Discussion Papers in Economics

No. 01/2018

Statistical arbitrage with optimal causal paths on high-frequency data of the S&P 500

Johannes Stübinger
University of Erlangen-Nürnberg

ISSN 1867-6707

Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute for Economics
https://www.iwf.rw.fau.de/research/iwf-discussion-paper-series/

Statistical arbitrage with optimal causal paths on high-frequency data of the S&P 500

Johannes Stübinger[a,1]

[a] *University of Erlangen-Nürnberg, Department of Statistics and Econometrics, Lange Gasse 20, 90403 Nürnberg, Germany*

Wednesday 17th January, 2018

Abstract

This paper develops the optimal causal path algorithm and applies it within a fully-fledged statistical arbitrage framework to minute-by-minute data of the S&P 500 constituents from 1998 to 2015. Specifically, the algorithm efficiently determines the optimal non-linear mapping and the corresponding lead-lag structure between two time series. Afterwards, this study explores the use of optimal causal paths as a means for identifying promising stock pairs and for generating buy and sell signals. For this purpose, the established trading strategy exploits information about the leading stock to predict future returns of the following stock. The value-add of the proposed framework is assessed by benchmarking it with variants relying on classic similarity measures and a buy-and-hold investment in the S&P 500 index. In the empirical back-testing study, the trading algorithm generates statistically and economically significant returns of 54.98 percent p.a. and an annualized Sharpe ratio of 3.57 after transaction costs. Returns are well superior to the benchmark approaches and do not load on any common sources of systematic risk. The strategy outperforms in the context of cryptocurrencies even in recent times due to the fact that stock returns contain substantial information about the future bitcoin returns.

Keywords: Finance, optimal causal path, statistical arbitrage, lead-lag structure, high-frequency trading, cryptocurrency.

JEL classification: C1, C5, C6, G1, G12

Email address: johannes.stuebinger@fau.de (Johannes Stübinger)

[1]The author has benefited from many helpful discussions with Sylvia Endres, Ingo Klein, Julian Knoll, and Sandra Romeis.

1. Introduction

Statistical arbitrage pairs trading is a market neutral strategy which has been developed by a group of quantitative analysts at Morgan Stanley in the mid-1980s (Vidyamurthy, 2004). Following Gatev et al. (2006), the approach identifies pairs of stocks that show a strong relation over a historical time period. In case of temporary anomaly, an arbitrageur goes long in the undervalued stock and goes short in the overvalued stock. If history repeats itself, prices converge to their long-term equilibrium and a profit is drawn.

The majority of literature uses classic similarity measures for finding co-moving securities (see Gatev et al. (2006), Do and Faff (2010), Do and Faff (2012), Huck and Afawubo (2015), Rad et al. (2016), and Stübinger and Endres (2018)). Specifically, these studies quantify the similarity between two time series $x = (x(1), ..., x(N)) \in \mathbb{R}^N$ and $y = (y(1), ..., y(N)) \in \mathbb{R}^N$ by the distance

$$d(x, y) = \sum_{i=1}^{N} d(x(i), y(i)), \tag{1}$$

where $d(x(i), y(i))$ describes the distance at fixed time i ($i \in \{1, \ldots, N\}$). By construction, the measure outlined in equation (1) is very sensitive to misalignments and time shifts (Ding et al., 2008). This drawback is eliminated by introducing a model that permits an elastic adjustment of the time axis in order to identify sequences that are similar but out of phase. For this purpose, the co-moving between the sequences $x = (x(1), ..., x(N)) \in \mathbb{R}^N$ and $y = (y(1), ..., y(M)) \in \mathbb{R}^M$ is specified by

$$c(x, y) = \sum_{i=1}^{I} c(x(n_i), y(m_i)), \tag{2}$$

where c describes the local cost measure and $I \in \{\max(N, M), \ldots, N + M - 1\}$. The concept of dynamic time warping provides an efficient technique for finding the most suitable non-linear mapping by minimizing the measure depicted in equation (2). In stark contrast to classic similarity measures, this method is in a position both to handle time series with different lengths and to be robust against amplitude change, migration, and noise of time series (Wang et al., 2012).

Due to its superior flexibility, dynamic time warping is applied in a wide range of research areas. Originally, it is used within the framework of spoken word recognition, i.e., the technique eliminates non-linear time shifts between two speech patterns caused by different speaking rates (Juang, 1984; Rath and Manmatha, 2003; Muda et al., 2010). In recent times, dynamic time warping is especially utilized in gesture recognition (Arici et al., 2014; Cheng et al., 2016), chemistry (Jiao et al., 2014; Dupas et al., 2015), and medicine (Rakthanmanon et al., 2012; Fu et al., 2017). Surprisingly, there exist only two academic studies in the context of statistical arbitrage trading. Chinthalapati (2012) adds a curvature energy term to the existing method and employs it to intraday-data of 97 selected stocks from NYSE on January 1st, 2006. Notably, the proposed directional trading represents no statistical arbitrage strategy in the sense of Avellaneda and Lee (2010). Kim and Heo (2017) use dynamic time warping for detecting similar patterns on daily prices of the KOSPI 100 index stocks from January 2005 to June 2015.

This paper enhances the existing research in several aspects. First, the manuscript contributes to the literature by introducing the optimal causal path algorithm, which determines the most suitable lag between two time series using a parameter-free procedure. The performance of the 3-step algorithm is demonstrated with the aid of a simulation study. Second, the essay develops a fully-fledged statistical arbitrage framework based on optimal causal paths. Top pairs are selected possessing the most stable lead-lag structure during the formation period. In the out-of-sample trading period, information about the returns of the leading stock are exploited to predict the future returns of the following stock. Third, the value-add of the proposed trading framework is assessed by benchmarking it with well-known quantitative strategies in the same area of research. Specifically, the paper considers statistical arbitrage trading variants on the basis of correlation, Manhattan distance, and lagged cross-correlation as well as an S&P 500 long-only benchmark. Fourth, this article presents the first academic contribution applying a large-scale empirical study of a sophisticated backtesting framework on minute-by-minute data of the S&P 500 constituents from January 1998 to December 2015. The strategy generates statistically and economically significant returns of 54.98 percent p.a. after transaction costs. The results are far superior in comparison to the benchmarks ranging from 2.19 percent for a naive buy-and-hold investment in the S&P 500 index to 33.72 percent for the algorithm adapted from lagged cross-correlation. Fifth, the manuscript proves the strategy's profitability in the context of cryptocurrencies in the sample period from 2012 to 2015. A deep-dive analysis shows that stock returns include substantial information about the bitcoin returns in the future. This result posits a severe challenge to the semi-strong form of market efficiency even in recent times.

The paper is organized as follows. In section 2, a detailed description of the theoretical concept is provided. Section 3 introduces the optimal causal path algorithm and conducts a simulation study. Section 4 specifies the study design of the back-testing framework. Empirical results and key findings are presented in section 5. Finally, section 6 concludes and provides suggestions for further research.

