Pairs Trading in the Foreign Exchange Rate Markets

Tan Hoang

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science
Quantitative Finance and Actuarial Science

Department of Tilburg School of Economics and Management

Tilburg University

2022

Supervisor:
Prof. dr. P. Cizek

Second Reader:
Prof. dr. B.J.M. Werker

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Abstract

Pairs trading strategies are market-neutral trading strategies which are used by investors to take hedged positions using co-movement of asset prices. We study the pairs trading strategy in the foreign exchange rate markets considering 26 different forex pairs using 1-hour intraday observations. We find that several pairs trading strategies in the forex markets are able to generate positive returns even after considering transaction costs. There are two formation methods used, one based on minimal distance and one based on cointegration of time-series. Trading models consist of thresholds for trading signals and we study the use of recurrent neural networks for forecasting signals that are supplied to the trading models. The results suggests that the trading model consisting of cointegration as formation method with thresholds trading signals is the most optimal strategy and generates a cumulative net return of 16.48% over a six year out-of-sample time horizon from 2015 to 2021. After filtering out the Swiss Franc and the Japanese Yen pairs from the dataset, we find that the same strategy performs better with a net cumulative return of 42.53% over the same out-of-sample period.

Contents

1 Introduction **1**
1.1 Literature Overview . . . 2

2 Data **5**

3 Methodology **6**
3.1 Pairs formation methods . . . 7
3.1.1 Distance method approach . . . 7
3.1.2 Cointegration approach . . . 8
3.2 Neural Networks . . . 10
3.2.1 Long Short-Term Memory Networks . . . 11
3.3 Optimizing trading parameters . . . 15
3.4 Trading Models . . . 17
3.4.1 Transaction costs and returns . . . 19

4 Empirical Results **21**
4.1 In-sample trading parameter optimization . . . 21
4.2 Out-of-sample trading performance . . . 27
4.2.1 LSTM performance . . . 29
4.2.2 Trading model robustness . . . 30

5 Conclusion **34**

References **36**

Appendices **39**

A Time-series plots H1 spot prices **40**

B Time-series plots H1 returns **49**

C Pair formation statistics **58**

List of Tables

2.1 Descriptive statistics for the return series using 1-hour closing bid prices for the 26 FX pairs from January 3 2006 until November 1 2021. 5

4.1 DM-THRESHOLD strategy: in-sample trading results. 23
4.2 COINT-THRESHOLD strategy: In-sample trading results . . 24
4.3 COINT-DIF-LSTM: In-sample results for the LSTM strategy using differenced return time-series for pairs trading. MSE denotes the average forecasting MSE during the trading periods in-sample. 25
4.4 COINT-REG-LSTM: In-sample trading results. MSE denotes the average forecasting MSE during the trading periods in-sample. 26
4.5 Out-of-sample cumulative trading results for each strategy. Note that no MSE is calculated for the DM and COIN strategies. 27
4.6 Out-of-sample pairs trading results using two different filters: (i) Swiss Franc and Japanese Yen pairs removed and (ii) only training the neural networks on data after 2010. 33

C.1 Statistics of the cointegration formation method. The min. and max. values refer to the minimum and maximum number of eligible significant pairs found in one formation period during the whole sample. 58

List of Figures

3.1 Example of a MLP structure using the previous L lags as input with two hidden layers of equal node size and output node. y_t could for example denote the one step ahead forecast output of the model. 10
3.2 Example of a LSTM structure using the previous L lags as input with one LSTM layer connected to a standard layer and output layer. Note that the red arrow illustrates how past information is passed through a LSTM layer. Note that this structure can happen at more layers but for illustration purposes of this example it is a single layer. 13
3.3 Schematic of one memory LSTM cell showing its interconnections in relation with its previous, next and output cells along with the modification gates. The red line shows the cell state which is passed through the LSTM layer memory cells. 14
3.4 In-depth schematic of one memory LSTM cell illustrating the gates and their connections through the cell. The ”sigm” and ”tanh” stand for the sigmoid and tanh activation functions, respectively. Lastly, the $\otimes$ and $\oplus$ denote concatenations using multiplication and addition, respectively. 14
3.5 The methodology used for finding optimal trading parameters using the training and test split. The Train and Test segments illustrate the length of the formation and trading period respectively. 16
3.6 Out-of-sample trading period example showing the normalized spread and the thresholds ($K = 1$) for the COINT-THRESHOLD strategy. The blue boxes show opening and closing of the long position taken. The red boxes show the opening and the closing of the short position taken. Note that the last short is closed due to reaching the end of the trading period. 19

4.1 Out-of-sample cumulative returns of the distance method over the period of Apr-2015 to Nov-2021. 28
4.2 Out-of-sample cumulative returns of the cointegration method over the period of Apr-2015 to Nov-2021. 28
4.3 Out-of-sample cumulative returns of the D-LSTM method over the period of Apr-2015 to Nov-2021. 28
4.4 Out-of-sample cumulative returns of the R-LSTM method over the period of Apr-2015 to Nov-2021. 29

4.5 LSTM fit on differenced return time-series of the cointegrated pair EURCHF/GBPNZD. 29
4.6 COINT-REG-LSTM fitting performance on CADJPY/GBPUSD. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period. 30
4.7 COINT-REG-LSTM fitting performance on EURJPY/USDJPY. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period. 30
4.8 COINT-REG-LSTM fitting performance on EURAUD/GBPNZD. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period. 31
4.9 COINT-REG-LSTM fitting performance on EURAUD/GBPCAD. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period. 31
4.10 COINT-REG-LSTM fitting performance on CADJPY/GBPUSD. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period. 32
4.11 COINT-REG-LSTM fitting performance on EURJPY/USDJPY. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period. 32

A.1 1-hour AUDCAD spot price 40
A.2 1-hour AUDCHF spot price 40
A.3 1-hour AUDJPY spot price 41
A.4 1-hour AUDUSD spot price 41
A.5 1-hour CADCHF spot price 41
A.6 1-hour CADJPY spot price 42
A.7 1-hour EURAUD spot price 42
A.8 1-hour EURCAD spot price 42
A.9 1-hour EURCHF spot price 43
A.10 1-hour EURGBP spot price 43
A.11 1-hour EURJPY spot price 43
A.12 1-hour EURNZD spot price 44
A.13 1-hour EURUSD spot price 44
A.14 1-hour GBPAUD spot price 44
A.15 1-hour GBPCAD spot price 45
A.16 1-hour GBPCHF spot price 45
A.17 1-hour GBPJPY spot price 45
A.18 1-hour GBPNZD spot price 46
A.19 1-hour GBPUSD spot price 46
A.20 1-hour NZDCAD spot price 46
A.21 1-hour NZDCHF spot price 47
A.22 1-hour NZDJPY spot price 47
A.23 1-hour NZDUSD spot price 47
A.24 1-hour USDCAD spot price 48
A.25 1-hour USDJPY spot price 48

B.1 AUDCAD hourly returns 49
B.2 AUDCHF hourly returns 49
B.3 AUDJPY hourly returns 50

B.4 AUDUSD hourly returns 50
B.5 CADCHF hourly returns 50
B.6 CADJPY hourly returns 51
B.7 EURAUD hourly returns 51
B.8 EURCAD hourly returns 51
B.9 EURCHF hourly returns 52
B.10 EURGBP hourly returns 52
B.11 EURJPY hourly returns 52
B.12 EURNZD hourly returns 53
B.13 EURUSD hourly returns 53
B.14 GBPAUD hourly returns 53
B.15 GBPCAD hourly returns 54
B.16 GBPCHF hourly returns 54
B.17 GBPJPY hourly returns 54
B.18 GBPNZD hourly returns 55
B.19 GBPUSD hourly returns 55
B.20 NZDCAD hourly returns 55
B.21 NZDCHF hourly returns 56
B.22 NZDJPY hourly returns 56
B.23 NZDUSD hourly returns 56
B.24 USDCAD hourly returns 57
B.25 USDJPY hourly returns 57

1. Introduction

Pairs trading originates back to the 80s where investment banks and hedge funds looked for market neutral trading strategies. Pairs trading can be considered as statistical arbitrage. It involves first finding a pair of assets that have been shown to move together historically. When the spread between the two assets is large enough, a trade is initiated: long the undervalued asset and short the overvalued asset. So a trader speculates on the reversion of prices of the two assets taking both long and short positions.

Prior literature has not fully studied the most liquid financial market that being the traditional currencies. Furthermore, existing research in pairs trading is heavily focused on equities, commodities and derivatives markets using daily data. So, the main focus in prior work is related to different markets and time-frames. This thesis will study the profitability of the pairs trading method in the forex markets on an hourly time-frame.

Therefore, the contribution of this thesis to the field of pair trading is two-fold: first, verify whether pairs trading in the current forex (FX) markets is profitable on an 1-hour intraday time-frame due to prior research suggesting declining profitability of the strategy in recent years. Second, we are going to study different trading strategies that could be applied to pairs trading. Two of these strategies involve the use of neural networks forecasts.

Furthermore, the distance method and cointegration method will be used as formation methods, where the distance method will serve as benchmark. Moreover, the cointegration method has been studied thoroughly and it has been shown by several authors that is an effective and profitable method for forming pairs.

In conclusion, this thesis studies pairs trading in forex markets by using intraday hourly data where the pairs will be formed using minimal distances and cointegration. Statistical thresholds and neural network forecasting will be used for trading models. Note that cointegration tests will be carried out using Engle-Granger's method.

Overall, we find that cointegration as formation method along with a threshold pairs trading model is profitable in recent years.

This thesis is structured as follows. First, a literature review is given in Section 1.1 followed by a brief overview of the data used in Section 2. In Section 3, the formation methods (Sections 3.1.1 and 3.1.2), neural networks (Section 3.3) and trading models (Section 3.4) will be discussed. Lastly, we will discuss the performance of the trading strategies in Section 4.

1.1 Literature Overview

Classical pairs trading strategies have been documented by Gatev, William and Rouwenhorst (2006) which has been one of the most cited papers in the field of pairs trading. Pairs trading has been studied in many different financial markets: equities, commodities, derivatives, currencies but also cryptocurrencies (Fil & Kristoufek, 2020; Lintilhac & Tourin, 2016).

Research suggests that pairs trading is sensitive to financial stability (Rad et al. 2016). Baronyan et al. (2010) have shown that pairs trading during the financial crisis in 2008 is more profitable compared to standard economic regimes and argued that less competition in the markets during financial turmoil is the key driver. Moreover, the performance of the strategy also depends on which time-frame the method is deployed. Stübinger and Endres (2017) have shown that using 1-minute data from S&P 500 oil companies yield 60.61% annualized return after transaction costs, which is substantially higher compared to the excess returns found using daily data in existing literature. Profitable pairs trading using high frequency data and Euclidean distance is suggested by the findings of Stübinger and Bredthauer (2017).

There are several formation methods to find suitable pairs with the distance method and cointegration method being the most researched ones.

First, the method of minimum distance studied by Gatev et al. (2006) is a logical approach for finding pairs. They found pairs by minimising the sum of squared differences between two normalized asset prices using daily CRSP US equities data. Their trading strategy simply involved waiting for the spread between the pair to go exceeds two standard deviations to initiate the trades. Yet, Do and Faff (2010) showed that the distance method became less profitable in recent years. Specifically, they show that in the sub-periods 1962-1988, 1989-2002 and 2003-2008 the mean excess returns generated by this pairs trading strategy was 1.24%, 0.66% and 0.35%, respectively.

Second, the method of cointegration has been studied by Vidyamurthy (2004) and is one of the pioneers of the method. The main concept of cointegration is to test whether a pair of assets have a stationary and mean-reverting time series. This test is generally done by using the Engle-Granger test. Indications of cointegration implies that these assets should form a pair in the trading period. Moreover, Rad et al. (2016) have studied the method of cointegration but also the Distance Method on US stocks. Interestingly, they found that these two methods perform better in times of increased market volatility. Chen et al. (2018) studied pairs trading in Chinese commodity futures markets using daily data and showed that the cointegration approach yield profitable returns in Chinese commodity markets and thus also in emerging markets. Caldeura and Moura (2013) used cointegration methods in Brazilian equity markets and found excess returns of 16.38% per year also using daily closing prices. In addition, cointegration methods and distance methods have been compared in stock markets. Results suggests that cointegration methods outperform and generate high excess returns, whereas distance methods generate insignificant excess returns, controlling for transaction and various costs (Huck & Afawubo, 2016). In conclusion, previous studies have shown the effectiveness of cointegration method.

Furthermore, other methods for the formation of pairs have been studied as well. Discussions on copula methods can be found in, Xie et al. (2016) and Liew and Wu (2013). As this copula method is still not used frequently in pairs trading studies, we will focus on the most effective formation methods that have been documented by several sources in the field of pairs trading, which is the distance and cointegration method.

Lastly, after the formation of pairs, different trading strategies can be deployed. One classical strategy is the strategy proposed by Gatev et al. (2006), where they opened trades if the spread of a pair exceeds K standard deviations and closed trades as soon as the prices would cross (when the spread is equal to zero). Second, Chen et al. (2018) have used a rolling window to test cointegration spread. The rolling window is a fixed time window used for formation of pairs prior to the trading period. Specifically, they enter a position if the spread is K standard deviations above the mean and close the position if the spread is K standard deviations below the long term mean, where the long term mean and standard deviations are derived from the rolling window. This strategy attempts to maximize return by holding trades for a longer period compared to the strategy by Gatev et al. (2006). Interestingly, they also studied a strategy using a stop-loss order to attempt to minimize losses, yet they found that the returns are worse in terms of return and risk trade off (Chen et al., 2008).

Introducing machine learning methods like Principal Component Analysis to reduce dimensionality has been studied by Sarmento and Horta (2020).

The use of (recurrent) neural networks have also been implemented in pairs trading strategies. Dunis et al. (2006) have shown profitability of pairs trading when using a multilayer perceptron (MLP), also known as neural networks, to predict the future spread returns. Their trading strategy consist of entering trades if the *predicted spread change* by the neural network is above some threshold filter K and close the position when the spread forecast is in between K and $-K$, similar to other classical pairs trading strategies. Note that they used in-sample data to determine the optimal threshold K.

Flori and Regoli (2021) and Fischer and Krauss (2018) have shown success of implementing recurrent neural networks for financial time-series prediction which could be supplied to all sorts of trading strategies, including the pairs trading strategy.

Furthermore, Kim and Kim (2019) adopted an advanced reinforcement learning algorithm called DQN-learning to pairs trading. The idea is to formulate pairs trading as a game where the agent has to optimize stop-loss and entry boundaries using deep neural networks. The agent interacts with the states, being the spread, and gets rewarded determining profit and stop-loss boundaries and gets punished for sub-optimal decisions. This strategy, however, will not be used in this thesis because the goal of this thesis is to verify whether pairs trading will work in FX markets in general. The latter strategy could be an extension for future research.

Lastly, there is a strategy class which involves using a stochastic spread method and it has been documented frequently in the field of pairs trading (Göncü & Akyildirim, 2016; Yang et. al, 2016; Elliott et al., 2005; Stübinger & Endres, 2017). The method involves modelling the co-movement of asset prices as a mean-reverting stochastic model like the Ornstein-Uhlenbeck process and entering long and short positions whenever the spread exceeds some bound.

Note that both the stochastic spread method and neural networks both attempt to forecast spread movements. Yet, it is decided that in this thesis the neural networks will be studied, as we believe that financial spread time series are complex which require models with a high number of parameters such as neural networks.

This thesis will apply and study four different strategies on forex pairs, where the benchmark strategy will be the one by Gatev et al. (2006) using a minimal distances for formation and thresholds for trading.

The second strategy will be similar to the industry standard mentioned in the literature above using cointegration for formation and thresholds for trading. Next, a recurrent network is used to predict the returns of the cointegrated spread time-series, similar to the approach used by Dunis et al. (2006). Lastly, a new strategy in the forex pairs trading field will be studied which involves using a recurrent neural network to predict the spread itself acting as a regression model and executing trades based on thresholds.

2. Data

The foreign exchange rates are retrieved from the openly available database of Dukascopy Bank SA. This thesis will use hourly closing bid prices, although the data contains open, high, low, close and volume data. The sample spans from January 3, 2006 until November 1, 2021 containing 94,713 hourly closing bid prices per foreign exchange rate pair of which a total of 26 rates will be used. This sample is chosen since it captures many periods of financial currency volatility. To name a few examples, the crisis in 2008, interventions by central banks, like the Swiss National Bank and Bank of Japan, and the COVID-19 stock crash. It should also be noted that only the observations when *all* foreign exchange rate pairs are traded are considered here. In other words, the data is removed when at least one pair does not have an observation for that date. Below are the descriptive statistics for all time series in the dataset. Moreover, plots of the spot prices and their differenced series can be found in Appendix A and B. One can clearly see that the crisis in 2008 and the financial shock caused by the COVID-19 pandemic affected all pairs. In the CHF pairs, we see volatile returns due to interventions by the central bank to regulate the value of the currency.

Table 2.1: Descriptive statistics for the return series using 1-hour closing bid prices for the 26 FX pairs from January 3 2006 until November 1 2021.

Forex pair	observations	mean	std. dev.	min.	max.
AUDCAD	94713	0.9553	0.0581	0.7242	0.9553
AUDCHF	94713	0.8247	0.119	0.5384	0.8247
AUDJPY	94713	84.5107	8.8567	55.655	84.5107
AUDUSD	94713	0.8304	0.1214	0.5552	0.8304
CADCHF	94713	0.8656	0.1346	0.6614	0.8656
CADJPY	94713	88.6017	9.5619	68.875	88.6017
EURAUD	94713	1.5294	0.1568	1.161	1.5294
EURCAD	94713	1.4538	0.0887	1.2134	1.4538
EURCHF	94713	1.2555	0.1914	0.9752	1.2555
EURGBP	94713	0.8289	0.0673	0.6566	0.8289
EURJPY	94713	128.7794	15.3509	94.209	128.7794
EURNZD	94713	1.7512	0.2008	1.3893	1.7512
EURUSD	94713	1.257	0.1273	1.0356	1.257
GBPAUD	94713	1.8585	0.2574	1.4389	1.8585
GBPCAD	94713	1.7639	0.1659	1.484	1.7639
GBPCHF	94713	1.5339	0.3311	1.113	1.5339
GBPJPY	94713	157.3459	30.0238	116.949	157.3459
GBPNZD	94713	2.1278	0.3113	1.6718	2.1278
GBPUSD	94713	1.5282	0.2137	1.1433	1.5282
NZDCAD	94713	0.8369	0.0768	0.6155	0.8369
NZDCHF	94713	0.7166	0.0712	0.5331	0.7166
NZDJPY	94713	74.0693	9.5029	44.257	74.0693
NZDUSD	94713	0.722	0.0732	0.4912	0.722
USDCAD	94713	1.1683	0.1364	0.9076	1.1683
USDCHF	94713	0.9976	0.0989	0.716	0.9976
USDJPY	94713	103.0905	12.739	75.679	103.0905

3. Methodology

Let us discuss the approaches that we use for the study of the pairs trading strategies in the foreign exchange rate markets. Training the trading models is carried out as follows. The full dataset is split into the following in-sample and out-of-sample set in a 60/40 ratio. The in-sample period spans from January 3, 2006 00:00:00 up to April 1, 2015 00:00:00 containing approximately 53,000 observations and the out-of-sample set spans April 1, 2015 01:00:00 up to November 1, 2021 23:00:00 containing approximately 41,000 observations. Then in the in-sample period, we use a rolling-window scheme where we do the formation and trading periods through the dataset and choose the parameters that maximize cumulative trading returns through out the whole in-sample period. These parameters are then fixed for the out-of-sample period for assessing the strategies.

