
94	BCH-BAT	0.010	0.067	BCH-BAT	−7.20	−3.95	BCH-DASH	0.010	0.046	BCH-DASH	−3.36	−2.30
95	ETH-BCH	0.043	0.010	BCH-ONT	−6.28	−4.47	ETH-BCH	0.051	0.034	BCH-DASH	−6.40	−2.29
96	ETH-ADA	0.047	0.010	ETH-ADA	−2.40	−3.82	ETH-ADA	0.081	0.010	ETH-ADA	−2.44	−2.53
97	ETH-ADA	0.046	0.033	ETH-ADA	−2.38	−3.49	ETH-BAT	0.042	0.053	ETH-ADA	−2.47	−2.47
98	ADA-ETC	0.016	0.022	ETH-ADA	−2.35	−2.79	ADA-BAT	0.039	0.010	ETH-ADA	−1.97	−3.08
99	ADA-TRX	0.085	0.010	ETH-ETC	−2.30	−3.22	XTZ-BAT	0.010	0.010	ETH-ATOM	−1.98	−2.15
100	ETH-LTC	0.015	0.074	ETH-BCH	−2.67	−2.22	ETH-BCH	0.010	0.099	ETH-ONT	−2.20	−3.13
101	ETH-LINK	0.073	0.091	LINK-ATOM	−2.57	−1.95	—	-	-	LINK-BCH	-2.466	−2.002
102	ETH-BNB	0.044	0.041	ETH-LTC	−2.96	−2.16	ETH-BNB	0.057	0.033	ETH-LTC	−2.48	−2.00
103	BNB-LINK	0.028	0.021	BNB-LINK	−3.53	−2.43	BNB-LINK	0.010	0.020	BNB-LINK	−3.85	−3.27
104	LINK-BCH	0.069	0.010	LINK-ADA	−2.79	−2.50	LINK-DASH	0.061	0.100	LINK-ONT	−2.49	−2.31

Table 10 Bivariate Archimedean copulas distributions

Name	Bivariate copula distribution $C(u_1, u_2)$ $(\overline{u_i} = 1 - u_i)$	Generator $\phi(t)$	Parameters
Clayton	$\left[\max\left(u_1^{-\theta} + u_2^{-\theta} - 1, 0\right)\right]^{-1/\theta}$	$\frac{1}{\theta}\left(t^{-\theta} - 1\right)$	$\theta > 0$
Gumbel	$\exp\left[-\left[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\right]^{1/\theta}\right]$	$(-\ln(t))^{\theta}$	$\theta \geq 1$
Frank	$-\theta^{-1}\ln\left[1 + \left(e^{-\theta} - 1\right)^{-1}\left(e^{-\theta u_1} - 1\right)\left(e^{-\theta u_2} - 1\right)\right]$	$-\ln\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right)$	$\theta \in \mathbb{R} \setminus \{0\}$
Joe	$1 - \left[\left(1 - u_1\right)^{\theta} + \left(1 - u_2\right)^{\theta} - \left(1 - u_1\right)^{\theta}\left(1 - u_2\right)^{\theta}\right]^{1/\theta}$	$-\ln\left(1 - (1 - t)^{\theta}\right)$	$\theta \geq 1$
BB1	$\left[1 + \left[\left(u_1^{-\theta} - 1\right)^{\delta} + \left(u_2^{-\theta} - 1\right)^{\delta}\right]^{1/\delta}\right]^{-1/\theta}$	$\left(t^{-\theta} - 1\right)^{\delta}$	$\theta > 0,\ \delta \geq 1$
BB6	$1 - \left[1 - \exp\left(-\left[\left[-\ln\left(1 - \overline{u_1}^{\theta}\right)\right]^{\delta} + \left[-\ln\left(1 - \overline{u_1}^{\theta}\right)\right]^{\delta}\right]^{1/\delta}\right)\right]^{1/\theta}$ $1 - \left[1 - \exp\left(-\left[\left[-\ln\left(1 - \overline{u_1}^{\theta}\right)\right]^{\delta} + \left[-\ln\left(1 - \overline{u_1}^{\theta}\right)\right]^{\delta}\right]^{1/\delta}\right)\right]^{1/\theta}$	$\left(\ln\left(1 - (1 - t)^{\theta}\right)\right)^{\delta}$	$\theta \geq 1,\ \delta \geq 1$
BB7	$1 - \left[1 - \left(\left(1 - \overline{u_1}^{\theta}\right)^{-\delta} + \left(1 - \overline{u_2}^{\theta}\right)^{-\delta} - 1\right)^{-1/\delta}\right]^{1/\theta}$	$\left(1 - (1 - t)^{\theta}\right)^{-\delta} - 1$	$\theta \geq 1,\ \delta > 0$
BB8	$\delta^{-1}\left[1 - \left(1 - \left[1 - (1 - \delta)^{\theta}\right]^{-1}\left[1 - (1 - \delta u_1)^{\theta}\right]\left[1 - (1 - \delta u_2)^{\theta}\right]\right)^{1/\theta}\right]$	$-\ln\left(\frac{1 - (1 - \delta t)^{\theta}}{1 - (1 - \delta)^{\theta}}\right)$	$\theta \geq 1,\ 0 < \delta \leq 1$