2. Theoretical concept

The concept of dynamic time warping aims at identifying the relation structure of two given time series $x = (x(1), ..., x(N)) \in \mathbb{R}^N$ and $y = (y(1), ..., y(M)) \in \mathbb{R}^M$. The underlying non-linear alignment between two temporal sequences is described with the aid of warping paths.

Following Keogh and Ratanamahatana (2005), a sequence of points $p = (p_1, \dots, p_I)$ with $p_i = (n_i, m_i) \in \{1, \dots, N\} \times \{1, \dots, M\}$ for $i \in \{1, \dots, I\}$ ($I \in \{\max(N, M), \dots, N+M-1\}$) is called warping path if the following three properties are satisfied:

1. Boundary condition: $p_1 = (1, 1)$ and $p_I = (N, M)$.
2. Monotonicity condition: $n_1 \leq n_2 \leq \cdots \leq n_I$ and $m_1 \leq m_2 \leq \cdots \leq m_I$.
3. Step size condition: $p_{i+1} - p_i \in \{(1, 0), (0, 1), (1, 1)\}, \quad \forall i \in \{1, \dots, I-1\}$.

It should be noted that the step size condition implies the monotonicity condition, which nonetheless is indicated for the sake of clarity. Let P be the set of all possible warping paths between the input time series x and y. The total cost of a warping path p ($p \in P$) is defined by

$$c_p(x, y) = \sum_{i=1}^{I} c(x(n_i), y(m_i)), \tag{3}$$

where c describes the local cost measure. As such, $c_p(x, y)$ characterizes the sum of differences between the realizations of x at time n_i and y at time m_i ($i \in \{1, \dots, I\}$). Typically, the cost measure is based on the Manhattan distance (Müller, 2007; Li and Clifford, 2012; Zhang et al., 2012) or the Euclidean distance (Vlachos et al., 2002; Senin, 2008; Coelho, 2012). The optimal warping path p^* between x and y depicts the lowest total cost among all possible warping paths:

$$p^* = \underset{p \in P}{argmin}\, c_p(x, y). \tag{4}$$

Calculating the total cost $c_p(x, y)$ for all possible warping paths $p \in P$ would yield to a complexity of exponential order. Therefore, the optimal warping path p^* is determined using dynamic programming, i.e., the underlying problem is divided into sub-problems. The corresponding solutions are stored for future reference leading to a lower time complexity $\mathcal{O}(NM)$. The total cost of p^* is defined as $c_{p^*}(x, y)$, i.e., the sum of all local costs of p^*. Figure 1 illustrates the local costs and the identified optimal warping path p^* given two time series. To visualize this, the sequence of points p^* runs along a "valley" of low cost (light colors) and avoids "mountains" of high cost (dark color).

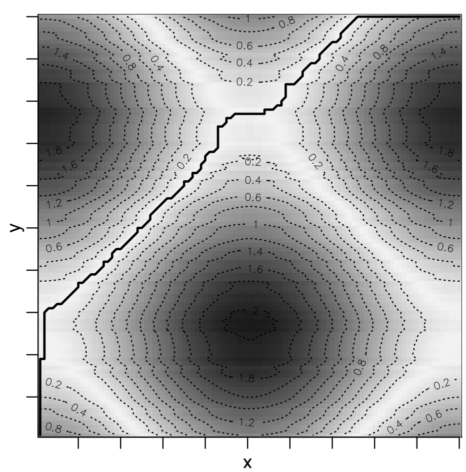


Figure 1: Local costs of two time series and the corresponding optimal warping path p^* (solid line). Regions of high cost (low cost) are indicated by dark colors (light colors).

In addition to the three conditions outlined above, academic research introduces global and local conditions on the warping path with the main purpose of speeding up the computational run time. Global constraints aim at limiting the deviation of a warping path from the diagonal – key representatives are given by the Sakoe-Chiba band (Sakoe and Chiba, 1978) and the Itakura parallelogram (Itakura, 1975) (see figure 2). Local constraints modify the step size condition by altering the set of steps or favoring specific step directions (see Myers et al. (1980), Myers and Rabiner (1981), Rabiner and Juang (1993), and Berndt and Clifford (1994)). Nonetheless, this manuscript avoids any restrictions on the warping path because global and local constraints both imply further parameter settings and generate insufficient results in the vast majority of domains (see Salvador and Chan (2007)).

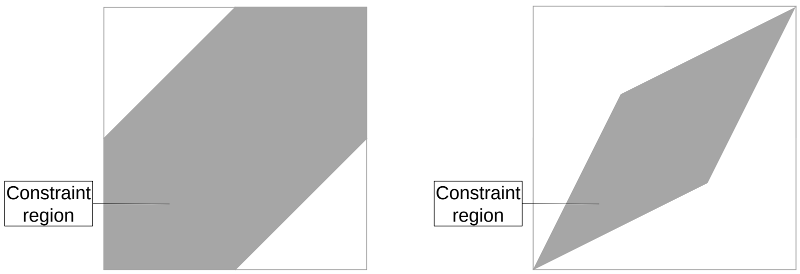


Figure 2: Sakoe-Chiba band (left) and Itakura parallelogram (right).

Nowadays, research studies either focus on optimizing the run time of dynamic time warping or center the development of a generalized model framework. Across all contributions, the setting of model parameters takes a central part – the criticism of arbitrariness and data snooping is omnipresent.

In the context of optimization, Keogh and Pazzani (2000) introduce a modification of dynamic time warping that exploits a higher level representation of time series data. Müller et al. (2006) and Salvador and Chan (2007) recursively project an alignment path computed at a coarse resolution level to the next higher level and then to refine the projected path. Al-Naymat et al. (2009) dynamically utilize the possible existence of inherent similarity and correlation between two time series. Prätzlich et al. (2016) introduce a memory-restricted alignment procedure that combines concepts from Müller et al. (2006) with the idea of using rectangular local constraint regions. Silva and Batista (2016) apply an upper bound estimation to prune unpromising warping alignments.

In the context of generalization, Sornette and Zhou (2005) generalize the optimal search by adding a Boltzmann factor proportional to the exponential of the global mismatch of this path. Zhou and Sornette (2006) test the introduced methodology on the dynamical time evolution of the lead-lag structure between two arbitrary time series. Meng et al. (2017) present a symmetric variant to determine the time-dependent lead-lag relation.

3. Optimal causal path algorithm

3.1. Methodology

This section presents a non-parametric approach, called "optimal causal path algorithm", which determines the optimal causal path and its corresponding lead-lag relation given two time series $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$. Without loss of generality, the description assumes $N \geq M$.

Step A determines the optimal causal path under the assumption of a constant lead-lag structure, i.e., the time series exhibit a fixed lag. First, a loop measures the total costs of the causal paths supposing lag l ($l \in \{0, \dots, M-1\}$). In case of $N = M$, the starting value $l = 0$ results in the well-known Manhattan distance. Each statement defines the considered causal path (n, m), where

$$n = (\underbrace{1, \dots, 1}_{\mathbb{R}^l}, \underbrace{1, \dots, N}_{\mathbb{R}^N}) \in \mathbb{R}^{N+l} \text{ and } m = (\underbrace{1, \dots, M}_{\mathbb{R}^M}, \underbrace{M, \dots, M}_{\mathbb{R}^l}, \underbrace{M, \dots, M}_{R^{N-M}}) \in \mathbb{R}^{N+l}. \quad (5)$$

To visualize equation (5), the sequence of points represents a diagonal shifted by the number of lags l and connected with the corners $(1, 1)$ and (N, M). The function $eval_A$ quantifies the total cost of the causal path (n, m). Second, the algorithm ascertains the lag l^* indicating the lowest total cost of all regarded causal paths with a constant lead-lag structure. The associated causal path (n_1, m_1) provides the initial setting for step B.

