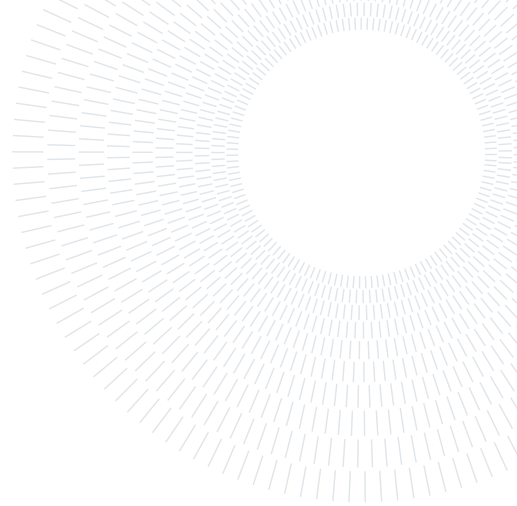




POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE



Statistical Arbitrage in Forex: A Study on Pairs Trading and Mean-Reverting Rates

TESI DI LAUREA MAGISTRALE IN
MATHEMATICAL ENGINEERING - INGEGNERIA MATEMATICA

Luca Piccinini, Student 10668909

Advisor:
Prof. Daniele Marazzina

Co-advisor:
Raffaele Zenti

Academic year:
2023-2024

Abstract: Following the collapse of the Bretton Woods system in 1971, currency values began to fluctuate freely, driven by supply and demand dynamics. In recent years, the forex market has emerged as a fertile ground for algorithmic and machine learning-based trading, also due to its high liquidity. Additionally, a growing body of literature highlights how certain exchange rates exhibit similar movements and, individually, tend to follow mean-reverting dynamics.

This study explores the exploitation of temporary disruptions in the comovements of exchange rate pairs, as well as the oscillatory behavior of individual rates, through the application of machine learning and statistical techniques. A key aspect of this approach is the careful selection of both exchange rate pairs and individual rates, which is achieved through innovative methodologies tailored to the different trading strategies.

Key-words: forex market; statistical arbitrage; pairs trading; mean reversion; algorithmic trading; copula; logistic regression; spread; Markov switching model

1. Introduction

The modern foreign exchange (forex) market, as we know it today, emerged after the collapse of the Bretton Woods system in 1971, transitioning the world from fixed exchange rates to a regime of freely floating currencies. This marked a critical milestone, as currency values began to fluctuate based on supply and demand dynamics. The introduction of electronic trading platforms between the end of the last century and the early 2000s democratized the access to forex market enabling retail participants to trade currencies directly from their computers.

In recent years, key innovations linked to computational developments have revolutionized trading techniques and this technological revolution has involved also forex markets. Algorithmic trading, which uses algorithms to automate trade execution based on predefined criteria, has become a dominant force. In this framework and thanks to the development of several sophisticated and automated trading techniques, the forex market has reached a staggering daily volume of \$7.5 trillion in April 2022 according to the Bank for International Settlements [53]. Today's trading techniques are increasingly shaped by machine learning, which can analyze vast datasets to identify patterns and generate signals in an automated way, and make extensive use of statistical and numerical procedures: among the other advantages, the employed algorithms are able to preclude the effects of human feelings [59]. The application of such automated strategies and methods quickly extended to the FX market, which has become a fertile ground for implementing algorithm-based trading strategies and machine learning techniques. However, this market presents distinct challenges, difficulties and opportunities compared to traditional equity markets.

The primary advantages of the FX market compared to others lie in its exceptional liquidity, particularly concerning major currency pairs, as previously highlighted. For instance, the forex spot market, which is the focus of this study, recorded a daily trading volume of \$2.1 trillion in 2022, always according to [53]. Moreover, its ease of access and continuous trading hours make it particularly convenient and efficient for retail investors and non-institutional traders to take positions.

Two additional characteristics of this market segment have significantly influenced the decision of this study to select the FX market as a proving ground for implementing and backtesting the proposed trading strategies. First, the high volatility of exchange rates and of their spreads and ratios provides opportunities for substantial gains, aligning well with the nature of the strategies presented. Second, short selling, a key kind of trade in this study, is remarkably straightforward in the FX market and is not subject to the same restrictions as in other markets, such as equity or futures. In this market, short selling is perfectly symmetrical to long positions, as it merely involves reversing the currency being sold and the currency being purchased. On the other hand, the high volatility in this field can be a significative challenge leading to massive drawdowns if the strategies and the investment horizon are not calibrated properly.

Despite the unique characteristics of the foreign exchange market, the followed approach is partially derived from the more extensively studied and well-established equity market. Several techniques are borrowed from equity trading and adapted to the specific features of this segment. Among the various trading methodologies proposed, this study places significant emphasis on the concept of pairs trading, a particularly profitable and easy to implement statistical arbitrage strategy. This approach takes advantage of the mispricing between two co-moving assets, by taking a long position in one and shorting the other, betting that the relationship will hold and that prices will converge back to an equilibrium level.

The rationale behind this strategy lies in the observation that equity stocks often move in tandem over extended periods due to shared exposure to common economic factors, similar responses to systematic risks, and the proximity of their underlying businesses (an aspect that can substantially enhance the profitability of the strategy, as noted by [22]). Similarly, currency exchange rates that exhibit co-movement over extended periods, driven by analogous responses to key monetary policies such as inflation measures [45] or by similar supply and demand dynamics, may experience temporary deviations from this relationship, creating profit opportunities for investors.

This study seeks to bridge the gap between the FX market and traditional equity trading strategies, with a particular emphasis on pairs trading and the integration of machine learning and statistical tools. These tools play a central role in both the automated selection of currency pairs to be traded and the generation of trading signals.

Section 2 presents in more details the innovations and contributions introduced by this study in the field of FX trading; *Section 3* briefly reviews the existing literature on forex and pairs trading, as well as the literature backing the theoretical foundations of the mathematical tools employed throughout this work; *Section 4* presents the problem formulation, the datasets utilized and the assumptions regarding profit realization; *Section 5* and *6* show the details of the methods employed to select the pairs of exchange rates and how they are traded; *Section 7* presents the results of applying the proposed methodologies to the dataset while in *Section 8* potential

improvements are discussed; finally, *Section 9* concludes.

2. Scope of the Study

As previously mentioned, the primary objective of this study is to extend existing methodologies in FX trading on long time horizons by drawing inspiration from well-established equity trading strategies. Specifically, it aims to demonstrate the feasibility of applying pairs trading techniques, an approach that capitalizes on the reversion of two co-moving assets to their long-run equilibrium following temporary deviations, and other statistical arbitrage methodologies in the context of the FX market. This approach has only recently drawn attention in this domain [76]; still, the existing applications of pairs trading in the FX framework are scarce. Furthermore, it aims to ensure that tests and results are as realistic as possible. To this end, the findings are presented both with and without transaction costs associated with the bid-ask spread, leading to more reliable and applicable conclusions under actual market conditions.

The key contributions of this study lie in two areas: the selection of currency pairs to be traded and the development of the actual trading mechanism. These contributions add to the previously mentioned innovation of using the FX market as a testing ground. It is worth emphasizing that both phases of the strategy are driven by automated and algorithmic processes, which eliminate the influence of human emotions and minimize biases associated with discretionary trading. These processes rely entirely on statistical methods and machine learning tools to ensure objectivity and consistency. Furthermore, the pairs are selected and traded in some cases analyzing their relative behavior, measured through their spread, and in others observing the individual properties and joint movements. The first scenario is more intuitive and understandable, the second one is more complex and sees the problem from a probabilistic point of view but keeps track of all the information regarding the rates.

In addition, the application of statistical arbitrage-based and machine learning techniques to the trading of individual foreign exchange rates are explored. While this deviates from the pairs trading framework, it remains grounded in the exploitation of statistically extreme values.

The following remaining part of this Section will provide a detailed discussion of the contributions and innovations introduced by this study and will serve as an initial overview of the methodologies employed in the research and deepened in the next Sections.

The first set of innovations introduced by this study regards the methods used for pair selection. In this context, it is crucial to identify pairs of exchange rates that exhibit similarities in their dynamics or frequent relative oscillations, enabling the exploitation of relative deviations from their shared behavior to generate profits betting on the return to it. A significant advancement over the approach presented in [76] lies in the consideration of exchange rate pairs that do not necessarily share a common currency. Specifically, some of the criteria presented in this research look at pairs such as *EURUSD - GBPCHF* as potential candidates, rather than restricting the analysis to pairs of the form $xUSD - yUSD$. Furthermore, in this study the pairs of exchange rates are considered eligible not solely based on strict econometric constraints: other selection criteria taking into account the similarity in the reaction to external variables as well as the historical potential for profitable trades, measured merely as the frequency of simultaneous extreme values, are considered.

This broader selection approach is motivated by several factors. From a statistical perspective, the primary focus of this study is not on the underlying causes of the co-movements between the assets under consideration but rather on the statistical properties of their relative behaviors and the profit opportunities arising from short-term deviations. From an economic point of view, on the other side, these rates may still respond similarly to shared risks, geopolitical developments, and monetary policies. For instance, currencies such as the Canadian Dollar (*CAD*) and the Australian Dollar (*AUD*) are well-known to be highly influenced by global commodity prices, including oil and minerals, particularly over a daily horizon, as shown by [41]. Consequently, such currencies may exhibit a shared appreciation or depreciation relative to other currencies, making pairs like $xCAD - yAUD$ (with $x \neq y$) eligible for selection. However, this approach introduces additional complexities in the interpretation of the results, as it expands the scope to include profits realized across multiple currencies.

Another key innovation in the selection process is the decision to move away from cointegration as the only criterion for determining pair eligibility. While cointegration remains a valuable tool, the decision to relax its application and not rely exclusively on it for pair selection has been made. This decision stems from the observation that a strict cointegration-based approach can become too restrictive, potentially discarding promising pairs whose spreads exhibit promising behavior.

Delving deeper into the selection methods, the most straightforward approach is firstly presented: the analysis of the spread, defined as the ratio between two exchange rates. In this context, the main innovation introduced by this study lies in the application of this methodology to the FX market and the corresponding trading strategies derived from it. The underlying mechanism of this selection criterion involves identifying pairs of exchange rates whose ratio exhibits sufficiently stationary and volatile behavior during the training (or formation) period. This

ensures both a tendency to revert to the mean, facilitating trade closure, and the presence of sufficiently extreme deviations to generate robust profitability. In this framework, some considerations regarding the need for the exchange rates of the pair to share a common currency must be done. A closely related methodology has also been employed to select individual exchange rates for a slightly different trading approach, which will be introduced and further examined in later sections.

Another more sophisticated and structured approach, though already explored in various forms within the domain of equity pairs trading, is the use of copulas. In this method, copulas are fitted to each pair of exchange rates during the initial segment of the formation period, after which a strategy based on the detection of extreme values is tested in the latter part of the selection period. The pairs that demonstrate the highest frequency of successful trade opportunities are then selected for execution during the actual trading phase. Despite the extensive application of copula-based methods in equity pairs trading, where they appear theoretically well-suited and one of the strongest tools, no prior application of copulas specifically to pairs trading within the FX market has been found. Furthermore, existing related literature tends to focus primarily on uncovering the underlying relationships between exchange rates rather than leveraging copulas as a tool for structuring trading strategies.

The final and most original contribution of this study in the domain of pairs selection is the innovative application of *Generalized Additive Models* (*GAMs*). Unlike their conventional use in forecasting, *GAMs* are employed to identify similarities between different exchange rates. While non-additive models, such as boosted trees, often outperform *GAMs* in predictive accuracy, *GAMs* offer exceptional interpretability, as highlighted in [62]. Given their well-documented transparency in mapping the relationships between exogenous variables and output responses, in this research *GAMs* are used to analyze the dependencies between each input variable and exchange rate movements. The exchange rate pairs exhibiting the highest degree of similarity in both the shape and behavior of these functional relationships are selected for trading. This approach ensures that the identified pairs are intrinsically linked by their shared response to common external dynamics, increasing the likelihood that they will continue to exhibit similar behavior in reaction to these forces. Notably, the exogenous data used in this analysis is minimal and easily accessible, consisting solely of treasury bond yields of the countries issuing the considered currencies across different maturities. This represents a promising advancement from the vast datasets typically employed in machine learning-based forecasting models, such as those in [28]. However, since the primary objective is not to enhance predictive accuracy but rather to examine how consistently different exchange rates react to relevant external factors, this restricted data selection aligns well with the goals of the study.

Finally, it is important to highlight the last selection method employed in this study in the pair rates framework, which integrates both the copula and *GAM* selection criteria. In this hybrid approach, the two methodologies work in conjunction to provide more robust and reliable scores for the exchange rate pairs obtained as the average between the scores of the criteria taken alone. These scores ultimately determine whether a given pair will be selected for trading, enhancing the overall robustness and effectiveness of the selection process.

The second set of innovations regards the trading strategies implemented in the FX market. The first strategy is relatively straightforward, relying solely on the statistical behavior of the spread and following strictly from the spread-based selection criterion. Specifically, when the spread reaches a statistically high level, a short position is taken on it, whereas when it is low, a long position is initiated. This approach capitalizes on mean reversion, assuming that deviations from the typical spread behavior will eventually correct over time.

The main innovations lie in the other more advanced trading techniques. In particular, a strategy based on copulas has been implemented. For each pair of exchange rates selected using one of the previously discussed methods, a copula is fitted during the training/formation period. During the trading phase, each new pair of exchange rate values is then mapped through the fitted copula, generating a corresponding pair of probabilities that indicate the extremeness of the observed values. This probabilistic framework provides a structured approach to identifying trading opportunities based on deviations from the established dependence structure between the two exchange rates.

In addition, a *Markov Switching Model* (*MSM*), which offers significant potential for enhancing trading strategies, is proposed as an innovation. The strength of this model lies in its ability to identify distinct behavioral regimes of the spread using data from the training period and to estimate the probability of each new observation belonging to a specific regime in the trading phase. In this study, two regimes are considered: one associated with high spread values and the other with low ones. The trading rules for opening and closing positions are then adapted based on the most probable prevailing regime.

While the application of *Markov Switching Models* has been primarily confined to equity trading, their capacity to detect shifts in the relationship between exchange rates makes them particularly promising in the FX market. This is especially relevant given the proved occurrence of structural breaks and multiple shifting regimes [82], which can disrupt the equilibrium identified during the training period and make the implemented trading

strategies ineffective. The challenge of identifying structural breaks in FX time series has been extensively studied, with [3] providing significant contributions toward addressing this issue.

The use of the Markov Switching Model is particularly justified by events such as the unpegging of the Swiss Franc from the Euro in January 2015, as it will be deepened in *Subsection 6.2* together with all the details regarding this trading strategy.

Finally, on the periphery of the study a logistic regression-based approach has been proposed for trading individual exchange rates, following the methodology presented in [58], where a similar technique was employed to generate trade signals on the S&P 500. The primary innovation here lies in the application of this algorithm to the FX market. While logistic regression has been previously utilized in FX trading [54], this study represents the first instance in which it is applied to such an extensive range of exchange rates and without incorporating any external data. Unlike other machine learning approaches [10], where several exogenous variables are used as inputs, this methodology relies exclusively on features derived directly from the historical time series of exchange rates. The objective of this method is to investigate whether profitability can be achieved by exploiting the less evident properties of stationarity and mean reversion of individual exchange rates, rather than focusing solely on pair-based strategies, exploiting anomalous and individual asset behaviors and by exploiting a machine learning approach independent from any external data. Trading individual rates rather than pairs amplifies the significance of each trade because as losses can no longer be offset by gains in the other component of the pair.

In the following sections, in particular where the results will be presented, it will become clear how the selection methods and trading strategies have been combined. For a more synthetic summary, Figure 11 can be helpful.

3. Literature Review

The existing literature on pairs trading within the foreign exchange market is relatively limited and primarily focuses on the use of cointegration as a detection tool. This approach is exemplified in the work of Goudvis [25], where the author examines a broad basket of currency exchange rates testing for cointegration and develops a trading algorithm that generates signals when deviations occur between cointegrated rates, achieving an average annual return of up to 6.5%.

A similar methodology is adopted in the recent study by Lemishko, Landi, and Caicedo-Llano [76]. The authors employ a purely cointegration-based approach to identify currency pairs exhibiting long-term equilibrium through the Engle-Granger test [66]. They then design a trading strategy that capitalizes on deviations from this equilibrium, yielding notable results in terms of profitability and return stability.

This section is structured into two parts: the first reviews the existing literature on forex trading, exploring analytical techniques beyond pairs trading and statistical arbitrage on individual exchange rates. The second subsection summarizes key contributions in pairs trading, mainly in equity and futures landscape, presenting the primary detection methods and trading strategies introduced in this domain.

3.1. FX Trading

Over the years, research on exchange rate dynamics and forecast has transitioned from traditional econometric and statistical approaches to modern machine learning techniques. While econometric and statistical models have primarily been explored from a theoretical and modeling perspective, their findings remain crucial for trading applications.

A landmark study in this field, despite its discouraging conclusions, is the work of Meese and Rogoff (1983) [65], which demonstrated that a random-walk model outperformed both structural and time-series models in predicting the exchange rates of major currencies during the 1970s. Their findings underscored the difficulty of forecasting exchange rate movements, thereby complicating trading strategies. Further modeling efforts were conducted by Boothe and Glassman (1987) [60], who showed that logarithmic daily returns of spot rates could be effectively modeled using t-Student distributions and mixtures of Gaussian distributions with time-varying parameters.

Interest in the mean-reverting properties of exchange rates gained traction in the 1990s. Jorion and Sweeney (1996) [61] found that a mean-stationary model outperformed a random walk, particularly over long horizons, and provided evidence of mean reversion in real exchange rates over the 1973–1993 period. Campa and Wolf (1997) [38] further examined whether mean reversion in exchange rates was driven by arbitrage; however, they found no conclusive evidence supporting this hypothesis.

Expanding on this research, Taylor and Peel (2000) [52] explored nonlinear time-series models to describe deviations of the dollar–sterling and dollar–mark exchange rates from levels predicted by simple monetary fundamentals. Their findings suggested nonlinear mean reversion, where the speed of reversion increased with the magnitude of deviation from equilibrium. Such properties can be leveraged in arbitrage and pairs trading

strategies that capitalize on expected reversion to the mean.

In more recent years, Evans and Lyons [51] demonstrated that order flow data, influenced by macroeconomic news, significantly impacted short-term exchange rate movements, explaining approximately 30% of price variance and offering valuable insights for order flow-based trading strategies. Additionally, Menkhoff et al. (2012) [44] examined the relationship between global foreign exchange volatility risk and the cross-section of excess returns derived from carry trade strategies, further enriching the understanding of forex market dynamics.

From a machine learning perspective, advancements in forex trading strategies have emerged more recently, driven by the increasing availability of data and computational power. These developments enable the identification of complex, nonlinear relationships within financial data, enhancing predictive capabilities and trading performance.

One of the earliest applications of machine learning in this domain involved Recurrent Neural Networks, as demonstrated by Kuperin et al. [80] and later by Kondratenko and Kuperin [78] in 2001 and 2003, respectively. Their studies showed that neural networks could achieve a forecasting accuracy sufficient to construct an automated and profitable trading strategy for the *GBPUSD* exchange rate. In 2005, Hryshko and Downs [2] introduced a genetic algorithm utilizing simple technical indicators to generate trading signals, marking another step toward algorithmic trading based on machine learning tools.

Nielsen (2018) [56] explored a broad spectrum of machine learning techniques, ranging from Ordinary Least Squares regression to Long Short-Term Memory (LSTM) neural networks, and applied them to high-frequency exchange rate data. His findings revealed that the performance of these models varied depending on the sampling frequency and model complexity. Notably, more complex models tended to underperform at higher frequencies due to overfitting, as they learned from noise rather than meaningful patterns. Although the study did not explicitly develop a trading algorithm, the quality of the forecasts suggests its potential applicability in algorithmic trading strategies.

Another significant approach was proposed by Nguyen and Xuan (2018) [77], who reframed forex trading as a supervised classification problem. They generated trading signals by extracting endogenous features from exchange rate time series and applying Support Vector Machines (SVMs) to classify and predict the best action for any given market condition. Similarly, SVMs, along with other classification algorithms such as Logistic Regression and Random Forest, were employed by Mabrouk et al. [54] to identify profitable trading opportunities.

In the space of reinforcement learning, Reho (2019) [69] achieved positive returns in trading the *EURUSD* exchange rate by leveraging reinforcement learning algorithms, further highlighting the potential of adaptive machine learning approaches in forex markets.

For a comprehensive overview of machine learning applications in forex trading from 2010 to 2021, the review by M. Junior et al. [47] provides an extensive collection of key studies in the field.

More recently, Lin and Liu [11] synthesized empirical research on exchange rate forecasting within machine learning, demonstrating that models incorporating interest rate policy information outperformed both random walk models and recursive linear regression approaches in predicting currency fluctuations.

3.2. Pairs Trading

In the field of pairs trading, a unifying contribution is the extensive review by Christopher Krauss [43], published in 2015. This work systematically categorizes various approaches to pairs trading, identifying five primary methodologies for selecting and trading suitable pairs or groups of assets for statistical arbitrage.

Building upon Krauss’s classification, a notable branch of research focuses on assessing the closeness of time series using specific distance metrics. The foundational study in this domain is that of Gatev et al. [18] (2006), in which stock pairs are selected based on their closeness measured using the sum of squared Euclidean distances between their price time series. This methodology was further refined by Do and Faff [5, 6] (2010, 2012), who incorporated an additional constraint requiring securities to belong to the same industry classification within the 48 Fama-French industries. Another development in this line of research is the use of the *Pearson* correlation coefficient as a measure of time series similarity, introduced by Chen et al. [27] (2012).

A second body of literature focuses on the role of cointegration in pair selection, employing both the *Engle-Granger* [66] and *Johansen* [40] tests. The main work in this category is provided by Vidyamurthy [79] (2004), who develops a univariate cointegration framework for pairs trading, though without empirical applications. Building on this foundation, Lin et al. [83] (2006) introduce a minimum profit condition for a cointegrated pair, further refining the selection process. Studies applying this concept to commodity futures pairs, such as those by Girma and Paulson [57] (1999), Dunis et al. [12] (2006), and Cummins and Bucca [50] (2012), demonstrate the profitability of this approach in futures markets. However, Faizullin [19] highlights critical challenges in relying too heavily on cointegration for pair selection, noting its sensitivity to various factors, including time period selection and the risk of disappearing cointegration relationships.

A third line of research emphasizes time series modeling to identify suitable trading pairs, focusing on the mean-reverting properties of asset spreads. The key metric for pair suitability in this framework is the strength of mean reversion. Among the most cited works in this area is the study by Elliott et al. [67] (2005). Bertram [7] (2010) further develops a statistical arbitrage trading model in which the spread follows an *Ornstein-Uhlenbeck* process, though he also cautions against applying a Gaussian OU process to non-Gaussian financial data. Other notable contributions include Bock and Mestel [49] (2009), who employ a Markov Switching Model for pairs trading, and Bogomolov [8] (2013), who introduces an innovative approach based on *renko* and *kagi* models.

A fourth research direction, identified in Krauss’s review, involves stochastic control techniques for optimizing trading strategies. The most impactful study in this domain is that of Jurek and Yang [39] (2007). Building upon their findings, Liu and Timmermann [37] (2013) propose a strategy based on modeling asset price dynamics through an error correction model. Other contributions in this area explore optimal stopping theory and liquidation strategies under loss conditions, such as the work of Kuo et al. [42] (2015).

The final category encompasses machine learning and statistical tools for pair selection. One prominent approach involves the use of copulas to generate trading signals. Following a formation period based on conventional selection criteria, copulas are used to classify price movements of the pairs as anomalous, triggering buy or sell signals. This methodology is extensively studied by Ferreira [20] (2008), Liew and Wu [68] (2013), and Stander et al. [81] (2013).

Another widely applied statistical tool is *Principal Component Analysis* (PCA), which facilitates pair selection by constructing synthetic portfolios based on factor models. Avellaneda and Lee [48] (2010) employ PCA to generate as many synthetic portfolios as there are factors in their underlying factor model.

Clustering techniques also play a role in pair selection. These methods identify groups of similar stocks suitable for strategies that capitalize on the persistence of shared dynamics after temporary deviations from equilibrium. A significant contribution in this area is the study by Cartea, Cucuringu, and Jin [1] (2023), who apply graph clustering algorithms to partition the correlation matrix of market residual returns into clusters, subsequently constructing mean-reverting statistical arbitrage portfolios that yield annualized returns exceeding 10%. Other clustering-based approaches, such as the correlation-based methodology proposed by Zhan et al. [26] (2015), further support the efficacy of this technique.

A purely neural network-based approach has been proposed by Ordonez, Pelger, and Zanotti [36] (2019). Their methodology first identifies stock pairs sharing similar asset pricing factors. Time series signals are then extracted using a convolutional transformer and subsequently used to formulate an optimal trading policy. Their results indicate a substantial outperformance relative to benchmark approaches.

Lastly, an innovative hybrid methodology has been proposed by Palomar [63] in his 2020 lecture series. While not introducing new individual techniques, Palomar’s contribution lies in the way he integrates multiple tools. His approach involves selecting pairs based on a cointegration test, modeling their spread as a mean-reverting process with zero mean, and accounting for time-varying parameters by applying a Kalman filter to the time dependent cointegration vector.

4. General Setting: Data and Assumptions

In this section some general considerations regarding the source and management of the data and some key assumptions regarding the logic and formalism followed throughout the entire later discussion when discussing about the opening of the trades and the realization of the profits are introduced.

4.1. Data Source: FX and Treasury Bond Yield

Regarding the datasets used in this study, daily observations of spot exchange rates over the period from January 4, 2010 ($t = 0$) to December 29, 2023 ($t = T$) are taken into account. Selecting such an extended and multifaceted period enables the proposed intuitions to be tested across diverse market conditions and financial scenarios. The dataset includes the spot exchange rates between ten of the most actively traded currencies: the American Dollar *USD*, Euro *EUR*, British Pound *GBP*, Swiss Franc *CHF*, Japanese Yen *JPY*, Chinese Yuan *CNY*, Australian Dollar *AUD*, New Zealand Dollar *NZD*, Canadian Dollar *CAD*, and Hong Kong Dollar *HKD*, collectively referred to as *CURR*. Days corresponding to weekends or market closures have been excluded from the dataset to ensure consistency in the analysis and to have valid values in each considered date.

As previously mentioned, all possible exchange rates between the ten selected currencies have been considered, resulting in $10 \times 9 = 45$ distinct time series. For each of these time series $\{XY\}_t$, where $X, Y \in \text{CURR}$ and $t \in [0, T]$, the corresponding inverse rate time series $\{YX\}_t$ has been computed as $YX_t = \frac{1}{XY_t} \forall t \in [0, T]$, leading to a total of 90 time series. From this point onward, unless explicitly required, the time dependencies in

the notation will be omitted and XY used to refer to the entire time series. The motivation behind incorporating inverse exchange rates may initially appear counterintuitive: given the statistical nature of this approach, it is essential to track the behavior of both XY and YX . Despite being analytically linked and conveying the same economic information under the same financial laws, their empirical distributions, as well as their responses to the selection methods and trading strategies employed, may differ. Given the nonlinear relationship between an exchange rate and its inverse, excluding one from the analysis could result in missing valuable and exploitable market behaviors.

The FX time series data utilized in this study are sourced from the *Yahoo Finance API*, with all published values referring to 00 : 00 *UTC*. This choice makes use of the 24-hour availability of the FX market to establish a standardized and contemporaneous observation time, which also serves as the trading execution time. By doing so, the trading strategies operate smoothly without requiring additional adjustments to address the otherwise arising issue of data asynchronicity.

The daily treasury bond yields of the countries issuing the currencies in *CURR*, used as exogenous variables in the *GAM* selection method, are sourced from *Investing.com* over the same time period. For both the exchange rate and bond yield datasets, data of the initial date are downloaded solely to obtain lagged values for subsequent dates. The selection of 2-year, 5-year, and 10-year bond yields is driven by economic considerations further elaborated in *Subsection 5.3*. Each of these yield horizons provides distinct information, aiding in the interpretation of exchange rate behavior. For the Eurozone, since each country has its own treasury bond yield, a weighted average of the yields from the largest European economies, Germany, France, Italy, Spain, and the Netherlands, has been used as an aggregate proxy. The weights are assigned based on the relative GDP sizes of these countries: Germany 4, France 3, Italy 2.5, Spain 1.5, and the Netherlands 1.

4.2. Management of Missing and Problematic Values

Both the FX and bond yield datasets contained missing values and inputs where decimal points were incorrectly placed evidently. The latter issue has been addressed manually: each erroneous value has been reviewed and corrected individually. The missing data, which was primarily observed in the FX dataset, has been handled by supplementing the gaps with values from alternative data sources when only a small fraction (approximately 10 out of 45) of the time series values were missing. However, when the missing data exceeded this threshold, the affected dates were excluded from both the training and trading periods of the strategies. This occurred on specific dates, such as April 15, 2011; December 4, 2012; October 8, 2013; July 11, 2017; November 16, 2017; and May 22, 2019, which were treated as market holidays.

Finally, some Japanese bond yield values recorded as *Nan* were manually replaced with 0. This decision was based on two key considerations. First, these problematic values occurred within extended periods where bond yields fluctuated around slightly positive and slightly negative levels, or where they remained flat at zero. Second, it is widely acknowledged that Japan's monetary policy has repeatedly reduced interest rates, even entering negative territory, in an effort to stimulate economic growth.