First, one needs to define formation and trading windows as hyper parameters for the strategies. Note that in convention with machine learning fields we denote the formation and trading windows also as train and test windows, respectively. The trading window is always half the length of the formation window and the trading period follow directly after the formation as in line with the approach of many other authors. The window sizes will range from 1000, 2000, 3000, 4000, 5000 and 6000 hourly observations. All of these windows are then applied to each strategy and we will use the one optimal window that maximizes the cumulative return over the in-sample period.

In the formation period we first min-max scale the prices and then find the top 3 pairs based on the formation methods. Also, during the formation period we train our models, for example to find the trading thresholds or train the neural networks. For the formation we will use two different formation methods, one based on minimal distances and one based on cointegration.

During trading periods we first normalize the spread of the pairs by subtracting the mean and dividing by the standard deviation found during formation. Then, we use two different types of trading signals, one based on thresholds and one based on neural network output for trading. Note that we only allow one active trade at a time and we close all position by the end of the trading period.

We will now discuss in depth the formal definitions of the formation and parameter optimization methods. Then we will conclude with a discussion of the trading models.

3.1 Pairs formation methods

In this thesis we define the spread S_t between foreign exchange rate i and j at time t as

$$S_t = P_t^{(i)} - P_t^{(j)}, \tag{3.1}$$

where $P_t^{(i)}$ and $P_t^{(j)}$ denote the spot prices for foreign exchange rates i and j. During the formation phase we min-max scale the spot prices. In the trading phase we normalize the spread into z-scores by subtracting the mean and dividing by the standard deviation to generate trading signals for the trading models.

The formation period is the first step in our trading model. In this period we attempt to find the best three co-moving asset price pairs using the distance method (Section 3.1.1) and cointegration method (Section 3.1.2). We min-max scale all foreign exchange rates where the minimum and maximum are found in the formation period. This ensures that all prices are on a scale from 0 to 1. Scaling the prices is needed due to high variation in price scales in for example JPY currency pairs which are on a scale of around 100 while most currency pairs trade around 1. The stability of the min-max scaling is assumed to be sufficiently stable for this study with the use of a rolling window scheme as explained in Section 3.3. One could argue that over long periods of time the scaling might not be able to correct for large shifts in the exchange rates. However, since price scaling is generally accepted to be part of pairs trading by other authors, this scaling is still applied in this thesis.

In addition, during the formation period we store the mean and standard deviation of the spread pair used for the normalization in the trading period which follows directly after the formation. Normalization is required for the thresholds in the trading strategies which are going to be covered in Section 3.4. This process is repeated for all forex pairs considered in this study. Therefore, per formation period we need to test $\binom{26}{2} = 325$ different pairs for potential candidates for trading.

3.1.1 Distance method approach

The distance method is the benchmark pairs formation method due to its simplicity. The approach used here will be identical to the one used by Gatev et al. (2006). In each formation period we find for each scaled FX rate a corresponding FX rate that minimizes the sum of squared deviations, hence the name. Let $d_{i,j}$ denote the distance between scaled foreign exchange rates i and j, then we have

$$d_{i,j} = \frac{1}{T}\sum_{t=0}^{T}(\tilde{P}_t^{(i)} - \tilde{P}_t^{(j)})^2, \tag{3.2}$$

where $\tilde{P}_t^{(i)}$ and $\tilde{P}_t^{(j)}$ denote the min-max scaled prices for foreign exchange rates i and j. T denotes the total amount of observations in the formation period.

Lastly, we pick the top 3 pairs with the minimal distances, in line the benchmark strategy by Gatev et al. (2006).

3.1.2 Cointegration approach

Weak Stationarity

First we need to define the concepts of a weakly stationary process in order to apply cointegration. A stochastic process X_t is weakly stationary (hereafter referred to as stationary) if the process has a finite and constant mean and variance. Stated more formally:

Definition 1. *Let X_t be a stochastic process. Then X_t is a weakly stationary process if*

$$\begin{aligned} \mu_X(t) &= \mu_X < \infty \ \ constant, \\ \sigma_X(t) &= \sigma_X < \infty \ \ constant, \\ Cov(X_t, X_s) &= Cov(X_{t+r}, X_{s+r}), \end{aligned} \tag{3.3}$$

for all integers r, s and t.

As an example, Figure A.24 suggests that the USDCAD spot price is not stationary as the mean for the years 2006 and 2008 is not the same as the mean price in 2020 to 2022. Note that the differenced version in Figure B.24 suggests a stationary process. The stationarity of this USDCAD spot price and returns are only suggested by means of illustration but one needs to formally test whether a series is stationary. Testing for stationarity can be done by carrying out the Augmented Dickey-Fuller (ADF) test (Dickey & Fuller, 1979). Consider the following model

$$\Delta y_t = \alpha + \phi y_{t-1} + \beta_1 \Delta y_{t-1} + \cdots + \beta_p \Delta y_{t-p} + \epsilon_t, \tag{3.4}$$

where Δ is the time difference operator so that $\Delta y_t = y_t - y_{t-1}$. The ADF-test tests whether Δy_t in (3.4) has a unit root. The Augmented Dickey-Fuller test statistic τ is defined as

$$\tau = \frac{\hat{\phi}}{s.e.(\hat{\phi})}, \tag{3.5}$$

where $\hat{\phi}$ is the estimated ϕ as in (3.4) and $s.e.$ denotes the standard error. Note that this test statistic has its own Dickey-Fuller distribution which causes the hypothesis test to be asymmetrical. So, we test the following hypothesis for stationarity

$$H_0 : \phi = 0 \quad H_1 : \phi < 0. \tag{3.6}$$

In the ADF test, the number of lags denoted by p is generally found empirically. Therefore, in this thesis we will resort to Information Criteria by minimizing the Akaike Information Criterion (AIC) which in essence penalizes the use of more parameters in competing models. For more discussions on the AIC, one can consult Akaike (1974). The AIC is defined as

$$AIC = 2k - 2\log\hat{\mathcal{L}}, \tag{3.7}$$

where k denotes the number of parameters which is the number of lags p in our case. $\hat{\mathcal{L}}$ denotes the maximum likelihood value which is obtained by maximizing the likelihood function

$$\mathcal{L}(\theta|x) = f_\theta(x), \tag{3.8}$$

where f denotes the density function of the random variable X with parameters θ.

We will use the Python `statsmodels` package function `adfuller` on default settings to carry out the tests at the significance level of 5%. Default settings of the package include:

- Constant only and no trend in the ADF regression in (3.4)
- AIC on 5% significance
- Maximum number of lags is $12 * (\frac{T}{100})^{\frac{1}{4}}$, where T denotes the size of the formation window
- ADF test on 5% significance.

Cointegration

Intuitively, cointegration means that the difference between two time-series X_t and Y_t is relatively constant. It is a way of measuring whether the difference of two time-series is stationary. We can test for cointegration by using the two-step Engle-Granger method (Engle & Granger, 1987). The first step in the method includes running a regression of Y_t on X_t and test whether the residuals are stationary. We will consider the window sizes of 1000, 2000, ..., 6000 hourly observations as formation and thus as testing periods for carrying out the cointegration tests prior to the trading phases. More formally, we run the following regression

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t. \tag{3.9}$$

Then in the second step we test whether the residuals $\hat{\varepsilon_t}$ are stationary using the ADF-test.

During the formation period, we only qualify pairs that have an ADF statistic τ which is smaller than the 5% critical value of the Dickey-Fuller distribution. Next, we will pick the top three pairs ranked by the smallest three ADF test statistics for the trading period, since the ADF-test is one-tailed. This approach is a slight adaptation of Gatev et al. (2006). Similar to the distance method approach, by filtering to just 3 pairs, we attempt to limit the Type I error (mistakenly rejecting a true null hypothesis) we include in the study.

3.2 Neural Networks

Neural networks refer to a class of machine learning methods that are complex function approximates. These models are also called black box methods since the user supplies the model with an input and gets the corresponding output, without knowing why the model produces this output.

Their usage is versatile with notable performance in classification to regression problems. For instance, neural networks find their uses in finance by fitting non-linearities which are inherently apparent in financial data like stock prices. In a pairs trading context, Dunis et al. (2006) fit the spread returns by using neural networks to predict the future spread returns that is then used for the entry triggers for the pairs trading strategies. In this thesis we will follow a similar approach, where we will use neural networks to forecast the spread change ΔS_t between two foreign exchange rates. In addition, trading signals from forecasted spot spread prices S_t from a neural network will also be used.

The most basic form of neural networks are Multilayer Perceptron (MLP) networks. These feed-forward networks consist of three layers: input, hidden and the output layer. In short, the input can be thought of as the explanatory variables and in our case we will be using lagged spread (change) values as explanatory variables. Next, each input is mapped using a weight and bias, which are similar to the coefficients and intercept of a standard regression, to a node in the next layer. This process is then repeated for each hidden layer that is in the neural network. Lastly, the information that is passed through the network is passed through an activation function which yields the output, $\hat{S}_t$ (or $\Delta\hat{S}_t$). A schematic of such network is depicted in Figure 3.1.

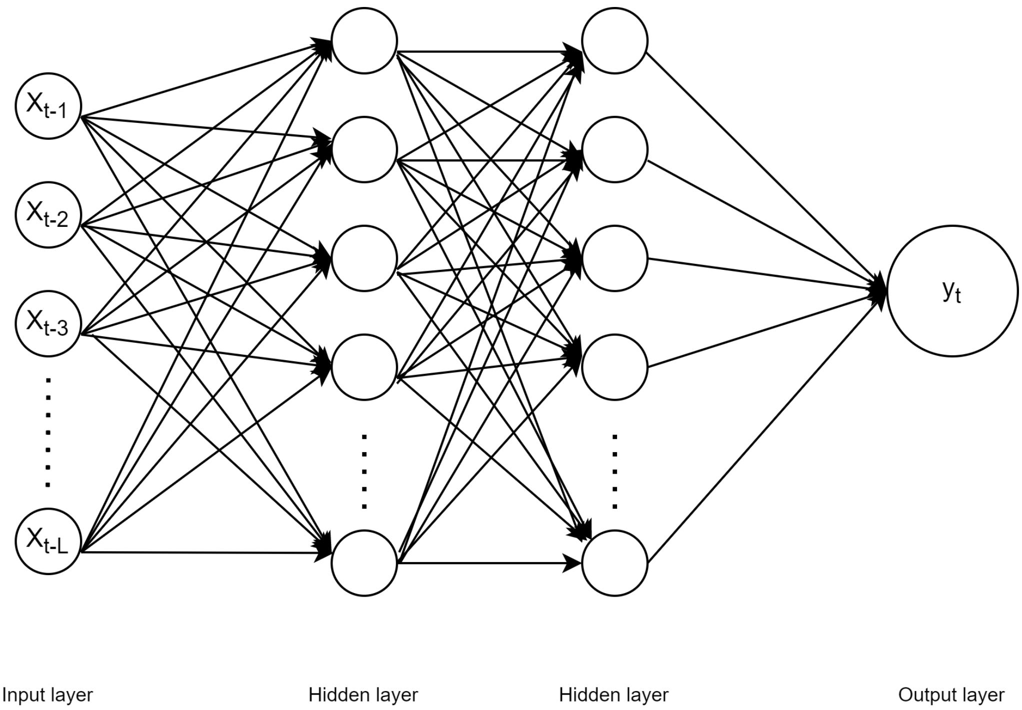


Figure 3.1: Example of a MLP structure using the previous L lags as input with two hidden layers of equal node size and output node. y_t could for example denote the one step ahead forecast output of the model.

Let us now delve deeper in what happens between the layers of a MLP. Formally, let j be the number of nodes in a hidden layer. Then, the output of each previous node is taken as a dot product with trainable weights vector W_j adjusted by a scalar bias b_j. Let $h^{(j)}$ denote the output of node j, then we have that

$$h^{(j)} = g^{(j)}(W_j x + b_j),$$

where x denotes the output vector of the previous layer and $g^{(j)}$ denotes the activation function of layer j. The activation function acts as an intermediary to produce the output from one node to the other and the choice of activation function highly depends on the type of modelling problem. For instance, in binary classifications problems, the sigmoid function

$$g(u) = \frac{1}{1+e^{-u}},$$

produces output between 0 and 1 (which then provides a probability distribution of both classes). A simple linear function

$$g(u) = \sum_j u_j,$$

can also act as an activation function. In this thesis we will only use the linear activation function for between hidden and output layers. Later in Section 3.2.1 we will use different activation functions in a more advanced class of MLP networks.

The training procedure of a neural network is as follows. In our trading strategies, we are dealing with a regression problem which is also known as supervised learning because we tell the network what the output should be. First, we initialize random weights and biases at each of the nodes. Then, we feed a batch of training samples through the network and choose a measure of error which generally is the mean squared error MSE. The optimization is then to minimize

$$MSE = \frac{1}{T}\sum_{t=0}^{T}(y_t - \hat{y}_t)^2, \tag{3.10}$$

where y_t denotes the observed value and $\hat{y}_t$ denotes the fitted value by our neural network.

Lastly, a so-called back-propagation algorithm is performed to adjust the weights and biases of the nodes to minimize this error. Essentially, this algorithm is a steepest descent algorithm which propagates the error backwards using partial derivatives, starting from the output layer into the hidden layers to the input. More in depth discussions and formal mathematics can be found in Svozil et al. (1997).

3.2.1 Long Short-Term Memory Networks

Long Short-Term Memory (LSTM) networks are an adaptation to the MLP networks originally developed by Hochreiter and Schmidhuber (1997). The general idea is that the LSTM layers have a mechanism that is able to carry information from previous time steps through the nodes of the hidden layers acting as a short-term memory, hence the name. Note that the LSTM networks are part of a larger class of Recurrent Neural Networks which use past information in its network. This property makes these models particularly attractive to apply in a time-series forecasting. Univariate based LSTM models tend to perform well in stock price prediction (Mehtab et al., 2021). In a pairs trading context, Chang et al. (2020) have trained a LSTM model to predict stock price spreads to use as input for pairs trading models, yielding positive returns.

Figure 3.2 illustrates an example of a LSTM network. Similar to the MLP regression networks, the LSTM networks used in this study will have the same activation layer as final output as a MLP network mentioned in Section 3.2. However, compared to the MLP network we have that the LSTM takes each input separately per node and there is a connection *between* the nodes in a LSTM layer. Note that this structure memory hidden layer structure can happen at more layers but for illustration purposes of this example model it is only a single layer.

Figure 3.3 shows roughly how such LSTM memory cell works. The information is passed through each node in a layer where the red trajectory depicts the short-term memory. This short-term memory is then modified by concatenating the previous cell's output with the current cell's input using a forget gate, input gate and output gate onto the next memory cell in the LSTM layer. Each of these gates have separate activation functions. Lastly, each node produces an output used for the next layer. In summary, the LSTM memory cell interacts with its previous cell, current input and its memory to produce output for the next cells and layer.

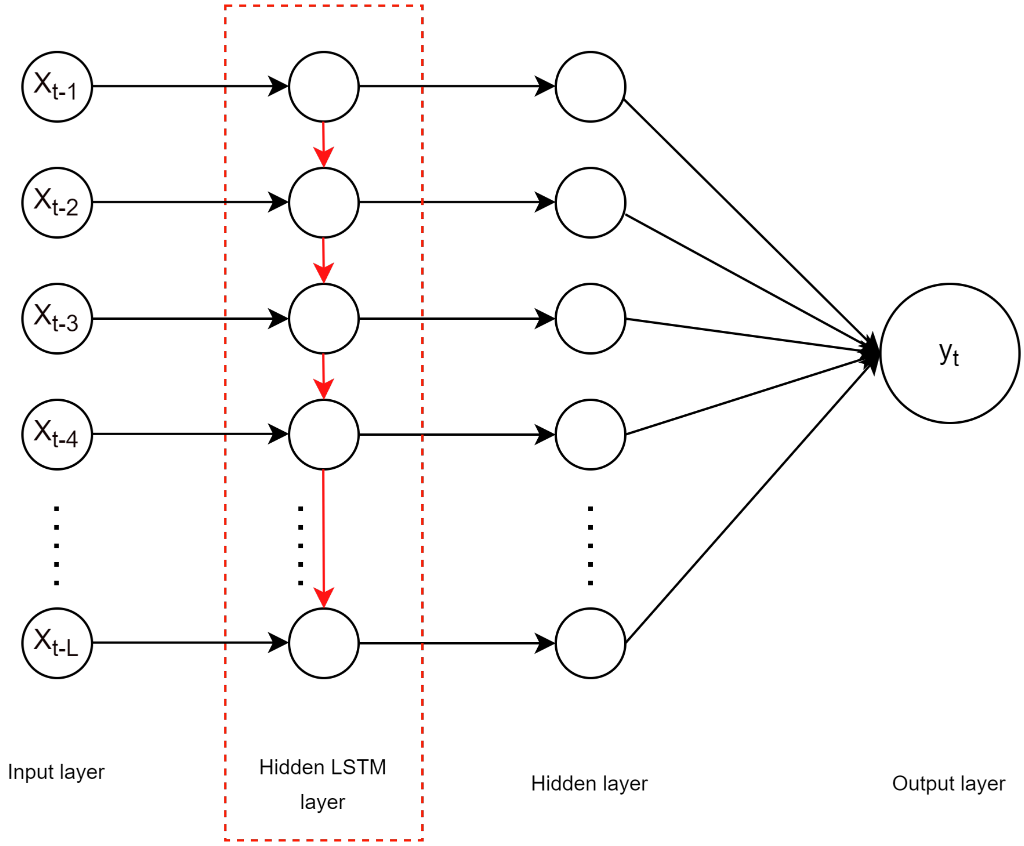


Figure 3.2: Example of a LSTM structure using the previous L lags as input with one LSTM layer connected to a standard layer and output layer. Note that the red arrow illustrates how past information is passed through a LSTM layer. Note that this structure can happen at more layers but for illustration purposes of this example it is a single layer.

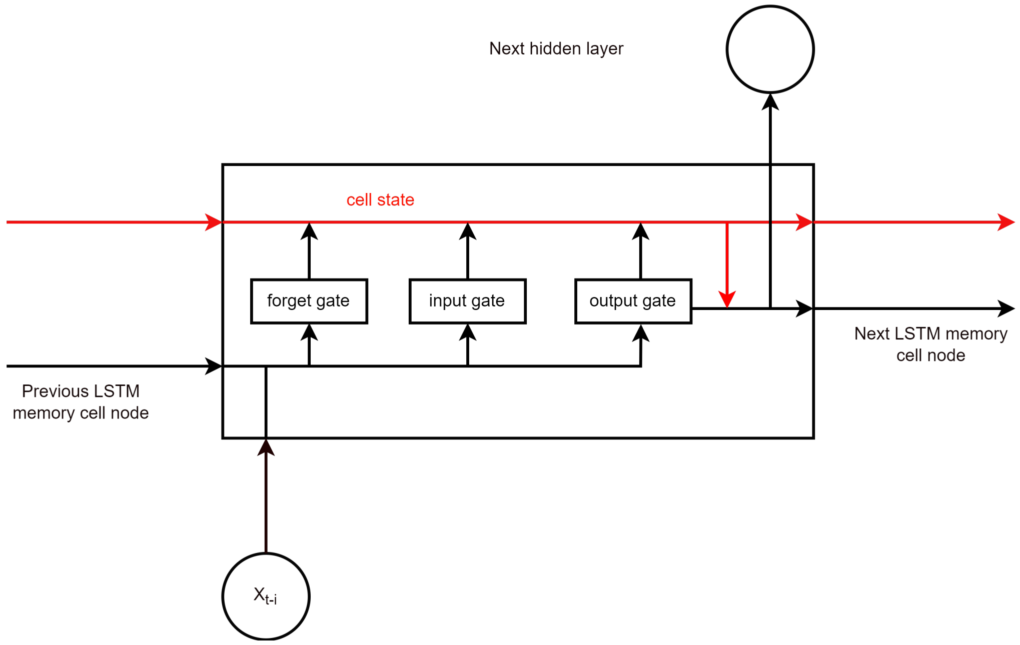


Figure 3.3: Schematic of one memory LSTM cell showing its interconnections in relation with its previous, next and output cells along with the modification gates. The red line shows the cell state which is passed through the LSTM layer memory cells.