Step B specifies the optimal causal path permitting a varying lead-lag structure. For this purpose, a loop enhances gradually the causal path with the objective of reducing the total cost. In each iteration step, the function $eval_B$ arranges the unrestricted elements of the current causal path (n_{h-1}, m_{h-1}) in descending order ($h \geq 2$). Then, $eval_B$ successively examines whether the fixed element combined with its neighborhood depicts a local optimal path. If there exists an admissible path in the vicinity of this element with lower cost than the current local path, then the new sequence of points substitutes the existing one. The loop ends when the updated path (n_h, m_h) equals the current path (n_{h-1}, m_{h-1}). This procedure guarantees that the algorithm provides the optimal causal path.

Step C determines the most suitable lag by calculating the arithmetic mean of all differences between the indices of the optimal causal path. The fluctuation around the optimal lag is defined as the corresponding standard deviation. The algorithm returns both the estimated lag and the appropriated deviation of the optimal causal path.

Algorithm 1 Optimal causal path algorithm

Input: Time series $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$ $(N \geq M)$ as well as local cost measure

Output: The optimal causal path, the corresponding estimated lag,
and the fluctuation of the unrestricted elements

Step A – Determine the optimal lag l^* assuming a constant lead-lag structure

$eval_A$: Function returning the total cost of a fixed causal path
for given time series x and y

$l = 0$;

loop

$n \leftarrow (1, \ldots, 1, 1, \ldots, N) \in \mathbb{R}^{N+l}$;

$m \leftarrow (1, \ldots, M, M, \ldots, M) \in \mathbb{R}^{N+l}$;

$c[l+1] \leftarrow eval_A(x[n], y[m])$;

$l \leftarrow l + 1$;

if $l = N$ **then break**;

end loop

$l^* \leftarrow argmin(c[1], \ldots, c[N]) - 1$;

Step B – Determine the optimal causal path permitting a varying lead-lag structure

$eval_B$: Function returning the causal path with local optimal paths
for given time series x and y

$h \leftarrow 1$;

$n_h \leftarrow (1, \ldots, 1, 1, \ldots, N) \in \mathbb{R}^{N+l^*}$;

$m_h \leftarrow (1, \ldots, M, M, \ldots, M) \in \mathbb{R}^{N+l^*}$;

loop

$h \leftarrow h + 1$;

$P \leftarrow eval_B(x[n_{h-1}], y[m_{h-1}])$;

$n_h \leftarrow P[, 1]$; $m_h \leftarrow P[, 2]$;

if $(n_h, m_h) - (n_{h-1}, m_{h-1}) = 0$ **then break**;

end loop

$n \leftarrow n_h$; $m \leftarrow m_h$

Step C – Determine lag and corresponding standard deviation of the optimal causal path

3.2. Simulation study

In this section, a simulation study with synthetic data is carried out in order to validate the optimal causal path algorithm. Following Sornette and Zhou (2005) and Zhou and Sornette (2006), two stationary time series $X = (X_t)_{t\in\{1,\dots,N\}}$ and $Y = (Y_t)_{t\in\{1,\dots,N\}}$ are constructed under the assumption that X leads Y by time lag l ($l \in \mathbb{N}_0$). Mathematically, the leading time series X is defined by the following autoregressive process:

$$X(t) = bX(t-1) + \nu(t),$$

where $b < 1$ and $\nu(t) \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_X^2)$. The stochastic process Y is given by

$$Y(t) = aX(t-l) + \varepsilon(t),$$

where $a \in \mathbb{R}$ and $\varepsilon(t) \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_Y^2)$. The parameter $f = \sigma_Y^2/\sigma_X^2$ specifies the amount of noise diminishing the dependence between X and Y.

The baseline parameter setting follows Sornette and Zhou (2005) and Zhou and Sornette (2006), i.e., we set $N = 100$, $l = 5$, $a = 0.8$, $b = 0.7$, $\sigma_X^2 = 1$, and $f = 1$. Furthermore, this manuscript defines the local cost measure c as the absolute difference between $x(n_i)$ and $y(m_i)$ ($i \in \{1,\dots,I\}$), see equation (3). We vary ceteris paribus the sample size N, the coefficient a, and the amount of noise f – the other conditions remain the same since they do not directly affect the dependency between both time series. Then, algorithm 1 is used to identify the optimal causal path, to estimate the lead-lag structure, and to calculate the corresponding total cost. Following McFadden and Train (2000), Ilzetzki et al. (2013), and Létourneau and Stentoft (2014), 1,000 repetitions for each parameter constellation are conducted. Figure 3 portrays the resulting boxplots of the average total costs $\bar{c}_{p^*}(x,y)$ (left column) and the estimated lags $\hat{l}$ (right column) for varying the parameters N, a, and f.

First of all, we observe that an increasing sample size N leads to lower average total costs $\bar{c}_{p^*}(x,y)$ – this fact is not surprising since the percentage of data pairs with lag l grows. Simultaneously, total range and interquartile range decrease close to zero indicating robustness and prediction accuracy. As expected, the estimated lag converges to the true value, e.g., $\hat{l}$ and l are identical in more than 97.5 percent of all cases for $N = 50$.

Furthermore, the average total costs $\bar{c}_{p^*}(x,y)$ decline for ascending parameter a due to the fact that the dependency between both time series gets stronger. Notably, the hit ratio of the estimated lag, i.e., the percentage with identical $\hat{l}$ and l, is above 90 percent even for a low-mid value of $a = 0.4$. We observe a symmetric boxplot in case of $a = 0$ because this parameter constellation implies no direct relation between x and y.

Finally, augmenting f causes rising average total costs $\bar{c}_{p^*}(x, y)$ with larger differences between maximum and minimum as well as upper and lower quartile. If σ_X^2 and σ_Y^2 are at a similar level, we find high precision of the estimated lags. An increasing amount of noise provokes that the median of the estimated lags $\hat{l}$ converges to zero and the corresponding ranges widen out.

Summarizing, the optimal causal path algorithm shows strong performance in the vast majority of parameter constellations with respect to robustness, efficiency, and feasibility.