4.3. Trading Assumptions

The first aspect concerns the process of opening and closing trades, as well as the realization of profits. When it is said that a long position on the XY rate is opened and the value of XY is v it means that $1X$ is bought paying vY . When a short position on the same rate and with the same value is opened, $1X$ is sold in exchange for vY . The profit or eventually the loss is then always realized in the second currency Y : if a long position on the rate XY is opened with initial value v_1 and closed with final value v_2 , then the final Y profit is $v_2 - v_1$. On the contrary, if a short position is opened and closed on the same rate and with the same values the final Y profit is $v_1 - v_2$. With this set of assumptions the rate XY is being somehow treated as an asset named X whose price is v in Y currency and whose trading size is always 1. As an example, if a long position on *EURUSD* is opened when its value is 1.0708 and closed when the value of the exchange rate is 1.0903 it means that a total profit of 0.0195 *USD* has been realized buying 1 *EUR* for 1.0708 *USD* and then selling it for 1.0903 *USD*. In this framework, the action of short selling is equivalent under any aspect to a long position in the inverse rate and for this reason it does not require special treatment or imply any limitations.

Furthermore, only a reduced and fixed number of exchange rate pairs are traded in each time period. Trading all possible pairs, or even just those that meet certain constraints, would significantly increase the computational complexity and make it difficult to compare the results across different methods and strategies due to the varying number of assets involved. To address this, the number of exchange rates or rate pairs to be traded is predefined in advance. This approach reduces computational effort and ensures that the analysis of results

are more comparable and consistent. The exchange rate pairs selected for trading are determined based on the score rankings obtained during the selection phase. This also ensures that the traded exchange rates, at the moment of the selection, are the ones that best show behaviors aligned with the trading strategies presented.

Additionally, with respect to the management of trade dates, the decision to fix the first trade date has been made to ensure that the results are comparable across different methods. Regardless of the selection method, trading strategy, or their respective hyperparameters, specifically the frequency used to select tradable pairs and rates and the length of the training period for algorithm training and asset behavior analysis, April 25, 2013 has been chosen as the common starting date for all trades. This date is selected because it provides sufficient data to train all trading strategies with the longest possible training period among the considered sizes. By standardizing the starting date, a consistent trading period for each selection and trading combination is ensured, facilitating a valid comparison of results.

Another consideration regarding the date management involves the assumption related to the realization of profits. When a trade is closed, its profit or loss is recorded and considered realized on the same day the trade is closed, and the opening of a new trade is considered to occur immediately. This study does not account for the actual settlement date of spot trades, which typically occurs two business days after the order date ($t + 2$) for most currency pairs. This choice has been made to treat all trades and operations equally even when some of them follow specific rules and conventions, such as *USDCAD*, where settlement happens the next business day ($t + 1$). Moreover, incorporating settlement dates would necessitate additional attention to US holidays, as these can influence the actual settlement date even when the *USD* is not directly involved in the currency pair.

4.4. Transaction Costs

The approach used to model transaction costs in this research is outlined. In order to ensure the results are as realistic as possible, this study accounts for a bid/ask spread for each trade, based on the reported middle spot rate. Since precise bid and ask values for each exchange rate are not available, these spreads are kept fixed across different dates and rates, which appeared as the most practical, despite rough, solution. Specifically, from the perspective of a mid-size institutional investor, additional costs (or reduced gain) of 1.5 basis points for exchange rates involving two between *USD*, *EUR*, *GBP*, and *JPY* have been assumed, while for all other currency rates, given that they do not involve highly exotic currencies, a moderately higher transaction cost of 5 basis points has been applied. These values may appear too conservative but they can balance for periods of market illiquidity due to financially extreme events that caused enormous and unforecastable spreads between ask and bid prices.

5. Selection Methods

The core concept underlying these algorithms is the computation of a similarity score between exchange rates that, in some cases, takes into account also the past profitability of the pair. Pairs of exchange rates with the highest similarity scores are subsequently considered for pair trading and actually pair traded if one (or both) of them reaches extreme values. The specific rules governing trade opening and closure will be discussed in *Section 6*; this section is dedicated exclusively to the criteria employed for score evaluation and pair selection. Nevertheless, certain details regarding the trading strategy associated with each selection method will be provided where necessary to enhance comprehension of the underlying logic and to contextualize it appropriately (refer to Figure 11 for a broader understanding of the relationship between selection methods and trading strategies). Finally, the last subsection focuses on the selection criterion applied to identify individual tradable rates, rather than pairs, which is exclusively used in the only trading algorithm that targets individual exchange rates instead of their relative dynamics.

Some general considerations regarding all selection methods and their connection to their respective trading strategies can be made at this point. The selection of currency pairs or individual exchange rates is performed periodically, with a relatively low frequency, typically on the scale of several hundred days. The decision to adopt extended selection periods and to maintain fixed tradable assets for longer durations has two reasons. Firstly, it accounts for data availability and computational feasibility. Secondly, it aligns with the broader research objective, which fundamentally differs from that of high-frequency trading strategies.

Shorter time horizons would necessitate the use of high-frequency data, which is generally noisier, falls outside the scope of the available dataset, and tends to exhibit price divergences that are closed almost instantaneously. The primary focus of this research is to take advantage of the creation and breakdown of long-term relationships between exchange rates, rather than exploiting momentary market inefficiencies. The latter approach would require an extensive dataset (potentially with minute-level granularity), specialized algorithms, and computational power capable of detecting optimal trade entry and exit points within seconds or minutes, far beyond the

capabilities of a standard personal computer. Such high-frequency strategies are typically leveraged by large institutional actors who can instantly identify and act upon price discrepancies before they are corrected by market forces and before competitors react to these market opportunities [73].

Conversely, the aim of this study is to identify prolonged and stable relationships between exchange rates that, for various reasons, experience disruptions over extended periods. These disruptions create opportunities for investors to anticipate a reversion to the previous equilibrium and generate profits accordingly. Such relationships typically evolve and reestablish themselves over long time frames and this explains the choice to keep long time horizons.

Another important consideration regarding the selection frequency concerns the relationship between the selection method and the trading strategy. The time horizon of the strategies, defined as the period after which all trades initiated using the most recent fit of the trading algorithm are closed, does not necessarily coincide with the asset selection frequency. This study allows for an overlapping period between the old and new selections: while the newly selected pairs become tradable, the previously chosen assets remain active, but only for closing existing trades. During this transition phase, new trades can only be initiated with the newly selected assets, while positions involving the previously chosen assets can only be closed. Once the overlapping period concludes, trading activity is exclusively conducted on the newly selected rates, with both trade openings and closings restricted to this updated selection.

This approach, illustrated in Figure 1, is primarily implemented to prevent the premature termination of trades that were initiated shortly before the new selection is made. By allowing these positions to remain open for a longer period, the strategy aims to fully capitalize on potentially profitable opportunities that might otherwise be closed prematurely, often with minimal gains, and to close them just when needed, allowing them to complete the reversion or reach the behavior that would be expected to close them with a profit.

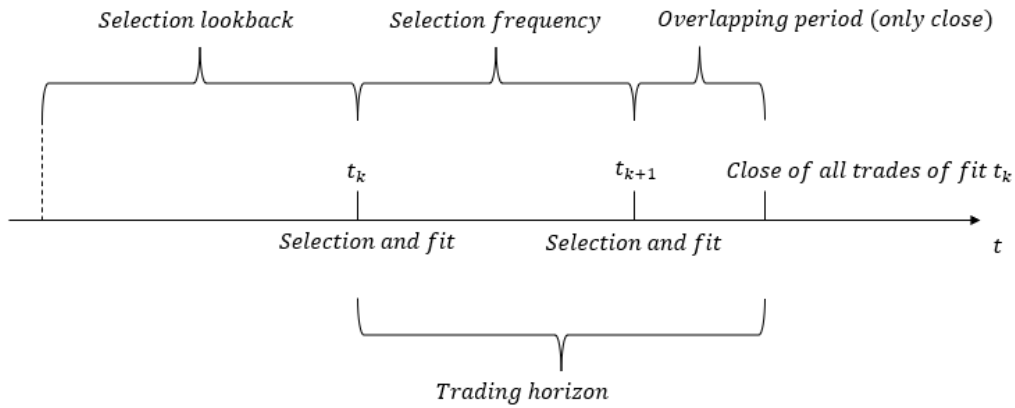


Figure 1: Approach to subsequent trading fitting and selections. In the overlapping period, trades involving rates selected in t_k can just be closed; the opened ones involve just rates selected in t_{k+1} .

As a matter of notation, the length of the historical period analyzed at each selection date to assign a score and rank all pairs of assets will be referred to as *lookback length*. This hyperparameter can provide varying weights to historical observations. A longer lookback length offers a larger sample size, a broader understanding of past events, and a ranking method more resilient to extreme occurrences; however, the resulting scores may be influenced by outdated events that are no longer relevant. Conversely, a shorter lookback length places greater emphasis on the most recent past, making the ranking more responsive to current market dynamics, but it may also be more susceptible to transient and non-permanent oscillations. Intuitively, *selection frequency* indicates the number of daily observations that occur after the selection of the assets before a new selection takes place. In some cases, the values of these quantities have been fixed, while in others they have been made variable to study different behaviors.

Finally, it is necessary to highlight that the criteria used to select the pairs of exchange rates are sometimes based on the past observation of their spread, meaning that a single indicator is used for each couple, while others are based on the individual and joint movements of the rates.

Regarding the initial universe of exchange rates and pairs, it is important to highlight that most selection methods also consider mixed pairs as eligible, i.e., pairs of the form (XY, ZT) , leading to a total of 3960 ($= \frac{90 \times 89 - 90}{2}$) pairs. Unless otherwise specified, the selection methods applied to exchange rate pairs will operate within this broader universe.

This approach is justified by the fact that these selection criteria primarily focus on the reciprocal behavior of

the exchange rates while retaining information about their individual dynamics and so providing information for the subsequent trading phase about both the involved rates trying to ensure profit in both the legs. This fundamental distinction sets them apart from the methodology discussed in *Subsection 5.1*. Furthermore, the possibility that exchange rates may exhibit similar dynamics, thus making them suitable for pairs trading even in the absence of a shared common currency, has already been discussed in *Section 2* and motivated by empirical evidence within this study.

5.1. Spread Selection

The first method introduced seeks to identify pairs of exchange rates whose spread, defined as their ratio following the approach proposed in [49], demonstrates the most pronounced mean-reverting behavior and profitability potential. Empirical results, documented in the scientific literature, indicate that the ratio of two financial assets can be assumed to follow a mean reverting process in several cases as documented in [17, 79]. The core strategy involves trading this spread by establishing a long position on asset 1 (the numerator) and a short position on asset 2 (the denominator) when the ratio is considered low, while adopting the opposite stance when the spread is considered high.

5.1.1 The Role of the Ratio

The purpose of this selection method, along with the trading strategies that leverage the spread's behavior to generate trading signals, is to determine whether it is possible to select and trade exchange rate pairs based on the behavior and distribution of a single dynamics, rather than relying on the study of their joint distributions. This approach aims to enable the selection of rate pairs and the daily decision-making regarding entry and exit positions without the need to analyze and model their mutual behavior, as all trading decisions would be based on a single derived value. While this objective may seem simplistic at first glance, it is, in fact, highly ambitious: it essentially assumes that a dynamic derived from the underlying exchange rates encapsulates all the necessary information for both pair selection and profitable trading decisions.

The financial interpretation of the ratio spread is, however, debatable. Although empirical studies [17, 49, 79] report that the ratio of two financial assets often exhibits mean-reverting behavior, an aspect of particular interest in this analysis, this measure lacks a precise and well-defined economic meaning. Conversely, a spread approach based on the difference $XY - \gamma ZT$ would be much more aligned with common practice but it would necessitate an estimation of the parameter γ to ensure an unbiased analysis, as assuming $\gamma = 1$ could lead to distortions due to the often vastly different scales of exchange rates. The use of the ratio spread addresses this issue by eliminating the need for such an adjustment and allowing the trades to take place on pairs made of just two exchange rates and not $1 + \gamma$, resulting in the trade of just two exchange rates. Even a prior normalization of both the time series would not be a viable solution because the distribution of their new difference would be different from the initial one without any Gaussianity assumption.

Most of all, adopting a difference-based spread approach would place the study within the realm of a purely cointegration-based methodology, a powerful yet highly rigid framework. The intent of this research is to move beyond such a restrictive approach, despite the significant research potential in further exploring cointegration-based pairs trading in the forex market, as established by earlier foundational works in the field.

5.1.2 The Starting Universe

This is the only selection method that imposes a restriction on the universe of possible pairs. Given that the quantity under analysis is the ratio between two exchange rates, only pairs sharing the same currency in which profit or loss is realized should be included; namely, spreads of the form (XY, ZY) . As a result, the total number of considered pairs is reduced to 720 ($= 90 \times 8$).

From a practical perspective, this selection criterion enhances the likelihood of achieving the desired profitability (whether the spread is computed as a ratio or as a difference). The object under observation is the spread of the exchange rates; any desired and actually realized movement of the spread does not necessarily imply that both components of the pair behave as expected but that at least one of them does and that it compensates also for the other if they are expressed in the same currency. If the two exchange rates composing the spread realize profits and losses in different currencies, this leads to the concrete and actual possibility that one of the legs of the trade results in a loss even if the spread evolves in the wanted direction. This is different from other selection methods as later examined.

Additionally, it is important to note that spreads (XY, ZY) and (ZY, XY) are treated as distinct entities despite containing the same exchange rates. Consequently, there exists a scenario in which one of these spreads is selected while the other is not. Empirical analysis suggests that the scores assigned to these spreads, using

the methodology outlined in this section, tend to be similar, eventually raising a diversification issue: the actual number of distinct spreads to which the portfolio is exposed is only half of the preselected ones, effectively doubling the strategy's exposure to the same underlying risks while halving the overall number of idiosyncratic risks.

5.1.3 Score Methodology

The objective is to identify a spread that exhibits frequent and rapid mean reversion while avoiding a pronounced trend behavior. The first property is essential to ensure that opened positions can be closed profitably within the trading horizon.

The second characteristic ensures the stability of the relationship between the two exchange rates. Here, "stability" is an intuitive notion, without a formalized technical definition, but it implies, among the others, a certain degree of similarity between the two rates. A strong directional trend in the spread would complicate the implementation of the associated trading strategies, as it would necessitate precise modeling and estimation of that trend.

Additionally, an ideal spread should exhibit substantial oscillations, frequently crossing its standard deviation thresholds. The stronger this property, the greater the likelihood of profitable trading opportunities, as it ensures sufficiently extreme values that can be exploited for entry and exit points.

Given all the sought properties, the computation of the score for the pair associated with the spread under analysis is straightforward. First, the spread is computed and Z-normalized for numerical reasons. Then, the spread undergoes a stationarity test using the *Augmented Dickey-Fuller* test [14, 15]: if the resulting *p value* exceeds 0.10, the pair is automatically discarded. The failure to reject H_0 , indicating the absence of stationarity around a constant mean, would make the estimation of drift and trend difficult, thus complicating the development of trading rules. Moreover, profit would be highly dependent on the precision of these estimations. Including non-stationary spreads around the mean would interfere with the creation of entry/exit rules and complicate any forecasts regarding mean reversion. Furthermore, in this case, in contrast to what will be discussed in *Subsection 5.5* for a completely different setup, stationarity also serves as a measure of similarity between the exchange rates, excluding the possibility of marked trends in the spread, divergences, or a systematic appreciation of one exchange rate relative to the other as already mentioned.

Then, the weighted average of three quantities is computed: the number of times the spread has crossed its mean value (*mean crossings*), the number of times it has crossed its one-standard-deviation thresholds (1σ *crossings*), and the number of extreme values, defined as observations falling outside the two-standard-deviation range, it has recorded (*extreme values*) (Equation (1)) during the lookback period. These quantities aim to serve as proxies for the sought features, ensuring fast and frequent reversion to the mean as well as the presence of sufficiently extreme values to open trades and generate consistent profits. An ideal behavior is exemplified in Figure 2. In particular, the number of mean-crossings as a proxy for the mean-returning property is inspired by the work of Donaldson and Mehra [35].

The focus is then on two aspects: the similarity between the rates and the profitability of the spread.

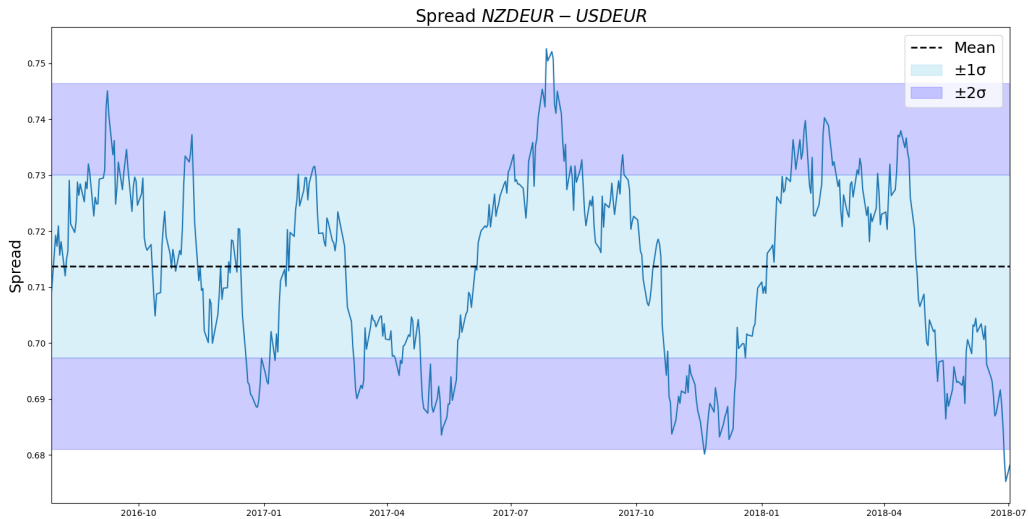


Figure 2: A frame of the behavior of the *NZDEUR - USDEUR* spread.

The sets of weights used reflect different approaches to the problem and each one stress the research for a

particular feature, respectively *meancrossings*, *1σ crossings* and *extreme values*. The first set of weights represents the most balanced approach, aiming to capture and evenly weight all mean-reverting features: $w = [0.2, 0.5, 0.3]$. The second set places greater emphasis on the occurrence of extreme values while favoring a fast mean-reverting property, optimizing for a more profit-oriented strategy: $w = [0.45, 0.1, 0.45]$. Lastly, the third set is designed to ensure the return-to-mean property, prioritizing sufficiently high oscillations while reducing reliance on extreme values that could introduce excessive risk to the portfolio: $w = [0.4, 0.4, 0.2]$.

Finally, the two original time series of exchange rates values are tested for cointegration using the *Engle-Granger* test [66]: in case of a p value less than 0.10, meaning a likely rejection of the H_0 hypothesis of no cointegration, the total score receives a 10% increment.

pair score

$$= \begin{cases} (w_1 \cdot \text{meancrossings} + w_2 \cdot 1\sigma \text{ crossings} + w_3 \cdot \text{extreme values}) \cdot (1 + 0.1 \cdot \mathbf{1}_{p_{EG} \leq 0.10}) & \text{if } p_{ADF} \leq 0.10 \\ \text{default discarded} & \text{if } p_{ADF} > 0.10 \end{cases} \quad (1)$$

In this study, different values of lookback length and selection frequency have been employed in combination with the aforementioned set of weights: $[\text{lookback length}, \text{selection frequency}] = [500, 250], [400, 200],$ and $[300, 150]$. These configurations represent varying approaches to the problem, ranging from a more historically dependent and slower-to-adapt scheme to a more responsive and higher-frequency framework for incorporating new information.

5.1.4 Use of Stationarity and Cointegration

The use of the stationarity criterion as a restrictive instrument to filter the eligible pairs can be questioned for its severity. As reported in Figure 26 in *Appendix A*, the percentage of pairs for which H_0 can be rejected never crosses 25% and for some periods it is below 5%. Despite that, it is useful to recall that this study deals with 720 pairs and that at each selection date just a little fraction of pairs are selected and then considered for trading. Thanks to that, the number of tradable assets remains sufficiently high even after a strict preselection. As a remark, the stationarity can be also considered as an helpful tool to estimate in a non traditional way the similarity between the rates as already mentioned.

The described application of cointegration can arise some concerns as it diverges from the traditional hypothesis testing framework, particularly regarding first-order integration and the estimation of the cointegration vector. In this practical setting, cointegration serves primarily as an empirical tool to boost the selection of exchange rate pairs that have exhibited comovement in the recent past, rather than entirely delegate to the spread the pair selection procedure. Its objective is to enhance the identification of exchange rate pairs with similar behavior while avoiding excessive penalization of those that, despite not being formally cointegrated, have demonstrated a promising spread dynamic in terms of oscillatory behavior.

This is achieved, as discussed in *Subsubsection 5.1.3*, by assigning a 10% higher score to pairs that are successfully tested for cointegration while still allowing the consideration of those that do not meet this criterion by default. As noted by G. Vidyamurthy [79] and referenced by C. Krauss [43], the key consideration for a pair trader is not necessarily the presence of cointegration but rather a spread characterized by strong mean-reversion properties, which are partially covered by the use of a stationarity test.

5.2. Copula Selection

The selection method based on copulas aims to identify, among all possible pairs of exchange rates, those in which the exchange rate values have exhibited a higher frequency of extreme deviations during the lookback period. This is achieved through the application of copulas, statistical tools that model the dependence structure between random variables by decomposing their joint distribution into marginal distributions and a copula function. This approach allows for the analysis of complex dependencies without imposing the assumption of normality of the data.

Copulas are particularly useful within the framework of this study for two main reasons. First, they effectively detect extreme values by capturing tail dependencies, which is something crucial for the purpose of this study. Second, the absence of the Gaussianity assumption aligns well with the empirical characteristics of the data under consideration. As illustrated by the way of example in Figure 27 and further validated in Table 1 (both in *Appendix B*), the levels of spot prices do not follow a normal distribution, instead exhibiting multimodal distributions.

5.2.1 Application to Exchange Rate Time Series

The application of copulas to exchange rate levels rather than daily returns is undoubtedly improper from a theoretical and econometric point of view. However, from an empirical perspective, it proves to be effective in achieving the intended objective. The primary role of copulas in this context is to identify simultaneous outlier values in the exchange rates belonging to the analyzed pairs, leveraging their ability to model tail dependencies. The detection of these extreme observations serves as the entry signal for a trade. Given that the primary focus is on classifying individual observations as extreme, the use of exchange rate levels instead of returns can be a pragmatic solution. This approach is supported by [64], where copulas are directly fitted to levels to capture the dependence structure between their marginal distributions and to determine when a spot rate can be classified as extreme relative to the value of another asset.

Moreover, spot exchange rates exhibit stronger mean-reverting and stationarity properties compared to equities, particularly over the extended time horizons considered in this study, as documented in [34]. Within the FX framework, the quantiles of the values remain valid and do not exhibit the flattening to 1 that would typically occur in the presence of a drift (refer to Figure 29 in *Appendix D* for a better explanation).

5.2.2 Score Methodology

Delving into the specifics of the selection method, a score is assigned to each pair of exchange rates, and the pairs with the highest scores are incorporated into the basket of tradable pairs. These pairs will be subject to pair trading in the subsequent trading period as soon as both legs of the pair reach extreme values. The strategy involves shorting the overvalued asset and taking a long position in the undervalued one, with the expectation that they will revert to their prior *status quo*, as it will become clearer in *Subsection 6.3*. The data are not normalized, as normalization or logarithmic transformations would yield identical copula fits. Since copulas operate on quantiles, any strictly monotonic transformation using an increasing function on the marginal data produces the same results.

The selection method based on copulas is intrinsically linked to its corresponding trading strategy. Specifically, pairs are chosen based on the tendency of their exchange rates to simultaneously exhibit extreme values in opposite directions, as determined by the fitted copula. The lookback period for each candidate pair is divided into training and test segments according to a 60%/40% rule. The test size is relatively large compared to standard practices; however, a sufficiently high number of days is required to carry out this type of procedure effectively. The training portion is used to fit a copula, with several Archimedean models being tested (*Gumbel* [55], *Frank* [23], *Clayton* [13], *Joe* [29]). The ones for which the fit is not possible due to numerical reasons are excluded a priori. The motivations for restricting the selection to this specific set of copulas will be discussed in detail later. The copula fit that minimizes the *Akaike Information Criterion* [4] is selected, and a simulated trading phase is conducted on the test portion of the lookback period. During this test phase, the number of trades that would have been initiated in a real trading scenario is recorded. A trade is triggered when the conditional probabilities, derived from the distributions of the quantiles, of the observed quantile pair on the current day satisfy one of the two predefined conditions.

$$\frac{\partial C(u_1, u_2)}{\partial u_2} = \mathbb{P}(U_1 \leq u_1 | U_2 = u_2) \leq p_{low} \wedge \frac{\partial C(u_1, u_2)}{\partial u_1} = \mathbb{P}(U_2 \leq u_2 | U_1 = u_1) \geq p_{up} \quad (2)$$

$$\frac{\partial C(u_1, u_2)}{\partial u_2} = \mathbb{P}(U_1 \leq u_1 | U_2 = u_2) \geq p_{up} \wedge \frac{\partial C(u_1, u_2)}{\partial u_1} = \mathbb{P}(U_2 \leq u_2 | U_1 = u_1) \leq p_{low} \quad (3)$$

where u_1 and u_2 represent the quantiles derived from observed data for the two exchange rates, mapped through their respective marginal *Cumulative Distribution Functions*. The corresponding quantile random variables are denoted as U_1 and U_2 , while $C(u_1, u_2)$ represents the fitted copula. The parameters p_{up} and p_{low} are the threshold probabilities for trade initiation, serving as the main hyperparameters of the model. Their role is to establish the criteria for classifying observed value pairs (or quantiles) as extreme. The specific values chosen for these parameters will be discussed in subsequent sections.

The relationships expressed in Equations (2) and (3) hold due to *Sklar's theorem* [72]. When these conditions are satisfied with appropriately selected threshold probabilities, they respectively indicate that the first exchange rate is undervalued while the second is overvalued, and vice versa. A trade is subsequently closed when both asset values revert from their extreme positions to a more stable region, characterized by probability values that cross again the 40%/60% thresholds in the quantile distribution. The conceptual framework underlying this logic is visually depicted in Figure 3.

The score assigned to each pair is determined by the number of successfully closed trades (n), computed following the previously introduced rules. This value is then multiplied by the average outlier probability, which is defined as the mean of the most extreme probabilities, that are the probabilities that triggered the opening of trades. This computation is formally summarized in Equations (4) and (5).

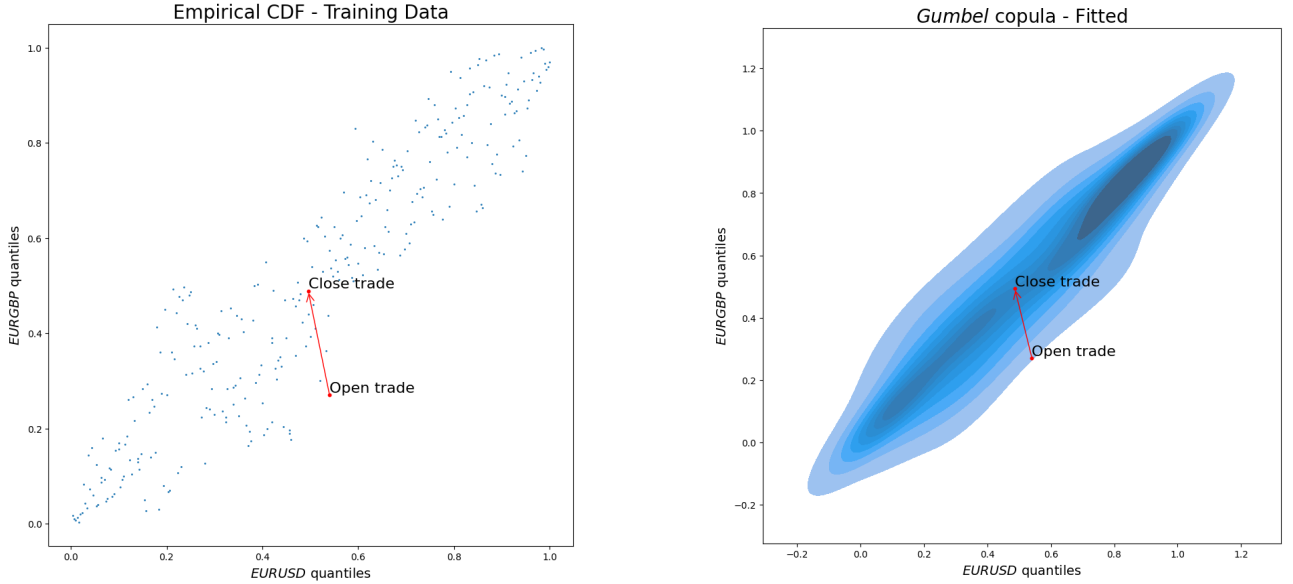


Figure 3: The logic followed to evaluate n , the number of trades that have been opened and successfully closed within the test period, for the pair $EURUSD - EURGBP$.

$$pair\ score = n \cdot \bar{P} \quad (4)$$

$$\bar{P} = \frac{1}{n} \cdot \sum_{i \in opened\ trades} (0.5 + \max(|0.5 - \frac{\partial C(u_1^i, u_2^i)}{\partial u_2}|, |0.5 - \frac{\partial C(u_1^i, u_2^i)}{\partial u_1}|)) \quad (5)$$

This score ranks both the tradability of the pair and the ability of its constituents to reach extreme values in probability. It serves as a proxy that combines the financial and statistical properties of the pair.