$$
\begin{aligned}
f_X(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \mathrm{e}^{-\frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]} \\
\boldsymbol{\mu} &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},
\end{aligned} \tag{34}
$$

where ρ is the correlation between random variables X_1 and X_2 and between $\sigma_1 > 0$ and $\sigma_2 > 0$. If $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$, then the density and distribution functions of the standard bivariate Gaussian distribution are obtained by

$$
\begin{aligned}
\phi_{X_1X_2}(x_1, x_2; \rho) &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}} \\
\Phi_{X_1X_2}(u_1, u_2; \rho) &= \int_{-\infty}^{u_1}\int_{-\infty}^{u_2} \phi_{X_1X_2}(x_1, x_2; \rho)\, dx_1\, dx_2
\end{aligned} \tag{35}
$$

Using Sklar's theorem, as shown in 13 and 35, the bivariate Gaussian copula is defined by

$$
C(u_1, u_2; \rho) := \Phi_{X_1X_2}(\Phi_{X_1}^{-1}(u_1), \Phi_{X_2}^{-1}(u_2); \rho) \tag{36}
$$

In the bivariate t distribution, $g(.)$ and k_n function in the formula 17 are defined by $g(x) := (1 + x/\nu)^{-(\nu+2)/2}$, $k_n := \Gamma(\frac{\nu+2}{2})/\left(\Gamma(\frac{\nu}{2})\nu\pi\right)$, and the probability density function of $\boldsymbol{T} = (T_1, T_2)$ is obtained by

$$
f_{\boldsymbol{T}}(\boldsymbol{t}; \nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\nu\pi|\boldsymbol{\Sigma}|^{1/2}} \left[1 + \frac{1}{\nu}(\boldsymbol{t} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{t} - \boldsymbol{\mu})\right]^{-(\nu+2)/2} \tag{37}
$$

where $\nu > 0$ is the degree of freedom parameter and $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the same as 34. If $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$, and knowing that $\Gamma\left(\frac{\nu+2}{2}\right)/\Gamma\left(\frac{\nu}{2}\right) = \frac{\nu}{2}$, the density and distribution function of the standard bivariate Student-t distribution are obtained by

$$\begin{aligned} f_{T_1T_2}(t_1, t_2; \nu, \rho) &= \frac{(1-\rho^2)^{-1/2}}{2\pi}\left[1 + \frac{t_1^2 - 2\rho t_1 t_2 + t_2^2}{\nu(1-\rho^2)}\right]^{-(\nu+2)/2} \\ F_{T_1T_2}(u_1, u_2; \nu, \rho) &= \int_{-\infty}^{u_1}\int_{-\infty}^{u_2} f_{T_1T_2}(t_1, t_2; \nu, \rho)\, dt_1\, dt_2 \end{aligned} \tag{38}$$

Then the bivariate Student-t copula is defined by

$$C(u_1, u_2; \nu, \rho) := F_{T_1T_2}(F_{T_1}^{-1}(u_1), F_{T_2}^{-1}(u_2); \nu, \rho) \tag{39}$$

Abbreviations

ADF	Augmented Dickey–Fuller
AIC	Akaike information criterion
API	Application programming interface
BLVT	Binance leveraged tokens
BUSD	Binance USD
CMI	Cumulative mispricing indices
CML	Canonical maximum likelihood
EG	Engle-Granger
IFM	Inference for the margins
MLE	Maximum likelihood estimation
RoMaD	Return over maximum drawdown
SETAR	Self-exciting threshold auto-regressive
SSD	Sum of squared distance
STAR	Smooth transition auto-regressive
TAR	Threshold auto-regressive

USDT TetherSupplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40854-024-00702-7.