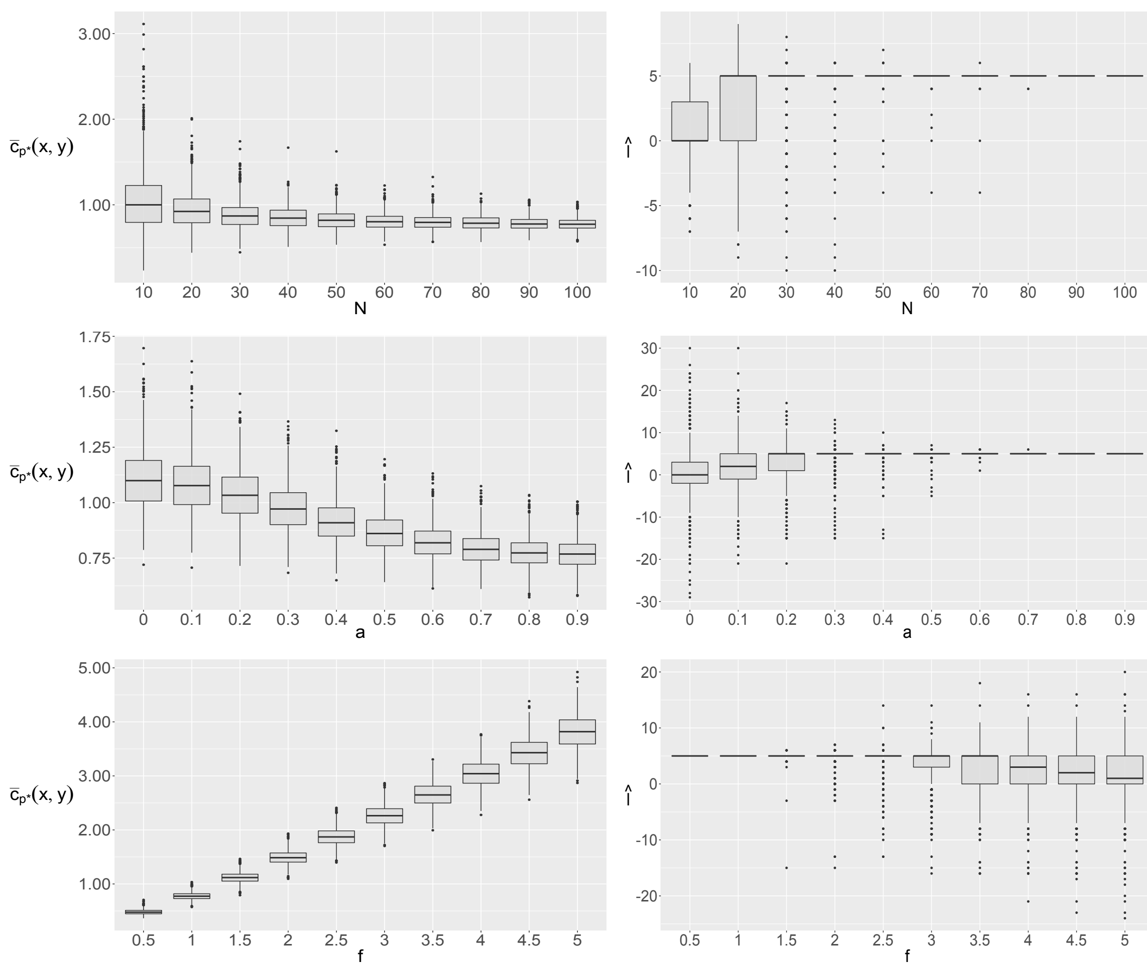


Figure 3: Boxplots of the average total costs $\bar{c}_{p^*}(x, y)$ (left column) and estimated lags $\hat{l}$ (right column) for varying the length of the time series N (first row), the coefficient a (second row), and the amount of noise f (third row).

4. Study design

The empirical back-testing framework is conducted on minute-by-minute prices on the S&P 500 index constituents from January 1998 to December 2015 (see subsection 4.1). Following Gatev et al. (2006), the data set is sliced into 4527 overlapping study periods, each shifted by one day. Each study period consists of a 1-day formation period (subsection 4.2) and a 1-day out-of-sample trading period (subsection 4.3). While the former trains the model and selects the most suitable pairs using pre-defined criteria, the latter trades the top pairs applying rule-based entry and exit signals.

4.1. Back-testing framework

The empirical application is performed on minute-by-minute data of the S&P 500 from January 1998 to December 2015. This highly liquid stock universe comprises the stocks of the 500 leading blue-chip companies which provide high-quality, widely accepted commodities and services. This data set serves as a crucial test for any potential capital market anomaly since the S&P 500 index covers 80 percent of total U.S. market capitalization (S&P Dow Jones Indices, 2015). Following Stübinger and Endres (2018), a 2-stage process is implemented with the aim of removing any survivor bias from the data. First, a constituent list for the S&P 500 stocks is obtained from QuantQuote (2016) from January 1998 to December 2015. The framework exploits this information by creating a binary matrix – rows characterize the trading days and columns specify all stocks having ever been listed in the index. Each element of this matrix indicates a “1” if the corresponding corporation is a constituent of the S&P 500 index at the associated day, otherwise a “0”. Second, the full archive of minute-by-minute stock prices from January 1998 to December 2015 is downloaded from QuantQuote (2016). The associated stock exchange is opened from 9.30 am to 4.00 pm Eastern time, Monday through Friday. Consequently, the minute-by-minute price time series of one stock involves 391 data points per day. Data are adjusted by stock splits, dividends, and further corporate actions. Performing these two steps, the study design is able to entirely replicate the S&P 500 constituency and the appropriated price time series.

The introduced methodology and all relevant evaluations are conducted in the statistical programming language `R` (R Core Team, 2017). The source code of computationally intensive tasks is implemented in `C++` and connected to `R`.

4.2. Formation period

The 391-minute formation period conducts both an in-sample training of all possible pair combinations and a selection procedure to find the most suitable pairs for the trading period. Typically, the S&P 500 index comprises 500 stocks, i.e., the strategy handles $500 \cdot (500 - 1)/2 = 124,750$ pairs per study period. For each pair, algorithm 1 is applied to the respective return time series. Outputs are the optimal lag and the corresponding fluctuation of the unrestricted part of the optimal causal path.

The model selects the top s pairs ($s \in \mathbb{N}$) exhibiting the most stable lead-lag structure during the formation period. To be more specific, the top s pairs with the lowest standard deviation around the specified lag are transferred to the trading period. Furthermore, two additional constraints are applied to secure a clear lead-lag relationship. The algorithm only considers pairs possessing non-zero lags and no lead-lag change during the formation period.

4.3. Trading period

The top pairs with lowest fluctuation around the specified lag are transferred to the 391-minute trading period (T_{tra}). If the assumption holds and algorithm 1 captures the correct lead-lag structure, then the strategy is in a position to predict the future returns of the following stock by exploiting the information about the leading stock. To be more specific, the algorithm generates trading signals for the following stock based on the development of the leading stock. Without loss of generality, the following lines assume that x leads y by l minutes.