Different couples of threshold probabilities have been used in combination with $lookback\ length = 500$ and $selection\ frequency = 250$. All these pairs satisfy $p_{up} + p_{low} = 1$, as the legs of the pair need to be treated symmetrically. In this study, the final decision was to choose $[p_{low}, p_{up}] = [0.05, 0.95], [0.125, 0.875], [0.20, 0.80]$. These choices represent different attitudes, ranging from a more conservative to a more dynamic and risk-taking approach, towards the detection and classification of extreme values, and they can influence the composition of the basket of rates that will be traded later.

5.2.3 Final Considerations

As previously mentioned, the fit of the copula on the first fraction of the lookback period is performed using four types of Archimedean copulas, and the one yielding the best AIC value is selected for the scoring phase. The decision to restrict the choice to this list is based on the strong ability of these copula models to capture upper and lower tail dependencies with varying intensities, which is the core of the analysis in this section, and the empirical similarity between the distribution of the quantiles and the shapes of the copula functions. The *Gaussian* copula has been excluded because it lacks tail dependencies, despite its widespread use due to its simplicity. Finally, although the *t-Student* copula would have likely been the most appropriate for capturing both upper and lower tail dependencies symmetrically and for its wide use with financial data, its use would have been computationally prohibitive due to excessive processing time. Additionally, an empirical analysis conducted on *USD* denominated pairs over the period between January 5, 2010 and January 30, 2013 highlights that just the 13.9% of these pairs would have adopted the *t-Student* copula if considered with an average improvement of the AIC value of 8%.

It is important to note that, among all the proposed selection pair methods, this one is the least related to the others. The concept of similarity introduced by this approach is solely concerned with the ability of the time series to exist simultaneously within an acceptable range of values after having taken extreme opposite values. This method is arguably the most profit-oriented selection criterion, as it aims to capture the profitability of the pair, but it is the least capable of properly describing the similarity between the exchange rates. In contrast to the selection criteria described in *Subsection 5.3*, which are entirely focused on detecting similarities between the time series from a financial perspective, and those in *Subsection 5.1*, where both similarity (measured through

the stationarity of the spread) and profitability (gauged by the oscillations of the spread) are considered, this approach prioritizes the potential for trading the pair. It does so by exploiting the ability of the pair's components to take extreme opposite values simultaneously and then revert to their normal behavior. In essence, this method assesses the suitability of the pairs for future trading by evaluating their past performance.

5.3. Generalized Additive Model Selection

Among all the proposed approaches for selecting tradable pairs, this is the most machine-learning-based and the one that, more than any other, seeks to quantify the actual similarity between the time series of spot exchange rates at each selection date. Qualitatively, its objective is to assess the similarity between exchange rates by comparing how alike their responses and connections are to the same type of exogenous variables, as Figure 4 exemplifies. To identify the relationship between each individual exogenous variable and the spot rate, *Generalized Additive Models* [75] have been employed. *GAM* are particularly well-suited for this task, as they can model predictors independently in a non-linear manner, generating shape functions that capture arbitrary patterns while remaining fully interpretable and explicit. In this framework *GAM* are not utilized for their conventional forecasting purpose.

5.3.1 Model Formulation

GAM ability to identify the non-linear patterns between input and response variables, the exchange rates, can be expressed as in Equation (6), useful also as a matter of notation:

$$g(XY) = c + \sum_{i=1}^k f_i(X_i) + \epsilon \quad (6)$$

where XY is the exchange rate between currencies X and $Y \in CURR$, X_i are the k exogenous variables, f_i are the identified non-linear smooth functions capturing the relationships between the exogenous variables and the output variable, c is the intercept term, g is the link function applied to the expected value of the response variable and ϵ are normally distributed errors.

The exogenous variables employed are derived from the treasury bond yields of the countries issuing the currencies in $CURR$ as the spreads between the yields of the two respective currencies. The exceptional status of *EUR* has already been discussed in *Section 4*. More specifically, the relationships are modeled following Equation (7):

$$\begin{aligned} XY = c &+ f_1^{XY}(L(XY)) + f_2^{XY}(BY_{2y}^X - BY_{2y}^Y) + f_3^{XY}(L(BY_{2y}^X - BY_{2y}^Y)) + f_4^{XY}(BY_{5y}^X - BY_{5y}^Y) \\ &+ f_5^{XY}(L(BY_{5y}^X - BY_{5y}^Y)) + f_6^{XY}(BY_{10y}^X - BY_{10y}^Y) + f_7^{XY}(L(BY_{10y}^X - BY_{10y}^Y)) \end{aligned} \quad (7)$$

where $L(\cdot)$ denotes the lag operator ($L(x_t) = x_{t-1}$) and BY_{ty}^X represents the t -year bond yield of currency X . Treasury bond yields are widely recognized as one of the most influential factors driving currency dynamics. In particular, the 10-year yield is closely monitored by currency traders as an indicator of the overall direction of a country's interest rates. Conversely, shorter-term yields often reflect current monetary policy stances or immediate economic conditions and can still provide valuable insights into the expected behavior of a currency. The notion of "expectation" regarding future exchange rate movements is the primary reason for incorporating the lagged values of yield spreads into the model. Assessing whether the relationship between exchange rate expectations, derived from treasury bond yields, and the exchange rate itself exhibits consistency across different currency pairs can be highly significant. Additionally, the lagged value of the exchange rate itself has been included to examine whether autocorrelation patterns, if present, share similar structures across different exchange rates or whether spot rates respond differently to past values depending on the specific currency pair. Consequently, each exchange rate is modeled solely as a function of the exogenous variables of the currencies involved in the rate.

The link function g has been omitted in Equation (7) because the entire modeling process has been carried out using a *Linear GAM*, where the link function is assumed to be the identity, rather than a *Gamma GAM*, which models the logarithm of the response variable. This choice is questionable: the assumption of Gaussian-distributed data underlying the *Linear GAM* is certainly inadequate in this context, as illustrated by the examples in *Appendix B*. However, the Gamma distribution also appears to be unsuitable for this case study. Additionally, *Gamma GAM* requires the response variable to be strictly positive. While this condition is theoretically met, the necessary normalization procedure, which is discussed later, results in data that can also take negative values.

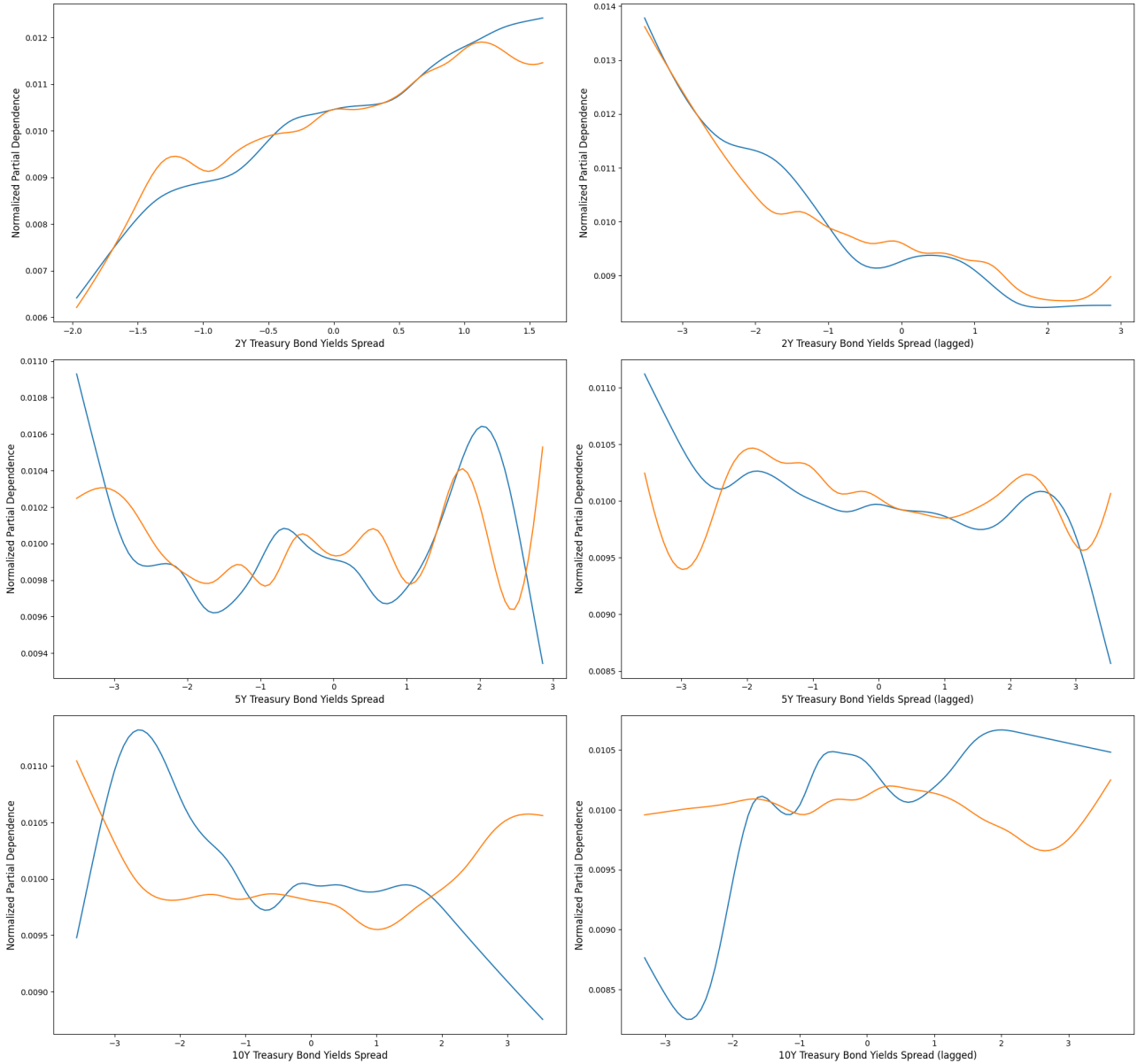


Figure 4: The similarity of the responses of *AUDUSD* (blue) and *AUDGBP* (orange) to the same kind of variables. Normalized values.

5.3.2 Score Methodology

This family of models must be fitted for each exchange rate at each selection date and particular attention must be given to the role of the smoothing parameter λ . This hyperparameter controls the degree of smoothness of each function f_i . A lower λ value results in a function that more closely interpolates the data, minimizing error but producing a less interpretable and less smooth function. For the sake of simplicity and computational efficiency, λ has been tuned while keeping it fixed across all smooth functions, and the functions f_i have been defined as *spline* terms due to their flexibility.

The calibration procedure consists of two phases for each exchange rate. First, the values of the exchange rate XY , the spread between treasury bond yields for currencies X and Y , and their respective lagged values during the lookback period are Z-normalized. The model specified in Equation (7) is then run for $\lambda = [0.001, 0.01, 0.1, 1, 10, 100, 1000]$. In the second phase, the appropriate λ value is selected based on an extended set of rules derived from the elbow method. Specifically, the chosen λ is the one that maximizes the distance between the *Mean Squared Error* curve as a function of λ and the line connecting the first and final points of the error function, assuming that the error function is convex and always remains below the connecting straight line. This approach ensures that, while keeping the error under control, the function is sufficiently smooth to capture an interpretable pattern between the exogenous and response variables. Any exceptional cases are discussed in *Appendix C*.

If the lowest $\sqrt{MSE}$ obtained during this tuning phase (not necessarily the one associated with the selected λ)

exceeds a threshold the exchange rate is automatically discarded. Different values of the tolerance have been adopted in combination with *lookback length*= 500 and *selection frequency*= 250: 0.10, 0.25 and 0.40, which correspond to 10%, 25% and 40% of the standard deviation of the Z-normalized response variable XY . In this case, the model's reliability using *GAM* is considered insufficient, and the rate cannot be considered for inclusion in any tradable pair.

The use of the elbow method is driven by the need to strike an optimal balance between smoothness, viewed as interpretability, and goodness of fit, interpreted as reliability. Minimizing the error or maximizing the *Effective Degrees of Freedom* (EDoF) would lead to overfitting the data, complicating comparisons with other exchange rates and giving high importance also to the behavior of some outlier observations. Conversely, minimizing the EDoF would necessitate large values of λ , resulting in an excessively smooth function that exhibits almost linear behavior, thereby reducing the model to a simple linear regression. Furthermore, the decision to prioritize *MSE* over R^2 as a measure of goodness of fit is driven by the observation that R^2 consistently delivers high values that are close to each other, making it less effective for distinguishing model performance.

By observing the resulting *p values*, the input variables used are generally significant at varying levels of confidence, typically ranging from 0.001 to 0.01. Some exchange rates exhibit exceptional behavior, and for these, certain exogenous variables may not be significant. However, such cases are rare. Moreover, to ensure a fair comparison between exchange rates, the same set of input variables must be applied consistently. Excluding specific variables in particular contexts would lead to an unbalanced procedure.

Some considerations regarding the procedure described above, particularly the normalization of the data, are necessary. First, the data are normalized for computational reasons, which facilitates a more accurate fit of the model due to numerical advantages, and for time saving motivations. Secondly, the data remain normalized for subsequent procedures: given that the detected patterns need to be compared, it is highly beneficial to maintain them within the same value range to better understand the similarities between the shapes of these functions. Another potential issue is the lack of a train/test split for the lookback period during the λ tuning process. The chosen λ value corresponds to the best fit across the entire lookback dataset. This decision is justified by the fact that the goal of the model is not forecasting, which is its typical application, but interpretation. The focus of this study is not on applying the fitted model to out-of-sample data; rather, the model needs to explain the relationships within the available observations. Therefore, the aim is to identify the best trade-off between accuracy and smoothness over the entire lookback period.

With the tuned values of λ for each exchange rate it is now possible to compute the actual similarity scores. The logic follows the approach already outlined: for each potential pair of exchange rates, a similarity score is calculated at each selection date, provided that both exchange rates have passed the test based on the lowest attainable $\sqrt{MSE}$. The entire model is then rerun on the complete available dataset using the tuned λ values, and the fitted f_i^{XY} for $i = 1, \dots, 7$ are obtained. For a given pair (XY, ZT) , the similarity score is computed as follows in Equation (8):

$$pair\ score = - \sum_{i=1}^7 d(f_i^{XY}, f_i^{ZT}) \quad (8)$$

where $d(\cdot, \cdot)$ measures the distance between the two functions: the lowest this value, the closest are these patterns. For this reason the opposite value is used to compute the similarity score. At this stage the involved functions are normalized such that they can be viewed as probability densities ($f_i^{XY, norm} \geq 0$ and $\int_{-\infty}^{\infty} f_i^{XY, norm} dx \stackrel{\text{discretized}}{=} \sum_{k=1}^N f_i^{XY, norm}(x_k) = 1 \ \forall \ i, XY$, where the function has been computed on a discretized grid of N points). This is accomplished thanks to the *softmax* normalization:

$$f_i^{XY, norm}(x_k) = \frac{e^{f_i^{XY}(x_k)}}{\sum_{k=1}^N e^{f_i^{XY}(x_k)}} \quad (9)$$

This type of transformation, thanks to the Z-normalization that has been performed on the data at the beginning of the computations leading to the data to be close to 0 and centered in it, ensure the conservation of the shape of the functions as it is possible to see from Figure 5. $d(\cdot, \cdot)$ is then computed as the *Jensen-Shannon divergence* [46] (10) between the two functions seen as normalized probability densities:

$$d(f, g) = \frac{1}{2} \cdot (d^{KL}(f, m) + d^{KL}(g, m)) \quad (10)$$

where $m = \frac{1}{2}(f + g)$ and $d^{KL}(\cdot, \cdot)$ is the *Kullback-Leibler divergence* [70]. To conclude for sake of precision, *lookback length* and *selection frequency* have been kept fixed to 500 and 250, respectively.

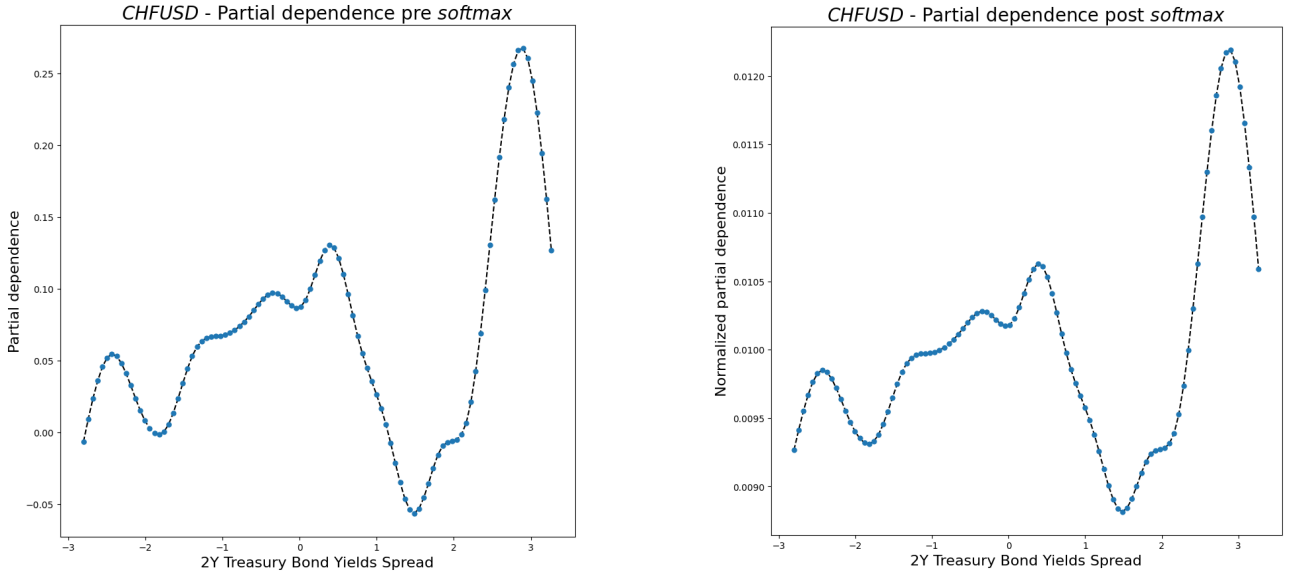


Figure 5: Example of the detected partial dependence between the normalized 2Y Treasury Bond Yields Spread of Switzerland and USA and the normalized *CHFUSD* spot rate pre and post *softmax* normalization.

5.4. Ensemble Selection: Copula + Generalized Additive Model

The purpose of this selection criterion is to combine the techniques presented in *Subsection 5.2* and *5.3* into an ensemble method that identifies tradable pairs through a voting system. At each selection date, the scores obtained from the two different methods are combined using a weighted average. The weights w_1 and $w_2 = 1 - w_1$, which determine the contribution of each method, have been fixed at three different values to test varying emphases on one method over the other. Specifically, the tested weight combinations are $[w_1, w_2] = [0.33, 0.67]$, $[0.5, 0.5]$, and $[0.67, 0.33]$. To ensure comparability and maintain the same magnitude across scores, they are individually normalized at each selection date using a MinMax scaler that maps the results to the range $[0, 1]$ (11):

$$pair\ score = w_1 \cdot pair\ score_{Copula}^{norm} + w_2 \cdot pair\ score_{GAM}^{norm} \quad (11)$$

where $pair\ score_{Copula}^{norm}$ and $pair\ score_{GAM}^{norm}$ are the scores obtained by the pair using respectively the copula and *GAM* score methodology and then normalized between 0 and 1.

This technique aims to evaluate whether integrating two distinct selection criteria, each grounded in structurally different methodologies, one based on reciprocal tradability and the simultaneous occurrence of extreme values (probabilistic perspective), and the other on the similarity of relationships with identical types of exogenous variables (financial perspective), enhances the identification of tradable pairs.

Not all the combinations have been tested. Recalling that both copula and *GAM* scores have been computed using three different levels of elasticity and tolerance (respectively, elasticity in the probability levels used to open the trades in the copula selection period and tolerance toward the *MSE* used to discard some rates when the *GAM* partial dependencies were analyzed), the ensemble pair score has been computed with different weights only between the original scores obtained with $[p_{low}, p_{up}, tolerance] = [0.05, 0.95, 0.10]$, $[0.125, 0.875, 0.25]$ and $[0.20, 0.80, 0.40]$.

5.5. Individual Rates Selection

The objective of this selection method is to identify exchange rates to be traded individually rather than within a pair framework, as described in the previous selection criteria. The individual rates are selected and viewed as tradable based on characteristics similar to those outlined in *Subsection 5.1*. However, unlike the previous methods, this approach evaluates the suitability of a single exchange rate for trading by analyzing its oscillatory and stable behavior around its average, capitalizing on its peaks and lows, and betting on its reversion to a stable regime. The underlying pattern is thus shifted from a pair of exchange rates to a single one, which must exhibit closely related features.

5.5.1 A Different Approach

The theoretical advantage of this approach is that, by avoiding the complexity of dealing with two assets whose relative behavior is represented by their spread, the trades become inherently safer. When a spread increases, it merely indicates that the value of one of the two exchange rates is increasing more than the other, without necessarily implying that one is rising while the other is declining. Although, as documented in the relevant section, the trading rules should ideally ensure a profit in both legs, this outcome is not guaranteed unless a selection criterion based on individual exchange rates is applied.

On the other side, moving away from a pairs trading approach increases the significance and impact of each individual trade. In a pairs trading framework, a wrong decision on one leg could be offset by a correct decision on the other. This safeguard is no longer present in this approach, which instead relies on the correctness of the forecast on an individual rate.

Additionally, this method is computationally far more efficient, as the necessary evaluations and score rankings are conducted directly on exchange rates rather than on all available spreads. This reduces the number of computations from 3960 to just 90. While this method operates on the periphery of the study, since it does not directly engage with the topic of pairs trading, it remains firmly situated within the broader framework of statistical arbitrage, seeking to exploit anomalous asset behaviors by betting on their return to a normal dynamic in a kind of outlier detection framework.

Another key aspect of this selection criterion is that it inherently limits exposure to individual exchange rates. While the three aforementioned methods allow an exchange rate to be eligible multiple times across different pairs, the logic here is structurally distinct, preventing overexposure to any single rate. This constraint can be interpreted as a risk management measure that enforces diversification within the trading portfolio. However, as a trade-off, it restricts the ability to capitalize on particularly promising rates behaviors.

5.5.2 Score Methodology

Delving further into the details, the selection process relies on ranking exchange rates based on certain metrics computed over the recent past preceding the selection date. These metrics are designed to identify assets that have historically exhibited profitable behavior. The primary characteristic sought is a dynamic that frequently reverts to its mean while also reaching sufficiently extreme values a reasonable number of times. This ensures that a significant number of trades can be opened and closed with consistent profits, leveraging both the extreme price deviations and the subsequent return to the mean.

Another crucial property is the absence of a strong and pronounced trend, which could otherwise complicate any forecasts regarding a reversion to the mean, as well as the formulation and execution of entry and exit rules. With the exception of the requirement for extreme values, which is primarily an empirical consideration to guarantee ample price movements and potential profits, these are the key characteristics of a mean-reverting process, as previously discussed in *Subsection 5.1*.

The use of mean crossings as a metric to assess the degree of mean reversion in a time series is inspired by the work of Donaldson and Mehra [35], where the authors introduce the *Average Crossing Time* as a measure of this property.

From a technical perspective, at each selection date, the score assigned to each exchange rate is computed by analyzing a variable time period in the past, preceding the selection date. Prior to these calculations, the time series is Z-normalized. The score itself is a weighted average of three components: the number of times the exchange rate has crossed its mean value (*mean crossings*), the number of times it has crossed its one-standard-deviation thresholds (*1σ crossings*), and the number of extreme values, defined as observations falling outside the two-standard-deviation range, it has recorded (*extreme values*).

Following this computation, the time series undergoes a stationarity test using the *Augmented Dickey-Fuller* test [14, 15]. If the resulting *p value* is lower than 0.10, the null hypothesis of a unit root in the time series can be rejected (under the alternative hypothesis of constant mean stationarity), leading to the automatic selection of the rate regardless of its previously computed score. While this does not constitute a definitive proof of stationarity, it at least ensures the absence of a unit root within the given confidence interval.

Finally, the set of selected rates is completed up to the predefined number of tradable rates by including those with the highest scores according to the earlier computations. The score attribution is synthetically reported in Equation (12).

$$\text{rate score} = \begin{cases} w_1 \cdot \text{mean crossings} + w_2 \cdot 1\sigma \text{ crossings} + w_3 \cdot \text{extreme values} & \text{if } p \geq 0.10 \\ \text{default selected} & \text{if } p < 0.10 \end{cases} \quad (12)$$

The weights used to compute the score are variable, allowing for the selection of exchange rates to emphasize different aspects of the desired behavior. Each set of weights has been tested in combination with the proposed

values of *lookback length* (500) and *selection frequency* (250). These weights correspond, respectively, to the number of mean crossings, the number of one standard deviation crossings, and the number of extreme values observed. The arrays of weights used are $w = [0.2, 0.5, 0.3]$; $w = [0.45, 0.1, 0.45]$ and $w = [0.4, 0.4, 0.2]$. For the discussion regarding the properties emphasized by each set of values, refer to *Subsection 5.1*.

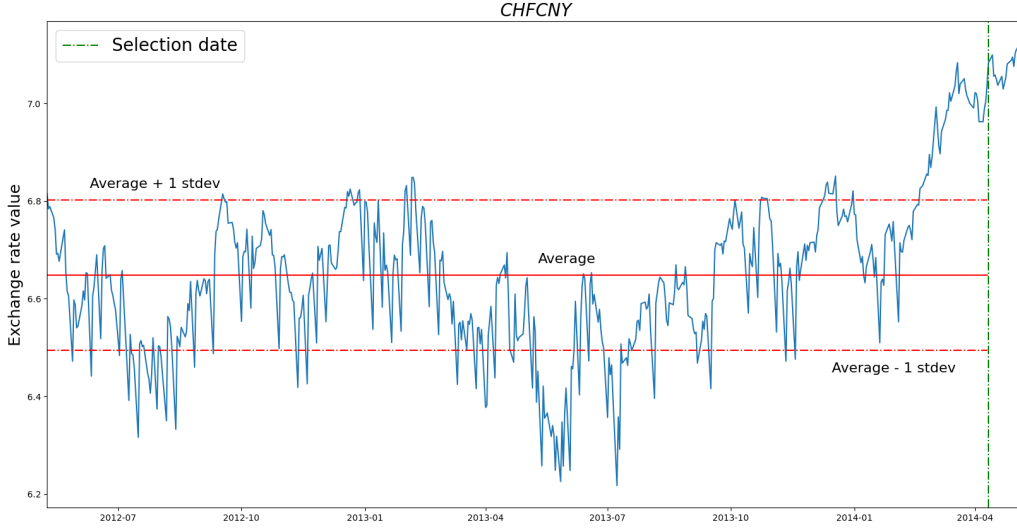


Figure 6: A frame of the behavior of the *CHFCNY* rate. Using an approach based solely on the ADF test, this rate would be discarded despite it clearly presents promising properties to be exploited, especially using long positions.

5.5.3 Role of Stationarity

It is crucial to emphasize that stationarity alone cannot serve as the sole criterion for asset selection. Firstly, relying exclusively on stationarity would significantly limit the number of exchange rates included in the tradable basket. Moreover, as illustrated in Figure 24 in *Appendix A*, the number of rates satisfying the stationarity condition fluctuates considerably over time. Additionally, the eligibility of a rate based on the stationarity criterion is highly sensitive to the specific time at which the test is conducted, as demonstrated in Figure 25 in *Appendix A*.

An approach solely based on the Augmented Dickey–Fuller (ADF) test would risk excluding promising exchange rates that exhibit frequent mean crossings and recurrent extreme values, characteristics that are well-suited for a buy-low/short-high trading strategy in subsequent periods. A clear example of a highly promising asset exhibiting numerous mean crossings and extreme values is the *CHFCNY* rate, as illustrated in Figure 6. In some cases, this restrictive approach would not even provide a sufficiently large set of tradable assets.

At the same time, stationarity is a desirable characteristic within this framework, as it ensures a more reliable foundation for mean reversion strategies. This allows for safer trading decisions based on the assumption that exchange rates exhibiting stationary behavior are more likely to revert to their mean over time. Consequently, such exchange rates are automatically prioritized in the selection process.

Notably, the percentage of exchange rates for which the ADF test rejects H_0 is not significantly different from the percentage of spreads for which the test rejects H_0 . This can be observed in Figure 26 in *Appendix A*. However, the key difference lies in the fact that, given the significantly larger number of spreads compared to individual rates, the pool of available tradable pairs remains sufficiently large within the scope of this study, whereas an overly restrictive criterion on individual rates could severely limit the number of eligible assets.

6. Trading Strategies

As previously mentioned, without further details, the trading strategies are recalibrated in conjunction with the selection of tradable (pairs of) exchange rates. This means that, at regular intervals, the trading strategies explored in this study are recomputed and adjusted based on the newly selected tradable assets. While this principle holds as a general rule, its specific implications vary depending on the algorithm considered and will be clarified in the subsequent sections.

In some cases, this recalibration simply involves applying the same established trading rules to a new basket of tradable assets. In other cases, it requires fitting a statistical or machine learning model to the most recent

available data in order to derive updated parameters that will govern trading decisions until the next recalibration. Generally, trading rules consist of a structured set of instructions that define entry and exit signals based on either statistical analysis, machine learning predictions, or the historical behavior of the selected assets.

It is also important to recall that these rules remain in effect for a period exceeding the actual interval between consecutive recalibrations. However, during the extended period, only the exit rules for closing existing trades are executable, ensuring that positions opened before the next recalibration can be managed effectively and closed just when needed and that new ones are initiated based just on the updated selection. If at the end of the overlapping period some trades are still open, they are closed.

Finally, it should be noted that, as with the selection methods, some algorithms are directly based on analyzing the behavior of the spread between the two rates, losing the individual properties and values of them, while others examine the joint movements of the two rates while consistently preserving their individual characteristics.