In Figure 3.4 we illustrate the technical workings of such LSTM cell.

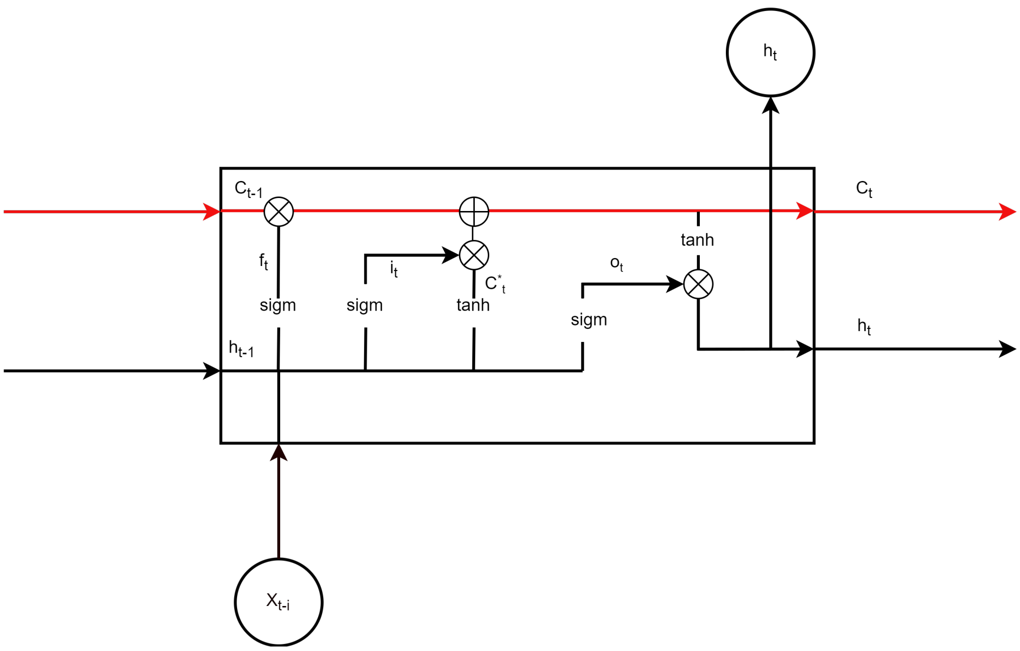


Figure 3.4: In-depth schematic of one memory LSTM cell illustrating the gates and their connections through the cell. The "sigm" and "tanh" stand for the sigmoid and tanh activation functions, respectively. Lastly, the $\otimes$ and $\oplus$ denote concatenations using multiplication and addition, respectively.

Each cell consists of the following gates:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{3.11}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{3.12}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{3.13}$$

$$C_t^* = tanh(W_C[h_{t-1}, x_t] + b_C) \tag{3.14}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t^* \tag{3.15}$$

$$h_t = o_t \cdot tanh(C_t). \tag{3.16}$$

The trainable weights are denoted by W_f, W_i, W_o, W_C and the trainable bias parameters are denoted by b_f, b_i, b_i, b_C. Also, σ denotes the sigmoid activation function and *tanh* denotes the *tanh* function. C_t is the cell state which contains the LSTM's memory over time. The information processed is then concatenated using addition or multiplication inside the cell. For complete formal derivations and mechanics of the LSTM network, one can consult Hochreiter and Schmidhuber (1997).

The weights are trained in a similar fashion as the MLP networks. After supplying the LSTM network a training batch, the weights are tuned using a similar back-propagation algorithm for each gate in the model minimizing error between the output and true value. The training is done based on the MSE criterion using the `adam` optimizer. The specific model structure and training criterion used will be further discussed in Section 3.3.

In this thesis, we attempt to use LSTM networks to fit financial time-series data as we suppose that these networks are able to capture complex time-series. Moreover, the structure of the model is constructed using the `Tensorflow keras` package.

3.3 Optimizing trading parameters

Firstly, one needs to determine the optimal window sizes for the formation and trading periods. Let ω denote the window size of the formation period. These windows are crucial for the pairs strategy's profitability, because too small or large ω may fail to capture intricacies of long-term or recent time-series behaviour of the spread. The formation window ω is used to perform the formation methods and find the required statistics like mean and standard deviation for trading. The optimization for the parameters will occur in the in-sample period of the dataset in order to prevent data-snooping when evaluating the strategies' returns.

Secondly, we need to construct trading thresholds K and find the most optimal threshold for trading. The construction of the threshold is done by normalizing the spread by calculating the z-score in the trading period by subtracting the mean and dividing by the standard deviation of the spread. These statistics are found in the formation period.

Following the approach of Kim and Kim (2019) and Wang et al. (2009), we consider multiple combinations of window sizes for the formation period and parameters K per window size. The ω will range from 1000, 2000, 3000, 4000, 5000 and 6000 hourly closing prices and we let K take the values 1, 2 and 3, as these are the most commonly used values for the threshold. Note that the trading window will follow directly after the formation period and the trading period is half of the length of the formation window size, as this is the general practice in the literature.

To the best of my knowledge, in most research focused on pairs trading, the criterion for selecting optimal window and trading parameters, is the cumulative return of the strategies and therefore this criterion will also be used here.

The formation window ω and corresponding trading windows will be so-called rolling windows. To be more precise, we shift the window by the length of the trading period in such a way that the trading periods are non-overlapping. Below in Figure 3.5, the whole mechanism of the methodology used is summarised, where the train windows denote the formation period and test windows illustrate trading phases. This notation is used in general literature. The following 60/40 dataset split is used:

- **In-sample**: January 3, 2006 00:00:00 - April 1, 2015 00:00:00, approx. 53,000 observations
- **Out-of-sample**: April 1, 2015 01:00:00 - November 1, 2021 23:00:00, approx. 41,000 observations.

The train and test (formation and trading) sets are constructed in a 2:1 ratio since the training period is twice as large as the testing period by construction.

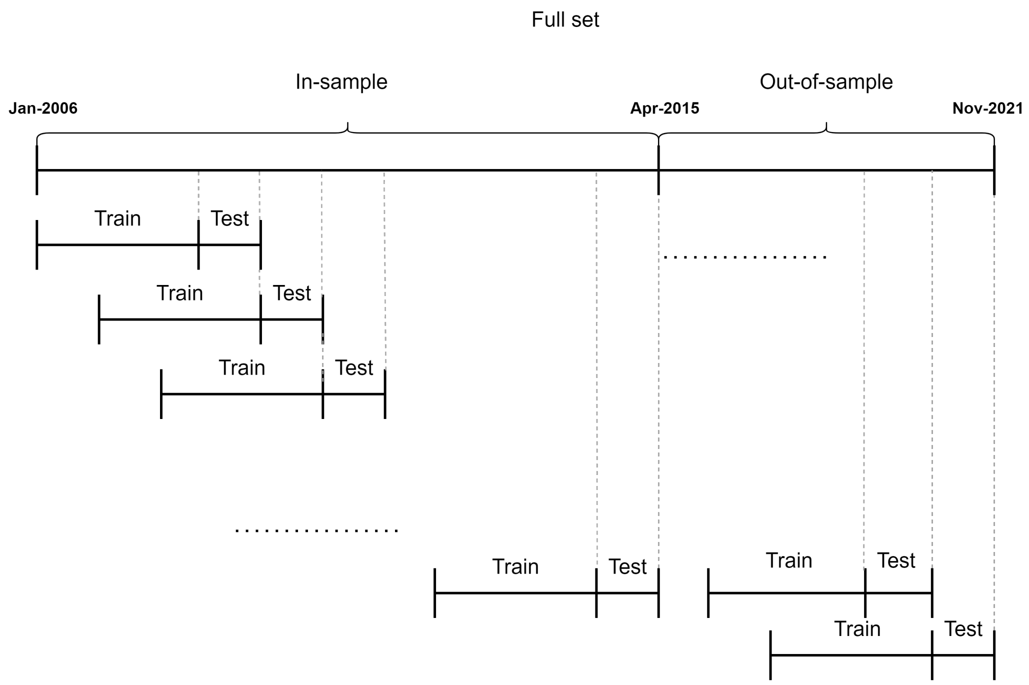


Figure 3.5: The methodology used for finding optimal trading parameters using the training and test split. The Train and Test segments illustrate the length of the formation and trading period respectively.

LSTM training and model structure

The training of the neural networks is similar to finding the trading thresholds. Firstly, we use cointegration as formation method. In this thesis we will use two types of regressions. One is based on the returns (difference) of the spread and the other regression is based on the spot spread.

Secondly, we assume that the spread of each cointegrated pair behaves similarly so that one LSTM model is able to predict all time-series regardless of the pair found in the formation period. So, in each formation period, we will train the network on cointegrated pairs only. This method relies on one main LSTM model used for trading all pairs considered in this thesis.

The LSTM network will use the past 24 hourly lagged spread (differences) of each formation pair to predict the next hourly spread (return) of that particular formation pair. Therefore, in each formation period we construct training samples consisting of 24 observations as input to forecast one hourly observation ahead. The first step in training consists of finding cointegrated pairs. Then we train the model on the cointegrated pairs only over the whole training period. After training, we will fit the model on the full formation period and calculate the mean and standard deviation of the model output. Next, in the trading periods we scale the model's output by this mean and standard deviation to get z-scores for the trading signals and then we can go into trading.

Lastly, the parameters for the window size ω and K are optimized using the highest net cumulative return. In-sample, formation window sizes ω will range from 1000, 2000, 3000, 4000, 5000 and 6000 observations and will be used along with thresholds values 1, 2 and 3 for K. Also here we let the trading period be half of the window size, and therefore a 2:1 train-test split will also be deployed here. The windows are shifted in the same procedure as in Figure 3.5.

The LSTM networks are constructed and trained using the Python `tensorflow keras` package using an Intel(R) Core(TM) i7-8750H CPU @ 2.21GHz and NVIDIA GeForce GTX 1050Ti GPU. The LSTM network uses 24 lags of the spread (change) as input and consists of a single LSTM layer of 64 nodes. The output node consists of a single node which is connected using a `Dense` layer (MLP conntected layer). The optimizer used is `Adam` with a learning rate of 0.001. Batch size is fixed at 32. The number of epochs is fixed at 50 per pair and the training samples are shuffled to reduce overfitting.

3.4 Trading Models

Lastly, let us cover the trading models used for pairs trading in the foreign exchange rate markets. In this thesis we use four different trading strategies that differ from formation method and trading signals:

- *Distance Method Threshold (DM-THRESHOLD)*: Open positions when the spread closes above (below) K standard deviations above (below) the long-term mean. Close positions if when the spread reverts to zero (when the spot prices cross), for the first time after the trade. Pairs are formed using minimum distance. This strategy is similar to the classical pairs trading strategy introduced by Gatev et al. (2006).

- *Cointegration Method Threshold (COINT-THRESHOLD)*: Open positions when the spread closes above (below) K standard deviations above (below) the long-term mean. Close positions if spread closes below (above) K standard deviations below the long-term mean. Pairs are formed using cointegration.

- *Cointegration Method differenced LSTM (COINT-DIF-LSTM)*: This strategy attempts to predict the spread *change* $\Delta\hat{S}_t$ using a recurrent neural network. Let $\tilde{P}_t^{(i)}$ and $\tilde{P}_t^{(j)}$ be the min-max scaled prices of foreign exchange rates i and j at time t of the formation pair, then the spread change ΔS_t is defined as

$$\Delta S_t = \frac{\tilde{P}_t^{(i)} - \tilde{P}_{t-1}^{(i)}}{\tilde{P}_{t-1}^{(i)}} - \frac{\tilde{P}_t^{(j)} - \tilde{P}_{t-1}^{(j)}}{\tilde{P}_{t-1}^{(j)}}.$$

  The forecasted spread change is normalized using the mean and standard deviation of the predictions by the network in the formation period. Then, the trader opens a long (short) position when the forecasted spread change, $\Delta\hat{S}_t$, is above (below) a threshold K $(-K)$. Positions are closed when the spread is in between $-K$ and K, similar to the approach by Dunis et al. (2006). The neural network is trained only on cointegrated pairs and is used for predicting all price spreads qualified for trading.

- *Cointegration Method Regression LSTM (COINT-REG-LSTM)*: This strategy attempts to predict the spread using a recurrent neural network. The strategy is similar to the COINT-THRESHOLD strategy but with the extra condition that the trader only enters a position if the LSTM network predicts that the spread is going to revert below (above) the threshold K in the next hour.

Note that in all three strategies, remaining open positions will be closed at the end of the trading periods. Lastly, we do not enter multiple positions, so we wait until the position is closed out by the trading rules or we hold until the end of the trading period. The analysis of the strategies also involves the min-max scaling of prices in the formation period and Z-score normalization for the trading period. The minimum, maximum, mean and standard deviation are derived in the formation period.

The general idea of the trading strategies in a trading phase is shown in Figure 3.6 where the COINT-THRESHOLD strategy is illustrated. The first blue rectangle illustrates the entry of a long position in the spread, since the normalized spread is below the trading thresholds K (which is 1 in this case). The following blue rectangle shows the time of long position close. Similarly, the first red box depicts a short position in the spread since we are now above the threshold. Finally, the position is closed in a profit since the spread went down and we close all positions at the end of the trading period. The distance method is slightly different since we close the position at the crossing of the prices, i.e. when the spread reverts to zero. For the trading strategies we will pick the top 3 pairs found by either minimum distance or the top 3 smallest test statistics to limit the amount of trading due to trading costs.

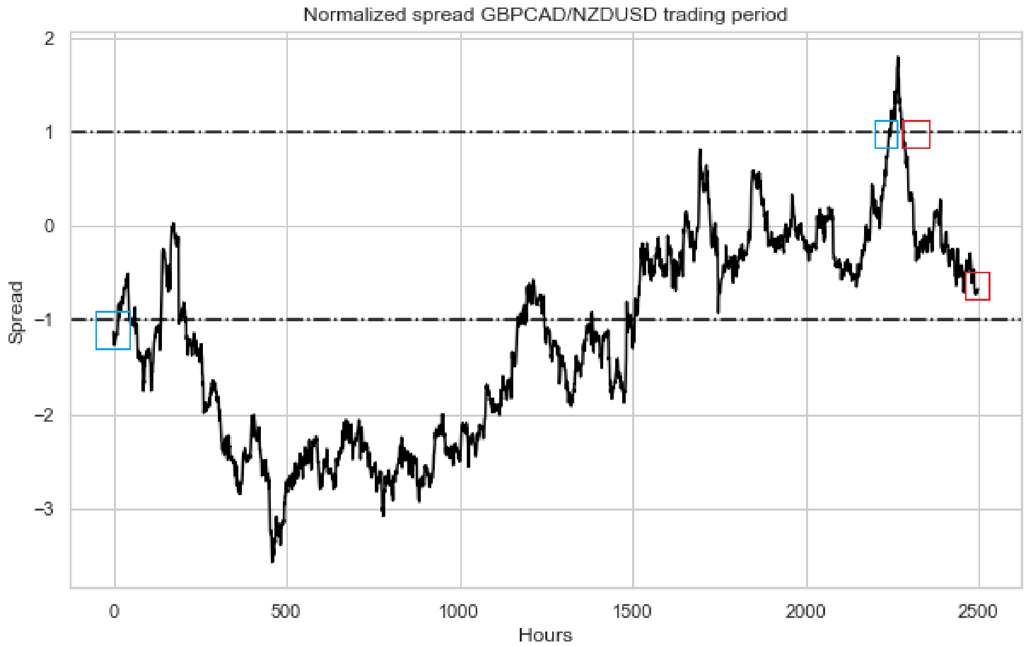


Figure 3.6: Out-of-sample trading period example showing the normalized spread and the thresholds ($K = 1$) for the COINT-THRESHOLD strategy. The blue boxes show opening and closing of the long position taken. The red boxes show the opening and the closing of the short position taken. Note that the last short is closed due to reaching the end of the trading period.

3.4.1 Transaction costs and returns

The trading strategies deployed here are based on intraday closing prices which implies a relatively high trading frequency. In addition, due to the nature of pairs trading, a trader needs to take *two* positions, one long and one short, which doubles the transactions costs per round lot. Also, the cost of the bid and ask spread needs to be taken into account for robust results. Bowen et al. (2010) show the sensitivity of intraday pairs trading returns to transaction costs. We will follow their estimates of 30 bps for two round trips. Then the return total return R on the pairs trading strategy in a trading period is

$$R = \sum_{n=1}^{N} r_n, \tag{3.17}$$

$$r_n = \theta \left(\frac{P^{(i)}_{n,close} - P^{(i)}_{n,open}}{P^{(i)}_{n,open}} - \frac{P^{(j)}_{n,close} - P^{(j)}_{n,open}}{P^{(j)}_{n,open}} \right) - 0.0030, \tag{3.18}$$

where $\theta = 1$ if we long the spread and $\theta = -1$ if we short the spread. $P^{(i)}_{n,open}$ and $P^{(i)}_{n,close}$ denote the spot prices of the n-th pairs trade at the time of opening the position and closing the position for FX pair i, respectively. We denote the total number of pairs trades taken in the trading period by N. Also, the position size is scaled such that the position size in each pair found in the formation is equal. In other words, we divide the returns in (3.18) by the number of pairs found, which is at most three for the cointegration method. Note that the distance method will always yield three pairs by construction.

4. Empirical Results

Firstly, let us discuss the results for the in-sample performance of the four strategies. As mentioned in Section 3, the in-sample period is used to find the best parameters for the strategies which are then deployed in the testing out-of-sample period for final assessment of the trading strategies. In this thesis, the optimal parameters are determined by the highest cumulative net return. The in-sample trading results are shown in the Tables 4.1 to 4.4 below. Pair formation statistics of the cointegration method can be found in Appendix C.

4.1 In-sample trading parameter optimization

On the one hand, from the DM-THRESHOLD strategy's results in Table 4.1, we can see that the only positive gross return of 0.0247 is generated with parameters $\omega = 4000, K = 1$. Interestingly, we find that after taking into account transaction costs, the strategy fails to generate positive returns in-sample and the highest cumulative net return of -0.0771 is achieved using the optimal trading parameters $\omega = 4000$ and $K = 3$. The difference is caused by the high transaction costs of the frequent trading that is done when setting low values for K.