Every incoming price of the leading time series at time t is used to calculate the corresponding minute-by-minute return x_t ($t \in T_{tra}$). The arbitrage strategy aims at capturing temporary divergences of x using a combination of economic threshold and market condition. First, the absolute minute-by-minute return has to exceed the transaction cost r ($r \in \mathbb{R}_0^+$) because a potential trade has to cover the expenses. Second, the approach accounts for the magnitude of x_t compared to the prevailing market condition, i.e., entry thresholds widen out in times of high market turmoil and vice versa. To receive a relative definition of high and low, the algorithm calculates the Bollinger bands based on the running mean level $\mu(t)$

and standard deviation $\sigma(t)$ of the returns of the past d minutes ($d \in \mathbb{N}$). The upper and lower band is obtained by adding (subtracting) k-times the time-varying standard deviation $\sigma(t)$ to (from) the historical equilibrium $\mu(t)$. Upon every entry signal, the framework buys 1 USD worth of the undervalued stock and shorts 1 USD worth of the overvalued stock. In line with Avellaneda and Lee (2010), market exposure is hedged trade-by-trade with appropriated capital expenditures in the S&P 500 index. Therefore, the constructed dollar-neutral portfolio represents a classical long-short investment strategy in the sense of Gatev et al. (2006).

From a technical point of view, the algorithm employs the following trading entry signals:

- $x_t > r$ and $x_t > \mu(t) + k \cdot \sigma(t)$, i.e., y is undervalued. Consequently, the trading strategy goes long in the stock of y and goes short in the S&P 500 index.
- $x_t < -r$ and $x_t < \mu(t) - k \cdot \sigma(t)$, i.e., y is overvalued. Consequently, the trading strategy goes short in the stock of y and goes long in the S&P 500 index.
- Otherwise, it is assumed that the stock of y will not show any meaningful mispricings in the future. Consequently, the trading strategy does not execute any trade.

Further entry signals are disregarded until the position is closed, so that at most one active position per pair is simultaneously permitted. The trade is closed if the trade return of the following stock exceeds the economic threshold – the time frame for this execution is a 99.5 percent confidence interval around the specified lag l. Also, active trades are closed when the trading periods ends or if one of the stocks of the respective pair is delisted from the S&P 500.

Following Miao (2014) and Stübinger and Endres (2018), a portfolio consists of the top 10 pairs ($s = 10$). The approach sets $d = 20$ to be in line with Bollinger (1992) and Bollinger (2001). Consistent with the high-frequency framework of Stübinger and Bredthauer (2017), the model chooses $k = 2.5$ in order to avoid high transaction costs due to excessive trading.

The trading framework follows Prager et al. (2012) and assumes 4 basis points per share per round-trip. This assumption is deemed feasible given our high turnover strategy in a highly liquid investment universe based on minute-by-minute data.

In accordance with Gatev et al. (2006), returns of the strategy portfolio are calculated by means of committed capital and actual employed capital. While the former divides the sum of net profits by the number of pairs that are selected for the trading period, the latter scales the portfolio payoffs by the number of pairs that are actually active during the trading period.

To assess the value-add of the trading strategy based on optimal causal paths (OCP), it is benchmarked with statistical arbitrage trading variants based on (1) correlation (COR), (2) Manhattan distance (MAN), (3) lagged cross-correlation (LCC), and (4) an S&P 500 buy-and-hold strategy (MKT) – all well-established quantitative strategies. Data and general framework are identical to OCP. The cornerstones of these classic strategies are briefly discussed below.

Correlation (COR). Following Chen et al. (2012), the co-movement of stock pairs is measured by Pearson's ρ (see Pearson (1895)). The top 10 pairs with the highest correlation coefficient are transferred to the trading period. Positions are put on at static upper and lower bands which are defined by the 2.5-standard deviation from historical mean. Trades are reversed, when the spread crosses the historical mean.

Manhattan distance (MAN). The second benchmark resembles COR and is motivated by the distance approach of Gatev et al. (2006). To ensure consistency, the selection criterion bears on the Manhattan distance, i.e., top pairs are determined exhibiting the smallest sum of absolute differences of their normalized prices during the formation period. Again, positions are opened at a 2.5-standard deviation trigger and reverted at the next crossing of the prices.

Lagged cross-correlation (LCC). In the spirit of Kim and Baginski (2016), the co-movement is quantified using lagged cross-correlation which represents a set of correlation coefficients for diverse time lags. The algorithm selects top pairs based on the highest lagged cross-correlation – the respective value provides the estimated lag between the given time series. The trading algorithm is identical to OCP. Summarizing, LCC is a reduced version of OCP since correlation does not necessarily imply causality (see Alexander (2001)).

S&P 500 buy-and-hold strategy (MKT). Last but not least, OCP is benchmarked to a naive S&P 500 buy-and-hold investment. The index is bought in January 1998 and held during the sample period. This passive strategy runs without any trading signals.

5. Results

Following Stübinger and Endres (2018)'s approach, this paper conducts a holistic performance analysis for the top 10 pairs of OCP from January 1998 to December 2015 compared to the benchmarks COR, MAN, LCC, and MKT. Specifically, the risk-return characteristics as well as trading statistics for each strategy are evaluated (subsection 5.1). In the following subsections, we focus on OCP and check its profitability in the context of cryptocurrency (subsection 5.2), investigate the exposure to common systematic sources of risk (subsection 5.3), and perform several robustness checks (subsection 5.4). Finally, the lead-lag structure and the portfolio composition are analyzed (subsection 5.5).

5.1. Strategy performance

Table 1 reports daily risk-return characteristics based on employed capital before and after transaction costs for the top 10 pairs per strategy from January 1998 to December 2015. Across all strategies, we observe positive returns after transaction costs ranging between 8 basis points per day for COR and 18 basis points per day for OCP compared to 2 basis points for the general market. From a statistical point of view, the returns after transaction costs are also significant with Newey-West (NW) t-statistics of at least 6.46. The S&P 500 long-only benchmark leads to a standard deviation of 1.26 percent, approximately 50 percent higher than the corresponding key figure of COR, MAN, LCC, and OCP. In stark contrast to the general market, all variants exhibit positive skewness which displays a desirable property for any potential investor (Cont, 2001). Kurtosis well above 3 suggests leptokurtic distribution – the extreme high value for OCP (631.89) is predominantly driven by one outlier. In line with Miao (2014), historical Value at Risk (VaR) measures are reported. Tail risk of all strategy variants is at a very low level by contrast with the S&P 500, e.g., the historical VaR 1% is -1.23 percent for OCP versus -3.50 percent for MKT. The strategy OCP produces the highest hit ratio, i.e., the percentage of days with non-negative returns, with 57.37 percent after transaction costs. Concluding, OCP achieves favorable return characteristics and risk metrics – this statement remains valid after transaction costs.