All considerations made in *Section 5* regarding the initial universe of tradable assets and the relevant exceptions remain valid and apply in this context as well. Consequently, each trading strategy will operate on the assets selected by the corresponding criteria, as outlined in Figure 11, and it inherits from them the considerations about the starting universe. Each trading strategy works on a basket of 30 exchange rates, namely 15 pairs.

6.1. Spread Behavior Trading

This strategy represents the most straightforward and intuitive approach among the pair trading strategies discussed. Its underlying mechanism is intuitively structured and directly aligns with the concepts introduced in *Subsection 5.1*, where exchange rate pairs were selected based on a mean-reverting spread exhibiting significant oscillations.

The scheme governing trade openings and the observed period follows the same structure described in Figure 1, while the employed *selection frequency* values are listed in *Subsection 5.1*. Throughout the trading period the spread of the chosen pairs is monitored on a daily basis. Recalling that a long (short) position on the pair (XY, ZT) corresponds to a long (short) position on the rate XY , which involves purchasing (selling) one unit of currency X in exchange for the equivalent amount of Y , and a short (long) position on the rate ZT , which results in selling (purchasing) one unit of currency Z in exchange for the equivalent amount of T , the rules governing trade openings, which can take place only before the succeeding selection date, are as follows:

- if $z_t < \mu - \delta_{open} \cdot \sigma$ and $z_t > \mu - \delta_{stoploss} \cdot \sigma$, open a long position on z
- if $z_t > \mu + \delta_{open} \cdot \sigma$ and $z_t < \mu + \delta_{stoploss} \cdot \sigma$, open a short position on z

where z_t is the value of the spread at time t and μ and σ are respectively the mean value and the standard deviation of the spread computed during the selection lookback period (whose used values are listed in *5.1*). In this manner, robust and consistent values and rules are maintained throughout the entire trading period, avoiding continuous updates. This approach should not pose any significant challenges or issues, as the spreads are selected based on their stationarity. The rules to ordinarily close the trades are the following:

- if $z_t > \mu - \delta_{close} \cdot \sigma$, close the long position on z
- if $z_t < \mu + \delta_{close} \cdot \sigma$, close the short position on z

Furthermore the trades can be closed prematurely in case of concerning deviations from the expected reversion towards the mean if there is significant exposure to the pair, serving as a stop-loss strategy:

- if $z_t < \mu - \delta_{stoploss} \cdot \sigma$ and $openedtrades > critical\ exposure$, close the long positions on z
- if $z_t > \mu + \delta_{stoploss} \cdot \sigma$ and $openedtrades > critical\ exposure$, close the short position on z

In case a persistent but temporary deviation from the average is misinterpreted as a divergence, the strategy will still be able to recover some of the losses faced due to a premature close entering again the trade as soon as the value of the spread returns to less dangerous values. The *and* condition in the stop-loss logic ensures that if the spread is temporarily diverging but few positions have been opened, the strategy is going to wait for further information without opening new trades but without closing existing ones.

During the overlapping period trades can just be closed and, at the end of it, all the trades that are still open are closed. An intuitive representation of the strategy is reported in Figure 7.

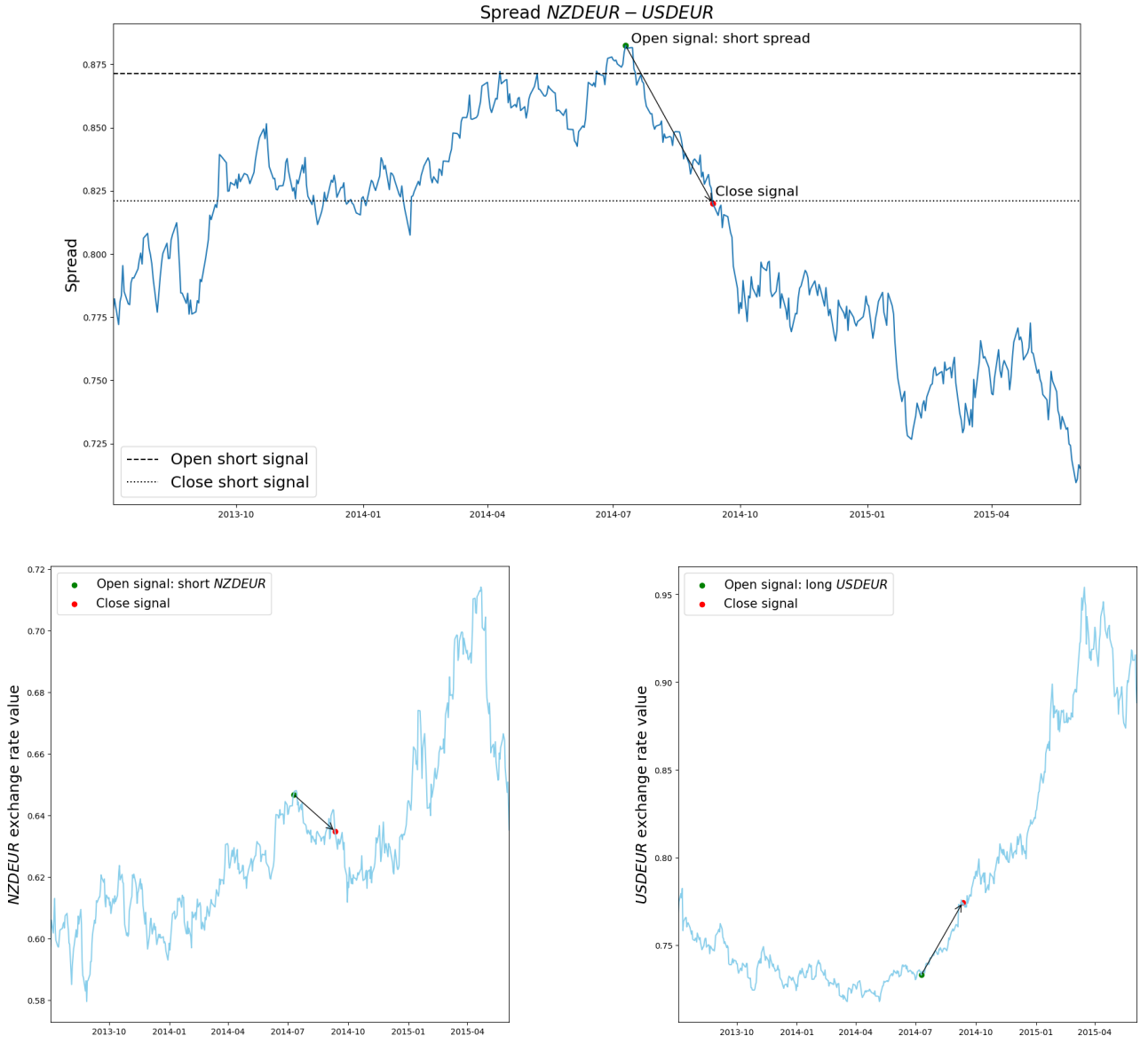


Figure 7: A frame of the strategy applied to the *NZDEUR* - *USDEUR* spread.

δ_{open} , δ_{close} and $\delta_{stoploss}$ represent the number of standard deviations from the average needed to respectively open, close and close for stop loss reasons the opened trades. Despite the values can be arbitrary chosen and this study will analyze different combinations, there are some constraints and implications linking these values. Obviously $\delta_{open} > \delta_{close}$, otherwise the strategy would not be profitable. The greater the difference $\delta_{open} - \delta_{close}$, the higher the potential profits to be realized; however, this would also result in fewer opened trades and a reduced number of standard trade closures. A smaller δ_{close} implies that the spread must be closer to the average for the trades to be closed. In this sense, δ_{close} serves as an indicator of the expected degree of mean reversion of the spread: the smaller its value, the stronger the expectation that the ratio will approach the average and complete the mean reversion process. In extreme cases, negative values would indicate that trades should only be closed once the mean reversion is fully completed and the spread has crossed the average. However, this is generally unnecessary, and not necessarily advantageous for the strategy, since substantial profits can still be achieved through partial reversion toward the mean.

Furthermore, $\delta_{stoploss} > \delta_{open}$, meaning that a suspicious behavior of the spread is identified only when it continues to reach increasingly extreme values relative to those used to open the trades.

Finally, the role of *critical exposure* is to mitigate potential losses by preventing excessive exposure to the pair. Firstly, opening too many trades could lead to severe consequences if the spread fails to revert to less extreme levels. Secondly, an excessively high exposure serves as a clear indication that the spread is struggling to revert to the mean. In such cases, prematurely closing the trades may still yield a marginal profit from the individual legs of the spread. A typical scenario involves the spread stabilizing and remaining nearly constant at levels between the opening values and the target closing values.

As previously mentioned, the selection frequencies used to determine the tradable pairs align with those specified in *Subsection 5.1* (250, 200, and 150 days). The corresponding trading periods, during which the spreads are monitored and at the end of which all trades are closed, span 300, 250, and 200 days, respectively. Meanwhile, the lookback period used to compute μ and σ is set to 500, 400, and 300 days, depending on the configuration. Each setup reflects a different level of responsiveness in interpreting new data. For each configuration, three distinct combinations of opening and closing thresholds have been tested: $[\delta_{open}, \delta_{close}] = [1.8, 0.7]$, $[1.4, 0.4]$, and $[1, 0.1]$. These configurations reflect different approaches to the flexibility in opening trades, ranging from more conservative to more aggressive entry conditions. Across all these configurations, the critical exposure level has been fixed at 100, while the stop-loss threshold has been set at $\delta_{stop,loss} = 3$.

6.2. Markov Switching Model Trading

This approach constitutes a significant advancement compared to the methodology introduced in *Subsection 6.1*, even though the fundamental rules governing the generation of open/close signals remain essentially the same and continue to leverage the mean reversion properties of the spreads. Consequently, in this case as well, the spreads are selected using the same methodology discussed in *Subsection 5.1*.

Pairs traders and the related literature [49] suggest that the mean of the price ratio between equity stocks tends to shift between different levels, often rendering traditional technical trading approaches ineffective in identifying profit opportunities. Similarly, as demonstrated by the example discussed in this section, currency exchange rates and their ratios can also experience significant regime shifts triggered by monetary policy decisions or major geopolitical events. Such shifts frequently lead to a pronounced change in market dynamics, resulting in the establishment of a new equilibrium around which the ratio resumes oscillating and mean reverting [17, 79]. The emergence of a new average level may be quickly reverted, with a rapid return to the original regime, or it may persist for years, potentially lasting until the end of the observed period. Financial markets, and exchange rates in particular, are highly susceptible to these types of sudden structural breaks [31]. Moreover, in addition to shifts in the average value, the amplitude of oscillations may also undergo significant changes.

Under this set of considerations, applying the same intuitive rules and logic outlined in *Subsection 6.1* becomes particularly challenging due to the difficulty in distinguishing between temporary and persistent spread deviations. The former are expected to revert, enabling the realization of profits, whereas the latter, if misinterpreted, can result in catastrophic losses. The estimation of the standard deviation of the spread, a critical component of the strategy, must also be conducted carefully, as it varies when the spread transitions to a different level. What was previously interpreted as a divergence from the average ratio level may now represent a transitional phase leading to the establishment of a new regime. This behavior poses significant challenges for the previously introduced approach but can also present substantial profit opportunities if the appropriate algorithm is employed.

A *Markov Regime-Switching Model (MSM)*, initially introduced in the seminal work of Goldfeld and Quandt [71] and later popularized by Hamilton [30], incorporating both switching means and switching variances, is capable of identifying the distinct phases of the ratio time series and to infer the probability of the out of sample ratio values being in each of them. This enables the application of different opening and closing rules based on the detected regime.

The key assumption of the model is that the current regime variable s_t , indicating in which regime the time series is currently living, is determined by a latent and unobservable variable. Thus, the inference of the predominant regime is based only on calculated state probabilities. The number of supposed regimes, as well as the order of the Markov Switching Model, can be arbitrary decided, but usually a two-state first-order model for the regime time series s_t is sufficient and the most widely used combination also in the literature.

The order of the model, in a pure Markovian flavor, is the number of past states involved in the computation of the probabilistic function governing the following one, essentially defining the length of the past dependence of the current regime. The main role of the model is the probabilistic estimation of the new state of the ratio in order to adjust the trading rules and to properly make the right decision each day. Misclassifying a state could lead, for instance, to interpreting the ratio as being above the average consequently opening a short position on it, while it is transitioning toward a new higher regime, implying that a long position should be opened.

A clear example of the consequences of monetary policy decisions on the creation of new stable regimes far from the prior one can be found in what happened to the *EURCHF* exchange rate in 2015. At that time, the Swiss National Bank ended its three-year policy of capping the *EURCHF* exchange rate at 1.20 [74]. This abrupt policy shift led to a significant departure from the previous regime, driving the rate to new stability levels after a steep adjustment. A method based solely on the average behavior of the spread between two exchange rates would likely fail to capture such a shift, particularly if this change occurred near the end of the formation period with only a small number of observations, which would not adequately influence the average value of the spread of the training period on which later trading decisions would be made. In contrast, the

Markov Switching Model is capable of identifying regime shifts and classifying subsequent observations as part of a new dynamic, distinguishing them from the prevailing pattern during the training period. This ability to detect multiple stability regimes makes the Markov Switching Model especially useful for accurately taking into account such dynamics and to refine trading decisions. Figure 8 and Figure 9 clearly highlight the ability of this model to detect a new and more volatile regime around lower values.

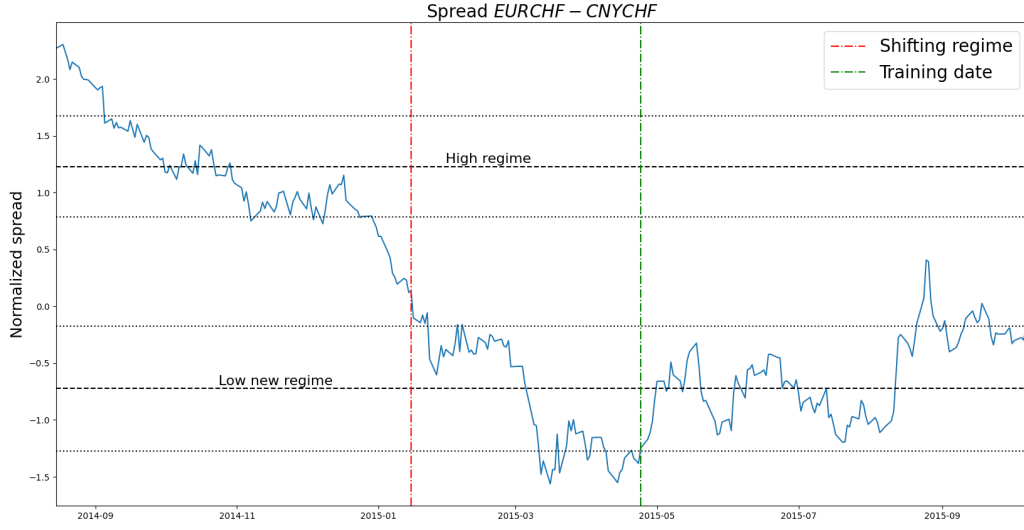


Figure 8: Behavior of the *EURCHF* - *CNYCHF* spread and detected regimes. Normalized values.

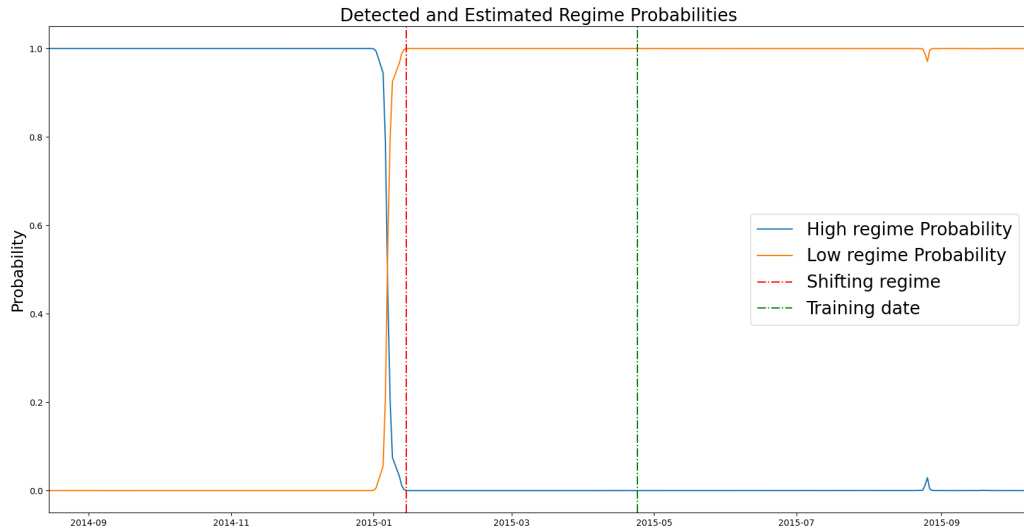


Figure 9: High and low regime in the training and trading phases.

Examining the algorithm in greater detail, Figure 1 largely illustrates the logic used to manage consecutive trading and selection periods. At each selection date, the basket of pairs identified in *Subsection 5.1* is considered for the entire subsequent trading window. Training is conducted over the full lookback period used for the selection, and the parameters derived from the model remain fixed throughout the trading period. The values employed are the same as those previously mentioned in *Subsection 5.1*. The same principle applies to the selection frequency, which also determines the frequency of model training. The trading period is slightly extended to account for the overlapping period.

The decision not to exclude the most recent observations from the training period is questionable. Empirical evidence suggests that in the short term the estimated probabilities of each state are sensitive to the specific moment at which the model is fitted. For this reason, one considered approach was the introduction of an offset value to filter out the latest observations, thereby enhancing the robustness of probability estimation during the trading period. However, it is crucial to emphasize that this sensitivity is precisely what makes the model so valuable in this context. Without it, the model would be unable to detect small regime shifts or integrate the most recent observations into its procedures, thereby improving the inference of out-of-sample

state probabilities.

A training period that is too long would incorporate an excessive amount of data, potentially learning from outdated macroeconomic, statistical and fiscal regimes, assigning insufficient weight to the most recent information and likely resulting in flat probability estimates, preventing the model from effectively distinguishing between different states and assigning all the new observations to the most recent observed regime. On the other side, a longer training period could capture cyclical and seasonal patterns, enhancing the model's ability to identify recurring behaviors. Conversely, a training period that is too short would overlook a significant amount of historical data.

Empirically it has been found that the train period should be more or less 1.5 times the length of the trade period where the fitted model is used. For this reason, the values used for the lookback period length during the selection of the spreads and the selection frequency (both defined in *Subsection 5.1*), reminding that the actual trading period is longer, are still reasonable and can be used for the train of the algorithm and subsequent trade.

At each selection date the model is run on the Z-normalized spread of the already discussed length: the order is fixed to one and the number of regimes to two which will correspond to an *high* and *low* regime. Thanks to the *Broyden-Fletcher-Goldfarb-Shanno* [9, 21, 24] algorithm, which has been empirically chosen as the most efficient one thanks to its quadratic complexity, the fitted model retrieves the parameters used in what follows for the trading rules and for the modeling of the underlying state's Markov chain. In particular it is able to retrieve the average values of the two regimes μ_{high} and μ_{low} and the respective standard deviations σ_{high} and σ_{low} .

Given each new value of the spread during the trading phase, after Z-normalizing it, it is necessary to compute the probabilities of it belonging to the *high* or *low* state, based on its current value and the preceding state. During training, the model infers the transition matrix that governs the transition from one state to the next. For each new observation, the Gaussian likelihoods under the two existing regimes are calculated, and, using the transition matrix and the filtered probabilities of the previous observation, the filtered probabilities for the current observation are derived.

It is important to note that, during the trading window, only the filtered probabilities can be computed, based on the observations available up to the current point. In contrast, the training procedure computes the smoothed probabilities on the train values: those probabilities are computed from all the available observations in the train values. This approach ensures better model calibration and more accurate results for the transition matrix.

Once for the new value of the spread the filtered probabilities $\mathbb{P}(s_t = high|s_{t-1})$ and $\mathbb{P}(s_t = low|s_{t-1})$ are computed a decision can be taken. The spread is logically supposed to be in the state that gives the highest filtered probability.

If $\mathbb{P}(s_t = low|s_{t-1}) > \mathbb{P}(s_t = high|s_{t-1})$ ($s_t = low$),

- if $z_t < \mu_{low} - \delta_{open} \cdot \sigma_{low}$, open a long position on z
- if $z_t > \mu_{low} + \delta_{open} \cdot \sigma_{low}$ and $\mathbb{P}(s_t = low|s_{t-1}) > \rho$, open a short position on z

If $\mathbb{P}(s_t = low|s_{t-1}) < \mathbb{P}(s_t = high|s_{t-1})$ ($s_t = high$),

- if $z_t > \mu_{high} + \delta_{open} \cdot \sigma_{high}$, open a short position on z
- if $z_t < \mu_{high} - \delta_{open} \cdot \sigma_{high}$ and $\mathbb{P}(s_t = high|s_{t-1}) > \rho$, open a long position on z

The addition of checks about the probability of being in a given state prevents from changing the trading positions the wrong way around. If the spread is currently living in the middle between the two detected regimes, the model is able to estimate in which state it is more likely to be and as a consequence which set of rules should be followed and with which parameters.

The closing rules are the same already presented in *Subsection 6.1*: when the spread starts to revert its behavior and to approach again the average the trades are regularly closed. This procedure is ruled by δ_{close} . Again, the trades can be closed prematurely in case of concerning deviations from the expected reversion towards the mean (governed by $\delta_{stoploss}$) in presence of significant exposure to the pair (*critical exposure*), serving as a stop-loss strategy. No trade can be opened if the spread is deep in extreme territory ($\delta_{stoploss}$).

After the next selection date, trades can just be closed, giving them the time to complete the expected behavior. At the end of the overlapping period, all the trades that are still open are definitely closed.

For what regards the relation between δ_{open} , δ_{close} and $\delta_{stoploss}$ the same considerations already made still hold. ρ is a probability threshold that assesses the required confidence in the state forecast: the higher it is, the surer the state must be in that regime in order to be considered safe to open a trade on the spread when the spread is between the average values of the two detected regimes.

Although the selection method described in *Subsection 5.1* has been employed, only a subset of the possible configurations has been tested. Specifically, the analysis focuses on configurations that utilize a training period of 500 samples, a selection frequency of 250 observations, and a corresponding trading period of 300

days. The set of δ thresholds for opening and closing trades align with those specified in *Subsection 6.1* and together with the parameter ρ serve as a proxy for risk appetite measured as the willingness to initiate new trades: $[\delta_{open}, \delta_{close}, \rho] = [1.8, 0.7, 0.8]$, $[1.4, 0.4, 0.7]$, and $[1, 0.1, 0.6]$. Across all these configurations, the critical exposure level has been fixed at 100, while the stop-loss threshold has been set at $\delta_{stop,loss} = 3$.

6.3. Copula Trading

As introduced in *Subsection 5.2*, the trading strategy based on copulas is inherently linked to the selection criterion utilizing the same tool. Additionally, the criterion outlined in *Subsection 5.3* has also been applied to select the pairs to be traded using this algorithm (*Subsection 5.4*). Since this algorithm relies exclusively on the analysis of simultaneous extreme values, incorporating a selection criterion that evaluates the similarities between time series based on their relationships with the same type of external variables can be beneficial.

Exchange rates exhibiting similar linkages to the same exogenous variables are more likely to revert together to their common behavior after experiencing extreme values. Consequently, integrating this technique into a trading framework that primarily interprets time series from a probabilistic perspective can enhance overall performance. Moreover, given the computational effort that would be required to fit copulas across all possible pairs during the trading phase, preselecting potentially more profitable pairs can improve efficiency and effectiveness more here than in other presented algorithms.

The lookback period used for pair selection is used to fit four copula models (*Gumbel, Frank, Clayton, Joe*) for each pair of exchange rates within the selected basket. The model that achieves the best *AIC* [4] score is then chosen for the subsequent trading period. If for none of them the fit is possible, then the chosen type of copula is the best fit of the selection phase and it is trained on the first section of the training period (using the same methodology of the selection phase). During the trading period, the fitted copula is employed to evaluate the probabilities given in Equation (13):

$$\mathbb{P}(U_1 \leq u_1 | U_2 = u_2) \text{ and } \mathbb{P}(U_2 \leq u_2 | U_1 = u_1) \quad (13)$$

for each new pair of observed quantiles with the usual formalism introduced in *Subsection 5.2*. Based on the observed values, a long or short position is taken on the pair. If exchange rate 1 is considered undervalued and exchange rate 2 overvalued (that is $\mathbb{P}(U_1 \leq u_1 | U_2 = u_2) \leq p_{low}$ and $\mathbb{P}(U_2 \leq u_2 | U_1 = u_1) \geq p_{up}$, where $p_{low} + p_{up} = 1$ and $p_{up} - p_{low}$ is sufficiently large), then a long position on the pair is opened. This means that rate 1 is purchased while rate 2 is shorted, and vice versa when the opposite condition holds. The values of p_{up} and p_{low} are treated as variables, as it has been done in the selection, and tested with different values.

These trades are subsequently closed when the observed values revert to a range of quantiles that can be considered "usual" from a probabilistic perspective. Typically, this occurs when $\mathbb{P}(U_1 \leq u_1 | U_2 = u_2)$ and $\mathbb{P}(U_2 \leq u_2 | U_1 = u_1)$ fall again within the 40%-60% range. The logic followed to open and close trades is exemplified in Figure 3.

Some considerations regarding the logic behind opening and closing trades are necessary. First, a size position threshold has been established in conjunction with all other hyperparameters to ensure that, once this limit is reached, the exposure to a given pair cannot increase further. This serves as a risk management measure to mitigate potential losses: if exposure to a pair reaches significant levels, meaning that, after several days, the pair's values have not reverted to the close-trade probability range, there is a risk that the values are diverging rather than reverting to their standard regime.

Another important consideration concerns the boolean logic used to open and close trades. The activation of an opening or closing procedure depends on conditions involving both exchange rates, and these conditions can be combined using either an *and* or *or* logic for each procedure. In this study, the *and* logic was chosen for trade openings: a position is initiated only if both exchange rates are classified as probabilistically extreme.

For the closing procedure, both logical approaches have been tested. The *and* logic is stricter, leading to fewer trades being closed during standard time (recalling that any remaining open trades are automatically closed at the end of the trading period), but it ensures a profit on both legs of the pair. Conversely, the *or* logic allows trades to be closed more frequently, guaranteeing a profit on at least one leg, though the other leg may potentially result in a loss.

Finally a last consideration regarding the overlapping period. Given a selection frequency fixed to 250, the length of the trading period has been fixed to 300 to allow for 50 days where existing trades can still be closed regularly but all the new operations are started with the new set of pairs.

The combinations of probability thresholds used to open the trades are inherited by the employed selection method ($[p_{low}, p_{up}] = [0.05, 0.95]$, $[0.125, 0.875]$ and $[0.20, 0.80]$) and each configuration is tested with both the *and* and *or* closing logic. The limit exposure has been fixed to 100.

6.4. Logistic Regression Trading (individual rates)

This machine learning-driven methodology seeks to leverage the mean-reverting characteristics of the selected exchange rates, identified through the techniques outlined in *Subsection 5.5*. This approach exploits their propensity to revert from extreme positions within a reasonable timeframe, generating profits from value oscillations and betting on the reversion of the behavior rather than on the continuation of a diverging trend.

The main tool of this approach is the exploitation of a *Multinomial Logistic Regression*, a generalization of the more common *Logistic Regression* already used in a related but fundamentally different flavor in [58]. An approach based on this type of machine learning tool recasts the exploitation of mean reversion properties into a multi class classification problem: each day, a decision must be made, whether to go long or go short on one of the preselected rates, or remain inactive, based on the predicted future movements of these rates, which have been purposely chosen for their potential to revert to their current values.

The management of the time periods is slightly different and more complex than what is reported in Figure 1. Each time a new trading window is opened, the basket of tradable assets is considered. Three quantities need to be fixed in this case: the *max feature lookahead*, the *forecast length* and the *train length*, as it is possible to see from Figure 10. The frequency used to reselect the rates and train the model is inherited by the linked selection method 5.5 and is equal to the trading period, meaning that no overlapping period is considered in this method. Reasons will be made clearer later.

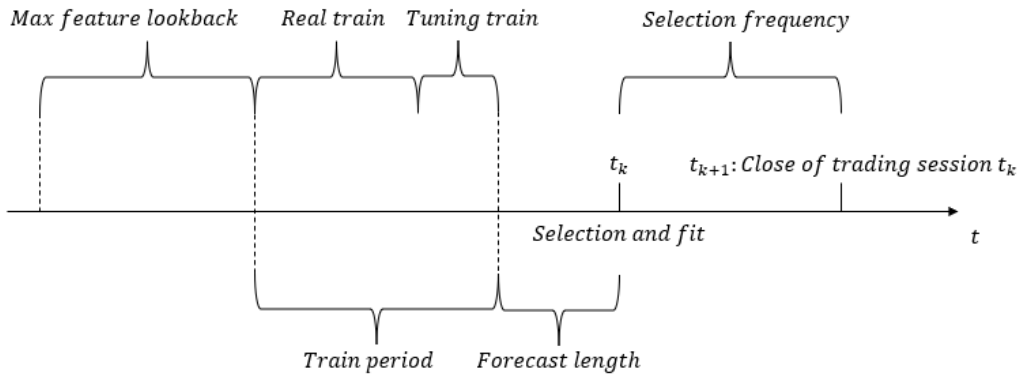


Figure 10: Approach to subsequent trading fitting and selections and involved time periods.