On the other hand, the COINT-THRESHOLD strategy shows better performance in-sample. To be more precise, both the gross of 0.461 and net returns of 0.352 are highest when setting the formation period to 5000 and K to 1. Also, compared to DM-THRESHOLD, the COINT-THRESHOLD strategy has a higher win rate of 69%.

For the neural network strategies let us first focus on the COINT-DIF-LSTM strategy. Although the strategy is performing well in terms of cumulative gross returns, which is comparable to the COINT-THRESHOLD strategy performance, more than the whole starting balance is lost for all parameter settings. As mentioned before, this is due to the high number of executed trades as can be seen in the number of trades in the table. Despite this, we will still choose the optimal trading parameters $\omega = 1000$ and $K = 3$, as in accordance with the rest of the strategies.

Note that the Mean Squared Error (MSE) column shows the average forecasting MSE for all traded pairs over the whole in-sample dataset.

Lastly, the COINT-REG-LSTM shows similar but slightly worse results compared to the COINT-THRESHOLD strategy. The optimal parameters are when $\omega = 5000$ and $K = 1$, with gross returns of 0.3899 and net returns of 0.2909. Interestingly, the win rate of 0.70 is slightly higher compared to the win rate of the COINT-THRESHOLD method.

Now that the parameters are optimized in-sample, one can now evaluate each strategy in an out-of-sample period for a more realistic assessment of the performance of the strategies covered in this thesis.

Table 4.1: DM-THRESHOLD strategy: in-sample trading results.

ω	K	gross return	net return	mean	std. dev.	skewness	trades	win rate	long wins	long loss	short win	short loss
1000	1	-0.1345	-0.6735	-0.2806	0.2026	-0.1056	539	0.57	157	116	148	118
	2	-0.0719	-0.4459	-0.1678	0.1154	-0.3977	374	0.5	101	92	87	94
	3	0.002	-0.283	-0.1072	0.0729	-0.119	285	0.49	76	74	64	71
2000	1	-0.21	-0.458	-0.1824	0.1181	-0.1859	248	0.52	65	64	65	54
	2	-0.1353	-0.3103	-0.127	0.0718	0.0398	175	0.5	44	45	43	43
	3	-0.0502	-0.1842	-0.0913	0.0419	0.473	134	0.53	37	30	34	33
3000	1	-0.3386	-0.4986	-0.1856	0.1396	-0.4214	160	0.51	42	48	39	31
	2	-0.1963	-0.3173	-0.0802	0.0913	-0.5036	121	0.45	32	40	23	26
	3	-0.1385	-0.2275	-0.0533	0.0597	-0.554	89	0.49	22	29	22	16
4000	1	**0.0247**	-0.1053	-0.0137	0.0391	-0.7755	130	0.58	36	36	39	19
	2	-0.0685	-0.1515	-0.0591	0.0553	0.7424	83	0.42	18	31	17	17
	3	-0.0151	**-0.0771**	-0.0188	0.0442	1.3182	62	0.47	20	19	9	14
5000	1	-0.3537	-0.4337	-0.1734	0.0924	-0.7361	80	0.41	20	28	13	19
	2	-0.3727	-0.4317	-0.2094	0.0878	0.0885	59	0.31	10	26	8	15
	3	-0.2728	-0.3198	-0.1498	0.058	-0.3464	47	0.34	9	21	7	10
6000	1	-0.3051	-0.3731	-0.2297	0.0768	0.5886	68	0.47	19	27	13	9
	2	-0.1546	-0.2066	-0.1463	0.0437	0.3261	52	0.44	14	22	9	7
	3	-0.2075	-0.2435	-0.1878	0.0522	1.3713	36	0.42	9	17	6	4

Table 4.2: COINT-THRESHOLD strategy: In-sample trading results

ω	K	gross return	net return	mean	std. dev.	skewness	trades	win rate	long win	long loss	short win	short loss
1000	1	0.3937	-0.1243	0.0041	0.0593	0.014	518	0.6	152	112	157	97
	2	0.1175	-0.2055	-0.0436	0.07	-0.1789	323	0.55	87	79	90	67
	3	0.2301	0.0111	0.0111	0.0397	-0.0403	219	0.55	64	49	56	50
2000	1	0.0009	-0.2501	0.0076	0.1432	0.5165	251	0.56	74	61	67	49
	2	-0.1618	-0.3168	-0.0894	0.1532	0.5385	155	0.57	47	37	42	29
	3	-0.2362	-0.3362	-0.1241	0.1326	0.4811	100	0.47	29	31	18	22
3000	1	0.0584	-0.0966	-0.0212	0.0842	0.4259	155	0.6	46	33	47	29
	2	0.0552	-0.0498	-0.0084	0.0779	0.8121	105	0.58	30	22	31	22
	3	-0.0419	-0.1079	0.0162	0.0628	0.95	66	0.53	15	16	20	15
4000	1	0.2886	0.1666	0.2337	0.1006	-0.1652	122	0.63	32	31	45	14
	2	0.0074	-0.0635	-0.008	0.054	0.1017	71	0.51	17	23	19	12
	3	0.0536	0.0026	0.0332	0.0418	0.7594	51	0.53	11	18	16	6
5000	1	**0.461**	**0.352**	0.0978	0.1211	0.1509	109	0.69	47	19	28	15
	2	0.1294	0.0754	0.0049	0.0572	-0.7784	54	0.56	17	15	13	9
	3	0.038	0.003	-0.0347	0.0633	-0.4554	35	0.57	11	8	9	7
6000	1	0.2874	0.2154	0.0783	0.1102	0.046	72	0.68	20	17	29	6
	2	0.1636	0.1216	0.0501	0.0735	-0.1169	42	0.57	10	13	14	5
	3	-0.0787	-0.1047	-0.1027	0.0633	0.5923	26	0.62	8	7	8	3

Table 4.3: COINT-DIF-LSTM: In-sample results for the LSTM strategy using differenced return time-series for pairs trading. MSE denotes the average forecasting MSE during the trading periods in-sample.

ω	MSE	K	gross return	net return	mean	std. dev.	skew	trades	win rate	long wins	long loss	short wins	short loss
1000	1.7556	1	**0.398**	-28.151	-14.0611	8.0863	-0.0091	28549	0.52	7447	6850	7277	6975
		2	0.2763	-6.4858	-3.2755	1.8607	0.0362	6762	0.52	1078	1036	2412	2236
		3	0.1198	**-1.7412**	-0.8761	0.5152	0.1191	1861	0.52	276	276	691	618
2000	2.14635	1	0.145	-24.976	-12.5017	7.275	0.0177	24735	0.51	6287	6252	6256	5940
		2	0.0527	-6.6293	-3.2536	1.9386	-0.0282	6640	0.51	1601	1554	1807	1678
		3	0.2553	-1.902	-0.9153	0.5332	-0.1417	2139	0.52	500	474	622	543
3000	2.15796	1	-0.2709	-16.5859	-8.3367	4.803	0.0234	16315	0.5	3936	3957	4261	4161
		2	-0.1372	-6.6662	-3.3927	1.9392	0.0399	6529	0.5	1546	1618	1747	1618
		3	-0.0989	-3.1	-1.5683	0.9019	-0.03555	3001	0.49	712	754	771	764
4000	1.5329	1	0.1619	-18.2881	-9.1468	5.2615	-0.0053	18450	0.51	5099	4935	4246	4170
		2	0.2357	-6.2023	-3.1259	1.7814	-0.0078	6438	0.51	1938	1814	1363	1323
		3	0.0458	-2.5142	-1.2344	0.7254	-0.0678	2560	0.51	749	703	555	553
5000	1.615	1	0.1892	-15.9618	-7.9266	4.6027	-0.0208	16151	0.51	4333	4134	3878	3806
		2	0.1764	-6.03561	-2.9799	1.7406	-0.027	6212	0.51	1703	1590	1493	1426
		3	0.2445	-2.4286	-1.1793	0.7137	-0.0853	2673	0.51	703	681	660	629
6000	1.58264	1	0.0344	-13.2836	-6.6139	3.8376	-0.0197	13318	0.5	3389	3366	3341	3222
		2	0.0904	-5.9026	-2.9116	1.6807	-0.0477	5993	0.52	1547	1528	1546	1372
		3	-0.0182	-3.0082	-1.5095	0.8611	-0.0016	2990	0.51	791	770	735	694

Table 4.4: COINT-REG-LSTM: In-sample trading results. MSE denotes the average forecasting MSE during the trading periods in-sample.

ω	MSE	K	gross return	net return	mean	std. dev.	skew	trades	win rate	long wins	long loss	short wins	short loss
1000	3.0843	1	0.3708	-0.0892	0.0096	0.0499	0.5729	460	0.61	143	97	138	82
		2	0.1226	-0.1924	-0.0339	0.0655	-0.3504	315	0.55	85	78	89	63
		3	0.2314	0.0124	0.0108	0.0383	-0.0772	219	0.55	64	50	56	49
2000	2.4279	1	0.1206	-0.1074	0.0859	0.1259	0.2848	224	0.58	68	54	62	40
		2	-0.1736	-0.3256	-0.0864	0.1568	0.3808	150	0.58	47	36	40	27
		3	-0.2197	-0.3177	-0.0976	0.1255	0.2836	98	0.48	29	31	18	20
3000	4.5519	1	-0.0229	-0.1599	0	0.0553	-0.3337	137	0.58	43	31	36	27
		2	0.0622	-0.0388	0.0214	0.0736	0.7765	101	0.58	29	21	30	21
		3	-0.0385	-0.1045	0.0314	0.0675	0.7828	66	0.52	14	17	20	15
4000	0.9846	1	0.3809	0.2749	0.2609	0.1272	-0.9112	106	0.63	29	24	38	15
		2	0.0505	-0.0175	0.008	0.0445	0.4803	68	0.51	17	21	18	12
		3	0.0602	0.0092	0.0395	0.0426	0.5214	51	0.55	11	18	17	5
5000	1.9319	1	**0.3899**	**0.2909**	0.0628	0.0909	0.5953	99	0.7	43	18	26	12
		2	0.1244	0.0704	-0.0036	0.0545	-0.7658	54	0.54	16	16	13	9
		3	0.0376	0.0026	-0.0466	0.0613	-0.2121	35	0.57	11	8	9	7
6000	1.8194	1	0.0819	0.0159	-0.0074	0.0637	-0.587	66	0.64	16	20	26	4
		2	0.1669	0.1249	0.0551	0.0606	0.114	42	0.57	10	13	14	5
		3	-0.0759	-0.1019	-0.0647	0.0789	0.0234	26	0.62	8	7	8	3

4.2 Out-of-sample trading performance

Now let us discuss the trading results in the period of April 2015 to November 2022 which acts as the out-of-sample period in this thesis. Table 4.5 and Figures 4.1 to 4.4 show the cumulative returns and other statistics of the trading results over the whole out-of-sample period. Note that the remaining statistics are calculated based on the net returns. Firstly, we see that the benchmark DM-THRESHOLD strategy is able to generate positive returns even after transaction costs of 0.0372. The COINT-THRESHOLD strategy performs the best, showing a 0.1648 return after transaction costs. The COINT-DIF-LSTM, does not perform well as it show negative gross returns. More importantly, after accounting for trading costs, this strategy loses more than the starting balance due to the high frequency of trade executions. This performance was also noted in the in-sample training phase. Lastly, the COINT-REG-LSTM is able to generate positive returns, but the strategy does not beat the COINT-THRESHOLD strategy. We can therefore conclude that the COINT-THRESHOLD method is the best performing strategy considered in this study.

As discussed in the literature review, pairs trading returns in stock markets can yield around 60% return annually (Stübinger & Endres, 2017) using 1-minute data. The performance of the pairs trading strategy considered in this thesis on an hourly data time frame are mediocre compared to their results. However, research have shown declining profitability of pairs trading in the stock markets using daily data of around 1% net return over a five year horizon. In that regard, the intraday pairs trading results in the forex markets show more optimism for pairs trading. In addition, the COINT-THRESHOLD pairs trading strategy is tested on recent data showing that pairs trading can still be profitable. Hedge funds and investors may indeed consider the pairs trading methods of this thesis to use as investment strategy but they need to keep in mind that the pairs trading returns can be volatile and its performance is not comparable with pairs trading in other financial markets. Furthermore, the pairs trading strategy could also be implemented in currency hedging strategies.

Table 4.5: Out-of-sample cumulative trading results for each strategy. Note that no MSE is calculated for the DM and COIN strategies.

Strategy	ω	K	gross return	net return	mean	std. dev.	skew	trades	win rate	long wins	long loss	short wins	short loss	MSE
DM-THRESHOLD	4000	3	0.0832	0.0372	0.0421	0.0276	-0.2124	46	0.57	17	11	9	9	NA
COINT-THRESHOLD	5000	1	0.2288	0.1648	0.1289	0.0343	-0.2357	64	0.59	22	12	16	14	NA
COINT-DIF-LSTM	1000	3	-0.0356	-1.5486	-0.6959	0.5029	-0.3592	1508	0.48	255	260	475	518	1.3598
COINT-REG-LSTM	5000	1	0.112	0.049	0.0495	0.0212	0.9296	63	0.59	22	12	15	14	0.7348

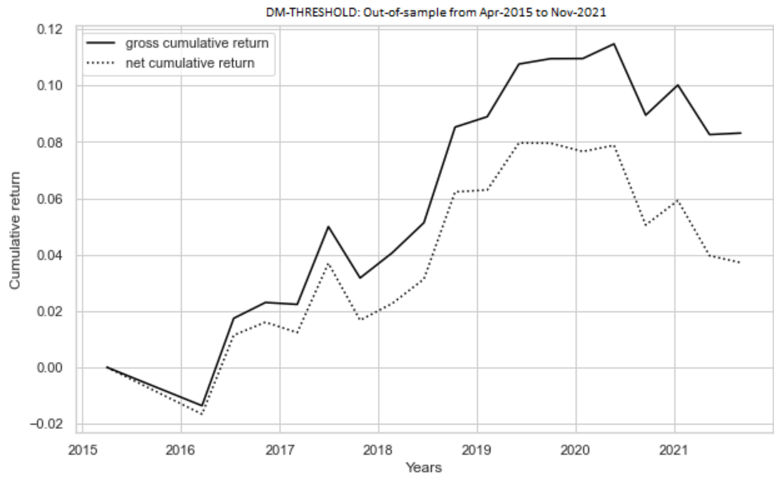


Figure 4.1: Out-of-sample cumulative returns of the distance method over the period of Apr-2015 to Nov-2021.

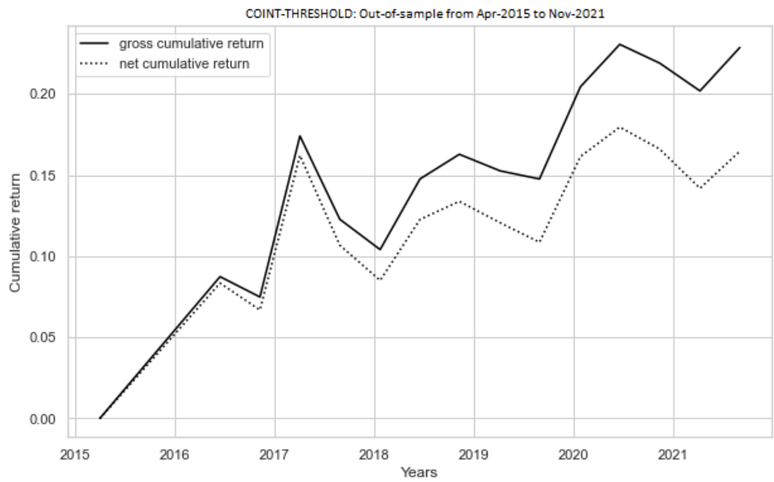


Figure 4.2: Out-of-sample cumulative returns of the cointegration method over the period of Apr-2015 to Nov-2021.

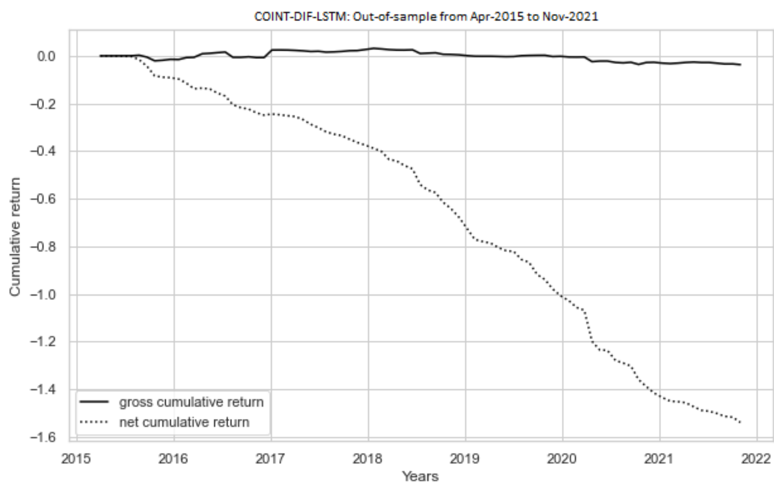


Figure 4.3: Out-of-sample cumulative returns of the D-LSTM method over the period of Apr-2015 to Nov-2021.

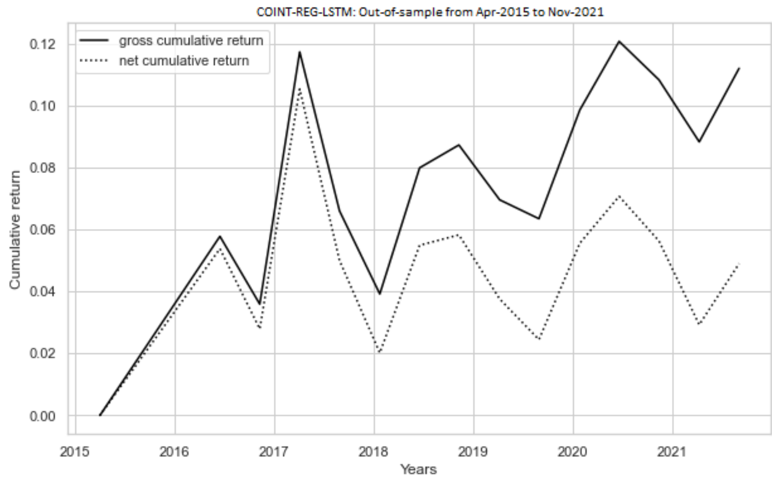


Figure 4.4: Out-of-sample cumulative returns of the R-LSTM method over the period of Apr-2015 to Nov-2021.

4.2.1 LSTM performance

As for both LSTM strategies, one could argue that these strategies would perform at least as well as the simple strategies such as the COINT-THRESHOLD strategy. The reason is that these models attempt to fit the spread behaviour using a high number of parameters versus two paramaters that are needed for the COINT-THRESHOLD strategy. However, we find sub-optimal results from these models.

On the on hand, the COINT-DIF-LSTM strategy does not perform well, due to the mediocre fitting of the network on the stationary returns. One can conclude that predicting financial returns is rather challenging which can be explained by the Random Walk Theory of asset prices. As Figure 4.5 illustrates, one can observe that the model is not able to fit the sporadic volatility of the observed spread returns.