	Before transaction costs				After transaction costs				
	COR	MAN	LCC	OCP	COR	MAN	LCC	OCP	MKT
Mean return	0.0027	0.0028	0.0038	0.0039	0.0008	0.0009	0.0012	0.0018	0.0002
Standard error (NW)	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002
t-Statistic (NW)	22.4292	32.6506	26.4764	26.7328	6.4584	11.4329	8.3270	12.2256	0.9487
Minimum	-0.0555	-0.0402	-0.0717	-0.0844	-0.0563	-0.0413	-0.0742	-0.0855	-0.0947
Quartile 1	-0.0009	0.0000	-0.0001	0.0002	-0.0026	-0.0016	-0.0027	-0.0017	-0.0056
Median	0.0024	0.0025	0.0028	0.0027	0.0005	0.0006	0.0002	0.0006	0.0006
Quartile 3	0.0060	0.0051	0.0064	0.0063	0.0039	0.0030	0.0038	0.0040	0.0061
Maximum	0.0775	0.1277	0.1342	0.3786	0.0755	0.1242	0.1312	0.3767	0.1096
Standard deviation	0.0079	0.0055	0.0089	0.0092	0.0076	0.0052	0.0088	0.0092	0.0126
Skewness	0.5814	3.2646	2.3371	15.8288	0.6226	3.5380	2.3077	16.1916	-0.2020
Kurtosis	8.0890	64.8701	29.5211	611.3627	8.7552	73.9596	29.4934	631.8882	7.5312
Historical VaR 1%	-0.0191	-0.0100	-0.0165	-0.0100	-0.0204	-0.0114	-0.0191	-0.0123	-0.0350
Historical CVaR 1%	-0.0265	-0.0146	-0.0282	-0.0184	-0.0277	-0.0160	-0.0308	-0.0206	-0.0503
Historical VaR 5%	-0.0082	-0.0041	-0.0059	-0.0038	-0.0098	-0.0057	-0.0086	-0.0060	-0.0196
Historical CVaR 5%	-0.0150	-0.0078	-0.0129	-0.0084	-0.0164	-0.0093	-0.0155	-0.0107	-0.0302
Maximum drawdown	0.1200	0.0445	0.1065	0.0900	0.3355	0.2265	0.6291	0.6596	0.6433
Share with return ≥ 0	0.6967	0.7541	0.7453	0.7782	0.5405	0.5695	0.5147	0.5737	0.5317

Table 1: Daily return characteristics and risk metrics for the top 10 pairs of COR, MAN, LCC, and OCP compared to an S&P 500 long-only benchmark (MKT) from January 1998 until December 2015. NW denotes Newey-West standard errors with 1-lag correction and CVaR the Conditional Value at Risk.

Table 2 depicts summary statistics about the trading frequency of COR, MAN, LCC, and OCP. Across all strategies, the number of pairs traded per 1-day period exceeds 7.86, a value well in line with Gatev et al. (2006) as well as with Stübinger and Bredthauer (2017). The average number of round-trip trades per pair is vastly different for COR (1.93) and MAN (2.30) compared to LCC (6.67) and OCP (5.32). This dissimilarity is potentially driven by the different trading strategies based on static bands (COR, MAN) and variable bands (LCC, OCP). This picture barely changes considering the trade duration – the average time pairs are open is approximately 0.3 days for the static variants and around 0.05 days for the dynamic approaches.

	COR	MAN	LCC	OCP
Average number of pairs traded per 1-day period	7.8615	9.8471	9.8184	9.4489
Average number of round-trip trades per pair	1.9281	2.2953	6.6708	5.3198
Standard deviation of number of round-trip trades per pair	3.4057	1.9657	2.9933	3.7426
Average time pairs are open in days	0.2769	0.3441	0.0243	0.0567
Standard deviation of time open, per pair, in days	0.3681	0.3702	0.0528	0.0987

Table 2: Trading statistics for the top 10 pairs of COR, MAN, LCC, and OCP per 1-day trading period.

Table 3 portrays annualized risk-return measures for all strategies. After transaction costs, OCP achieves 54.98 percent – classic trading strategies and a naive buy-and-hold strategy are clearly outperformed. As expected, COR, MAN, LCC, and OCP achieve substantial lower standard deviations than the general market resulting in Sharpe ratios between 1.50 for COR and 3.57 for OCP. Notably, only considering the downside risk reinforces this tendency: Sortino ratio, i.e., returns are scaled by their downside deviation, is at 10.05 for OCP compared to 6.05 for MAN, 4.50 for LCC, 2.73 for COR, and 0.15 for MKT. The results based on committed and employed capital are at a similar level – this fact is not surprising since the top pairs open in the vast majority of all cases (see table 2).

	Before transaction costs				After transaction costs				
	COR	MAN	LCC	OCP	COR	MAN	LCC	OCP	MKT
Mean return	0.9811	1.0274	1.5910	1.6617	0.2066	0.2647	0.3372	0.5498	0.0219
Mean excess return	0.9412	0.9865	1.5388	1.6082	0.1823	0.2392	0.3102	0.5186	0.0012
Standard deviation	0.1253	0.0871	0.1412	0.1465	0.1214	0.0831	0.1405	0.1454	0.2001
Downside deviation	0.0658	0.0346	0.0603	0.0439	0.0756	0.0438	0.0750	0.0547	0.1438
Sharpe ratio	7.5117	11.3265	10.9002	10.9758	1.5019	2.8772	2.2085	3.5671	0.0060
Sortino ratio	14.9202	29.6858	26.3883	37.8238	2.7318	6.0456	4.4960	10.0504	0.1520
Committed capital									
Mean return	0.6997	1.0022	1.5205	1.5498	0.1500	0.2579	0.3151	0.5193	0.0219
Sharpe ratio	6.7428	11.1915	10.6985	10.5628	1.3277	2.8349	2.1095	3.4746	0.0060

Table 3: Annualized risk-return measures for the top 10 pairs of COR, MAN, LCC, and OCP compared to an S&P 500 long-only benchmark (MKT) from January 1998 until December 2015.

Following Do and Faff (2010) and Bowen and Hutchinson (2016), a sub-period analysis is performed in order to analyze the performance of the strategies over time. For this purpose, figure 4 describes the development of an investment of 1 USD after transaction costs (first row) compared to the general market (second row).

The first sub-period ranges from January 1998 to June 2003 and defines the growth and collapse of the dot-com bubble. In stark contrast to the S&P 500, the trading strategies show a steady growth up, even in times of high market turmoil. Thus, it is not surprising that annualized returns of OCP exceed 120 percent at a Sharpe ratio of 8.80 after transaction costs. The second sub-period ranges from July 2003 to December 2008 and describes the time of moderation and the global financial crisis. We observe that the strategies are not affected by changing market regimes due to the long-short portfolios we are constructing – a favorable effect for investors. After transaction costs, OCP produces annualized returns of 59.70 percent compared to 24.43 percent for COR, 16.42 percent for MAN, and 4.53 percent for LCC. The third sub-period ranges from January 2009 to December 2015 and characterizes the period of regeneration and comebacks. Annualized returns vary between -2.53 percent for COR and 14.66 percent for OCP compared to 10.58 percent for the general market. All strategies, however, depict declining performance results since January 2012 – this fact is confirmed by the majority of academic research, e.g., Clegg and Krauss (2017) and Stübinger and Endres (2018). Summarizing, the trading strategy OCP outperforms classic approaches in a multitude of comparisons – complexity pays off. Therefore, detailed evaluations of OCP are conducted in the following subsections.