The reported time periods will assume a real meaning while the applied procedure is explained. First of all the train dataset is computed for each exchange rate. Features and response class variables are obtained for each daily observation in the *train period*. For each observation of the value of the rate under analysis, momentum and maximum drawdown, both of them dimensionless quantities, are computed for different time lengths H as features as described in Equation (14):

$$mom_{t,XY,H} = \frac{XY_t}{XY_{t-H}} - 1 \quad dd_{t,XY,H} = \frac{XY_t}{\max_{i=1,\dots,H}(XY_{t-i})} - 1 \quad (14)$$

H has been fixed to 30, 60, 90, 120, 180, 270, 300, 360 for momentum and 15, 60, 90, 120 for the maximum drawdown, resulting in a total of 12 original features. The lengths of these time frames are based on common practice and broadly match what would be expected to be pertinent historical horizons on which to evaluate short and long-term momenta, losses and stop-loss triggers strategies. The maximum H value is 360 and it is equal to *max feature lookahead*, representing the number of past observations outside the *train period* needed to compute the features.

Using momentum over multiple time frames matches the classic approach of using it as a useful information for investment decisions, while adding drawdown details is conceptually equivalent to enabling the use of stop losses signals in a classic approach where stop losses signals are typically triggered after a drawdown threshold is violated. The intent therefore is that by providing momentum and drawdown information over multiple time frames, the algorithm will learn which of these features are relatively more important than others as market conditions evolve and stress each one accordingly to its role in future prediction. During the training procedure, the model learns which information are the most useful ones in order to predict the class variable.

For each observation in the *train period*, the class to which it belongs is computed using a forward looking perspective (*forecast length* fixed to 100 days). This explains the early expiration of the *train period* before

the fitting date: the computation of the class of the last train observation needs the values of the rates up to the day before the fitting date.

The 100-day length was chosen as a trade-off: a longer forecast window would require using outdated data, while a shorter one would not provide enough time for the exchange rate to revert.

Class 1 is assigned to observations under the average yearly value which are going to grow more than a given threshold within the *forecast length* future, meaning that a long position on the rate should be opened; oppositely, class -1 is assigned to observations expected to decrease more than this threshold within this fixed amount of time and that are above the yearly average, suggesting the opening of a short position. The average value is computed on a yearly moving window with forward look: this will allow the algorithm to learn relative positioning of each value. Class 0 is assigned to all the other observations. The purpose of the class labels is to measure the positioning of the value of the rate in a mean-reversion perspective with respect to the future movements. The gain/drop threshold has been fixed to 5% for reasons made clearer in *Subsection 7.4*.

It is important to note that in the fitting instant all the needed information are available and no anticipation bias is present: this precisely justifies the inclusion of a gap period between the conclusion of the training dataset and the beginning of actual operations.

This approach aims to forecast the behavior of the rate under analysis after *forecast lengths* days in the future. This fixed investment horizon length, together with the style of early closing and risk management applied to it and explained in what follows, is justified by the similar approach used by M. Bock and R. Mestel in their work [49]. This also explains the absence of the overlapping period that was instead extremely helpful in the other trading strategies: in this case, the trades are closed after a fixed amount of days, the *forecast length*, which corresponds precisely to the period used to compute the class variables required for the train of the Multinomial Logistic Regression, and so the time period required for the rate to ideally realize the forecast gain or drop.

Delving into the details of the trading algorithm, the Logistic Regression is trained for each rate at the selection date using the train dataset made of the momentum and drawdown features computed for different time horizons (X) and the classes (y) computed using the behavior of the rate a fixed amount of days later, which is equal to the trade signal suggesting the position that should be taken on the rate. The train dataset is further divided in a first part (75% of the period) used to actually train the algorithm while the other one is used for the tuning of the hyperparameters. The hyperparameters involved are the polynomial degree and the regularization parameter λ .

In an attempt to identify non-obvious, complex and non linear patterns between the features and the forecast class, polynomial combinations up to a specified polynomial degree of momenta and drawdowns are created and added as new features. This can enhance the ability of the algorithm to detect the real drivers of y trying to exploit machine learning ability to learn from and adapt to changing market conditions using more numerous and more complex combinations of features than is practically feasible in a classic approach. If future performance is assumed to be a function of present values and of their difference with respect to the past, the new polynomial features will correspond to some of the lower-level polynomial terms in the Taylor series expansion of that function.

In this way the total number of features can rise quickly as the specified degree of polynomials increases. This in turn could lead to overfitting the training set data and result in poor strategy performance (in particular when the number of features exceeds the number of observations for high degree of the polynomial combinations and for short training dataset). In order to address this issue, the regularization parameter λ , which enables the algorithm to minimize the relative influence of some features, is included. This can avoid overfitting and stress the importance of some features above others. Consequently, this technique can be designed to use far more features than a classic counterpart without necessarily overfitting data. This further justifies the use of a tuning dataset.

Different combinations of λ (0.1, 1, 10, 100, 1000, 10000) (skewed towards higher values because a stronger regularization to prevent overfitting is claimed) and of the polynomial degree (2, 3, 4) are used to train the Multinomial Logistic Regression on the first part of the train dataset (which is, after the polynomial transformation, demeaned and rescaled between -1 and 1 to prevent unintended bias in the resulting weights) and to derive the vector of weights θ by minimizing the cost function Equation (15):

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k y_j^{(i)} \log \hat{y}_j^{(i)} + \frac{\lambda}{2m} \sum_{j=1}^k \sum_{l=1}^n \theta_{jl}^2 \quad (15)$$

where m is the number of observations used, n is the number of features, k is the number of classes (3 in this framework), $y_j^{(i)}$ is 1 if the correct class for observation i is j and 0 otherwise and $\log \hat{y}_j^{(i)}$ is the log of the predicted probability for the correct class. Once the set of weights for each combination of hyperparameters is obtained on the first fraction of training dataset, it is tested on the tuning dataset. The combination of hyperparameters able to reach the best performance in terms of a customized $F\beta$ score function on the second

fraction of the *train period* is the chosen one for the trading session of the rate under analysis. $F\beta$ score use is particularly relevant when dealing with imbalanced class distribution, that is what would be expected in this context, where several days are likely to be classified as class 0.

The choice to evaluate the performance according to a customized version of the $F\beta$ score aims to give different importance to misclassification errors according to their financial impact. Specifically, the objective is to assign the greatest penalty to the -1 and 1 classes when they are incorrectly classified as 1 and -1, respectively. Taking a long (short) position when the value will actually drop (rise) represents the most detrimental scenario. A slightly less severe punishment has been assigned to 0 observations classified as -1 or 1. Finally, the least severe weight is assigned to scenarios where -1 and 1 observations are classified as 0: this does not result in any direct loss but merely leads to an opportunity cost due to the missed potential. The score formula is reported in Equation (16) and (17):

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{(\beta^2 \cdot \text{Precision}) + \text{Recall}} \quad (16)$$

where $\beta = 1$ to give balanced focus between Precision and Recall.

$$\text{Precision} = \frac{\sum_i w_i \cdot TP_i}{\sum_i w_i \cdot (TP_i + FP_i)} \quad \text{Recall} = \frac{\sum_i w_i \cdot TP_i}{\sum_i w_i \cdot (TP_i + FN_i)} \quad (17)$$

where $i = -1, 0, 1$ and w_i is the i th row of the weight matrix (18):

$$W = \begin{pmatrix} 1 & 3 & 7 \\ 5 & 1 & 5 \\ 7 & 3 & 1 \end{pmatrix} \quad (18)$$

built such that the aforementioned misclassification errors are more or less penalized.

Once that the best couple of hyperparameters has been found, the model is trained again using the whole train dataset polynomial transformed, demeaned and rescaled between -1 and 1. Then, for each new observed value of the rate in the subsequent trading phase the new set of features x_i is computed using momenta and drawdowns and the trained model is applied to it. In case a -1/1 label is forecast, a short/long position is opened on the rate and closed after *forecast length* days. The opened position, anyway, is daily checked: if the forecast increase/drop threshold is reached before the natural closing, the trade is closed prematurely. Furthermore, if a rate on which a long (short) position is opened is facing a certain drop (increase) the operation is closed in advance to limit the losses, expecting a divergent behavior of the rate or a shift to another regime instead of a reversion to the mean. The emergency threshold used to close prematurely the trade is higher than the expected profit used for the train: this is because moderate and temporarily losses can occur during a long horizon investment strategy like this one.

The frequency at which the model is refitted is inherited by the values used for rate selection 5.5 and it is equal to 250 days. The training period is fixed at 400 observations, 300 for the real training and 100 for the tuning. This decision ensures a sufficiently large dataset for model training, particularly given that a portion of the data is set aside for hyperparameter tuning, while avoiding excessive reliance on old values due to the presence of the *forecast period*.

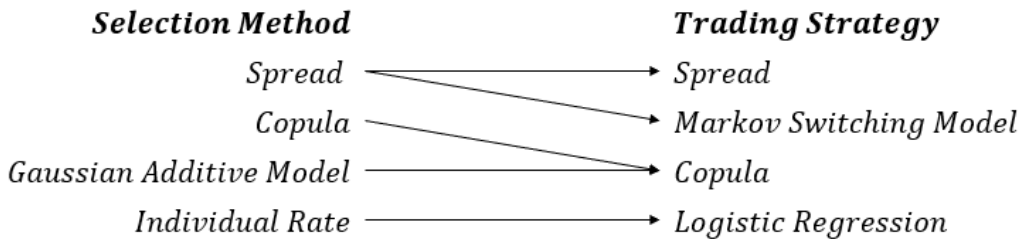


Figure 11: Selection methods and trading strategies where they have been employed.

7. Discussion of the Results

Given the large number of combinations of strategies and selection methods examined and the realization of the profits in ten different currencies, for the sake of clarity only the most important results will be reported. Furthermore, it is possible to identify some regularities in the performance and risk profile of each strategy

according to the set of hyperparameters used both during the selection procedure and during the trading period.

The evaluation metrics employed to assess the performance of the algorithms cover both return-based measures and risk quantification indicators. Each strategy, fed by a specific selection method, ultimately generates a total profit in each involved currency, referred to as *Total Currency Profit* in the tables. Additionally, the number of executed trades required to achieve this profit is reported, alongside the average return per trade for each currency.

Risk-related measures include the annualized volatility, the highest and lowest recorded returns, and the percentage of negative returns relative to the total trades. Moreover, the *Max Absolute Drawdown* is computed as the largest absolute loss incurred after a peak. This measure is preferred over the more conventional *Max Drawdown*, whose interpretation can be problematic in this context. Specifically, since the portfolio values initially start from zero, an early negative profit would drive the relative maximum drawdown towards negative infinity, making it unsuitable for meaningful comparison.

A key metric for evaluating the robustness of currency-specific profits is the relative impact of drawdowns. This metric, referred to as *Drawdown Impact*, is calculated as the average of the ratios between the *Max Absolute Drawdown* for each currency and its corresponding *Total Currency Profit*.

Finally, the *Drawdown Duration* is defined as the number of days required for the *PnL* to go from the highest peak to the lowest point realizing the *Max Absolute Drawdown*.

The analysis of the results from the performance standpoint is dual: from the return perspective, the computation of the average return per trade serves to quantify the quality of the executed trades and their effectiveness in generating a positive impact on the portfolio. This measure evaluates the effectiveness of selecting optimal entry and exit points.

From a profit-oriented perspective, the *Total Profit* quantifies the strategy’s capacity to generate an absolute gain, reflecting its ability to close the considered time window with a positive portfolio growth. This metric assesses the strategy’s potential to identify profitable opportunities and capitalize on them, emphasizing the strategy’s overall performance rather than the precision of individual trade executions. It highlights the strategy’s proficiency in making correct trade decisions more frequently than not.

The *Total Profit* (often referred to as *Profit USD* in the charts) is derived from the aggregated profits realized across the ten considered currencies, referred to as the *Total Currency Profits*. For simplicity, the profits from each portfolio are converted into *USD* using the most recent available exchange rates, with these values being summed to produce a singular metric representing the strategy’s quality from a profit perspective. However, the quality of this metric is debatable due to its sensitivity to the exchange rates on December 29, 2023.

It is important to recognize that the strategy yielding the highest average return does not necessarily emerge as the most profitable when assessed from the total profit perspective.

This section is structured as follows: first, an in-depth analysis of each individual strategy is conducted, with particular emphasis on the impact of hyperparameters on return, profit and risk metrics. Next, a comparative study is performed across different strategies, taking into account both the adopted hyperparameter values and the selection methods. Throughout the analysis, any problematic aspects and limitations are also identified and discussed.

7.1. Spread Behavior Results

The optimal strategy in terms of return generates an average return of 0.84% for each completed trade. Additionally, this strategy guarantees a positive profit across all currencies, with negative returns occurring in only approximately 30% of cases, indicating that the strategy yields a profitable outcome in 70% of instances. For nearly half of the involved currencies, the minimum return in absolute terms is smaller than the maximum return, suggesting the presence of a longer right tail in the return distribution. In certain currencies, such as *CHF*, the impact of the absolute drawdown is substantial, resulting in a total profit roughly equal in magnitude to the drawdown value. Conversely, for currencies like *USD* and *CAD*, the drawdown constitutes only about 10% of the total profit.

The aforementioned strategy, which yields the best results, is the one that rebalances the basket of tradable pairs most frequently (every 150 days). It adopts a highly conservative weight vector during the score computation phase ($w = [0.4, 0.4, 0.2]$) and employs the parameters $[\delta_{open}, \delta_{close}] = [1.8, 0.7]$ as trading rules.

Overall, within this framework, distinct patterns emerge: the top-performing strategies, in terms of return per trade, are those that frequently reselect the asset basket. This indicates that the speed at which new information is incorporated plays a crucial role. Additionally, the choice of the weight vector significantly impacts the quality of the results. The most favorable outcomes across various trading rules are typically achieved when the most conservative weight vector ($w = [0.4, 0.4, 0.2]$) is employed. This underscores the fact that it is not necessary to select spreads taking extreme values; the key factor is the frequent return to the mean and the consistent

crossing of the 1σ threshold when the objective is to optimize the quality of individual trades. The importance of the mean-reverting property is further emphasized by the fact that the poorest-performing strategies are those utilizing a weight vector of $w = [0.2, 0.5, 0.3]$, which places the least emphasis on mean reversion. Moreover, with regard to the open and close rules, the combination $[\delta_{open}, \delta_{close}] = [1.8, 0.7]$ emerges as the most effective, while $[\delta_{open}, \delta_{close}] = [1, 0.1]$ performs the worst, *ceteris paribus*. This suggests a prudent approach: the optimal strategy is to initiate trades when the spread is sufficiently far from the average, ensuring a higher likelihood that the spread is near its furthest point from the mean and that the reversion dynamics are about to start. Subsequently, the strategy capitalizes on this opportunity without waiting for the full reversion to occur. Interestingly, the set of trading rules that yields the best results requires the largest minimum spread movement between the open and close of the trade.

An exception to these clear patterns is the second-best performing strategy, which rebalances with the highest frequency but employs a weight vector of $w = [0.45, 0.1, 0.45]$ and the open and close parameters $[\delta_{open}, \delta_{close}] = [1.4, 0.4]$. This strategy's selection criterion features the selection of spreads that take frequently extreme values and characterized by a strong mean reversion property, coupled with a more moderate approach in terms of the trade rules.

From the total profit perspective, the considerations differ significantly. The highest profit is achieved when assessing the quality of the spreads using the weight vector $w = [0.45, 0.1, 0.45]$ during the selection phase, with the selection process occurring every 200 days, reflecting a medium frequency. However, the trading rules associated with this strategy are the least effective from a return-per-trade perspective, employing the parameters $[\delta_{open}, \delta_{close}] = [1, 0.1]$.

A general pattern can be inferred, indicating that the highest profits are realized when $w = [0.45, 0.1, 0.45]$ is used as the weight vector for score evaluation during selection. This choice favors the selection of spreads that frequently attain extreme values beyond the 2σ threshold from the mean. Furthermore, reinforcing this risk-taking orientation thesis, the highest profits are consistently realized when $[\delta_{open}, \delta_{close}] = [1, 0.1]$ is employed as the trading rule, with all other parameters held constant. The combinations of these two patterns makes the aforementioned strategy the most effective profit-wise.

The volatility associated with this strategy is notably high, at 67%, and the lowest recorded returns reflect its higher-risk profile, as evidenced by the maximum absolute drawdown values. Of particular note is the impact of the drawdown in the *USD* portfolio, where the maximum absolute drawdown is six times larger than the final total profit.

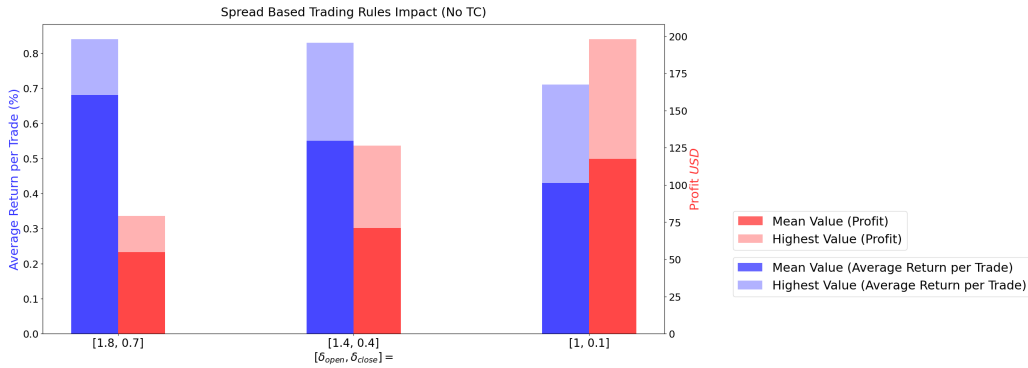


Figure 12: Average return per trade and final profit of each set of trade rules as mean of each group (with varying selection vector weights and selection frequency inside each of them) and the highest values in each group.

On the transaction costs front, the impact increases as the number of trades rises. Consequently, strategies utilizing $[\delta_{open}, \delta_{close}] = [1, 0.1]$ as the trading rule, which allow for a larger number of trades, experience a more significant effect. This impact is evident both in the average return per trade and in the total profit. When considering transaction costs, the best strategies in terms of total profit remain unchanged. However, the same cannot be said for the strategies that excel in terms of average return per trade. In this case, the optimal performance is achieved by the strategy that rebalances the asset basket every 250 days (the lowest frequency), with a weight vector of $w = [0.2, 0.5, 0.3]$ and trading rules $[\delta_{open}, \delta_{close}] = [1.8, 0.7]$. Notably, the trading rules in this strategy are identical to those of the best strategy without transaction costs, while the frequency and weight vector completely reverse the earlier conclusions.

In conclusion, it is essential to draw a clear distinction between the considerations related to the average return per trade and those pertaining to the total profit, as they assess different qualities of the strategies, as

previously noted. Throughout all the various settings, opening more trades results in a lower average return, reduced minimum returns, and higher maximum absolute drawdowns and volatility (52% versus 67% comparing the best strategies return and profit wise). However, it ultimately leads to higher profitability from an absolute profit perspective (Figure 12). Although the returns from the additional trades are marginally lower, causing the average return to decrease, they still contribute to the overall profitability of the strategy, despite being more susceptible to transaction costs.

7.2. Markov Switching Model Results

The impact of transaction costs is the least significant among the strategies considered, resulting in an average reduction of 0.25% in the gross average return per trade and a 16.2% decrease in total profit. This characteristic ensures that positive returns and relevant profits are still achievable even when the bid/ask spread is taken into account with almost all hyperparameter settings.

The best performance in terms of return is achieved by using the weight vector $w = [0.2, 0.5, 0.3]$ during the selection phase, which completely reverses the pattern observed in the purely spread-based trading algorithm. Meanwhile, $[\delta_{open}, \delta_{close}, \rho] = [1.8, 0.7, 0.8]$ consistently emerges as the optimal set of trading rules once the selection method's hyperparameters have been fixed. This reinforces, as emphasized in *Subsection 7.1*, the importance of caution when opening trades. Opening trades too early, specifically when the spread is still too close to the mean, and waiting for a complete mean reversion can result in more frequent losses. It is more crucial to identify entry points that are sufficiently distant from the mean, betting on the mean-reversion property of the spread, which is assumed to be stronger the further the spread is from the mean. Positions should then be closed as soon as a satisfactory gain has been achieved, without waiting for the complete reversion of the process, as is the case with the combination $[\delta_{open}, \delta_{close}] = [1, 0.1]$.

It is also important to highlight the role of ρ , the security level used to verify whether the spread is indeed in the forecasted regime before opening a trade. The best performance is attained when ρ is set to the highest value considered, 0.8, supporting the argument for prudence in trade execution (Figure 20).

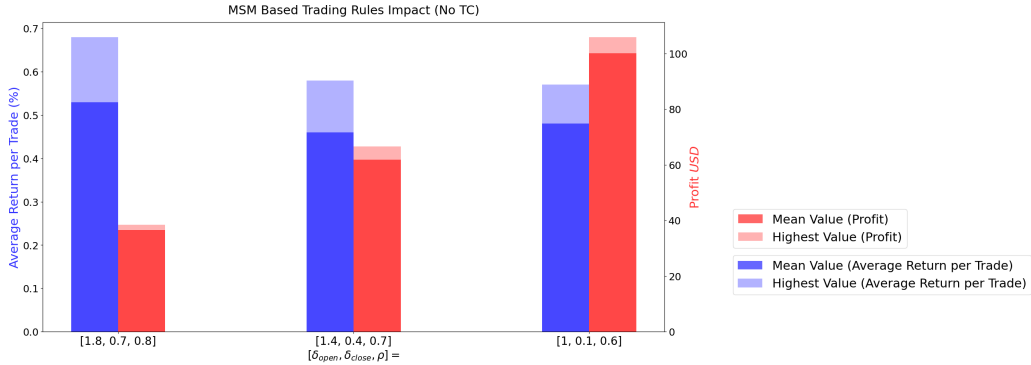


Figure 13: Average return per trade and final profit of each set of trade rules as mean of each group (with varying selection vector weights inside each of them) and the highest values in each group.

When these considerations are combined, they ensure a high quality of average return per trade (0.68% pre-transaction costs, 0.49% post-transaction costs), moderate volatility of returns (39%, refer to Figure 20), and positive profits in nearly all the currency portfolios. The percentage of negative returns is among the lowest across all the strategies considered, with an impressive 27% in the *EUR* portfolio. Furthermore, the lower bound of negative returns is among the least severe, with values of -7.66% for *NZD* and -7.93% for *EUR*.

From the profit perspective, as highlighted in the previous *Subsection 7.1*, the results differ and, to some extent, are inverted. This outcome is somewhat expected, given that the two metrics assess performance quality from different viewpoints. The highest profits are still realized when $w = [0.2, 0.5, 0.3]$ is used as the weight vector for scoring attribution but, as previously noted in *Subsection 7.1*, the strategies that generate the highest profits are those that are more flexible and less cautious when opening trades: $[\delta_{open}, \delta_{close}, \rho] = [1, 0.1, 0.6]$, followed by $[\delta_{open}, \delta_{close}, \rho] = [1.4, 0.4, 0.7]$.

The best strategy from a profit perspective combines these selections for weights and trading rules. This approach expands the set of opportunities and enables the exploitation of more opportunities, even though their marginal contributions gradually decrease. While this leads to lower average returns, higher transaction costs (which are more impactful as the number of trades increases), and higher volatility (48% versus 39%), the overall total profit increases (refer to Figure 20). However, the maximum drawdowns severely affect the final

profit, particularly in the *JPY* and *CNY* portfolios.

7.3. Copula Results

The first key result to highlight is the near-total absence of dependence, during the selection process, on the type of copula with respect to the selection date. In other words, the joint distribution of the same pair of exchange rate time series appears to be best described by the same type of copula throughout time.

First, it is important to highlight the substantial impact of transaction costs within this framework, with a few notable exceptions discussed at the end of this subsection. None of the combinations are able to achieve a positive average return per trade once transaction costs are factored in. As discussed later, this could be attributed to an overly simplistic and conservative modeling of the bid/ask spread; nevertheless, it remains a significant point to consider. Additionally, this outcome may be influenced by the relatively high number of trades initiated under this strategy, although there does not appear to be a strict correlation between the impact of transaction costs and the varying levels of tolerance, expressed as probability thresholds, used to trigger trade openings and that have an intrinsic link with the number of opened trades. The significant effect of transaction costs is also evident when examining the absolute profits: in some cases, what would have been slightly positive final profits are instead pushed into negative territory due to the transaction costs.

In general, the best results in terms of average return per trade are achieved when employing the *or* exit logic. The only exceptions arise when using the least stringent tolerance levels during the selection phase and for trade openings and in the computation of the *GAM* score, but only when copula and *GAM* scores are combined without prioritizing the copula-based score. This distinct pattern underscores that, in most cases, it is sufficient for just one of the two exchange rates to complete the expected movement; the other may still be in the process of reverting, yet not to the extent that it offsets the gains of the first. Thus, it is better to capitalize on the behavior of the first one betting on the fact that the other one, selected in order to revert its behavior simultaneously to the first one, is not too far from the complete reversion. This aligns with a key selection criterion: the frequency of such co-movements during the selection period.

From the perspective of tolerance levels, which influence both the selection phase and the probability thresholds for opening trades, it is evident that the lowest tolerance levels, imposing the strictest selection and trading rules, yield the worst results. Conversely, a moderate or highly tolerant approach tends to achieve similar results on average.

Additionally, the balance between the two selection methodologies significantly impacts performance. Specifically, a well-balanced combination of the *GAM* and copula scores, or a selection process slightly biased toward copula-based scoring, tends to produce superior outcomes.

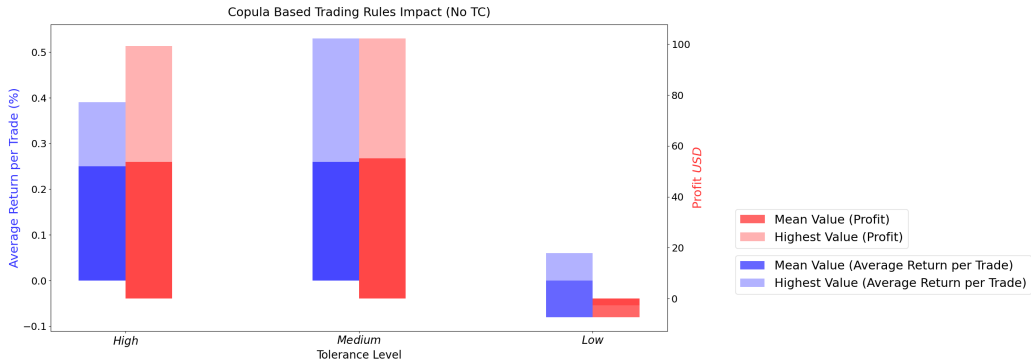


Figure 14: Average return per trade and final profit (mean and highest values) of the strategies grouped by tolerance. Recall that *High*, *Medium*, *Low* tolerance levels correspond to $[p_{up}, p_{low}, tolerance_{GAM}] = [0.8, 0.2, 0.4]$, $[0.875, 0.125, 0.25]$, $[0.95, 0.05, 0.1]$.

The best-performing strategy in terms of average return per trade adopts a moderate tolerance approach during the selection phase (*GAM* tolerance = 0.40 and $[p_{up}, p_{low}] = [0.875, 0.125]$, which are also the probability thresholds applied in the trading phase). Additionally, it employs a selection score biased toward the copula score and utilizes an *or* exit logic, as the identified patterns highlight. Despite achieving an overall positive average return per trade, the strategy generates a positive final profit in only seven out of the ten currency portfolios.

Examining the maximum drawdowns, it is highly likely that the negative final profits in the remaining cases stem

from an abrupt closure of trades, resulting in substantial losses. This underscores the necessity of implementing more refined liquidation rules and introducing stricter stop-loss criteria to mitigate excessive downside risk.

From the final profit perspective, the best-performing strategies align with those identified in terms of average return per trade (refer to Figure 14). Specifically, the top three strategies remain unchanged, all consistently generating positive profits. Notably, the highest-performing strategy achieves a level of profitability comparable to that of the Markov switching model and, surprisingly, exceeds it once transaction costs are accounted for, an exception to the generally significant impact of transaction costs in this framework. The most profitable strategy adopts a medium level of tolerance during the trading phase, prioritizes the copula-based selection score (as observed in the best strategy for return), but differs by employing an *and* exit logic.

Still from the profit perspective, the *and* exit logic emerges as the superior approach, contradicting prior observations. It outperforms the *or* exit logic in eight out of nine cases, both before and after transaction costs are factored in. In particular, the most profitable strategy yields a 14.9% higher profit compared to its *or* counterpart, as Figure 15 shows. This suggests that when employing copula-based algorithm, a more cautious approach should be taken in determining trade exit points.

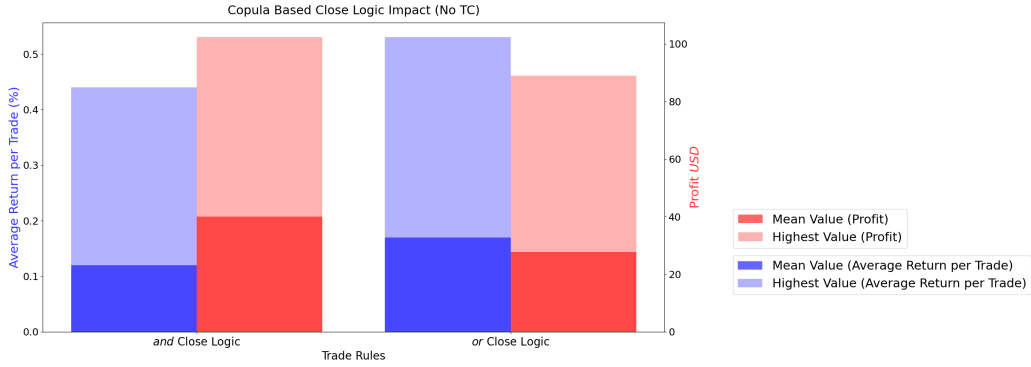


Figure 15: Average return per trade and final profit of the strategies that use an *and* and *or* exit logic and the highest values for both.