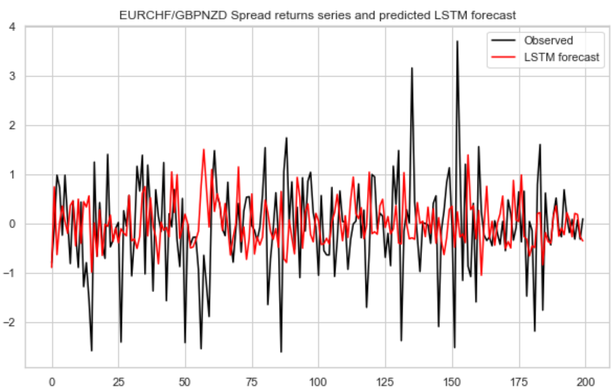
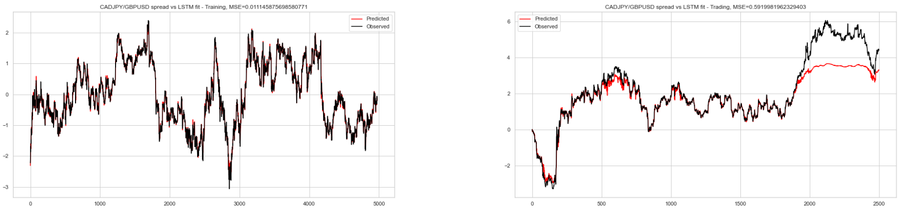


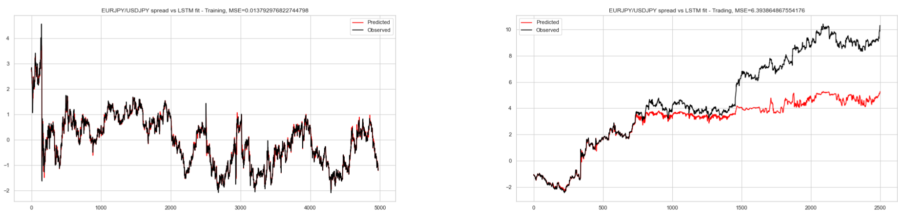
Figure 4.5: LSTM fit on differenced return time-series of the cointegrated pair EURCHF/GBPNZD.

On the other hand, the COINT-REG-LSTM fit against the observed spread in the formation periods show that for some pairs, the spread behaves in a different regime and therefore the network does not fully capitalize on this new behaviour. This can be seen in Figures 4.6 and Figure 4.7. The left and right illustrations show the COINT-REG-LSTM fit in the formation and trading periods, respectively. One can clearly see that in some trading periods there is a form of regime shift in time-series behaviour that the network fails to capture. In the training periods, the model seems to overfit. The Figures show that the spread moves outside the threshold into a high level of standard deviation which the model does not expect. All the figures below are from the out-of-sample dataset. We also see that these shifts tend to happen in JPY pairs.



(a) (b)

Figure 4.6: COINT-REG-LSTM fitting performance on CADJPY/GBPUSD. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period.



(a) (b)

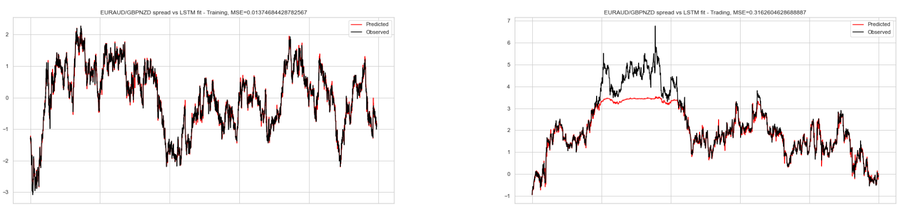
Figure 4.7: COINT-REG-LSTM fitting performance on EURJPY/USDJPY. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period.

4.2.2 Trading model robustness

In Section 4.2.1 we noted that the trading models might be sensitive to the data. More precisely, we suppose that the trading models are being affected by outliers due to significant interventions by central banks or financial crises. Therefore, we will remove the Swiss Franc and the Japanese Yen because it can be seen in the time-series plots in Appendix A and B that the time-series of these currency pairs contain outliers. Similarly, we will evaluate the trading models after they have been trained on data after 2010 while keeping the out-of-sample data set the same as before. Note that mainly the training of both LSTM trading models will be affected by this filter and not the DM-THRESHOLD and COINT-THRESHOLD strategies. The latter two strategies will therefore not be evaluated on the financial crisis filter data after 2010. Lastly, for the robustness analysis we will fix the optimal trading parameters as found on the full dataset and evaluate the robustness on the out-of-sample period.

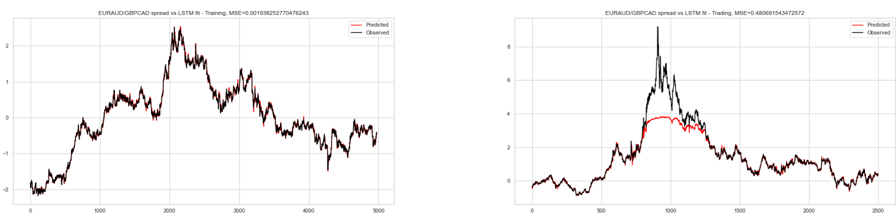
Swiss Franc and Japanese Yen pair filter

We attempt to remove CHF and JPY pairs from the data in this Section. The underlying reason could be significant interventions by the central banks that can be seen in the returns plots of the dataset (Appendix B). So, we trained the COINT-REG-LSTM model on the same data but filtered out all CHF and JPY pairs. The Figures 4.8 and 4.9 still show that the model is still not robust to regime shift of the time-series as these Figures are comparable with the fitting on the full dataset. However, Table 4.6 show the results of the robustness filters, which show that on average the LSTM forecast is better in terms of MSE. Interestingly, we find that the the DM-THRESHOLD, COINT-THRESHOLD and the COINT-REG-THRESHOLD strategies perform significantly better than on the full dataset. Furthermore, the MSE for the LSTM strategies is also lower, which suggests that the strategy is sensitive to volatile prices of foreign exchange rates. But, the returns of the COINT-REG-LSTM strategy are still lower than the COINT-THRESHOLD strategy which still suggests that the use of the LSTM network might not be necessarily better compared to using a simple threshold parameter.



(a) (b)

Figure 4.8: COINT-REG-LSTM fitting performance on EURAUD/GBPNZD. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period.



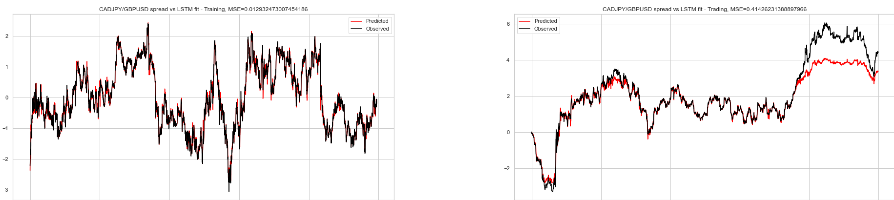
(a) (b)

Figure 4.9: COINT-REG-LSTM fitting performance on EURAUD/GBPCAD. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period.

Financial crisis filter

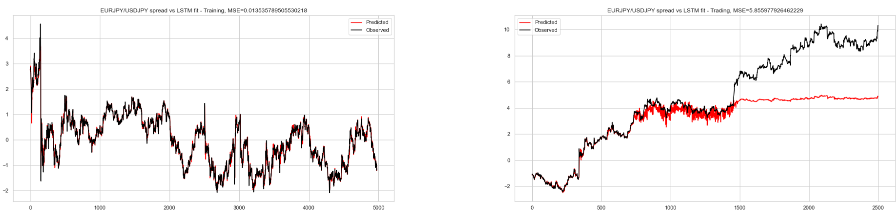
Lastly, let us investigate whether the training procedure of the LSTM models are affected by financial turmoil while keeping the trading parameters unchanged. To be exact, we start training the models using shorter in-sample horizon spanning 00:00:00 January 2010 until 01:00:00 April 1, 2015 which contains roughly 30,000 observations. The out-of-sample is not altered to the methodology mentioned in Section 3.3. In other words, we keep the trading parameters as found in Section 4.1 and re-estimate the LSTM parameters on this filtered dataset.

Interestingly, we find that the LSTM network behaves in the same way as in the full dataset. Financial crises do not seem to affect the fitting performance of the neural networks. Instead, from all of these results mentioned in Section 4.2.1, we suppose that the nature of this fitting behaviour is mostly explained by the model structure and not necessarily the data.



(a) (b)

Figure 4.10: COINT-REG-LSTM fitting performance on CADJPY/GBPUSD. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period.



(a) (b)

Figure 4.11: COINT-REG-LSTM fitting performance on EURJPY/USDJPY. Exhibit (a) illustrates the fitting during formation and Exhibit (b) illustrates the fitting in the trading period.

From the results in Table 4.6, one can observe a number of interesting findings. First, filtering out specific pairs that are highly volatile due to interventions seems to improve the trading results for all trading models. The returns for the DM-THRESHOLD strategy are similar to the returns achieved by the strategy on unfiltered data. For the COINT-THRESHOLD strategy we can see significant improvements with more than double the cumulative net return of 0.4253 compared to the return achieved in the unfiltered dataset. Both the COINT-THRESHOLD and the COINT-REG-LSTM strategies have a better win rate of 73% compared to the full dataset. The COINT-DIF-LSTM model still applies takes too many costly trades and therefore the performance is comparable to the performance in the standard dataset. The MSE for the COINT-DIF-LSTM increases slightly, with similar results as in the full sample. Furthermore, the COINT-REG-LSTM strategy improves by roughly 28 percentages points in net cumulative return. Note that the forecasting error drastically decreases from 0.7348 to 0.0816 after removing CHF and JPY pairs for the COINT-REG-LSTM model and hence better trading performance. In conclusion, the filtering of volatile pairs suggests an improvement in the cointegration method for formation but not necessarily for the distance method.

Secondly, altering the training set to filter out the financial crisis of 2008 seems to decrease returns of both strategies. The MSE for the COINT-DIF-LSTM increases slightly while the MSE for the COINT-REG-LSTM decreases. There seems to be no clear advantage of filtering out the financial crisis in our dataset. This phenomenon is partly suggested by the use of a smaller training set while keeping the out-of-sample dataset the same size (unfiltered) for comparing measures.

Table 4.6: Out-of-sample pairs trading results using two different filters: (i) Swiss Franc and Japanese Yen pairs removed and (ii) only training the neural networks on data after 2010.

CHF and JPY filter														
	ω	K	gross return	net return	mean	std. dev.	skew	trades	win rate	long wins	long loss	short wins	short loss	MSE
DM-THRESHOLD	4000	3	0.0838	0.0418	0.042	0.0386	-0.3786	42	0.55	14	12	9	7	NA
COINT-THRESHOLD	5000	1	0.5003	0.4253	0.3146	0.0904	-0.2569	75	0.73	26	9	29	11	NA
COINT-DIF-LSTM	1000	3	-0.0354	-1.5424	-0.787	0.4486	0.0131	1398	0.5	395	415	298	290	1.589
COINT-REG-LSTM	5000	1	0.3992	0.3292	0.2237	0.0812	-0.6421	70	0.73	24	10	27	9	0.0816
Financial Crisis filter														
COINT-DIF-LSTM	1000	3	-0.0027	-1.2017	-0.6117	0.345	0.0502	1198	0.49	287	319	300	29	1.8574
COINT-REG-LSTM	5000	1	0.0256	-0.0354	-0.0256	0.0299	0.7136	61	0.57	20	12	15	14	0.5398

5. Conclusion

Interest in exploiting financial market inefficiencies has been a popular field of study. In particular, market-neutral investment strategies are studied due to their risk return profile. One of these strategies is the pairs trading strategy. The pairs trading strategy involves finding a co-moving asset price pair, followed by taking a position in the spread between the two assets when anomalies are found, speculating on a reversion of the spread.

In this thesis we studied the strategy applied in the foreign exchange rate markets. For the formation method we have used the distance and cointegration method. Additionally, we studied four different trading strategies, two of which used LSTM recurrent neural networks. One strategy used the distance method as formation and threshold boundaries to enter trades and we the close position when prices crossed. The second strategy used the cointegration method for formation and uses thresholds to open and close positions. Also, one LSTM strategy focused on fitting returns which tends to be used in other studies in the field of pairs trading. Lastly, the other LSTM strategy attempted to fit the spread itself. All trades taken in these strategies were closed if the trading period ended and at most one trade was open during trading.

The objective of this thesis was to verify whether pairs trading in the foreign exchange rate markets can be profitable and to study the performance of different trading models in the pairs trading framework.

Firstly, the results show that some of these strategies are able to yield a positive net return over a six year out-of-sample time horizon of 2015 to 2021. This indicates that pairs trading strategies can be deployed in the forex markets to some extent. Yet, the performance of the pairs trading strategies considered in this thesis are still under performing compared to the pair strategies in other financial markets. Conclusively, the use of the pairs trading strategy in the foreign exchange rate markets as an investment strategy might be sub optimal and requires further parameter optimization in order to compete with existing strategies.

Secondly, we have found that the COINT-THRESHOLD strategy works best, yielding a net return of 0.1648. The COINT-REG-LSTM strategy performed sub optimally with a net return of 0.049 followed by the DM-THRESHOLD strategy net return of 0.0372. Lastly, the COINT-DIF-LSTM strategy was not even able to generate profits before transaction costs. Moreover, this strategy lost more than the starting balance after trading due to entering a high number of trades without covering the transaction costs per trade. Lastly, filtering out the Swiss Franc and the Japanese Yen pairs improved results of the trading models significantly. The removal resulted in a net cumulative return of 0.4253 for the COINT-THRESHOLD model, outperforming the other strategies with the same filtering. The CHF and JPY pairs were filtered out due to the volatile time-series resulting from interventions from their central banks. Filtering out the financial crisis from the training set did not improve the trading models. So, the results from this robustness test suggests that currencies considered in pairs trading models should be carefully selected in the dataset for optimizing the trading models as we have found the relevance of the choice of exchange rates to be traded.

Future research can extend these trading models in several ways. These strategies could be studied on different time frames with larger or smaller than the hourly frequency considered in this thesis. Also, the pairs formation method could be optimized by imposing alternative ways of qualifying pairs for trading. This thesis simply took the top three smallest ADF statistics as it is the general practice in similar studies. The trading strategies could be improved by using different trading entry and exit criteria, like stop-loss orders to prevent large losses. Additionally, the neural network strategy could be improved by using a collective group of networks to prevent overfitting, as well as venturing into different network structures.

References

Akaike, H. (1974), A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19 (6)*, 716–723.

Baronyan, S., R., Boduroglu, I., I. & Sener, E. (2010). Investigation of Stochastic Pairs Trading Strategies Under Different Volatility Regimes, *The Manchester School, 78(1)*, 114-134.

Bowen, D., Hutchinson, M. & O'Sullivan, N. (2010). High Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution and Patterns in Returns. *The Journal of Trading, 5(3)*, 31-38.

Caldeira, J. F., Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. *Brazilian Review of Finance, 11*, 49–80.

Chang, V., Man, X., Xu, Q. & Hsu, C. (2020). Pairs trading on different portfolios based on machine learning, *Expert Systems, 38(3).*

Chen, D., Cui, J., Gao Y. & Wu, L., (2018). Pairs trading in Chinese commodity futures markets: an adaptive cointegration approach, *Accounting & Finance, 57(5)*, 1237-1264.

Dickey, D. A. & Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*, 427–31.

Do, B. & Faff, R. (2010) Does Simple Pairs Trading Still Work? *Financial Analysts Journal, 66(4)*, 83-95.

Dunis, C. L., Laws, J., & Evans, B. (2006). Modelling and trading the gasoline crack spread: A non-linear story. *Derivatives Use, Trading & Regulation, 12 (1)*, 126–145.

Elliott, R. J., van der Hoek, J. & Malcolm, W. P., (2005). Pairs trading. *Quantitative Finance 5 (3)*, 271–276.

Engle, R. F. and Granger, C. W. J. (1987) Co-integration and error correction: representation, estimation and testing. *Econometrica, 55*, 251–76.

Fil, M., & Krisoufek, L. (2020). Pairs Trading in Cryptocurrency Markets.

IEEE, 8, 172644 - 172651.

Fischer, T. & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270(2)*, 654-669.

Flori, A. & Regoli, D. (2021). Revealing Pairs-trading opportunities with long short-term memory networks. *European Journal of Operational Research, 295(2)*, 772-791.

Gatev, E., Goetzmann, W.N. & Rouwenhorst, K.G. (2006) Pairs trading: performance of a relative-value arbitrage rule. *Review of Financial Studies, 19(3)*, 797–827.

Göncü, A. & Akyildirim, E. (2016). A stochastic model for commodity pairs trading. *Quantitative Finance 16(12)*, p. 1843-1857.

Hochreiter, S. & Schmidhuber, J. (1997). Long Short-term Memory, *Neural Computation 9(8)*, 1735-80.

Huck, N. & Afawubo, K. (2016). Pairs trading and selection methods: is cointegration superior? *Applied Economics, 47(6)*, 755-768.

Liew, R., Q., & Wu, Y. (2013). Pairs trading: A copula approach. *Journal of Derivatives & Hedge Funds 19*, 12-30.

Lintilhac, P., S. & Tourin, A. (2016). Model-based pairs trading in the bitcoin markets. *Quantitative Finance, 17(5)*, 703-716.

Kim, T. & Kim, H., Y. (2019). Optimizing the Pairs-Trading Strategy Using Deep Reinforcement Learning with Trading and Stop-Loss Boundaries. *Complex*, 3582516:1-3582516:20.

Mehtab, S., Sen, J. & Dutta, A. (2021). Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models, *Communications in Computer and Information Science, 1366*, 88-106.

Rad, H., Low, R.K.Y. and Faff, R.W. (2016). The profitability of pairs trading strategies: distance, cointegration,and copula methods. Working paper, University of Queensland.

Sarmento, S., M. & Harto, N. (2020). Enhancing a Pairs Trading strategy with the application of Machine Learning. *Expert Systems with Applications, 158*, 113490.

Stübinger, J. & Bredthauer, J. (2017). Statistical Arbitrage Pairs Trading with High-frequency Data. *International Journal of Economics and Financial Issues, 7(4)*, 650-662.

Stübinger, J. & Endres, S. (2017). Pairs trading with a mean-reverting jump–diffusion model on high-frequency data. *Quantitative Finance, 18(10)*, 1735-1751.

Svozil, D., Kvasnicka, V. & Pospichal (1997). Introduction to multi-layer feed-forward neural networks, *Chemometrics and Intelligent Laboratory Systems 39(1)*, 43-62.

Vidyamurthy, G. (2004) *Pairs Trading: Quantitative Methods and Analysis,* John Wiley & Sons.

Wang, J., Rostoker, C. & Wagner, A. (2009). A high performance pair trading application. *2009 IEEE International Symposium on Parallel Distributed Processing.* 23rd IEEE International Symposium on Parallel and Distributed Processing, Rome, Italy. IEEE.

Xie, W., Liew, R., Q., Wu, Y. & Zou, X. (2016). Pairs Trading with Copulas. *The Journal of Trading Summer 2016, 11(3)*, 41-52.

Yang, J., Tsai, S., Shyu, S. & Chang, C. (2016). Pairs trading: The performance of a stochastic spread model with regime switching-evidence from the S&P 500. *International Review of Economics & Finance, 43*, 139-150.