Supplementary file 1.

Supplementary file 2.

Acknowledgements
The authors gratefully acknowledge the financial support provided by the grant GAČR 22-19617 S "Modeling the structure and dynamics of energy, commodity, and alternative asset prices."

Author contributions
MT conceptualized the study, designed the methodology, and carried out the data gathering and coding. Additionally, MT conducted the analysis and interpretation of the findings. JW also contributed significantly to the manuscript, participating in the conceptualization and analysis stages, and providing valuable input during the review and editing process. Finally, all authors diligently read and approved the final manuscript, ensuring its accuracy and quality.

Funding
This research was conducted with the financial support of the grant GAČR 22-19617 S "Modeling the structure and dynamics of energy, commodity, and alternative asset prices."

Availability of data and materials
The dataset used and analyzed in the present study was acquired through the Binance API and is available as electronic supplementary material.

Declarations

Competing interests
The authors declare that they have no Conflict of interest.

Received: 9 August 2023 Accepted: 11 November 2024
Published online: 13 January 2025

References

Al-Yahyaee KH et al (2020) Why cryptocurrency markets are inefficient: the impact of liquidity and volatility. North Am J Econ Finance 52:101168
Andrew B (2020) Deep reinforcement learning pairs trading with a double deep Q-network. In: 2020 10th Annual computing and communication workshop and conference (CCWC). IEEE, pp. 0222–0227
Bağcı M, Soylu Pınar K, Kıran S (2024) The symmetric and asymmetric algorithmic trading strategies for the stablecoins. Comput Econ 1–22
Bertram WK (2010) Analytic solutions for optimal statistical arbitrage trading. Phys A 389(11):2234–2243
Binance Crypto Derivatives (Nov. 3, 2022). https://www.binance.com/en/support/faq/crypto-derivatives?c=4&navId=4
Bogomolov T (2013) Pairs trading based on statistical variability of the spread process. Quant Finance 13(9):1411–1430
Borri N, Shakhnov K (2022) The cross-section of cryptocurrency returns. Rev Ass Pricing Stud 12(3):667–705
Chang V et al (2021) Pairs trading on different portfolios based on machine learning. Expert Syst 38(3):e12649
Chen H et al (2019) Empirical investigation of an equity pairs trading strategy. Manage Sci 65(1):370–389
Cherubini U et al (2011) Dynamic copula methods in finance. Wiley, Hoboken
Claudia Czado (2019) Analyzing dependent data with vine copulas. Springer, Berlin
Clegg M, Krauss C (2018) Pairs trading with partial cointegration. Quant Finance 18(1):121–138
Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit- root. J Am Stat Assoc 74(366a):427–431
Do B, Faff R (2010) Does simple pairs trading still work? Financ Anal J 66(4):83–95
Do B, Faff R (2012) Are pairs trading profits robust to trading costs? J Financ Res 35(2):261–287
Elliott RJ, Van Der Hoek J, Malcolm WP (2005) Pairs trading. Quant Finance 5(3):271–276
Enders W, Siklos PL (2001) Cointegration and threshold adjustment. J Bus Econ Stat 19(2):166–176
Engle RF, Granger CWJ (1987) Co-integration and error correction: representation, estimation, and testing. Econom J Econom Soc 55(2):251–276
Fang F et al (2022) Cryptocurrency trading: a comprehensive survey. Financ Innov 8(1):13
Ferreira L (2008) New tools for spread trading. Futures 37(12):38–41
Fil M, Kristoufek L (2020) Pairs trading in cryptocurrency markets. IEEE Access 8:172644–172651
Gatev E, Goetzmann WN, Geert Rouwenhorst K (2006) Pairs trading: performance of a relative-value arbitrage rule. Rev Financ Stud 19(3):797–827
George C, Berger RL (2021) Statistical inference. Cengage Learning, Patparganj
George Kapetanios, Yongcheol Shin, Andy Snell (2006) Testing for cointegration in nonlinear smooth transition error correction models. Econom Theory 22(2):279–303
Gordon Gudendorf, Johan Segers (2010) Extreme-value copulas. Copula theory and its applications. Springer, Berlin, pp 127–145
Haddad K, GholamReza, Talebi H (2023) The profitability of pair trading strategy in stock markets: evidence from Toronto stock exchange. Int J Finance Econ 28(1):193–207
Hansen BE, Seo B (2002) Testing for two-regime threshold cointegration in vector error-correction models. J Econom 110(2):293–318
Harris L (1997) Decimalization: a review of the arguments and evidence". Unpublished working paper, University of Southern California
Huck N (2015) Pairs trading: does volatility timing matter? Appl Econ 47(57):6239–6256
Jurek JW, Yang H(2007) Dynamic portfolio selection in arbitrage. EFA 2006 Meetings Paper
Kakushadze Z, Willie Yu (2019) Altcoin-Bitcoin Arbitrage. Bull Appl Econ 6(1):87–110
Kapetanios G (2005) Unit-root testing against the alternative hypothesis of up to m structural breaks. J Time Ser Anal 26(1):123–133
Kapetanios G, Shin Y, Snell A (2003) Testing for a unit root in the nonlinear STAR framework. J Econom 112(2):359–379
Klement EP, Mesiar R, Pap E (2002) Invariant copulas. Kybernetika 38(3):275–286
Krauss C (2017) Statistical arbitrage pairs trading strategies: review and outlook. J Econ Surv 31(2):513–545
Krauss C, Stübinger J (2017) Non-linear dependence modelling with bivariate copulas: statistical arbitrage pairs trading on the S &P 100. Appl Econ 49(52):5352–5369
Leung T, Nguyen H (2019) Constructing cointegrated cryptocurrency portfolios for statistical arbitrage. Stud Econ Financ 36(3):581–599
Liew RQ, Yuan W (2013) Pairs trading: a copula approach. J Deriv Hedge Funds 19(1):12–30
Lintilhac PS, Tourin A (2017) Model-based pairs trading in the bitcoin markets. Quant Finance 17(5):703–716
Masood Tadi, Irina Kortchemski (2021) Evaluation of dynamic cointegration-based pairs trading strategy in the cryptocurrency market. Stud Econ Finance 38(5):1054–1075
Mudchanatongsuk S, Primbs JA, Wong W (2008) Optimal pairs trading: a stochastic control approach. In: 2008 American control conference. IEEE, pp. 1035–1039
Nelsen RB (2007) An introduction to copulas. Springer Science & Business Media, Berlin
Perlin MS (2009) Evaluation of pairs-trading strategy at the Brazilian financial market. J Deriv Hedge Funds 15:122–136
Phillips PCB, Ouliaris S (1990) Asymptotic properties of residual based tests for cointegration. Econom J Econom Soc 58(1):165–193
Pritchard BPA (2018) "Digital asset arbitrage". PhD thesis. Fundação Getulio Vargas's Sao Paulo school of business administration