7.4. Logistic Regression Results

The first crucial aspect to highlight concerns the impact of transaction costs. Given that this algorithm executes the highest number of trades, it experiences a significant difference between the profits before and after accounting for transaction costs. This difference is relevant also in terms of reduction of the average return per trade. However, despite this drawback, the strategy’s robust and consistent performance enables it to maintain positive average returns in approximately half of the analyzed cases, even under a conservatively modeled bid/ask spread.

The return performance of the strategy varies based on two key factors: the weight vector used to compute each exchange rate’s score during the selection and the stop-loss threshold applied to prematurely close trades during the trading phase. The expected profit within the next 100 days was fixed at 5% since empirical testing suggested this as the optimal level. Adjusting the expected profit threshold too high risks missing profitable opportunities. Conversely, setting it too low essentially reduces the model to random labeling due to its attempt to predict movements that are too small.

Both degrees of freedom show clear patterns. Unlike the pair selection approach based on spread properties (Subsection 5.1), the highest return performance in this context is achieved using the most balanced weighting vector, $w = [0.2, 0.5, 0.3]$, while the worst results come from using $w = [0.45, 0.1, 0.45]$ (see Figure 16). Favoring exchange rates with extreme values appears ineffective, whereas selecting those that frequently cross the 1σ threshold proves advantageous. A similarly clear trend emerges with stop-loss thresholds: once the weight vector is fixed, the best-performing strategy sets the stop-loss exit signal to 0.15, the lowest considered value. This underscores the resilience of exchange rates and the importance of trusting their mean-reverting behavior. Additionally, these findings highlight the potential of Logistic Regression as a classification tool in this framework. If the model predicts a 5% increase (or decrease) in an exchange rate, its signals should be trusted even if the rate initially moves in the opposite direction by up to 10%. Further testing of higher stop-loss values is warranted, as empirical observations suggest that a stop loss at the 20% level may be excessively soft.

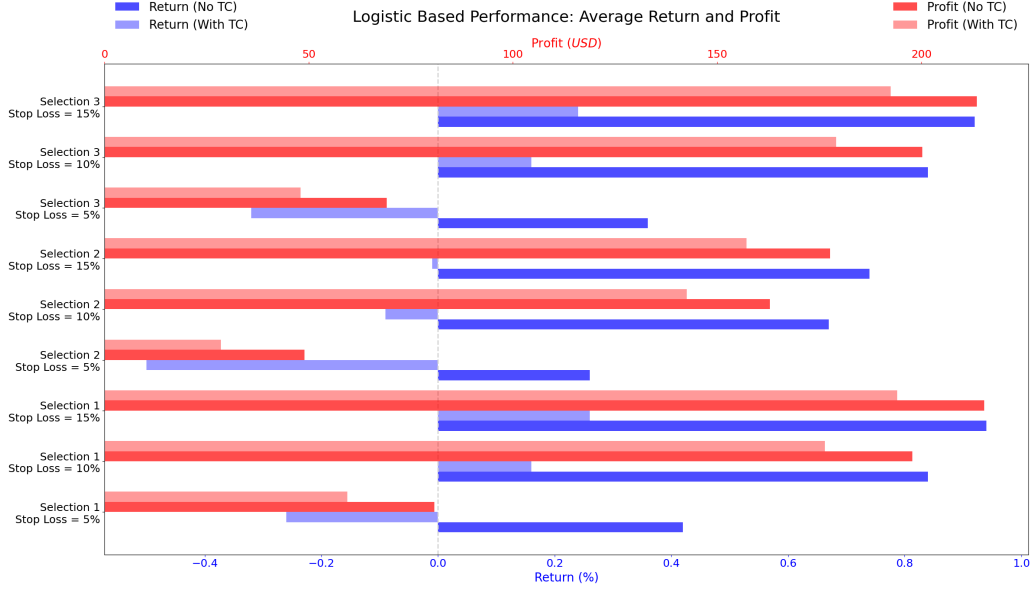


Figure 16: Reduction of average return and profit of the logistic regression-based strategies after accounting for transaction costs. Selection 1,2,3 correspond to $w = [0.2, 0.5, 0.3]$, $[0.45, 0.1, 0.45]$, $[0.4, 0.4, 0.2]$.

As a result, the best-performing strategy selects the exchange rates using $w = [0.2, 0.5, 0.3]$ and applies a stop-loss threshold of 15%, achieving an average return per trade of 0.94%, the highest in the entire study. Even after accounting for transaction costs, the strategy remains firmly profitable, with an average return per trade of 0.26%. Similar performance levels are observed when using $w = [0.4, 0.4, 0.2]$ while maintaining the same stop-loss threshold of 15%.

For each currency, fewer than 40% of the trades result in negative returns, meaning that more than half of the bets are correct. Naturally, the lowest returns are more negative compared to those obtained using different stop-loss criteria, stabilizing around -15%. This indicates that some trades were prematurely closed due to an apparent divergence from the expected behavior. On the other hand, the highest returns are remarkably large: this strategy is able to close some trades with extraordinary returns such as a +19.20% in *HKD*, a +19.17% in *CNY*, a +18.67% in *NZD*, and a +16.19% in *CHF*.

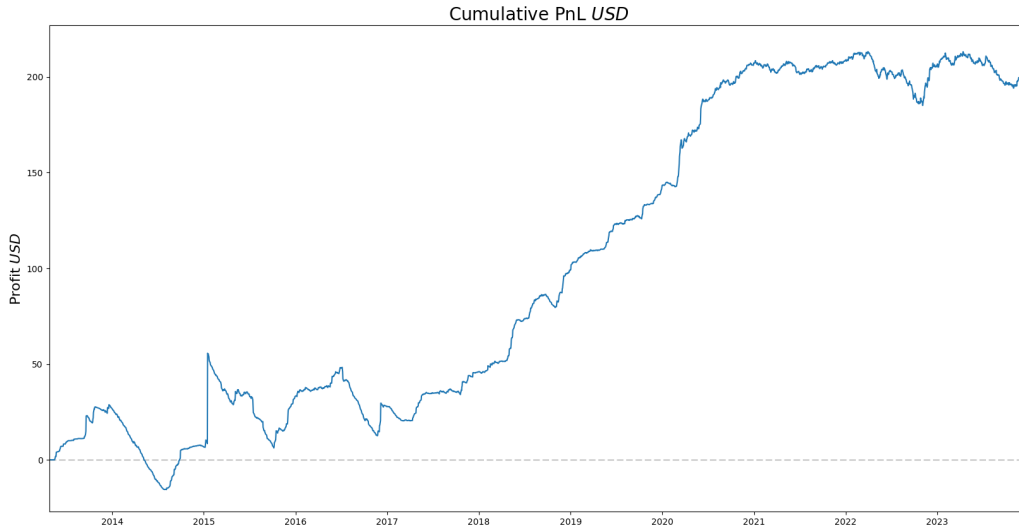


Figure 17: *PnL* of the most profitable strategy: cumulative value of the portfolios, *USD* expressed.

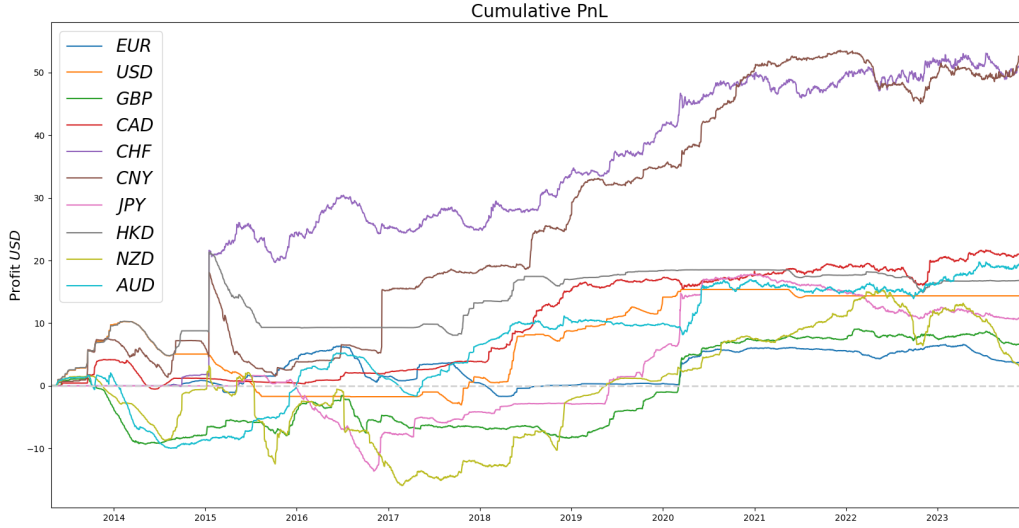


Figure 18: PnL of the most profitable strategy: individual portfolio values USD expressed.

Despite these strong performances, the maximum drawdowns remain significant, highlighting the impact of incorrect bets on the final total profit. The volatility in the PnL behavior is further confirmed by the high volatility values, reinforcing the need for refined risk management techniques.

Surprisingly, and in contrast to all other analyzed cases, the observed regularities regarding the impact of hyper-parameters on return performance remain valid when evaluating the strategies from the total profit perspective (Figure 17 and Figure 18 for the PnL behavior). Once again, setting the stop-loss threshold to 15% results in minimal differences in total profit when selecting the basket of exchange rates using either $w = [0.2, 0.5, 0.3]$ or $w = [0.4, 0.4, 0.2]$, with only a slight variation even after accounting for transaction costs. However, the first weighting scheme exhibits slightly lower volatility levels while experiencing a more pronounced maximum absolute drawdown.

From the prediction perspective, the best-performing strategy correctly classifies 70% of the observations, with only 2.7% resulting in severe misclassifications (i.e., incorrect classification between long and short positions) out of a total of 83,490 observations across all days and rates (see Figure 19). The observations classified as 0 instead of -1 or 1 are errors of minor importance, as previously discussed in *Subsection 6.4*, since they only result in a missed opportunity cost. These classification metrics are very similar to those of its twin strategy, which uses $w = [0.4, 0.4, 0.2]$.

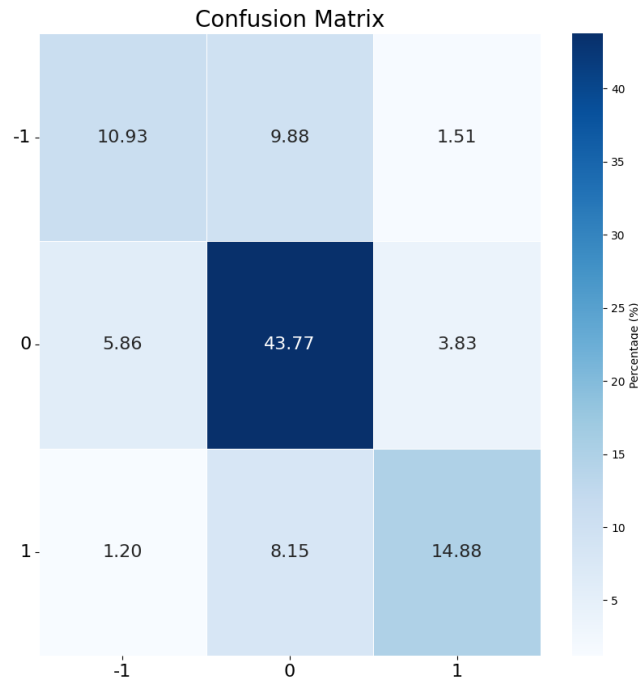


Figure 19: Confusion Matrix of the discussed strategy.

7.5. Comparative Analysis

In this subsection, the best strategies, first from a trade return perspective and then from a profit perspective, of each method will be discussed using the name of the method itself, without specifying the combination of hyperparameters employed. A detailed specification of the values used to achieve the best version of each model is provided in *Appendix E*.

The first aspect that needs to be outlined is the scarce performance of the copula trading algorithm compared to the others from an average return perspective; in particular, when transaction costs are taken into account, none of its versions are able to provide positive average returns. In terms of return performance, the other side of the coin is represented by the logistic regression-based trading algorithm, which delivers robust returns across all combinations of hyperparameters used. It is the absolute best algorithm return-wise among the considered ones if transaction costs are not accounted for.

It must be said that, despite being the best performer, the returns realized by the logistic regression strategy suffer from high levels of volatility and, most importantly, are severely impacted by transaction costs. Figure 16 clearly presents this reduction. After accounting for transaction costs, the best strategy return-wise is the purely spread-based approach. The reason lies in the high number of trades opened, approximately three times more than the Markov switching model, for instance. Opening more trades likely implies executing trades with lower profit margins, making transaction costs a decisive factor.

From this perspective, the Markov switching model and the spread-based trading algorithms emerge as the most reliable and robust. They consistently provide positive returns both with and without transaction costs, and their peak gross performances are not far from those of the logistic regression-based approach. Moreover, the proportion of negative returns is significantly lower than that observed in the copula and logistic regression-based algorithms (35% versus 42% of copula-based and 39% of logistic regression-based in their best return wise settings).

Between the spread-based and the Markov switching model algorithms the differences reflect the distinct approaches they follow. Despite both relying on the analysis of the ratio dynamics between exchange rates, their differing levels of confidence required before entering a trade lead to notable performance variations. The best-performing in an average return perspective version of the spread-based algorithm achieves a higher average return per trade than the peak performance of the Markov switching model (0.84% versus 0.68%). However, this comes at the cost of higher risk, as indicated by the annualized volatility values: 52% for the spread-based strategy compared to 39% for the Markov switching model, which records the lowest volatility among the top-performing strategies. This clearly illustrates a classic return/volatility trade-off scenario (identifiable also using Total Profit as performance and Drawdown impact as risk metrics as Figure 20 and Figure 21 show).

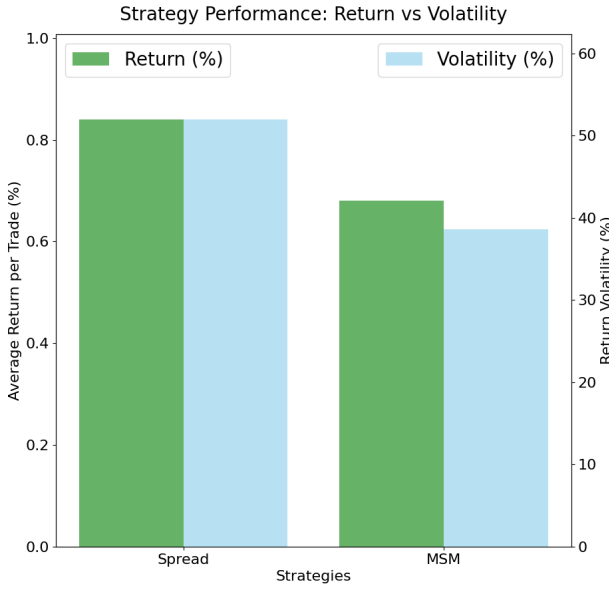


Figure 20: Return/volatility trade-off between the best return wise purely spread-based and *MSM* algorithms.

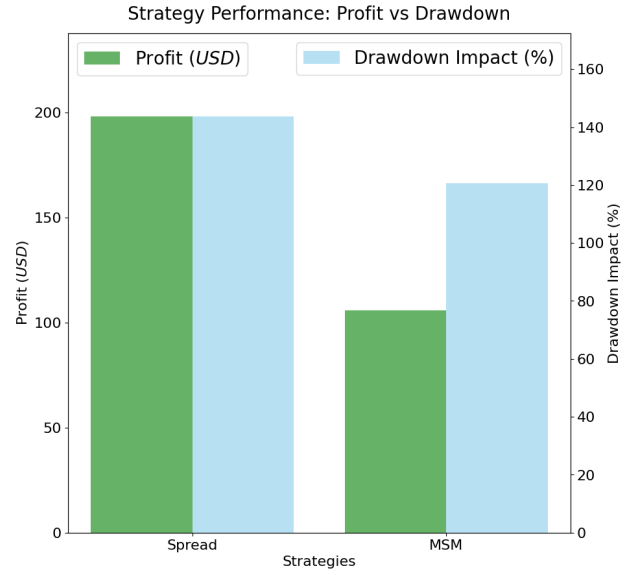


Figure 21: Total profit/drawdown impact trade-off between the best profit wise purely spread-based and *MSM* algorithms.

This aligns perfectly with the concept of the Markov switching model algorithm, whose primary objective is to identify the current regime state and open a trade only when a sufficiently high confidence level in the inferred probability is reached. This cautious approach is reflected in the number of trades executed, which is 40% lower than its purely spread-based counterpart. Additionally, it results in a narrower difference between the highest and lowest recorded returns for each currency.

The comparison between these two algorithms is also insightful from the perspective of the scoring method used to select exchange rates that yield the best returns. The spread-based algorithm prioritizes selecting spreads with strong mean-reverting properties and moderately pronounced oscillations. Conversely, the Markov switching model-based strategy achieves its best returns by using a weighting scheme that favors spreads exhibiting extreme and moderately extreme values over those displaying strong mean reversion. This distinction arises from the differing trading rules: the Markov switching model strategy generates trading signals based on the distance from the high and low regime averages, rather than the overall mean. As a result, the mean-reverting property is interpreted merely as the spread's ability to switch between regimes, making it less relevant than the crossing of standard deviation thresholds.

From the Maximum absolute drawdown perspective, the logistic regression-based algorithm does not achieve the expected results. Despite delivering a solid level of returns, its intended purpose was to reduce both the frequency and magnitude of losses by trading one rate at a time without considering the relative behaviors between rates. While the frequency of negative returns is comparable to that of the spread-based algorithms (33% for the purely spread-based, 35% for the *MSM*, 39% for the logistic regression based), the magnitude of the Maximum absolute drawdown is significantly higher.

The real disappointment of the study from a return perspective is the copula-based trading algorithm, which fails to generate positive returns when transaction costs are considered and, unlike the other strategies, results in deeply negative final profits in three out of ten currencies. Additionally, it experiences the second-highest levels of return volatility, only behind the logistic regression-based algorithm, and the proportion of the Maximum absolute drawdown relative to the final total profit is among the highest, again immediately behind only the logistic regression-based strategy. However, the latter is able to compensate for a large number of negative returns with more consistent positive returns.

Despite these weaknesses, the preference for the *or* exit logic provides valuable insights for future research. It suggests that integrating the analysis of individual exchange rates alongside the spread, combined with flexible exit rules and with the already discussed need for conservative open rules, could enhance the performance of spread-based algorithms, at least from a return perspective.

Finally, from the *Total Profit* perspective, some further comparisons need to be made. First of all, the best calibration of each method remains the same even after considering transaction costs. Secondly, among the best representatives of each method, the Markov switching model-based one experiences the highest percentage reduction in final profit when transaction costs are accounted for (19% reduction versus 11% for the other

methods). This contradicts the previous observations about the impact on average return for each strategy, but, as noted, these metrics assess different aspects of performance. The logistic regression-based strategy is the most profitable, closely followed by the purely spread-based one. The Markov switching model and copula-based strategies realize only half of the total profit of the former two.

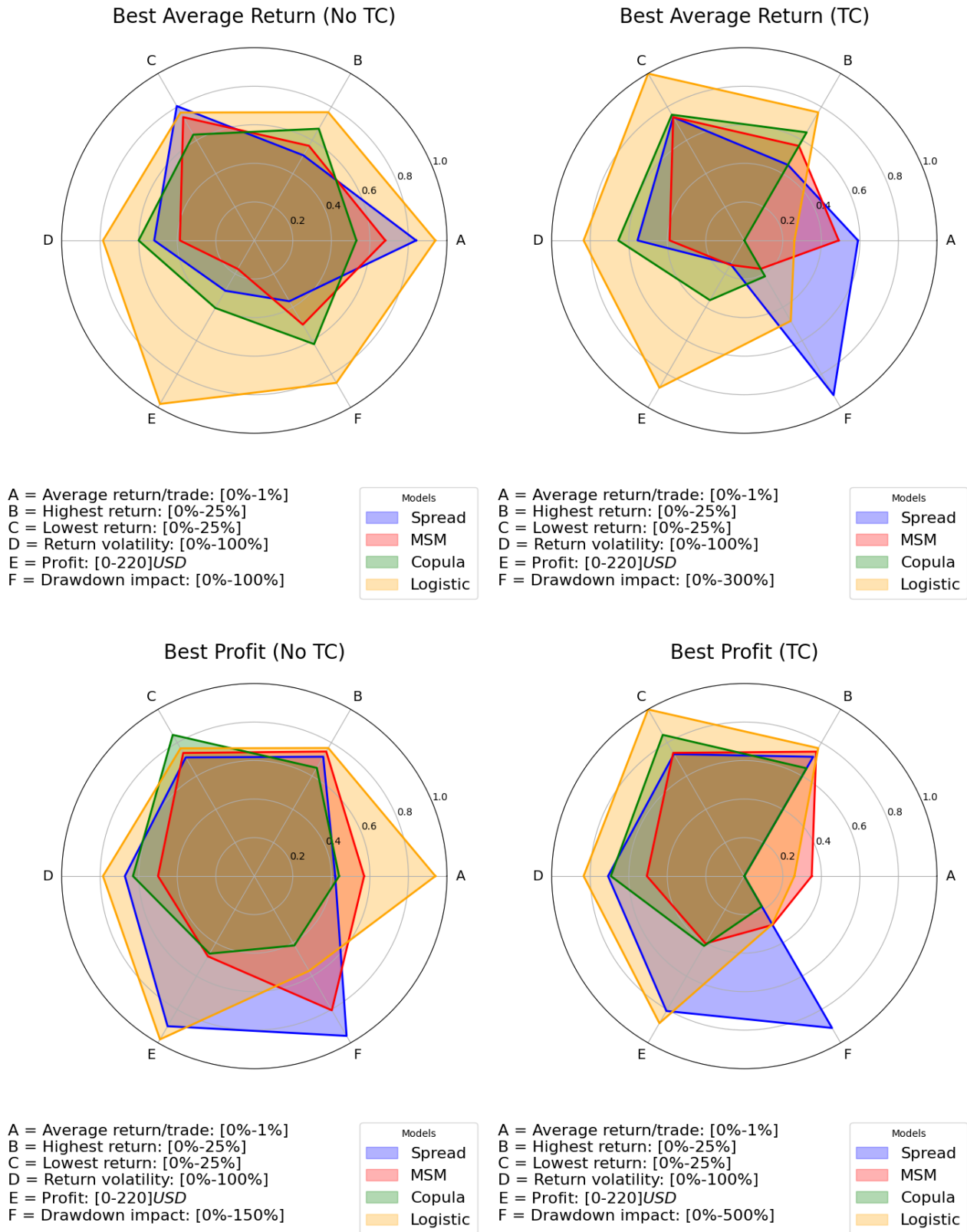


Figure 22: Key metrics of the best strategies for each model from a return and profit perspective, with and without transaction costs).

What has been said about the comparison between the Markov switching model and purely spread-based strategies regarding the returns still holds here regarding the profits. Despite realizing almost double the profit, the purely spread-based approach exhibits a volatility of 67%, compared to 48% for the Markov switching-based one. The trade-off still holds when considering the Drawdown impact as a risk metric, as Figure 21 shows. The logistic regression-based strategy remains the most volatile, with a volatility of 78%.

Finally, an important observation, more useful for future research but still noteworthy, is that from a purely return perspective, the best performance are achieved by being cautious when opening the trades (as seen in the regularities detected in both the purely spread-based strategies and the Markov switching model-based one) and less conservative when closing them (as indicated by the better performance of the *or* exit logic in the copula and of $\delta_{close} = 0.7$ as spread closing rule). However, from a profit perspective, these patterns seem to be inverted: higher profits are realized with more tolerant opening policies and more conservative closing rules (applicable to both copulas, with the preference for the *and* exit logic, and spread-based strategies, with the better profit performance realized by strategies using $\delta_{close} = 0.1$).

The charts above Figure 22 show the key metrics of the best strategy for each model from a return per trade and profit perspective both accounting and not for transaction costs. The scale of each metric is reported in the legend. The lowest returns are expressed as their absolute values, while negative average returns per trade are represented as 0. Figure 23 below shows the performance over risk metrics of each strategy measuring the performance both as profit and return over volatility. Negative returns or profit are again represented as 0. The hyperparameters used for each method are reported in the captions of the tables in *Appendix E*.

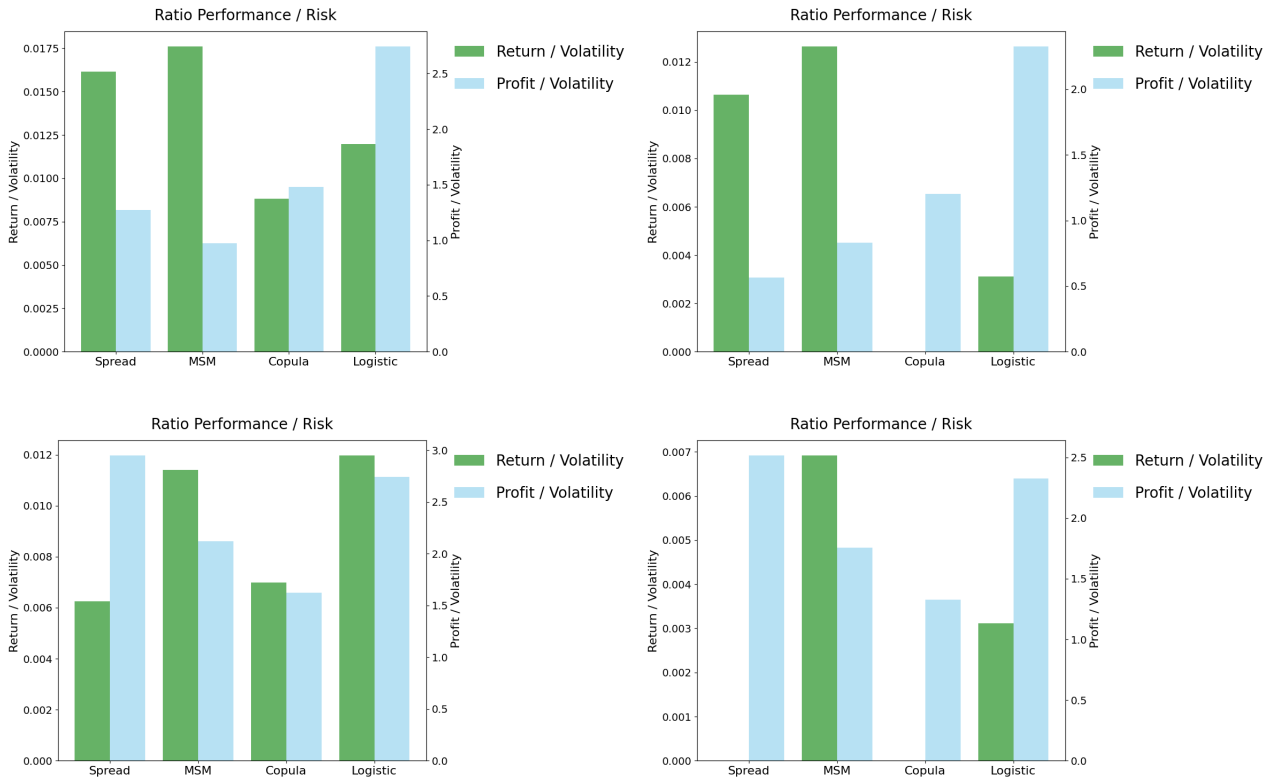


Figure 23: Performance / Risk ratios for the best strategy of each model (clockwise from top left: best strategies return perspective, no transaction costs; return perspective, transaction costs; profit perspective, no transaction costs; profit perspective, transaction costs).

8. Future Research

This field is open to multiple streams of research and future improvements. In this section, several possible refinements directly linked to the findings presented in this work are discussed, both to address some of the problematic aspects and to further enhance the potential of the methods.

The first issue to address concerns the spread. As already extensively discussed, the information embedded in the spread, both during the selection and trading phases, is generally insufficient to properly describe the individual exchange rates, and its use could lead to a loss in one of the two legs. This problem becomes even

more pronounced when pairs denominated in different currencies are considered. Despite these limitations, the behavior of the spread can still offer valuable information regarding the reciprocal behavior of the time series, providing insights into their profitability and similarity, especially when measured by their oscillations and stability. Therefore, integrating the spread with a more in-depth analysis of the individual time series could prove beneficial.

Some initial steps in this direction were taken in this work by incorporating an improper cointegration analysis during the selection phase. Further advancements, both in selecting eligible spreads and during the trading phase, could involve forecasting the individual exchange rates using econometric models, alongside an approach that incorporates an analysis of the distance between the time series of the exchange rates. In this regard, promising metrics to measure the distance, apart from the traditional *Euclidean Distance*, which is problematic in this context due to the curse of dimensionality, include the *Chi Squared Distance* and the *QQ Distance*.

Additional improvements in spread analysis could involve incorporating other metrics to assess its mean-reverting behavior, such as the *Hurst Exponent* [32] and the *Half-Life* value.

Furthermore, a deeper structuring of the trading rules could be implemented. In particular, given that not all the spreads, even among the selected ones, exhibit the same type and intensity of oscillatory behavior, it could be valuable to introduce trading rules based on the specific mean reversion properties of each spread. This would allow for taking more or less risky positions depending on the historical propensity of each spread to reach such extreme values and to revert from them. This approach should be complemented with a time-varying set of rules that can dynamically adapt to the most recent past, allowing the strategy to be more responsive to changing market conditions.