Appendices

Appendix A. Time-series plots H1 spot prices

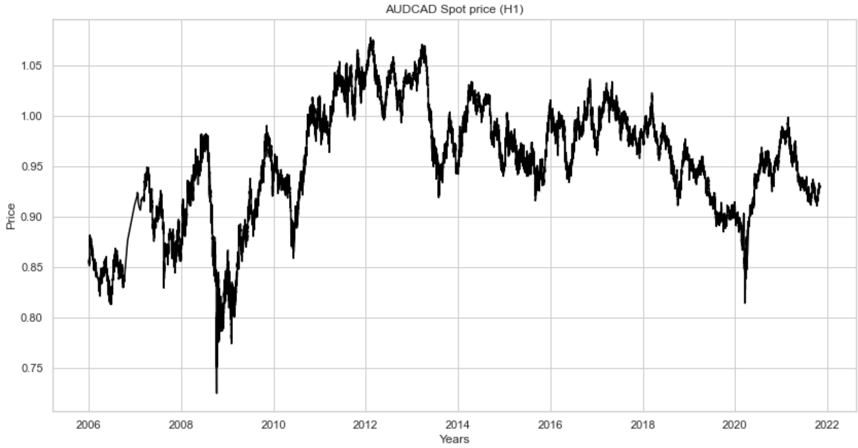


Figure A.1: 1-hour AUDCAD spot price

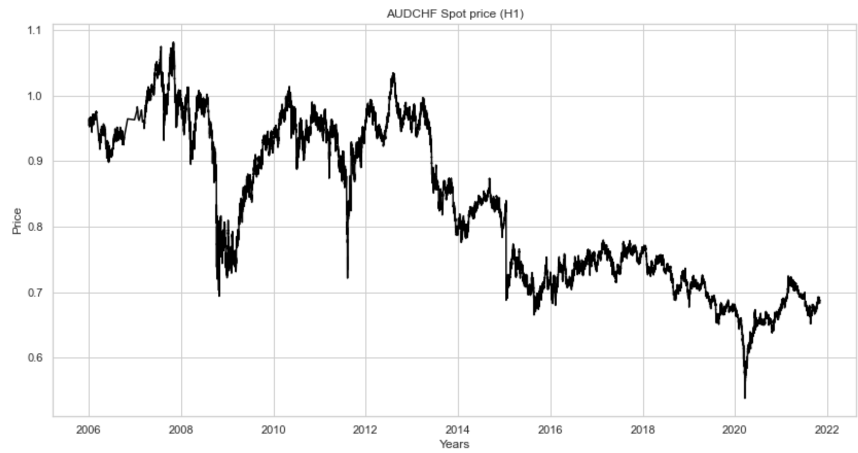


Figure A.2: 1-hour AUDCHF spot price

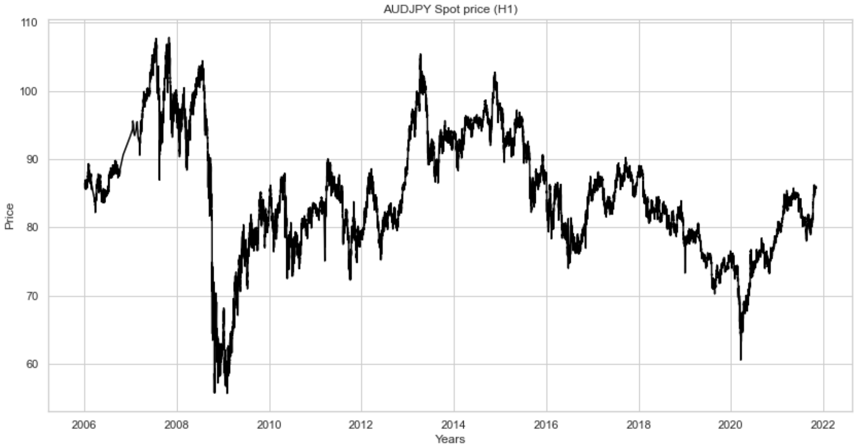


Figure A.3: 1-hour AUDJPY spot price

Figure A.4: 1-hour AUDUSD spot price

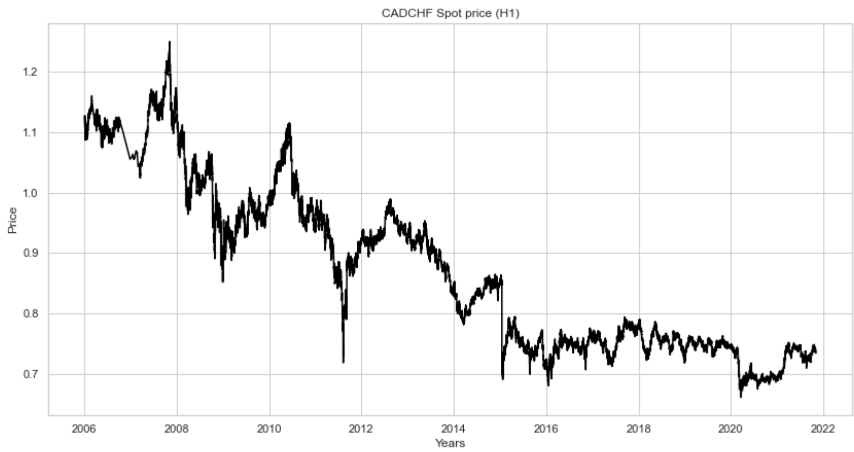


Figure A.5: 1-hour CADCHF spot price

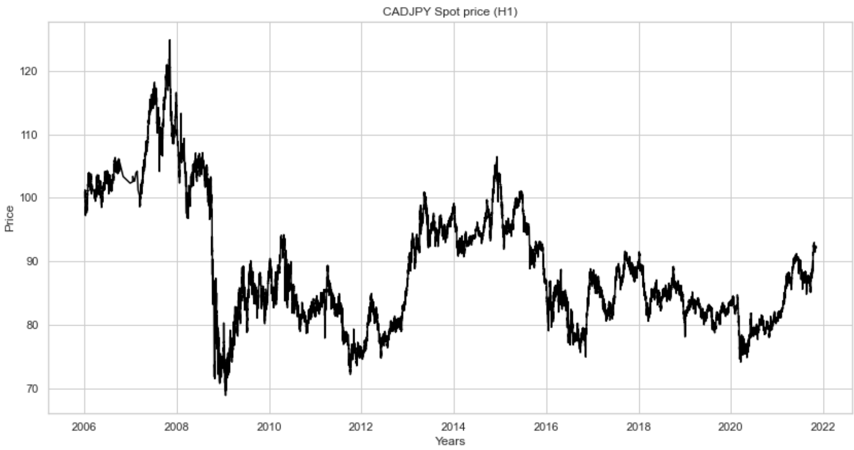


Figure A.6: 1-hour CADJPY spot price

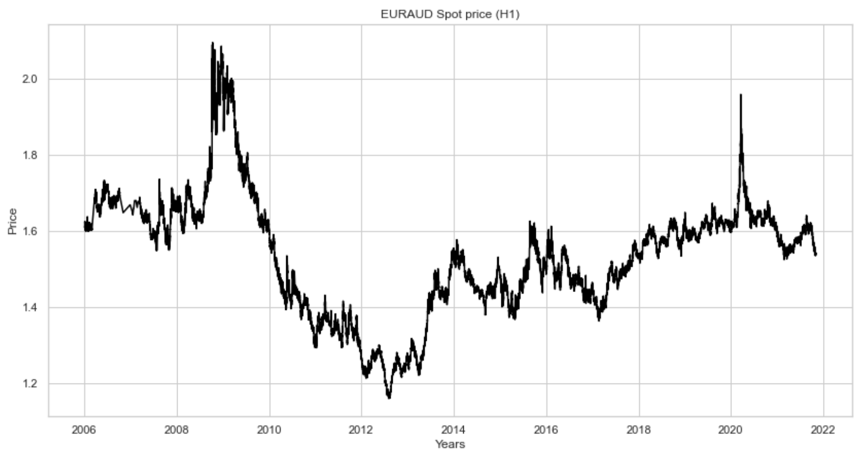


Figure A.7: 1-hour EURAUD spot price

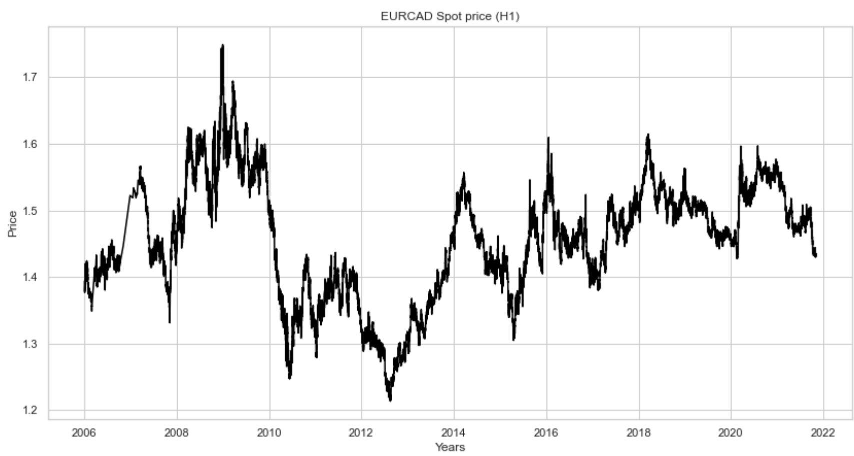


Figure A.8: 1-hour EURCAD spot price

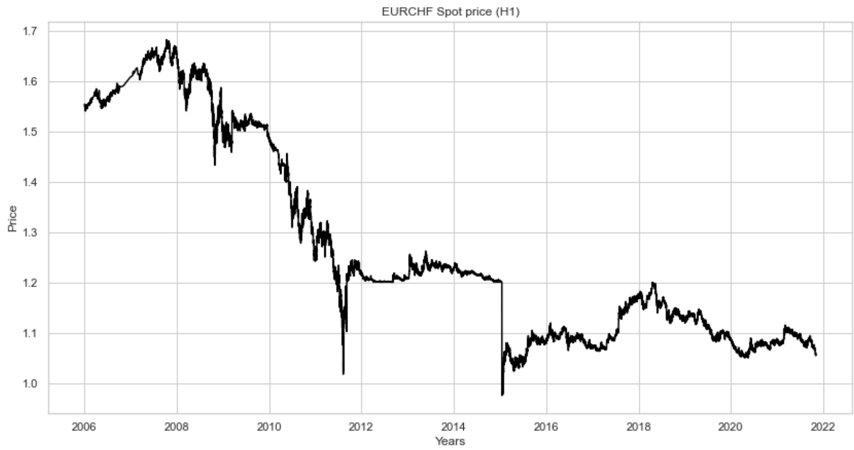


Figure A.9: 1-hour EURCHF spot price

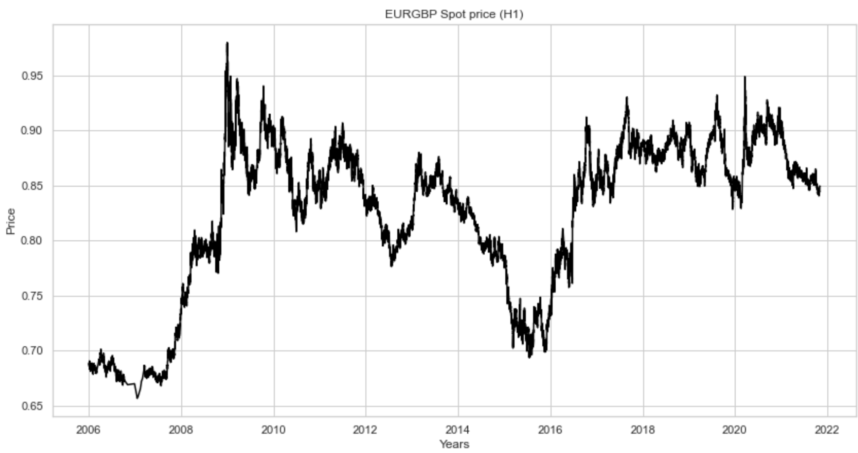


Figure A.10: 1-hour EURGBP spot price

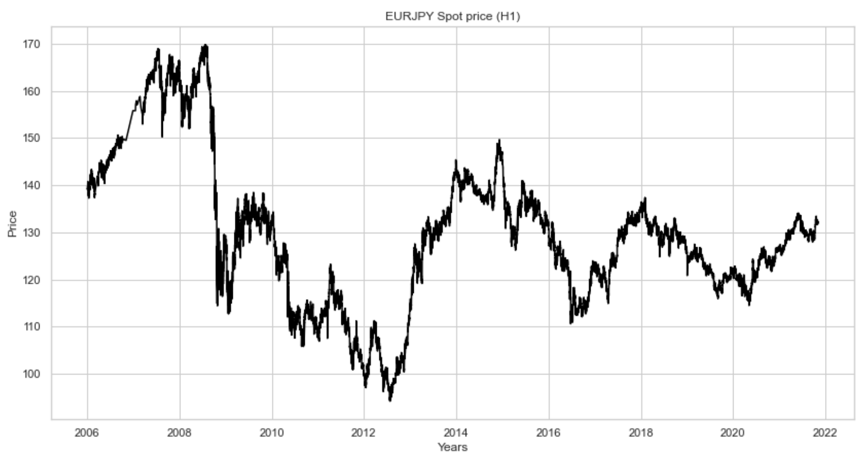


Figure A.11: 1-hour EURJPY spot price

Figure A.12: 1-hour EURNZD spot price

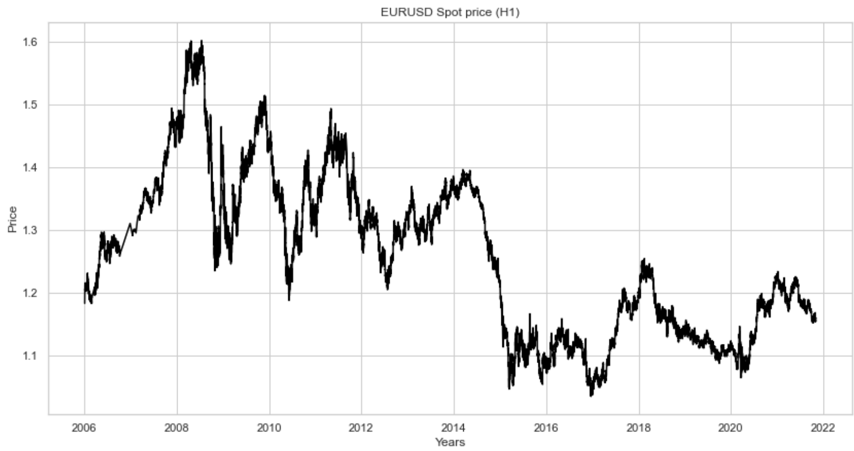


Figure A.13: 1-hour EURUSD spot price

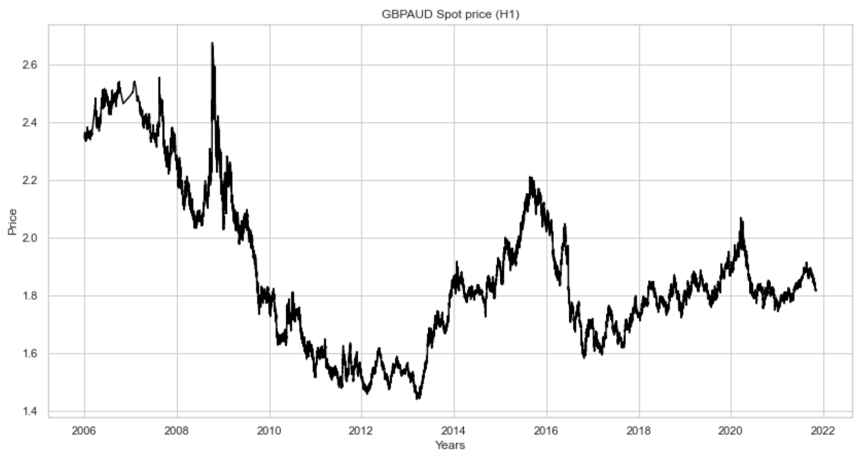


Figure A.14: 1-hour GBPAUD spot price

Figure A.15: 1-hour GBPCAD spot price

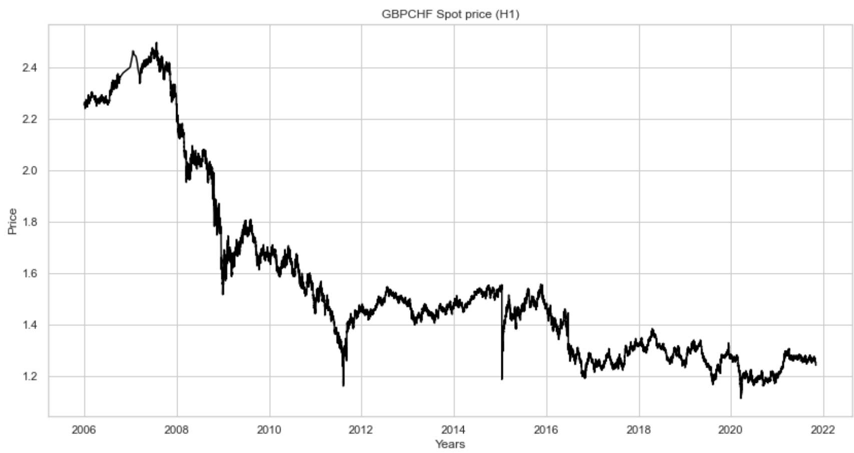


Figure A.16: 1-hour GBPCHF spot price

Figure A.17: 1-hour GBPJPY spot price

Figure A.18: 1-hour GBPNZD spot price

Figure A.19: 1-hour GBPUSD spot price

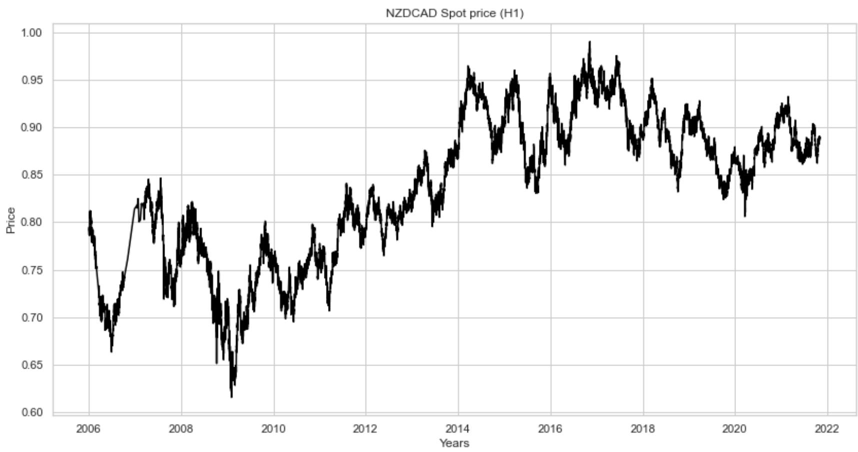


Figure A.20: 1-hour NZDCAD spot price