Rad H, Low RKY, Faff R (2016) The profitability of pairs trading strategies: distance, cointegration and copula methods. Quant Finance 16(10):1541–1558
Sarmento SM, Horta N (2020) Enhancing a pairs trading strategy with the application of machine learning. Expert Syst Appl 158:113490
Silva FABS, Ziegelmann FA, Caldeira JF (2023) A pairs trading strategy based on mixed copulas. Q Rev Econ Finance 87:16–34
Søren Johansen (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econom J Econom Soc. https://doi.org/10.2307/2938278
Stander Y, Marais D, Botha I (2013) Trading strategies with copulas. J Econ Financ Sci 6(1):83–107
Teräsvirta T (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. J Am Stat Assoc 89(425):208–218
Tourin A, Yan R (2013) Dynamic pairs trading using the stochastic control approach. J Econ Dyn Control 37(10):1972–1981
van den Broek L, Sharif Zara (2018) "Cointegration-based pairs trading framework with application to the Cryptocurrency market". Bachelor Thesis, Erasmus University Rotterdam
Vidyamurthy G (2004) Pairs trading: quantitative methods and analysis, vol 217. Wiley, Hoboken
Xie W, Liew RQ et al (2016) Pairs trading with copulas. J Trading 11(3):41–52
Xie W, Wu Y (2013) Copula-based pairs trading strategy. Asian Finance Association (AsFA)
Zivot E, Andrews DWK (2002) Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. J Bus Econ Stat 20(1):25–44

Publisher's Note