Finally, an interesting progression would be the identification of multiple levels of deviation from the average of the spread, similar to what is done in the equity market. By expanding beyond the three levels identified in this work, it would be possible to develop a richer set of trading rules, with one for each newly identified region of deviation. This would further enhance the flexibility and responsiveness of the strategy to different market environments.

A deeper structuring of the trading rules could also be beneficial in other settings, particularly regarding the stop loss strategy. While the stop loss strategy was one of the main focuses of the analysis for the logistic regression-based trading strategy, for the other strategies, the parameters and thresholds used to prematurely exit a trade have been somewhat overlooked. Both the critical exposures and the standard deviation thresholds used for premature exits should be analyzed more deeply, as they could have a significant impact on the overall performance of the strategy. As mentioned earlier, the use of pair-specific values based on historical behavior could prove advantageous in this context as well.

In this regard, another key improvement would be the application of refined liquidation rules at the end of the trading period for trades that are still open. This would prevent all trades from being closed at once, as previously discussed in *Subsection 7.3*, and could lead to better final outcomes by allowing more flexibility in managing open positions.

Another stream of development refers to the entire selection concept. So far, a symmetric selection of two exchange rates has been adopted. However, an asymmetrical treatment of the pairs, where more importance and exposure are given to the more promising rate, could be an interesting direction to explore.

In the same direction, the identification of clusters of exchange rates to be traded together, exploiting temporary inner divergences, could provide further enhancements. By grouping rates that exhibit similar behaviors over certain periods, one could take advantage of short-term divergences within the clusters. A promising approach to clustering exchange rates has been proposed by Fenn et al. [16], and integrating such a method could offer a more refined and dynamic selection process.

On the *GAM* selection side, the use of larger datasets could significantly improve the computation of the distances between the exchange rates. One of the key drivers of the rates' dynamics is known to be the trading volume. Incorporating volume data, as well as volatility indicators and currency indices, into the *GAM* framework could provide a more comprehensive view of the market's behavior. By feeding these additional variables into the model, the computation of distances between the rates would become more refined and accurate, ultimately leading to a better selection of the pairs to be traded.

Furthermore, the issue regarding the temporal depth used in training the algorithm or in observing the behavior of the involved time series has been widely discussed throughout this work. The main problem lies in the breakdown of the temporal structure and temporal dependence and in the risk of using outdated data. A more refined approach would incorporate time information by assigning different weights to observations based on their temporal distance. This approach would minimize the emphasis on selecting the right training period lengths and, as a result, the trading actions taken would likely be more accurate.

There is particularly significant room for this kind of development in the case of the copula-based model. A study on the evolution of copulas over time in the FX market was conducted by Kiselak et al. [33], suggesting that a

time-varying copula approach could enhance model performance. Building upon this, it would be interesting to develop not only a time-varying copula but also to modify the copula structure itself to assign varying weights to observations according to their distance from the present moment. This would allow the model to better capture the changing dynamics of the FX market and improve its overall predictive accuracy.

Always regarding the time management, an interesting side of research would be the increase in the fitting and training frequency of the algorithms used. In particular, it was observed in *Subsection 7.1* a significant increment of performance when the frequency was increased and the past lookback reduced. Due to the computational complexity of some of the employed algorithms, it was not possible to further increase the frequency, but an interesting development could be the employment of models fast enough to be fit with high frequency and that require a small amount of past observations, following the same flavor identified in the results of the spread-based trading.

Additionally, an important improvement in the Markov switching model-based approach could be tried. It consists of the possible improvements that the identification of structural breaks, in conjunction with the application of the algorithm, could provide. Despite being thought with the same logic, this would allow the model to be fit with a training time period known for sure to face a switch in the dominant regime, eventually leading to a better identification of the regimes in the out-of-sample probabilities estimation.

The biggest concern that needs to be taken care of is, for sure, the impact of the transaction costs. Two directions can be pursued to deal with that: the first one is a better modeling of the bid/ask spread based on historical data to enhance the backtesting reliability. The second one is a more precise and punctual policy of trade openings to limit the number of trades and the effects of the transaction costs.

Furthermore, as it is possible to infer from the reported values, the annualized volatility of the returns reaches concerning levels with all the tested strategies. A more refined approach to trade openings should ensure better performance on the risk profile side as well.

A key area for future improvement, although relatively straightforward, lies in conducting a more comprehensive and detailed analysis of all possible hyperparameter combinations. As highlighted throughout this study, the choice of hyperparameters, both in terms of asset selection and signal generation, has a significant impact on the performance of the strategies. An interesting direction for future research would involve systematically calibrating the models to identify the optimal combination of hyperparameters that yields the best performance during the backtesting process.

Finally, future improvements should build upon the aspects that have proven effective in this study. The observations regarding the necessity for frequent selection of tradable assets and the regular adjustment of the tools to fine-tune the trading rule parameters should serve as a foundation for further development. Additionally, the considerations related to the scoring methodology, particularly in terms of the weights assigned to the properties of the spread, should be explored and refined in future research.

9. Conclusions

This study has aimed to bridge the gap between equity pairs trading, anomalous behavior detection, and forex trading, yielding mixed results. Additionally, selection methods commonly employed in the equity framework have been adapted to the FX market, while alternative approaches have also been proposed.

On one hand, several of the implemented strategies demonstrate strong performance, yielding robust and positive returns. On the other hand, certain combinations of selection methods and trading algorithms fail to generate returns resilient to transaction costs. In particular, the copula trading algorithm and its *GAM* selection method appear to require a broader input dataset.

Nevertheless, the results remain promising. Some combinations of selection techniques and trading strategies already produce favorable outcomes. Enhancing the modeling of transaction costs and incorporating the improvements outlined in *Section 8* could further refine these methods, mitigating volatility and reducing the impact of the bid/ask spread.

It is essential to emphasize that pairs trading is a profitable strategy. Despite the suboptimal performance of the copula-based trading algorithm, whose limitations have been thoroughly discussed in *Section 7.3*, highlighting the need for more refined selection methods and liquidation strategies, the spread-based algorithms demonstrate the ability to generate positive returns. Furthermore, the introduction of the Markov switching model to characterize spread dynamics proves effective in reducing the strategy's risk profile, though at the cost of some returns and profits. This confirms that the spread between exchange rates denominated in the same currency carries sufficient information to enable efficient trading of the two components.

The true strength of spread-based algorithms lies in their capacity to capitalize on price divergences between

assets without relying on explicit predictions of their individual behaviors. As a result, they achieve a more favorable balance of returns and risk compared to the logistic regression-based algorithm, which focuses on the behavior of individual exchange rates. However, the logistic regression-based strategy's ability to identify and execute a high volume of potential trades makes it the most profitable in aggregate terms, even if it is less efficient on a per-trade basis.

This study has sought to emphasize the dual perspectives through which the performance of these trading strategies can be evaluated, as well as the various factors that influence them. The approach taken to maximize the average return differs fundamentally from the one aimed at maximizing total profit. In some instances, these approaches are not only distinct but also inherently opposed, addressing fundamentally different strategic objectives. Expanding the set of trades, when executed judiciously, can yield substantial gains and increased total profit. However, this expansion typically results in a lower average return per trade, indicating diminishing marginal benefits. Ultimately, an excessive increase in the number of trades risks surpassing a critical threshold, beyond which total profit may begin to deteriorate.

This study has also examined the optimal combinations of hyperparameters for each strategy, highlighting their impact on overall performance and identifying distinct patterns. Some of these patterns are universally shared across different selection methods and trading algorithms, while others are unique to specific approaches. By analyzing these relationships, the study provides a deeper understanding of how parameter tuning influences strategy effectiveness, offering valuable insights for refining future implementations.

Finally, several potential directions for future research have been proposed, spanning both selection methodologies and trading algorithm enhancements. Among these, one of the most critical directions is a more in-depth examination of hyperparameter influence, alongside an expansion of the degrees of freedom considered in the strategy design, particularly by analyzing the impact of parameters that have so far remained fixed. A more precise and refined calibration of these values could significantly improve performance, especially when combined with trading rules that are more tailored and pair-specific, allowing for a better adaptation to the unique behaviors of the traded currency pairs.

References

- [1] Q. Jin A. Cartea, M. Cucuringu. Correlation matrix clustering for statistical arbitrage portfolios. *4th ACM International Conference on AI in Finance*, 2023.
- [2] T. Downs A. Hryshko. A machine learning approach to intraday trading on foreign exchange markets. *Lecture Notes in Computer Science*. 3578. 588-595, 2005.
- [3] L. S. Hook A. Nassir A. Zarei, M. Ariff. Identifying multiple structural breaks in exchange rate series in a finance research. *Social Sciences and Humanities*, 2015.
- [4] H. Akaike. A new look at statistical model identification. *IEEE Transactions on Automatic Control*, AC-19(6), 716-723, 1974.
- [5] R. Faff B. Do. Does simple pairs trading still work? *Financial Analysts Journal*, 66(4):83-95, 2010.
- [6] R. Faff B. Do. Are pairs trading profits robust to trading costs? *Journal of Financial Research*, 35(2):261-287, 2012.
- [7] W. K. Bertram. Analytic solutions for optimal statistical arbitrage trading. *Physica A: Statistical Mechanics and its Applications*, 389(11):2234-2243, 2010.
- [8] T. Bogomolov. Pairs trading based on statistical variability of the spread process. *Quantitative Finance*, 13(9):1411-1430, 2013.
- [9] C. G. Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(3), 76-90, 1970.
- [10] M. Herman C. G. Rojas. Foreign exchange forecasting via machine learning.
- [11] T. Liu C. H. Lin. Machine learning in enhancing the performance of prediction and trading strategies in fx markets. 2024.
- [12] B. Evans C. L. Dunis, J. Laws. Trading futures spreads: An application of correlation and threshold filters. *Applied Financial Economics*, 16(12):903-914, 2006.

- [13] D. G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1), 141-151, 1978.
- [14] W. A. Fuller D. A. Dickey. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427-431, 1979.
- [15] W. A. Fuller D. A. Dickey. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4), 1057-1072, 1981.
- [16] P. J. Mucha M. McDonald S. Williams N. F. Johnson N. S. Jones D. J. Fenn, M. A. Porter. Dynamical clustering of exchange rates. 2010.
- [17] K. G. Rouwenhorst. E. G. Gatev, W. N. Goetzmann. Pairs trading: performance of a relative value arbitrage rule. *Review of Financial Studies*, 19(3):797-827, 2006.
- [18] K. G. Rouwenhorst E. Gatev, W. N. Goetzmann. Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies*, 19(3):797-827., 2006.
- [19] R. V. Faizullin. Problems of the use of cointegration pairs for pairs trading. *BULLETIN OF HIGH TECHNOLOGY*, 2022.
- [20] L. Ferreira. New tools for spread trading. *Futures: News, Analysis and Strategies for Futures, Options and Derivatives Traders*, 37(12):38-41, 2008.
- [21] R. Fletcher. A new approach to variable metric algorithms. *Computer Journal*, 13(3), 317-322, 1970.
- [22] S. Franco. Pairs trading profitability and style investing. Master's thesis, Universidade Católica Portuguesa, 2014.
- [23] M. Frank. On the simultaneous associativity of $f(x, y)$ and $x+y-f(x, y)$. *Aequationes mathematicae*, 19(1), 194-226, 1979.
- [24] D. Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109), 23-26, 1970.
- [25] Jeroen Goudvis. A cointegration trading strategy for currency exchange rates. Master's thesis, Erasmus School of Economics, 2020.
- [26] A. Rea H. C. J. Zhan, W. Rea. An application of correlation clustering to portfolio diversification. 2015.
- [27] F. Li H. Chen, S. J. Chen. Empirical investigation of an equity pairs trading strategy. *SSRN Electronic Journal*, 2012.
- [28] Z. Jiang S. N. Abdullah Deshinta Arrova Dewi H. Ding*, X. Zhao. Eur/usd exchange rate forecasting based on information fusion with large language models and deep learning methods. 2024.
- [29] Y.J. Lee W. Zhuang H. Joe, T. K. Hu. Generating random variates with specified correlation. *Journal of Statistical Computation and Simulation*, Vol 57(3), pp247-265, 1997.
- [30] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357-384, 1989.
- [31] J. D. Hamilton. Time series analysis. *Princeton University Press, Princeton*, 1994.
- [32] H. E. Hurst. The long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116, 770-799, 1951.
- [33] N. Ishimura Y. Yoshizawa L. Salazar M. Stehlík I. Kupka, J. Kiselak. Time evolutions of copulas and foreign exchange markets. *Information Sciences, Volume 467, Pages 163-178*, 2018.
- [34] D. Song I. O. Asensio. With cross-border investments, realize that currencies will mean revert. Technical report, Silicon Valley Bank, 2021.
- [35] R. Mehra J. B. Donaldson. Average crossing time: an alternative characterization of mean aversion and reversion. *National Bureau of Economic Research*, 2019.
- [36] G. Zanotti J. G. Ordonez, M. Pelger. Deep learning statistical arbitrage. 2019.

- [37] A. Timmermann J. Liu. Optimal convergence trade strategies. *Review of Financial Studies*, 26(4):1048-1086, 2013.
- [38] H. C. Wolf J. M. Campa. Is real exchange rate mean reversion caused by arbitrage? *NBER Working Paper No. w6162*, 1997.
- [39] H. Yang J. W. Jurek. Dynamic portfolio selection in arbitrage. *Working paper, Harvard University*, 2007.
- [40] S. Johansen. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica Vol. 59, No. 6*, 1991.
- [41] M. Williams K. Chan, Y. Tse. The relationship between commodity prices and currency exchange rates: Evidence from the futures markets. *Commodity Prices and Markets, East Asia Seminar on Economics*, 2009.
- [42] D. Nguyen E. Perkerson K. Thompson Q. Zhang K. Kuo, P. Luu. Pairs trading: An optimal selling rule. *Mathematical Control and Related Fields*, 5(3):489-499, 2015.
- [43] C. Krauss. Statistical arbitrage pairs trading strategies: Review and outlook. *IWQW Discussion Papers, No. 09/2015*, 2015.
- [44] L. Sarno A. Schrimpf L. Menkhoff, M. Schmeling. Carry trades and global foreign exchange volatility. *Journal of Finance*. 67. 681-718, 2011.
- [45] X. M. Li. *How Do Exchange Rates Co-move? A Study on the Currencies of Five Inflation-targeting Countries*. PhD thesis, School of Economics and Finance, Massey University (Albany), New Zealand, 2010.
- [46] J. Lin. Divergence measures based on the shannon entropy. *IEEE TRANSACTIONS ON INFORMATION THEORY. VOL. 37*, 1991.
- [47] O. Appiah C. N. Bombie M. A. Junior, P. Appiahene. Forex market forecasting using machine learning: Systematic literature review and meta-analysis. *Journal of Big Data*, 2023.
- [48] J.-H. Lee M. Avellaneda. Statistical arbitrage in the us equities market. *Quantitative Finance*, 10(7):761-782, 2010.
- [49] R. Mestel M. Bock. A regime-switching relative value arbitrage rule. *University of Graz, Institute for Banking and Finance*, 2008.
- [50] A. Bucca M. Cummins. Quantitative spread trading on crude oil and refined products markets. *Quantitative Finance*, 12(12):1857-1875, 2012.
- [51] R. K. Lyons M. D. D. Evans. How is macro news transmitted to exchange rates? *Journal of Financial Economics*, 2008.
- [52] D. Peel M. Taylor. Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals. *Journal of International Money and Finance*, 2000.
- [53] Monetary and Economic Department of Bank for International Settlements. Triennial central bank survey: Otc foreign exchange turnover. 2022.
- [54] Y. Chihab N. Mabrouk, M. Chihab. A trading strategy in the forex market based on linear and non-linear machine learning algorithms.
- [55] R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics, 2006.
- [56] L. G. Nielsen. Machine learning for foreign exchange rate forecasting. Master's thesis, Imperial College London, 2018.
- [57] A. S. Paulson P. B. Girma. Risk arbitrage opportunities in petroleum futures spreads. *Journal of Futures Markets*, 19(8):931-955, 1999.
- [58] S. He P. Beaudan. Applying machine learning to trading strategies: Using logistic regression to build momentum-based trading strategies. 2019.

- [59] H. Sravya C. Amarnath Reddy J. Medida P. Bharath Varma, N. Kasheety. Algo-trading using statistical learning and optimizing sharpe ratio and drawdown. *International Journal of Recent Technology and Engineering (IJRTE)*, 2021.
- [60] D. Glassman P. Boothe. The statistical distribution of exchange rates: Empirical evidence and economic implications. *Journal of International Economics*, 22, 297-319, 1987.
- [61] R. J. Sweeney P. Jorion. Mean reversion in real exchange rates: evidence and implications for forecasting. *Journal of International Money and Finance*, 1996.
- [62] N. Hambauer S. Zilker M. Kraus P. Zschech, S. Weinzierl. Gam(e) changer or not? an evaluation of interpretable machine learning models based on additive model constraints. *ECIS 2022 Research Papers*. 106, 2022.
- [63] D. P. Palomar. Pairs trading lectures. *MSc in Financial Mathematics, The Hong Kong University of Science and Technology (HKUST)*, 2020.
- [64] H. Pei. Copula for pairs trading: A unified overview of common strategies. Technical report, Hudson and Thames, 2021.
- [65] K. Rogoff R. A. Meese. Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economics* 14 (1983) 3-24, 1983.
- [66] C. W. J. Granger R. F. Engle. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 1987.
- [67] W. P. Malcolm R. J. Elliott, J. Van Der Hoek. Pairs trading. *Quantitative Finance*, 5(3):271-276, 2005.
- [68] Y. Wu R. Q. Liew. Pairs trading: A copula approach. *Journal of Derivatives and Hedge Funds*, 19(1):12-30, 2013.
- [69] G. Reho. Fx trading with reinforcement learning: an application of fitted q iteration (fqI). Master's thesis, Politecnico di Milano, 2019.
- [70] R. A. Leibler S. Kullback. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86, 1951.
- [71] R. E. Quandt S. M. Goldfeld. A markov model for switching regressions. *Journal of Econometrics*, 1(1), 3-15, 1973.
- [72] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8, 229-231, 1959.
- [73] A. Stan. Computational speed and high-frequency trading profitability: an ecological perspective. *Electron Markets* 28, 381-395, 2018.
- [74] Press release, Zurich, 15 January 2015.
- [75] R. Tibshirani T. Hastie. Generalized additive models. *Statistical Science*, 1(3), 297-310, 1986.
- [76] J. Caicedo-Llano T. Lemishko, A. Landi. Cointegration-based strategies in forex pairs trading. 2024.
- [77] V. Xuang T. Nguyen. Forex trading using supervised machine learning. *International Journal of Engineering and Technology*, 2018.
- [78] Y. A. Kuperin V. V. Kondratenko. Using recurrent neural networks to forecasting of forex. 2003.
- [79] G. Vidyamurthy. *Pairs Trading: Quantitative Methods and Analysis*. Wiley, Hoboken, 2004.
- [80] I. V. Soroka Y. A. Kuperin, L. Dmitrieva. Neural networks in financial markets dynamical studies. *Working paper series of the Center for Management and Institutional Studies. School of Management. Saint-Petersburg State University. No 2001-12*, 2001.
- [81] I. Botha Y. Stander, D. Marais. Trading strategies with copulas. *Journal of Economic and Financial Sciences*, 6(1):83-107, 2013.
- [82] U. G. Erlandsson Y. W. Cheung. Exchange rates and markov switching dynamics. 2004.
- [83] C. Gulati Y.-X. Lin, M. McCrae. Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *Journal of Applied Mathematics and Decision Sciences*, 2006(1):1-14, 2006.

A. Stationarity Test

This Section of the Appendix provides the supporting material used to determine when the Augmented Dickey–Fuller test should be applied strictly or in conjunction with other features. It is an extension of *Subsection 5.1* and of *Subsection 5.5*.

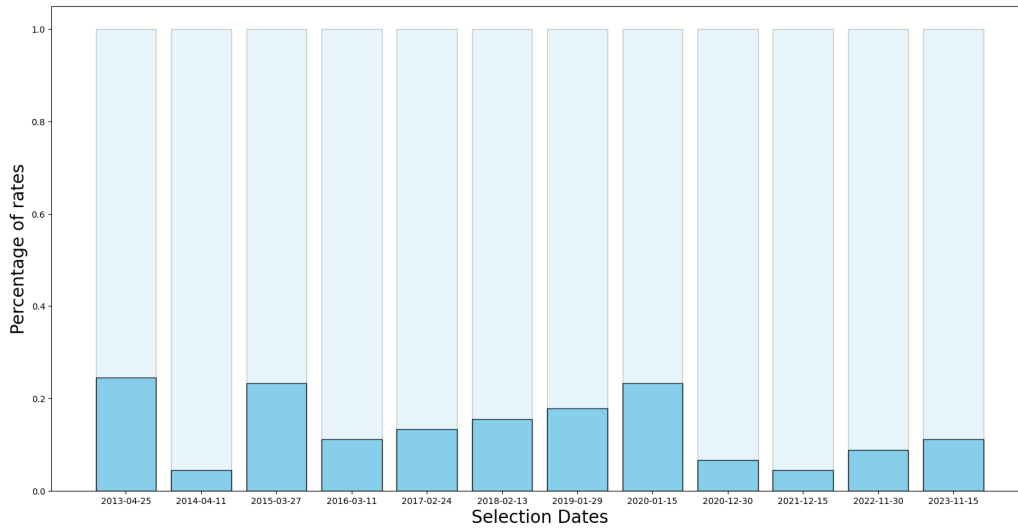


Figure 24: Percentage of rates for which the ADF refuses H_0 for *lookback length*=500 and *selection frequency*=250.



Figure 25: Two 30 days periods where the ADF test has been applied to a 500 long sample of the *EURUSD* exchange rate. The results of the test are highly sensitive to small time shifts.

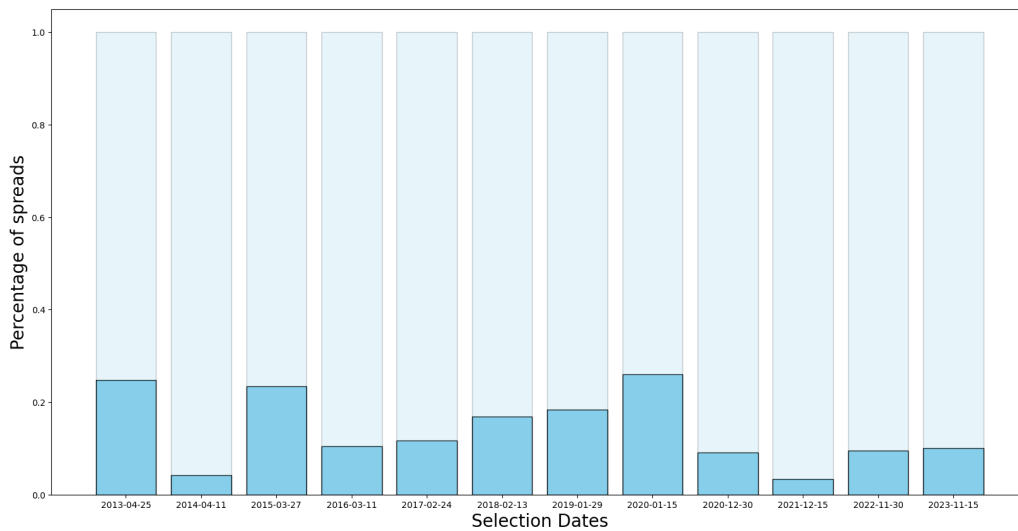


Figure 26: Percentage of spreads for which the ADF refuses H_0 for *lookback length*=500 and *selection frequency*=250.

B. Gaussianity of the Data

This Section of the Appendix is reserved to the material used to motivate the absence of gaussianity in the data. It is an extension of *Subsection 5.2*.

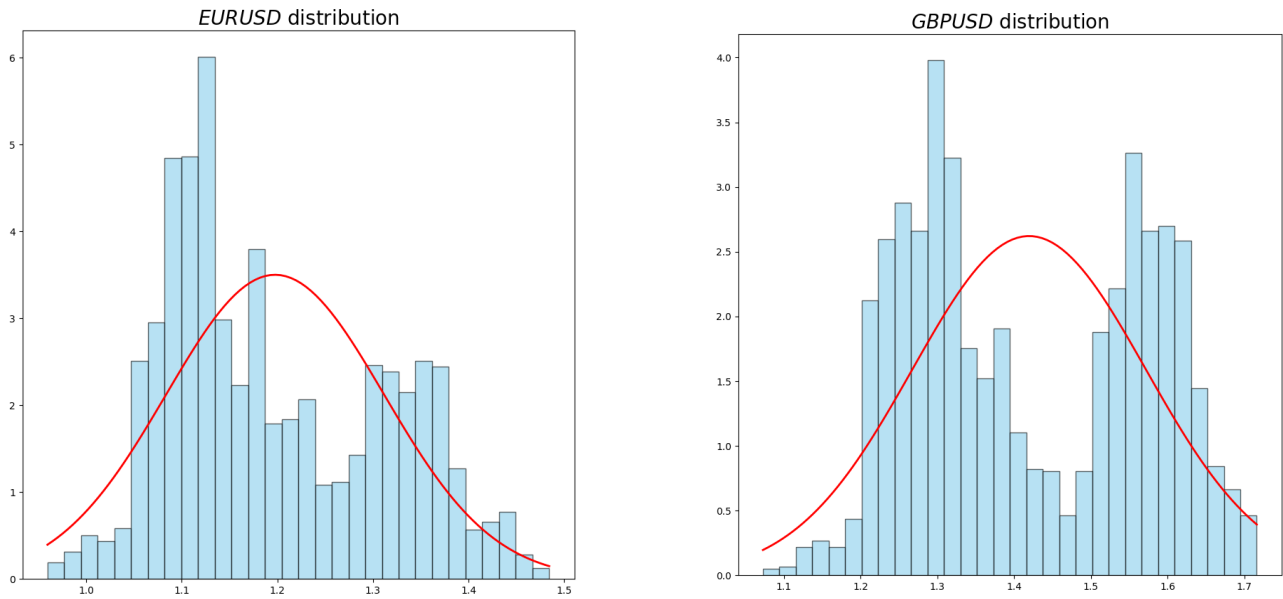


Figure 27: Distribution of the levels of the spot rates of *EURUSD* and *GBPUSD*. It is evident the absence of gaussianity and the presence of multimodal distributions, especially for what regards the second rate.

	Skewness	Kurtosis	p-value
EURUSD	0.44	-0.92	0.00000
GBPUSD	0.11	-1.37	0.00000

Table 1: Some metrics by way of example. The p-value is the one of the Shapiro-Wilk Test.

C. GAM λ Calibration

This Section of the Appendix is reserved to the explanation of the rules followed in *Subsection 5.3* to choose the right value of the smoothing parameter λ . The represented charts are purely qualitative and the considered range for λ is wider than the actually used one. If the error curve is convex and under the intercept line for the majority of the domain, then the elbow point, that is the most distant point of the error curve from the straight line, is also the chosen one (cases 1, 3, 5). If the error curve is concave and above the intercept line for the majority of the domain, then the chosen point is the one before the elbow point (cases 2, 4, 6): because of the particular shape of the curve in these cases, a decrease in the value of λ can lead to a significant drop in the value of the error without compromising too much the smoothness of the function.

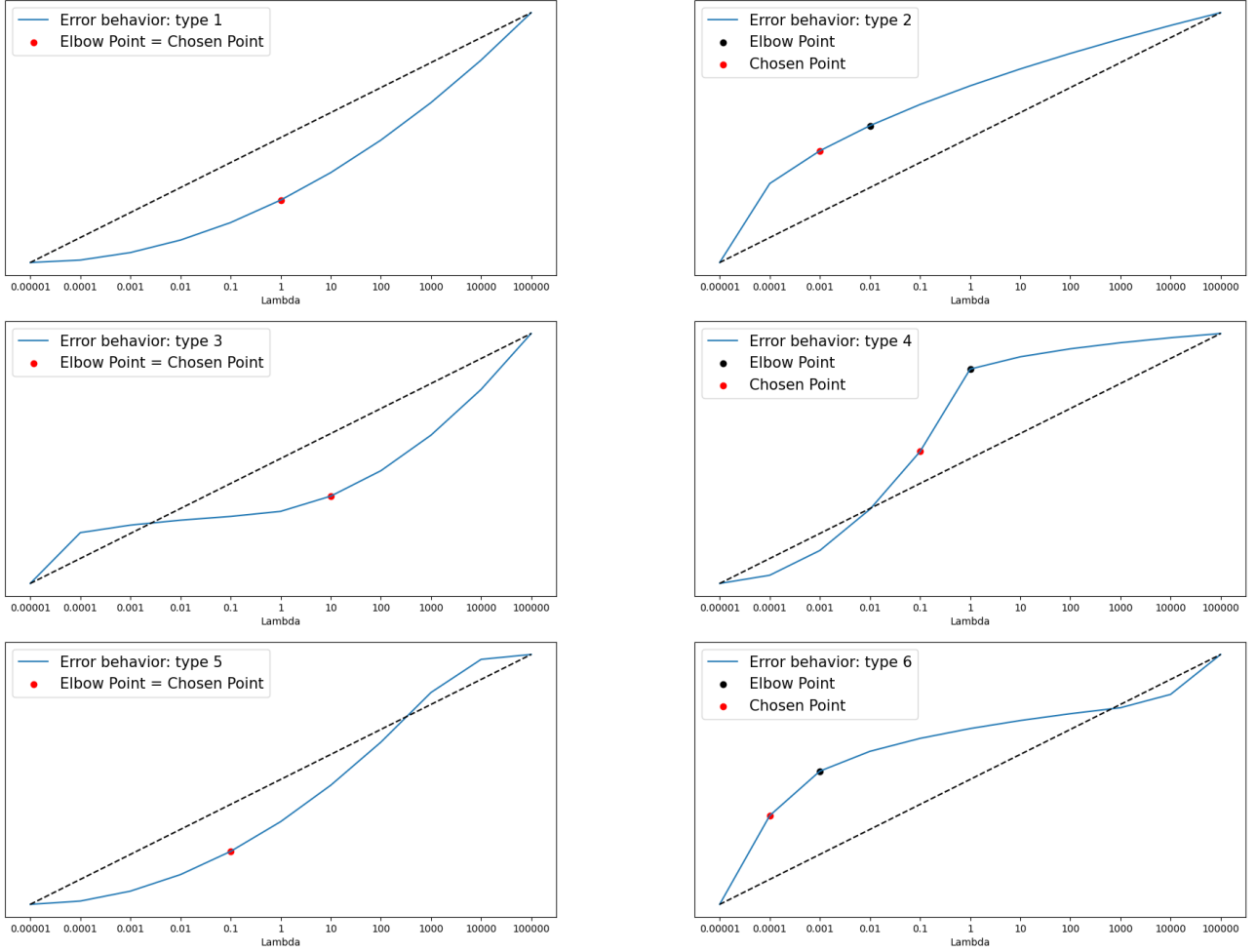


Figure 28: Six behaviors of the error as a function of λ and relative choice.