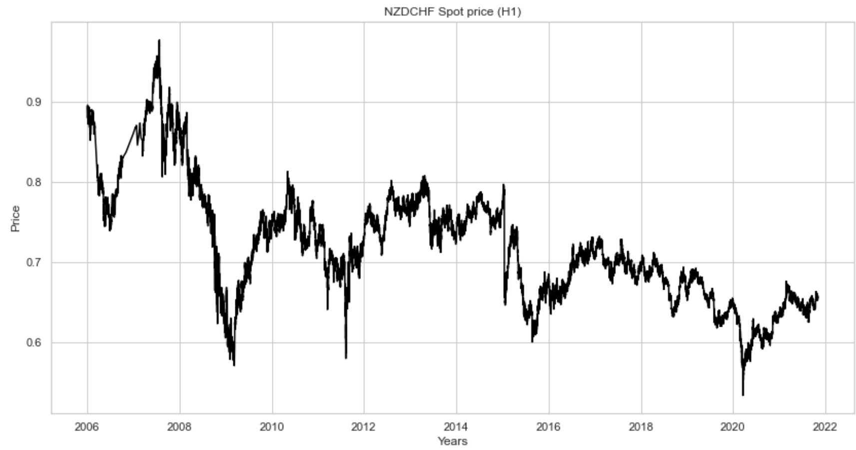


Figure A.21: 1-hour NZDCHF spot price

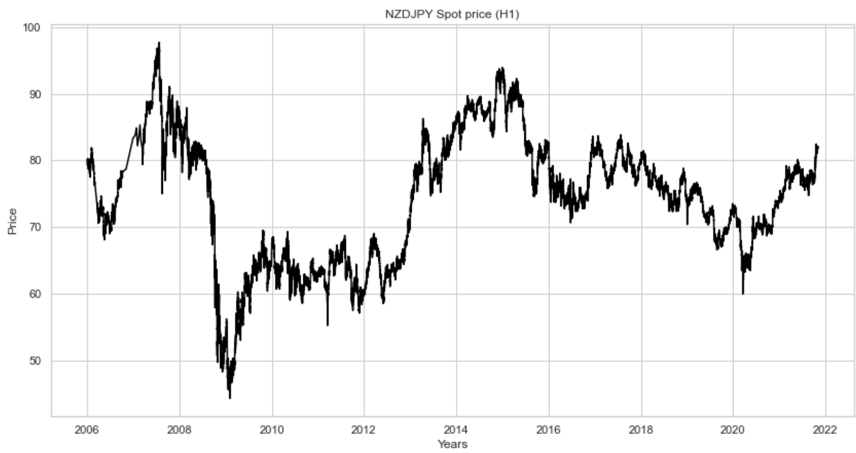


Figure A.22: 1-hour NZDJPY spot price

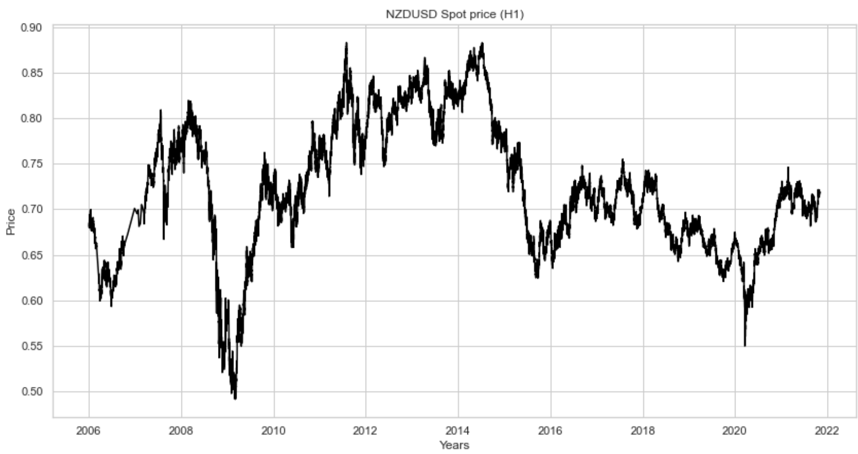


Figure A.23: 1-hour NZDUSD spot price

Figure A.24: 1-hour USDCAD spot price

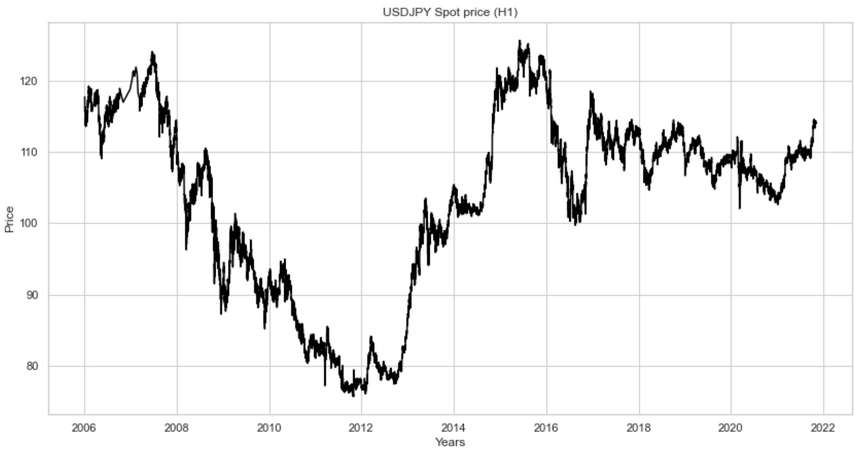


Figure A.25: 1-hour USDJPY spot price

Appendix B. Time-series plots H1 returns

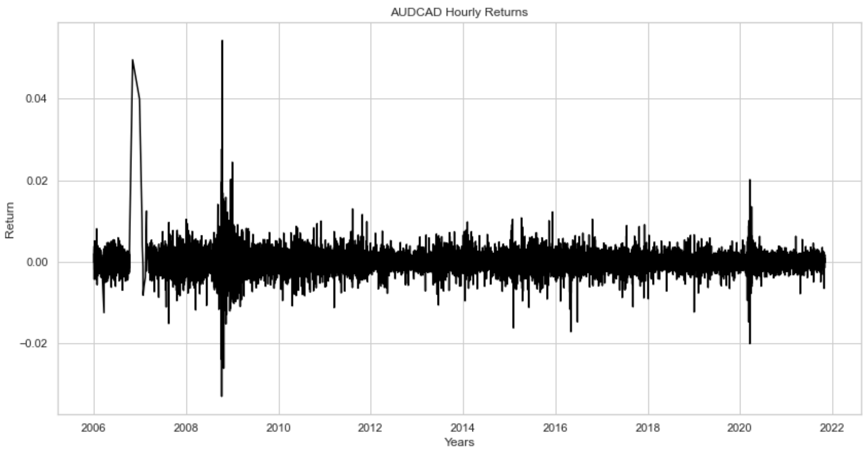


Figure B.1: AUDCAD hourly returns

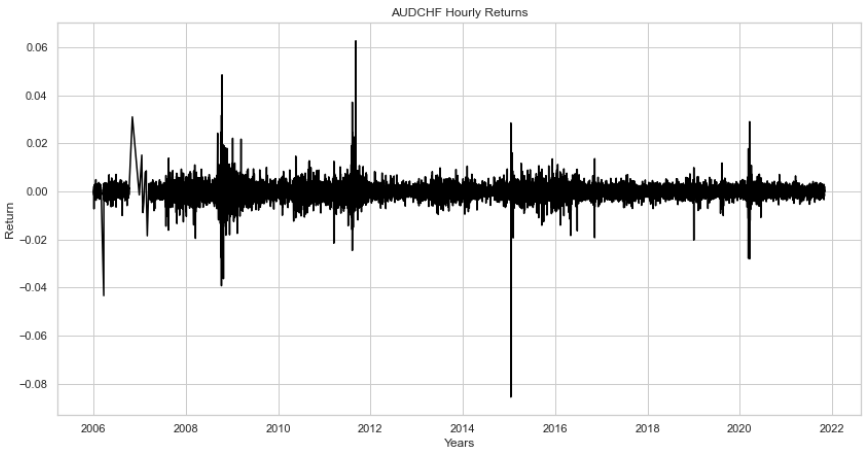


Figure B.2: AUDCHF hourly returns

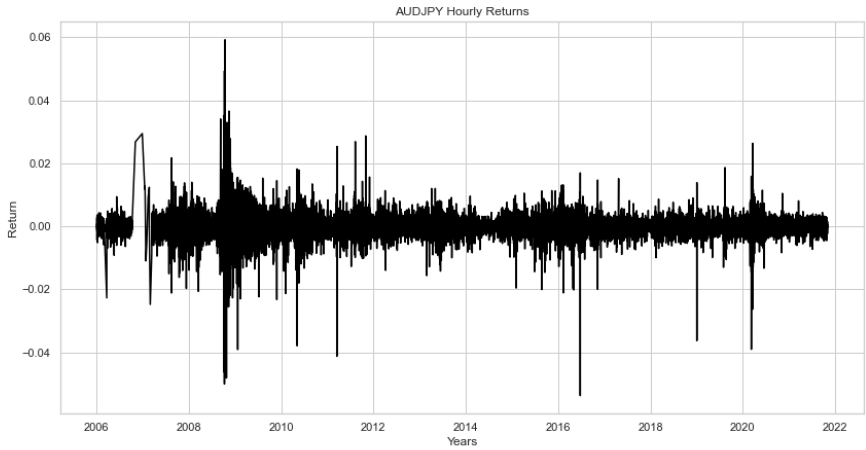


Figure B.3: AUDJPY hourly returns

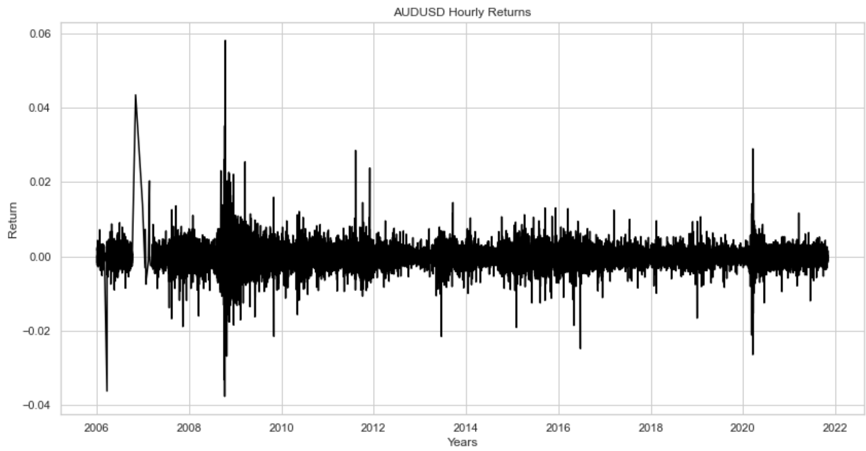


Figure B.4: AUDUSD hourly returns

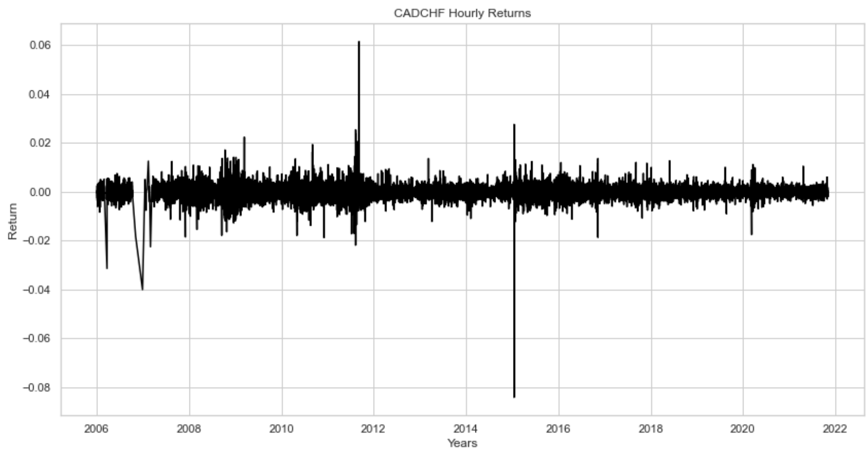


Figure B.5: CADCHF hourly returns

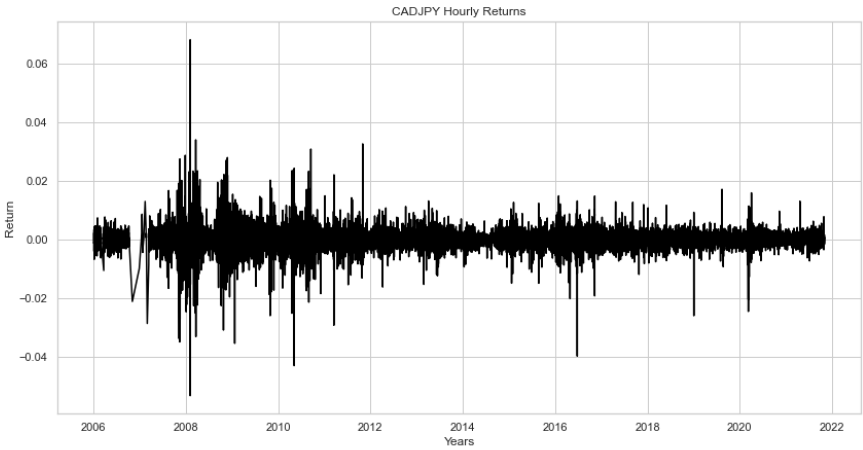


Figure B.6: CADJPY hourly returns

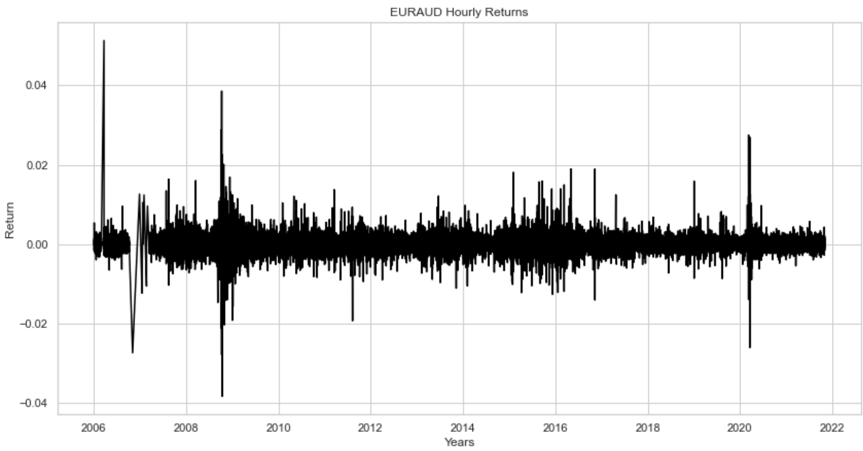


Figure B.7: EURAUD hourly returns

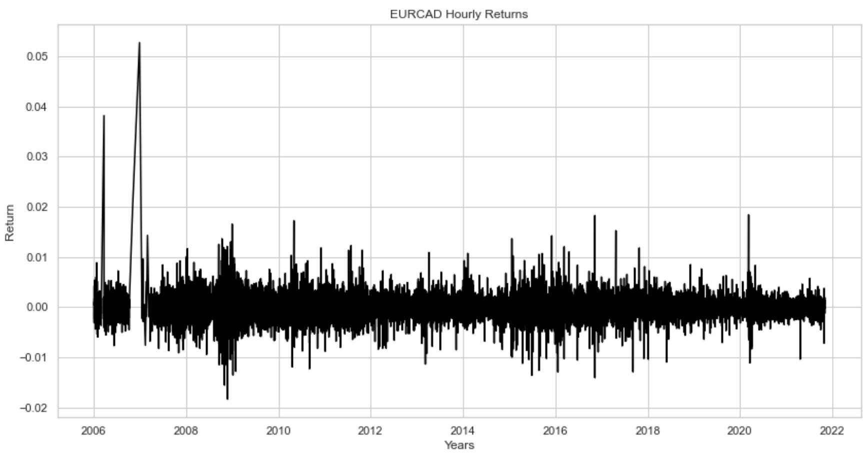


Figure B.8: EURCAD hourly returns

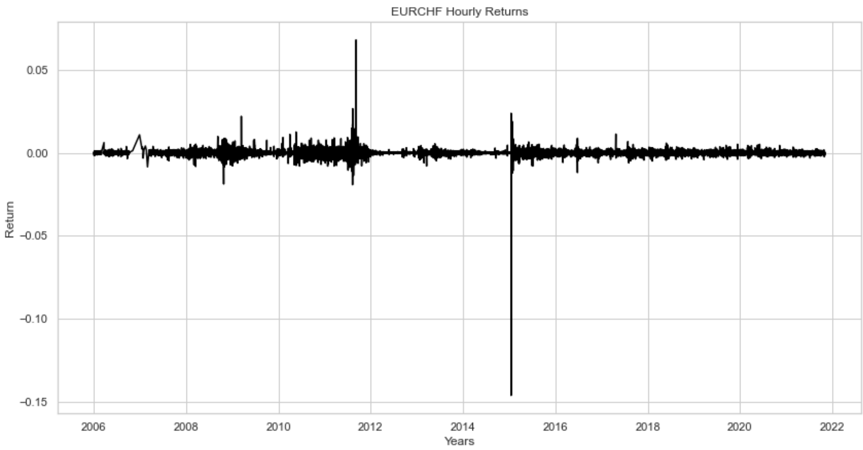


Figure B.9: EURCHF hourly returns

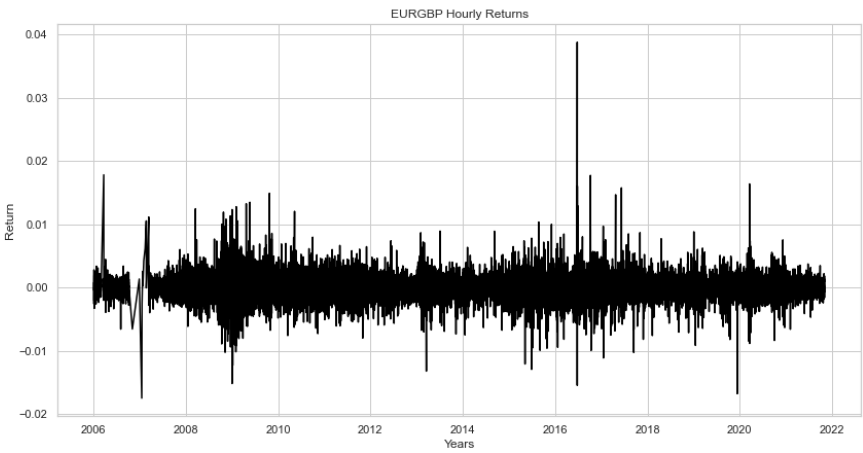


Figure B.10: EURGBP hourly returns

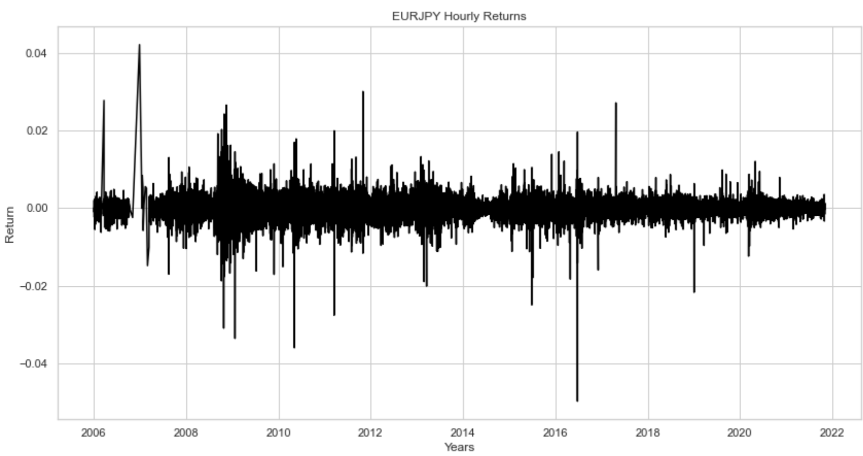


Figure B.11: EURJPY hourly returns

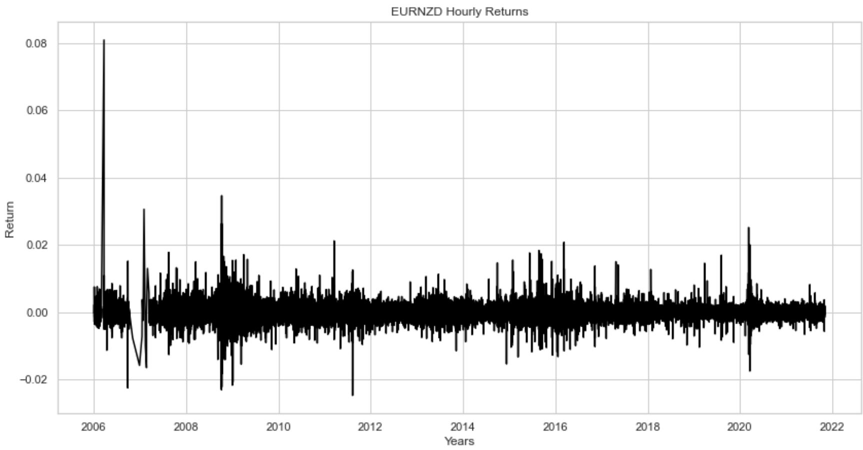


Figure B.12: EURNZD hourly returns

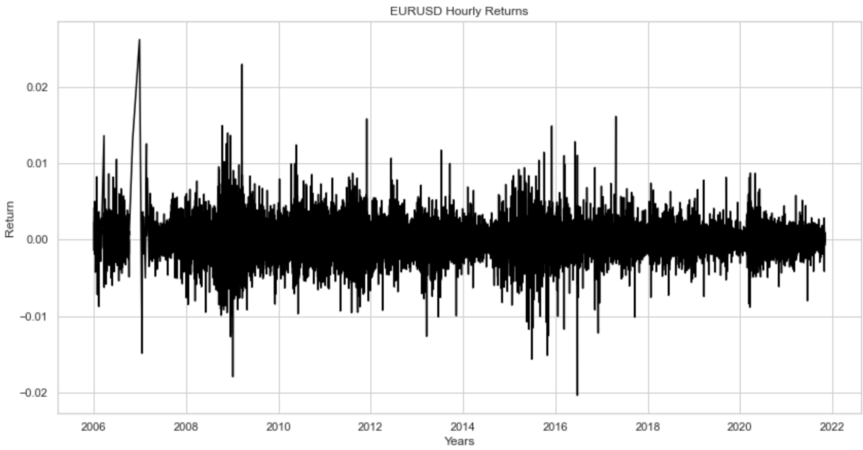


Figure B.13: EURUSD hourly returns