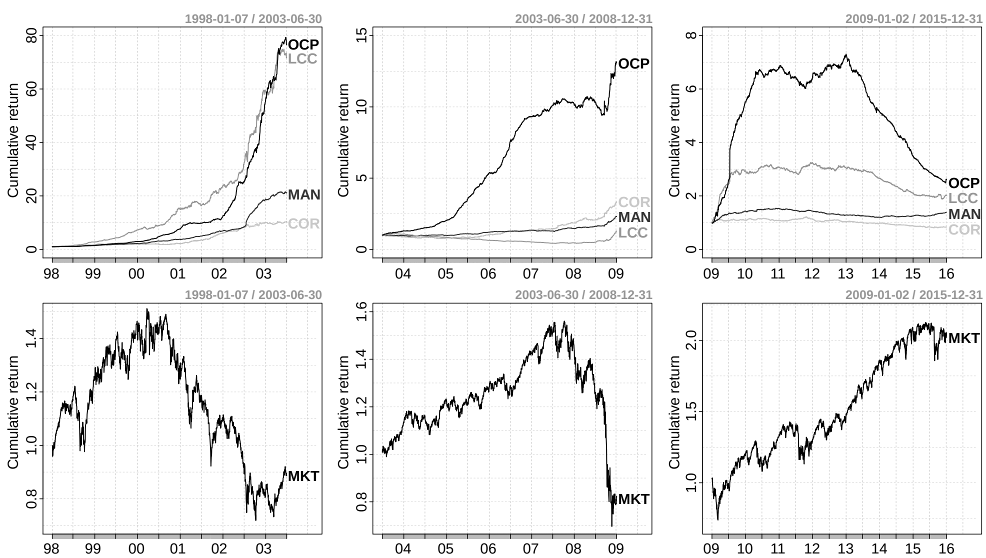


Figure 4: Development of an investment of 1 USD after transaction costs for the top 10 pairs of COR, MAN, LCC, and OCP in the first row compared to the S&P 500 index (MKT) in the second row. The time period from 1998 until 2015 is divided into three sub-periods (1998-01/2003-06, 2003-07/2008-12, 2009-01/2015-12).

5.2. Investment strategy based on bitcoins

This subsection demonstrates the profitability of OCP even in recent times by applying the outlined strategy in the context of cryptocurrencies. Following Narayanan et al. (2016) and Chohan (2017), a cryptocurrency represents an instrument of exchange that uses cryptography to control the transactional flow and the creation of additional units.

Key representative of cryptocurrencies are bitcoins which are introduced by a person or group under the pseudonym of Satoshi Nakamoto in 2008. Nakamoto (2008) develops a solution to the double-spending problem applying a peer-to-peer network to ensure the chronological order of transactions. The development of the bitcoin price (see figure 5) and the seminal paper by Nakamoto (2008) characterize the trigger for an ever-expanding interest in this field up to the present. Until today, this study has been cited over 2200 times, with more than 700 additional citations in 2017 on Google Scholar. Baek and Elbeck (2015), Kristoufek (2015), and Bouoiyour et al. (2016) investigate the most frequently claimed drivers of bitcoin prices, e.g., standard fundamental factors, political risk, and regulatory moves.

In the following, we apply the alternative investment strategy OCP_{BIT}, i.e., the trading algorithm in section 4 is extended by the condition that the bitcoin price characterizes the second stock of each pair.

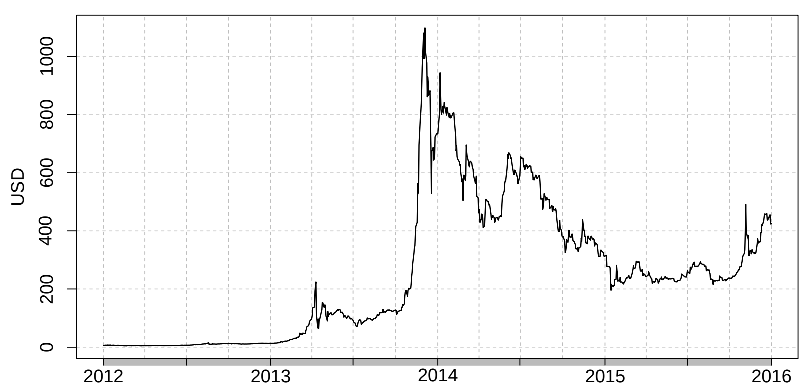


Figure 5: Bitcoin price from January 2012 to December 2015.

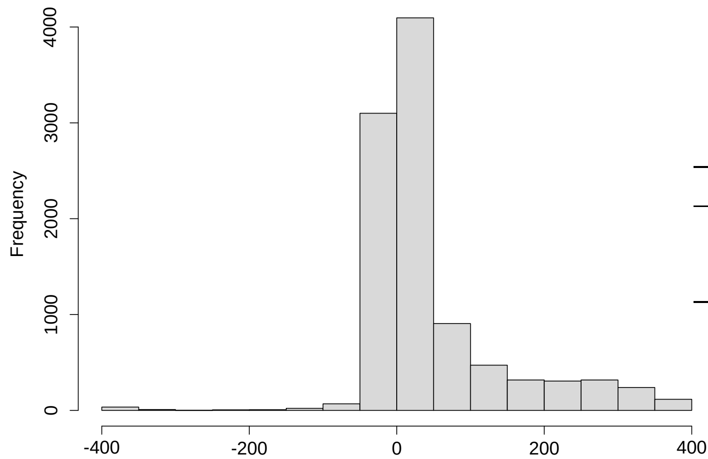
Table 4 exhibits annualized risk-return measures for the top 10 pairs of OCP_{BIT} from January 2012 until December 2015 compared to OCP, the bitcoin price (BIT), and the S&P 500 index (MKT). The top 10 pairs of OCP_{BIT} strongly outperform with annualized returns after transaction costs of 170.37 percent compared to -20.07 percent for OCP, 60.88 percent for

BIT, and 12.01 percent for MKT. Across all strategies, the mean returns almost equal the mean excess returns due to the fact that the risk free rate is close to zero during the considered sample period. Interestingly, standard deviation of BIT is 4-times to 20-times higher than OCP, OCP_{BIT}, and MKT – a desirable property since a stock market may be efficient during normal times (Kim et al., 2011). The Sharpe ratio is above 6 in case of OCP_{BIT} – the excess return clearly overcompensates the risk.

	Before transaction costs		After transaction costs			
	OCP	OCP_{BIT}	OCP	OCP_{BIT}	BIT	MKT
Mean return	0.4395	3.6493	-0.2007	1.7037	0.6088	0.1201
Mean excess return	0.4395	3.6490	-0.2007	1.7035	0.6087	0.1201
Standard deviation	0.0647	0.2635	0.0642	0.2679	1.0340	0.1280
Downside deviation	0.0276	0.0717	0.0454	0.0798	0.7538	0.0891
Sharpe ratio	6.7970	13.8475	-3.1271	6.3580	0.5887	0.9384
Sortino ratio	15.9312	50.9014	-4.4251	21.3423	0.8077	1.3480

Table 4: Annualized risk-return measures for the top 10 pairs of OCP and OCP_{BIT}, BIT, and MKT from January 2012 until December 2015.