D. Copula and Time Series

This section of the Appendix provides supporting material to justify the application of copulas to rate time series rather than to their daily return versions, as discussed and implemented in *Subsection 5.2* and *Subsection 6.3*.

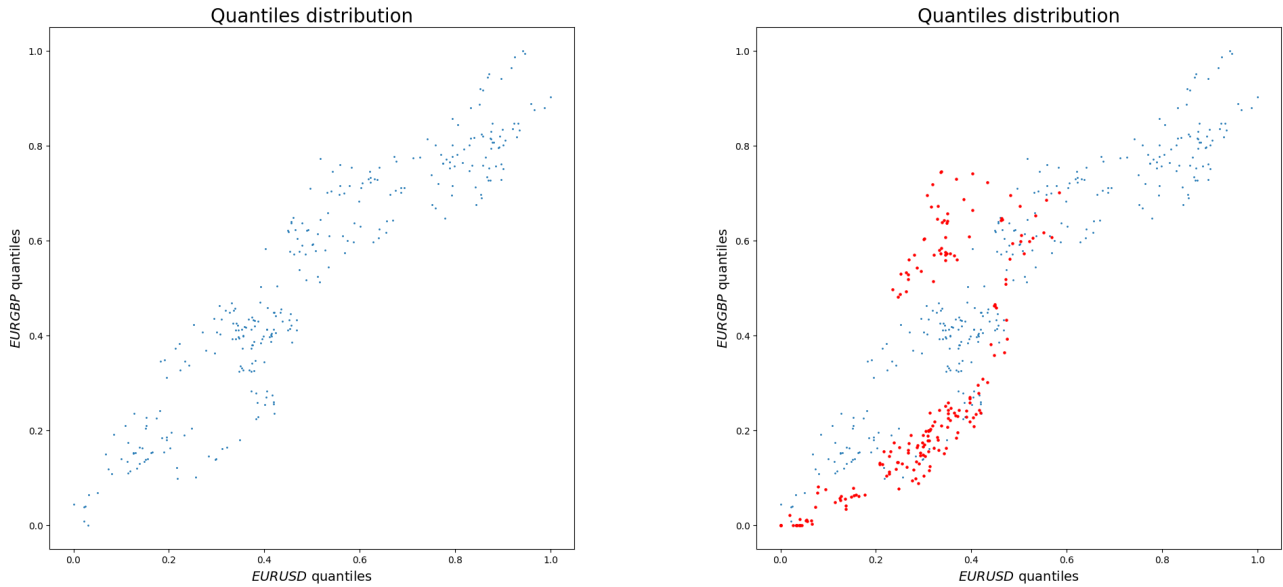


Figure 29: Distribution of the quantiles for the *EURUSD* - *EURGBP* pair between May 25, 2011 and July 17, 2012 (blue) and new observed quantiles between July 18, 2012 and April 24, 2013 (red).

Unlike equity stock time series, the absence of a strong drift in the near future within the exchange rate market legitimizes the application of copulas to the original time series. If a significant drift were present, new values would be compressed toward the boundaries of the $[0, 1]^2$ space, rendering them uninformative from a copula perspective. However, the specific characteristics of stationarity of this market segment allow for this approach.

E. Results

This section of the Appendix provides supporting and broader results whose purpose is to extend what was presented in *Subsection 7*. The reported results refer to the best strategies for each model both from the average return per trade and *Total Profit* perspective. The set of hyperparameters used is reported in the caption of each table.

E.1. Spread Behavior Results

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	2.35	8.64	7.00	13.22	3.05	46.33	1136.42	45.51	6.84	11.42
Number of Trades	876	1246	2034	1094	1576	2414	2230	1332	708	1786
Average Profit per Trade	0.0027	0.0069	0.0034	0.0121	0.0019	0.0192	0.5096	0.0342	0.0097	0.0064
Max Abs Drawdown	-1.14	-0.99	-1.53	-1.41	-2.78	-25.81	-565.80	-10.56	-1.32	-3.72
Drawdown Duration	264	6	403	227	264	519	637	256	264	391
Average Return	0.77%	0.74%	1.12%	0.84%	0.96%	0.52%	1.07%	0.92%	0.46%	0.76%
Return Volatility	53.72%	51.77%	46.76%	54.85%	45.16%	50.04%	57.25%	50.71%	50.28%	57.64%
Highest Return	9.40%	11.73%	12.37%	8.61%	8.73%	11.00%	12.74%	10.76%	7.01%	12.57%
Lowest Return	-18.40%	-12.55%	-20.12%	-8.54%	-18.72%	-9.96%	-12.50%	-11.99%	-9.95%	-10.95%
% Negative Return	31.05%	31.54%	28.47%	40.77%	28.36%	38.40%	27.94%	33.71%	35.88%	39.70%
Return per Trade	0.84%									
Volatility	51.98%									
Total Profit <i>USD</i>	66.26									
Drawdown Impact	36.42%									
% Negative Return	33.34%									

Table 2: Best strategy return perspective, no transaction costs. *train length=300, selection frequency=150, $w = [0.4, 0.4, 0.2]$, $[\delta_{open}, \delta_{close}] = [1.8, 0.7]$.*

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	6.49	1.80	3.57	3.40	9.19	-102.74	1527.65	42.11	4.46	-0.41
Number of Trades	984	610	1252	1006	914	1412	1630	820	556	904
Average Profit per Trade	0.0066	0.0030	0.0029	0.0034	0.0101	-0.0728	0.9372	0.0514	0.0080	-0.0005
Max Abs Drawdown	-1.06	-2.60	-2.53	-3.93	-2.01	-118.56	-99.22	-13.62	-1.95	-9.11
Drawdown Duration	685	1386	214	1388	319	1540	303	1010	203	1264
Average Return	0.59%	0.59%	0.55%	0.90%	0.67%	-0.75%	1.14%	0.98%	1.05%	0.73%
Return Volatility	47.67%	54.80%	50.08%	41.40%	69.70%	58.34%	55.82%	49.72%	45.11%	71.45%
Highest Return	6.61%	6.96%	11.31%	7.71%	9.26%	5.86%	10.39%	8.80%	9.00%	10.52%
Lowest Return	-8.06%	-11.77%	-16.55%	-8.22%	-12.68%	-18.47%	-9.52%	-9.13%	-9.22%	-11.75%
% Negative Return	41.97%	43.11%	38.66%	32.21%	43.98%	46.67%	38.40%	47.32%	35.25%	46.68%
Return per Trade	0.59%									
Volatility	55.46%									
Total Profit <i>USD</i>	31.28									
Drawdown Impact	277.88%									
% Negative Return	41.41%									

Table 3: Best strategy return perspective, transaction costs. *train length=500, selection frequency=250, $w = [0.2, 0.5, 0.3]$, $[\delta_{open}, \delta_{close}] = [1.8, 0.7]$.*

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	29.30	2.34	19.85	20.66	28.30	138.61	4995.11	139.18	29.10	-3.28
Number of Trades	2890	2298	4038	2602	3010	3084	5236	2720	3200	3582
Average Profit per Trade	0.0101	0.0010	0.0049	0.0079	0.0094	0.0449	0.9540	0.0512	0.0091	-0.0009
Max Abs Drawdown	-1.80	-13.27	-2.39	-11.85	-4.29	-30.61	-450.10	-81.87	-5.76	-21.92
Drawdown Duration	280	1134	132	198	822	12	50	198	217	1110
Average Return	0.58%	0.20%	0.27%	0.09%	0.95%	0.44%	0.64%	0.65%	0.37%	-0.11%
Return Volatility	68.80%	67.87%	60.43%	69.54%	59.16%	69.54%	70.95%	70.32%	68.74%	65.79%
Highest Return	13.51%	17.28%	10.76%	14.44%	9.70%	17.89%	14.19%	17.13%	13.33%	15.04%
Lowest Return	-14.47%	-12.03%	-17.05%	-10.91%	-15.25%	-12.74%	-15.93%	-11.88%	-16.14%	-17.79%
% Negative Return	35.78%	48.87%	39.72%	40.32%	30.90%	41.15%	33.90%	43.75%	46.19%	52.74%
Return per Trade	0.42%									
Volatility	67.20%									
Total Profit <i>USD</i>	198.05									
Drawdown Impact	143.70%									
% Negative Return	40.85%									

Table 4: Best strategy profit perspective, no transaction costs. *train length=400, selection frequency=200, $w = [0.45, 0.1, 0.45]$, $[\delta_{open}, \delta_{close}] = [1, 0.1]$.*

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	26.83	0.36	16.28	18.06	25.29	135.52	4990.69	136.46	25.90	-6.86
Number of Trades	2890	2298	4038	2602	3010	3084	5236	2720	3200	3582
Average Profit per Trade	0.0093	0.0002	0.0040	0.0069	0.0084	0.0439	0.9531	0.0502	0.0081	-0.0019
Max Abs Drawdown	-2.10	-14.27	-2.62	-12.44	-4.59	-30.79	-450.23	-82.46	-6.29	-24.26
Drawdown Duration	280	1134	132	198	822	12	50	198	217	1110
Average Return	0.14%	-0.13%	-0.21%	-0.18%	0.56%	0.40%	0.64%	0.61%	-0.60%	-1.84%
Return Volatility	71.16%	67.39%	62.46%	70.35%	60.17%	69.72%	70.96%	70.30%	79.69%	81.59%
Highest Return	13.48%	17.25%	9.83%	14.37%	9.09%	17.88%	14.19%	17.12%	13.27%	14.98%
Lowest Return	-14.86%	-12.06%	-17.20%	-11.45%	-15.35%	-12.75%	-15.93%	-11.89%	-16.60%	-18.28%
% Negative Return	40.59%	53.92%	44.13%	42.85%	33.85%	42.32%	34.51%	44.67%	52.03%	65.97%
Return per Trade	-0.05%									
Volatility	70.77%									
Total Profit <i>USD</i>	177.97									
Drawdown Impact	455.35%									
% Negative Return	44.96%									

Table 5: Best strategy profit perspective, transaction costs. *train length=400, selection frequency=200, $w = [0.45, 0.1, 0.45]$, $[\delta_{open}, \delta_{close}] = [1, 0.1]$.*

E.2. Markov Switching Model Results

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	3.45	-0.86	3.27	7.03	3.49	27.57	886.26	24.50	4.95	6.73
Number of Trades	990	458	876	966	818	1038	1900	824	880	994
Average Profit per Trade	0.0035	-0.0019	0.0037	0.0073	0.0043	0.0266	0.4665	0.0297	0.0056	0.0068
Max Abs Drawdown	-0.50	-2.01	-0.40	-0.81	-1.02	-28.85	-166.68	-7.55	-1.88	-0.67
Drawdown Duration	390	2705	469	276	14	136	314	716	31	37
Average Return	0.82%	0.62%	0.88%	0.83%	0.74%	0.60%	0.49%	0.94%	0.35%	0.69%
Return Volatility	33.49%	51.73%	38.43%	28.87%	50.01%	42.28%	38.66%	37.80%	35.94%	32.71%
Highest Return	6.45%	7.13%	12.24%	6.50%	8.48%	14.17%	7.73%	6.95%	5.94%	9.86%
Lowest Return	-7.93%	-9.96%	-16.40%	-5.52%	-12.58%	-18.46%	-10.46%	-9.12%	-7.66%	-10.46%
% Negative Return	27.68%	42.58%	34.47%	32.09%	31.66%	30.92%	40.47%	33.50%	41.70%	35.11%
Return per Trade	0.68%									
Volatility	38.64%									
Total Profit <i>USD</i>	37.59									
Drawdown Impact	50.40%									
% Negative Return	35.12%									

Table 6: Best strategy return perspective, no transaction costs. *$w = [0.2, 0.5, 0.3]$, $[\delta_{open}, \delta_{close}, \rho] = [1.8, 0.7, 0.8]$.*

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	2.72	-1.16	2.43	6.07	2.67	26.53	884.84	23.67	4.07	5.74
Number of Trades	990	458	874	966	818	1038	1900	824	880	994
Average Profit per Trade	0.0027	-0.0025	0.0028	0.0063	0.0033	0.0256	0.4657	0.0287	0.0046	0.0058
Max Abs Drawdown	-0.5416	-2.2938	-0.4530	-0.8313	-1.0417	-28.9570	-166.7800	-7.7253	-1.9733	-0.7064
Drawdown Duration	314	2705	770	276	14	136	314	716	343	37
Average Return	0.51%	0.41%	0.41%	0.62%	0.40%	0.57%	0.49%	0.91%	0.08%	0.45%
Return Volatility	33.53%	50.56%	39.41%	29.44%	49.99%	42.27%	38.67%	37.77%	36.68%	32.96%
Highest Return	6.27%	6.96%	11.31%	6.39%	7.84%	14.15%	7.73%	6.93%	5.53%	9.81%
Lowest Return	-7.97%	-9.98%	-16.55%	-5.60%	-12.68%	-18.47%	-10.46%	-9.13%	-7.71%	-10.51%
% Negative Return	36.16%	44.32%	44.62%	36.34%	43.15%	32.85%	41.53%	33.98%	47.16%	38.93%
Return per Trade	0.49%									
Volatility	38.79%									
Total Profit <i>USD</i>	32.2									
Drawdown Impact	50.96%									
% Negative Return	39.69%									

Table 7: Best strategy return perspective, transaction costs. $w = [0.2, 0.5, 0.3]$, $[\delta_{open}, \delta_{close}, \rho] = [1.8, 0.7, 0.8]$.

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	18.41	5.30	8.42	16.93	18.65	34.71	326.25	103.89	15.59	6.18
Number of Trades	3016	1318	2792	3030	3090	2880	4902	2034	2850	3060
Average Profit per Trade	0.0061	0.0040	0.0030	0.0056	0.0060	0.0121	0.0666	0.0511	0.0055	0.0020
Max Abs Drawdown	-1.39	-2.88	-1.00	-4.46	-2.63	-95.86	-1937.90	-16.47	-5.65	-10.54
Drawdown Duration	51	343	19	207	30	6	1441	343	273	119
Average Return	0.62%	0.55%	0.74%	0.65%	0.56%	0.41%	0.51%	0.90%	0.31%	0.61%
Return Volatility	44.42%	46.87%	42.25%	38.80%	51.21%	48.14%	61.51%	39.86%	55.59%	52.90%
Highest Return	8.05%	7.13%	12.24%	13.14%	8.65%	18.67%	16.08%	7.59%	12.25%	14.27%
Lowest Return	-9.33%	-12.29%	-16.40%	-13.01%	-12.58%	-18.46%	-15.99%	-10.74%	-12.06%	-12.08%
% Negative Return	33.89%	34.22%	32.74%	35.28%	39.13%	33.33%	42.19%	36.73%	43.86%	36.63%
Return per Trade	0.57%									
Volatility	50%									
Total Profit <i>USD</i>	105.89									
Drawdown Impact	120.70%									
% Negative Return	37.32%									

Table 8: Best strategy profit perspective, no transaction costs. $w = [0.2, 0.5, 0.3]$, $[\delta_{open}, \delta_{close}, \rho] = [1, 0.1, 0.6]$.

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	16.19	4.33	6.02	13.84	15.56	31.83	322.44	101.86	12.74	3.12
Number of Trades	3016	1318	2804	3026	3090	2880	4902	2034	2850	3060
Average Profit per Trade	0.0054	0.0033	0.0021	0.0046	0.0050	0.0111	0.0658	0.0501	0.0045	0.0010
Max Abs Drawdown	-1.4106	-3.1337	-1.0260	-4.7360	-2.7555	-96.0842	-1939.4731	-16.7610	-6.2087	-10.7690
Drawdown Duration	51	343	19	207	282	6	1441	343	306	119
Average Return	0.27%	0.31%	0.26%	0.40%	0.18%	0.38%	0.50%	0.87%	0.04%	0.35%
Return Volatility	46.80%	47.42%	43.56%	39.55%	52.50%	48.13%	61.50%	39.93%	56.49%	52.82%
Highest Return	8.01%	6.96%	11.31%	13.05%	8.54%	18.65%	16.07%	7.58%	12.16%	13.78%
Lowest Return	-9.36%	-12.31%	-16.55%	-13.70%	-12.68%	-18.47%	-15.99%	-10.75%	-12.74%	-12.13%
% Negative Return	40.02%	40.36%	43.22%	41.31%	45.28%	37.05%	43.29%	37.36%	49.16%	42.68%
Return per Trade	0.35%									
Volatility	50.61%									
Total Profit <i>USD</i>	88.83									
Drawdown Impact	146.41%									
% Negative Return	42.29%									

Table 9: Best strategy profit perspective, transaction costs. $w = [0.2, 0.5, 0.3]$, $[\delta_{open}, \delta_{close}, \rho] = [1, 0.1, 0.6]$.

E.3. Copula Results

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	4.57	-6.95	7.52	-25.69	37.95	106.12	2305.46	59.94	-10.20	34.12
Number of Trades	1339	1800	2162	1485	1699	2477	1241	2107	1124	2026
Average Profit per Trade	0.0034	-0.0039	0.0035	-0.0173	0.0223	0.0428	1.8577	0.0284	-0.0091	0.0168
Max Abs Drawdown	-1.40	-9.30	-3.00	-25.69	-0.80	-81.91	-287.74	-40.13	-15.42	-2.17
Drawdown Duration	1356	3840	852	3891	99	60	228	245	405	43
Average Return	1.63%	-0.58%	0.68%	-0.85%	1.51%	0.89%	0.43%	0.14%	0.38%	0.94%
Return Volatility	44.34%	56.84%	48.27%	68.17%	75.51%	46.42%	87.30%	56.38%	87.55%	37.24%
Highest Return	9.44%	9.08%	9.49%	12.75%	16.22%	8.79%	10.43%	12.49%	16.73%	9.04%
Lowest Return	-4.73%	-11.11%	-15.84%	-12.67%	-10.65%	-11.03%	-11.83%	-11.25%	-12.67%	-5.30%
% Negative Return	35.32%	51.22%	38.71%	55.42%	37.14%	32.54%	49.56%	46.42%	40.75%	36.03%
Return per Trade	0.53%									
Volatility	60.07%									
Total Profit <i>USD</i>	89									
Drawdown Impact	62.08%									
% Negative Return	41.66%									

Table 10: Best strategy return perspective, no transaction costs. $[p_{up}, p_{low}, tolerance_{GAM}] = [0.875, 0.125, 0.25]$, or close logic, $[w_{GAM}, w_{copula}] = [0.33, 0.67]$.

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	3.73	-8.25	5.77	-27.17	36.25	103.64	2304.80	57.84	-11.32	32.09
Number of Trades	1339	1800	2162	1485	1699	2477	1241	2107	1124	2026
Average Profit per Trade	0.0028	-0.0046	0.0027	-0.0183	0.0213	0.0418	1.8572	0.0274	-0.0101	0.0158
Max Abs Drawdown	-1.80	-10.48	-3.24	-27.17	-1.05	-82.13	-287.86	-40.23	-16.02	-2.28
Drawdown Duration	2314	3840	852	3891	420	60	228	245	405	43
Average Return	0.10%	-1.10%	0.11%	-2.47%	0.51%	0.83%	0.43%	-0.01%	0.06%	0.54%
Return Volatility	41.81%	62.42%	52.03%	71.20%	100.49%	46.53%	87.32%	58.62%	86.60%	46.64%
Highest Return	6.60%	8.91%	9.29%	6.98%	16.16%	8.77%	10.43%	12.47%	16.18%	8.97%
Lowest Return	-8.08%	-13.60%	-15.88%	-15.86%	-18.82%	-11.05%	-11.84%	-11.90%	-13.18%	-13.56%
% Negative Return	51.68%	55.78%	45.10%	66.94%	40.44%	33.95%	49.88%	47.03%	42.53%	39.78%
Return per Trade	-0.04%									
Volatility	65.46%									
Total Profit <i>USD</i>	78.69									
Drawdown Impact	64.43%									
% Negative Return	46.32%									

Table 11: Best strategy return perspective, transaction costs. $[p_{up}, p_{low}, tolerance_{GAM}] = [0.875, 0.125, 0.25]$, or close logic, $[w_{GAM}, w_{copula}] = [0.33, 0.67]$.

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	3.65	3.98	11.76	-26.47	31.99	54.95	1155.19	124.82	1.79	41.55
Number of Trades	1310	1777	2132	1465	1691	2459	1179	2063	1088	2022
Average Profit per Trade	0.0028	0.0022	0.0055	-0.0181	0.0189	0.0223	0.9798	0.0605	0.0016	0.0205
Max Abs Drawdown	-1.79	-4.30	-2.61	-26.47	-2.29	-95.45	-860.64	-6.39	-1.41	-2.17
Drawdown Duration	1453	65	420	3891	400	64	64	325	3106	43
Average Return	1.28%	-0.03%	1.05%	-1.04%	1.01%	0.28%	-0.72%	0.43%	0.32%	1.24%
Return Volatility	43.17%	65.34%	51.11%	66.23%	78.13%	49.41%	112.65%	61.81%	58.92%	39.97%
Highest Return	9.44%	12.69%	11.30%	9.18%	16.22%	9.37%	10.43%	12.74%	11.65%	10.92%
Lowest Return	-5.63%	-17.47%	-10.31%	-12.67%	-11.18%	-11.03%	-21.17%	-17.46%	-9.08%	-5.30%
% Negative Return	36.79%	45.64%	38.41%	57.13%	41.40%	40.38%	47.24%	43.43%	45.68%	30.91%
Return per Trade	0.44%									
Volatility	63%									
Total Profit <i>USD</i>	102.28									
Drawdown Impact	62.40%									
% Negative Return	41.99%									

Table 12: Best strategy profit perspective, no transaction costs. $[p_{up}, p_{low}, tolerance_{GAM}] = [0.875, 0.125, 0.25]$, and close logic, $[w_{GAM}, w_{copula}] = [0.33, 0.67]$.

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	2.82	2.69	10.03	-27.93	30.30	52.49	1154.56	122.76	0.70	39.52
Number of Trades	1310	1777	2132	1465	1691	2459	1179	2063	1088	2022
Average Profit per Trade	0.0022	0.0015	0.0047	-0.0191	0.0179	0.0213	0.9793	0.0595	0.0006	0.0195
Max Abs Drawdown	-2.23	-4.68	-2.79	-27.93	-2.46	-95.92	-860.91	-6.73	-1.76	-2.28
Drawdown Duration	2525	770	420	3891	400	64	64	662	590	43
Average Return	-0.19%	-0.52%	0.50%	-2.68%	0.01%	0.23%	-0.72%	0.29%	0.00%	0.84%
Return Volatility	44.59%	73.12%	56.01%	74.16%	101.52%	50.41%	112.67%	64.97%	58.71%	47.72%
Highest Return	6.60%	12.67%	11.23%	6.98%	16.16%	9.36%	10.43%	12.73%	11.21%	10.84%
Lowest Return	-8.08%	-20.55%	-10.35%	-15.86%	-18.82%	-11.05%	-21.17%	-18.78%	-9.56%	-13.56%
% Negative Return	56.87%	50.48%	42.87%	68.33%	45.30%	40.67%	47.33%	44.16%	48.99%	35.91%
Return per Trade	-0.12%									
Volatility	69.31%									
Total Profit <i>USD</i>	92.08									
Drawdown Impact	90.89%									
% Negative Return	46.85%									

Table 13: Best strategy profit perspective, transaction costs. $[p_{up}, p_{low}, tolerance_{GAM}] = [0.875, 0.125, 0.25]$, and close logic, $[w_{GAM}, w_{copula}] = [0.33, 0.67]$.

E.4. Logistic Regression Results

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	4.33	14.98	6.88	28.77	43.50	387.11	1607.61	131.89	11.74	34.27
Number of Trades	2569	1865	3629	2554	3546	3511	1583	2053	6360	4207
Average Profit per Trade	0.0017	0.0080	0.0019	0.0113	0.0123	0.1103	1.0155	0.0642	0.0018	0.0081
Max Abs Drawdown	-6.94	-13.31	-7.37	-5.09	-5.75	-102.06	-1616.55	-102.04	-26.17	-13.02
Drawdown Duration	675	1345	1918	110	424	264	491	971	776	237
Average Return	0.13%	1.05%	0.39%	1.02%	2.00%	2.09%	0.25%	0.83%	0.19%	1.40%
Return Volatility	78.18%	70.82%	78.32%	69.27%	80.66%	78.29%	85.52%	89.72%	83.90%	68.45%
Highest Return	7.24%	7.05%	14.16%	7.10%	16.19%	19.17%	8.55%	19.20%	18.67%	8.05%
Lowest Return	-16.74%	-14.05%	-19.15%	-15.69%	-17.45%	-15.20%	-13.57%	-15.02%	-17.61%	-13.13%
% Negative Return	44.57%	37.86%	45.16%	34.42%	35.03%	32.01%	45.80%	42.33%	42.42%	35.06%
Return per Trade	0.94%									
Volatility	78.57%									
Total Profit <i>USD</i>	215.39									
Drawdown Impact	85.27%									
% Negative Return	39.22%									

Table 14: Best strategy return and profit perspective, no transaction costs. $w = [0.2, 0.5, 0.3]$, $stop\ loss=15\%$.

	<i>EUR</i>	<i>USD</i>	<i>GBP</i>	<i>CAD</i>	<i>CHF</i>	<i>CNY</i>	<i>JPY</i>	<i>HKD</i>	<i>NZD</i>	<i>AUD</i>
Currency Profit	2.35	13.27	4.00	26.21	39.95	383.60	1606.43	129.83	5.37	30.06
Number of Trades	2569	1866	3629	2557	3549	3513	1583	2054	6362	4210
Average Profit per Trade	0.0009	0.0071	0.0011	0.0103	0.0113	0.1092	1.0148	0.0632	0.0008	0.0071
Max Abs Drawdown	-7.41	-13.87	-8.78	-5.20	-6.23	-102.41	-1616.92	-102.42	-27.94	-13.58
Drawdown Duration	675	1345	1918	110	424	264	491	971	777	250
Average Return	-0.60%	0.88%	-0.69%	-0.73%	1.58%	1.99%	0.25%	0.71%	-0.37%	0.11%
Return Volatility	79.98%	70.76%	84.03%	92.79%	83.53%	78.53%	85.54%	90.10%	85.59%	80.54%
Highest Return	7.09%	6.92%	10.02%	6.69%	16.10%	19.16%	8.55%	19.18%	18.59%	7.02%
Lowest Return	-18.96%	-14.18%	-24.05%	-24.98%	-17.59%	-15.22%	-13.57%	-15.04%	-17.69%	-19.15%
% Negative Return	48.70%	39.44%	48.94%	45.17%	37.62%	32.39%	45.80%	42.84%	47.08%	44.77%
Return per Trade	0.26%									
Volatility	83.44%									
Total Profit <i>USD</i>	194.02									
Drawdown Impact	144.60%									
% Negative Return	43.51%									

Table 15: Best strategy return and profit perspective, transaction costs. $w = [0.2, 0.5, 0.3]$, $stop\ loss=15\%$.

Abstract in lingua italiana

In seguito al crollo degli accordi di Bretton Woods nel 1971, i valori dei tassi di cambio tra le valute hanno iniziato a fluttuare liberamente, guidati dalle dinamiche di domanda e offerta. Negli ultimi anni, il mercato dei tassi di cambio si è affermato come un terreno fertile per il trading algoritmico e per l'utilizzo di tecniche di machine learning anche grazie alla sua elevata liquidità. Inoltre, un crescente corpus di letteratura evidenzia come alcuni tassi di cambio tendano a muoversi similmente e, individualmente, tendano a seguire dinamiche di mean reversion.

Il presente studio si propone di sfruttare temporanee deviazioni nei comovimenti delle coppie di tassi di cambio, nonché il comportamento oscillatorio dei singoli tassi, attraverso l'applicazione di tecniche statistiche e di machine learning. Un aspetto chiave di questo approccio è un'accurata selezione sia delle coppie di tassi di cambio sia dei tassi individuali, ottenuta mediante metodologie innovative adattate alle diverse strategie di trading.

Parole chiave: mercato delle valute; arbitraggio statistico; trading di coppie; ritorno alla media; trading algoritmico; copula; regressione logistica; spread; modello markoviano di cambio di regime