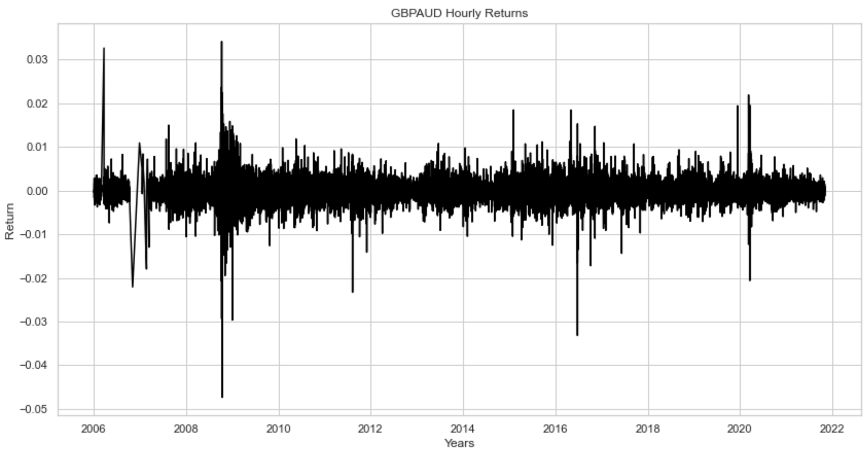


Figure B.14: GBPAUD hourly returns

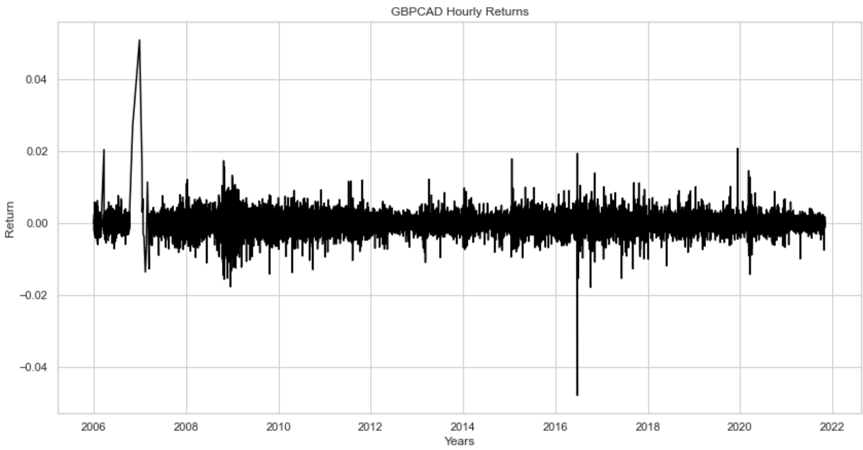


Figure B.15: GBPCAD hourly returns

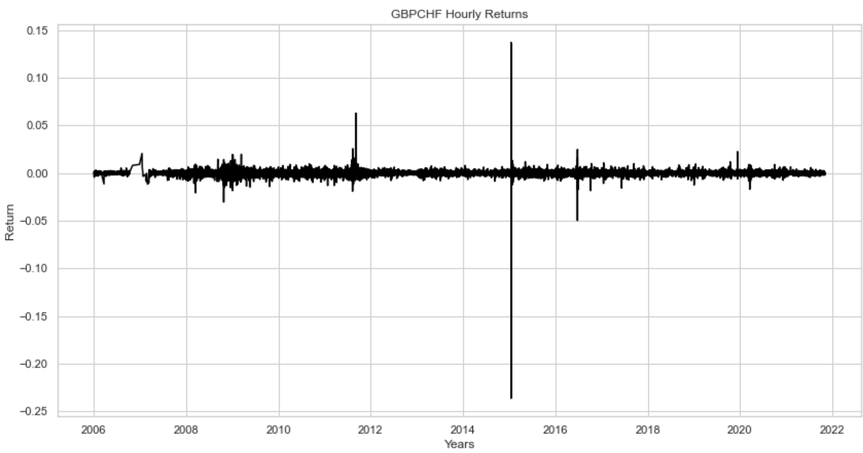


Figure B.16: GBPCHF hourly returns

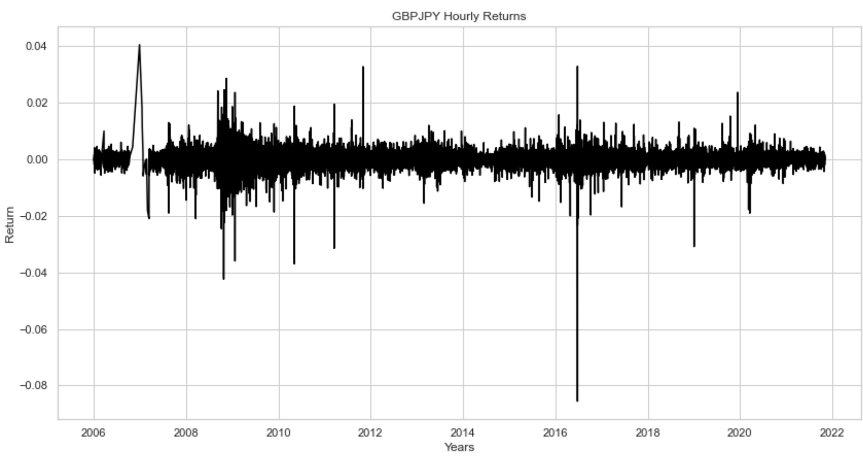


Figure B.17: GBPJPY hourly returns

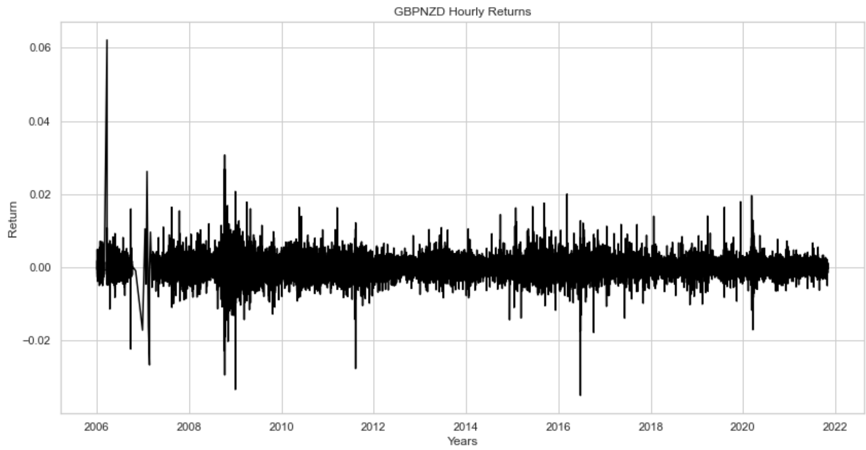


Figure B.18: GBPNZD hourly returns

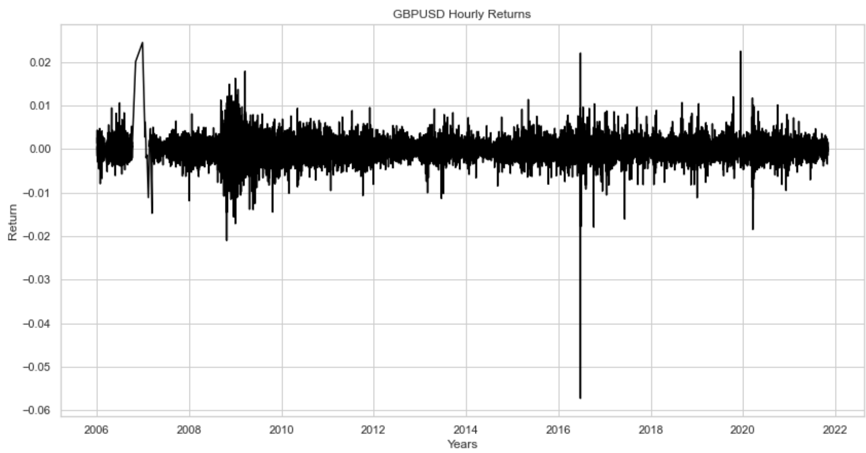


Figure B.19: GBPUSD hourly returns

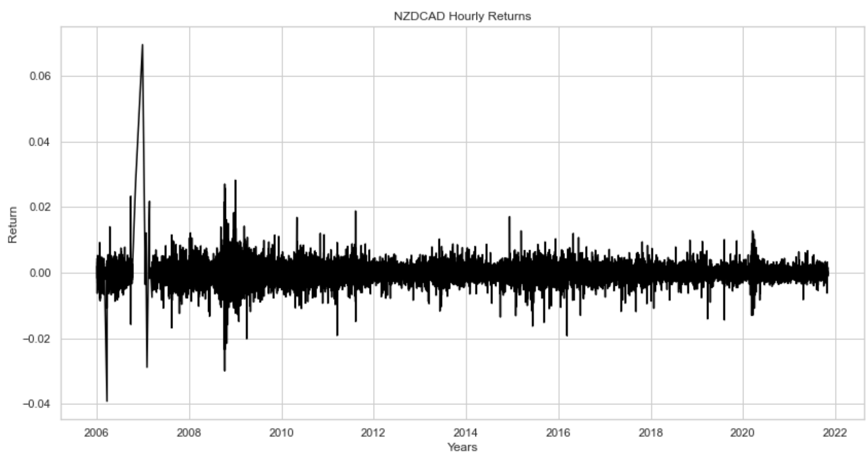


Figure B.20: NZDCAD hourly returns

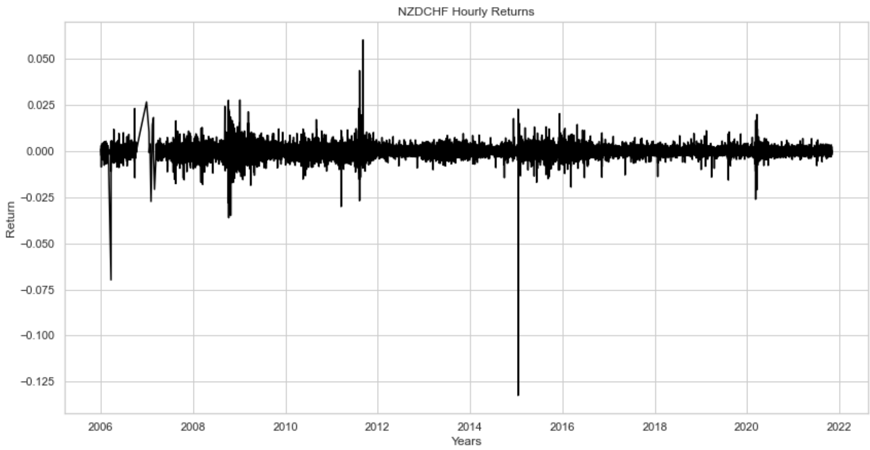


Figure B.21: NZDCHF hourly returns

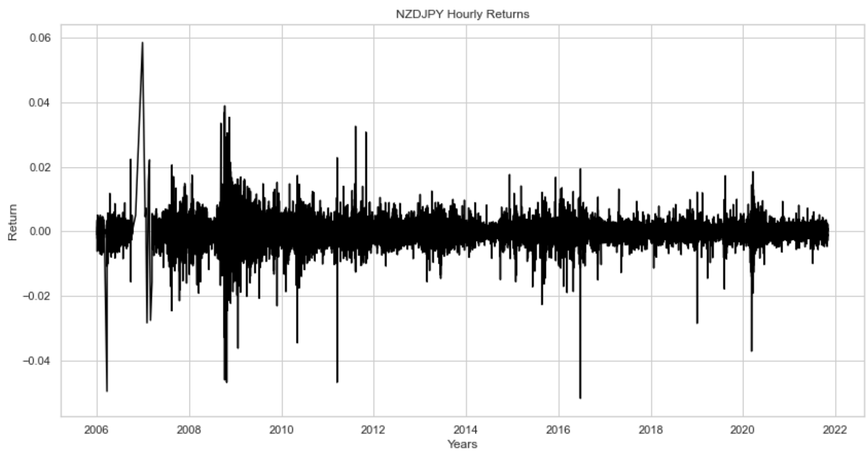


Figure B.22: NZDJPY hourly returns

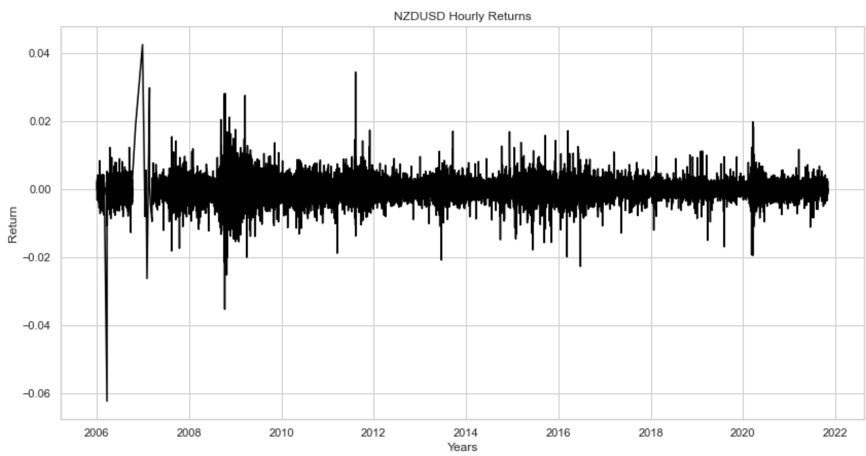


Figure B.23: NZDUSD hourly returns

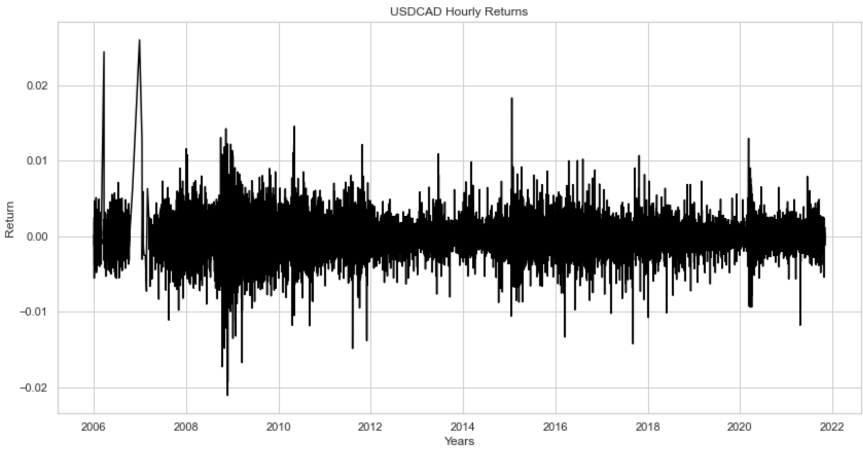


Figure B.24: USDCAD hourly returns

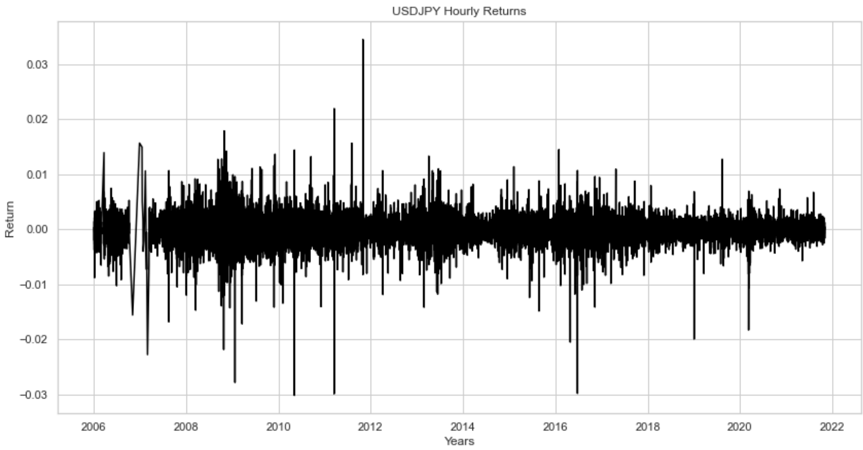


Figure B.25: USDJPY hourly returns

Appendix C. Pair formation statistics

Table C.1: Statistics of the cointegration formation method. The min. and max. values refer to the minimum and maximum number of eligible significant pairs found in one formation period during the whole sample.

ω	Total number of pairs formed	mean	std. dev.	min. number of pairs	max. number of pairs
1000	10154	54	35	3	163
2000	5352	58	36	1	163
3000	3736	61	39	4	163
4000	2564	57	36	4	174
5000	2021	58	43	11	197
6000	1805	62	50	9	189