In view of the clear outperformance of OCP_{BIT}, we analyze the portfolio composition on a more granular level. Figure 6 presents the histogram and descriptive statistics of the specified lags for the top 10 pairs of OCP_{BIT} from January 2012 to December 2015. A positive lag indicates that the partner stock leads the "bitcoin stock" and vice versa. First of all, we observe a clear asymmetry of the histogram – the vast majority of pairs shows a positive lag suggesting that the "bitcoin stock" follows the selected partner stock. This statement is confirmed by the descriptive statistics – on average the partner stock leads the "bitcoin stock" by 46.83 minutes. The corresponding median amounts 11.00 minutes. This finding indicates that the selected stocks contain remarkable information about the prospective bitcoin returns. In contrast to OCP, the strategy OCP_{BIT} is in a position to make capital out of this fact. Summarizing, OCP_{BIT} poses a severe challenge to the semi-strong form of market efficiency even in recent times.



	Lag
Mean return	46.83
Standard deviation	95.06
Median	11.00

Figure 6: Histogram of specified lags for the top 10 pairs of OCP_{BIT} from January 2012 to December 2015. A positive (negative) lag indicates that the "bitcoin stock" follows (leads) the corresponding partner stock.

5.3. Common risk factors

Table 5 evaluates the exposure of OCP after transaction costs to systematic sources of risk. Following Knoll et al. (2017), three types of regression are employed. The Fama-French 3-factor model (FF3) by Fama and French (1996) captures systematic risk exposure to general market, small minus big capitalization stocks (SMB), as well as high minus low book-to-market stocks (HML). The Fama-French 3+2-factor model (FF3+2), as outlined in Gatev et al. (2006), extends the first model by a momentum factor and a short-term reversal factor. The Fama-French 5-factor model (FF5) by Fama and French (2015) appends two additional factors to FF3, namely portfolios of stocks with a robust minus weak profitability (RMW5) and with a conservative minus aggressive (CMA5) investment behavior. All data related to these models are downloaded from Kenneth R. French's website.[2]

Irrespective of the regression model applied, daily returns after transaction costs exhibit significant alphas of 0.17 percent – slightly higher than the raw returns. As expected, FF3 and FF3+2 show no loading on the market – FF5 indicate a marginal but statistical significant positive effect. Loadings on SMB, HML, Momentum, Reversal, SMB5, HML5, RMW5, and CMA5 are statistically not significant and close to zero – this fact is not surprising since the strategy constructs dollar-neutral portfolios. Concluding, OCP produces

[2]Thanks to Kenneth R. French for providing all relevant data for these models on his website.

statistically significant and economically remarkable returns after transaction costs, outperforms classic arbitrage trading strategies and indicates no loading on any common sources of systematic risk.

	FF3	FF3+2	FF5
(Intercept)	0.0017***	0.0017***	0.0017***
	(0.0001)	(0.0001)	(0.0001)
Market	0.0158	0.0143	0.0267*
	(0.0108)	(0.0119)	(0.0125)
SMB	0.0154	0.0168	
	(0.0217)	(0.0218)	
HML	−0.0150	−0.0204	
	(0.0204)	(0.0219)	
Momentum		−0.0101	
		(0.0153)	
Reversal		−0.0030	
		(0.0154)	
SMB5			0.0221
			(0.0234)
HML5			−0.0335
			(0.0232)
RMW5			0.0354
			(0.0303)
CMA5			0.0368
			(0.0371)
R^2	0.0008	0.0009	0.0014
Adj. R^2	0.0001	-0.0002	0.0003
Num. obs.	4527	4527	4527
RMSE	0.0092	0.0092	0.0092

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: Exposure to systematic sources of risk after transaction costs for the daily returns of the top 10 pairs of OCP from January 1998 until December 2015. Standard errors are depicted in parentheses.

5.4. Robustness checks

Whenever strategies generate remarkable returns it arouses the suspicion of data snooping. Therefore, a series of robustness checks is conducted to demonstrate the value-add of the strategy outlined in section 4.

First, the performance of OCP is contrasted with 2,500 random bootstraps of monkey trading. To be more specific, top pairs are randomly selected. As expected, the average daily returns after transaction costs amount -0.0010 compared to 0.0018 for OCP. This finding is well in line with Gatev et al. (2006) and Stübinger et al. (2016).

Second, the robustness of OCP is evaluated in light of market frictions. Therefore, a one-minute-waiting rule is applied to deal with bid-ask bounces. After transaction costs, the delayed execution of OCP achieves annualized returns of 12.23 percent from 1998 to 2011 and -34.04 percent from 2012 to 2015. The strategy OCP_{BIT} with a one-minute-waiting rule produces returns of 10.08 percent p.a. during the second time span after transaction costs.

Third, the input parameters are motivated by the literature – the trading threshold is set to 2.5 standard deviation ($k = 2.5$), the length of the moving average to 20 minutes ($d = 20$), and the number of top pairs to 10 ($s = 10$). In table 6, the parameters k, d, and s are varied in two directions and annualized mean return as well as Sharpe ratio are reported. After transaction costs, the input parameter $k = 2.5$ generates the most promising risk-return relation. Higher values can generally be found at higher levels of d – this result is well in line with Stübinger et al. (2016). Sharpe ratio increases for a larger number of top pairs because portfolio standard deviation declines. Concluding, the initial setting of k, d, and s hits not the optimum in light of annualized return and Sharpe ratio but trading results remain meaningful irrespective of the parameter constellation.

		Return			Sharpe ratio		
	k \ d	10	20	60	10	20	60
Top 5	2	0.4905	0.4985	0.5820	2.5216	2.9294	3.5444
	2.5	0.3792	0.4975	0.5483	2.9741	3.5028	3.7178
	3	0.0774	0.3743	0.4447	0.8405	3.0795	3.6153
Top 10	2	0.5210	0.5768	0.6251	3.0166	3.2021	3.3834
	2.5	0.4192	0.5498	0.6149	2.9233	3.5671	3.7968
	3	0.1026	0.4287	0.5334	1.2946	2.4034	3.7975
Top 20	2	0.5239	0.5825	0.6454	3.1983	3.7908	3.9150
	2.5	0.4042	0.5472	0.6258	3.7457	4.0901	4.2927
	3	0.1043	0.4155	0.5314	1.4128	3.3895	4.6025

Table 6: Yearly returns and Sharpe ratios after transaction costs for the k-times of the standard deviation of OCP, the number of days to use in the moving window (d), and a varying number of target stocks (s) from January 1998 until December 2015.

5.5. Analysis of lead-lag structure and portfolio constituents

Figure 7 reports the absolute lag and correlation for the top 10 pairs over time. Overall, we observe antidromic developments of determined lag and correlation, i.e., if one variable increases, the other decreases and vice versa. To be more specific, the specified lag is approximately 120 minutes from 1998 until 2001. Since American financial markets are decimalized from September 2000 to April 2001, the lag decreases to approximately 20 minutes at the end of 2002 – an outlier is observed at the beginning of 2002. The correlation exhibits a positive trend with some temporarily downside fluctuations in 2002 and 2011.

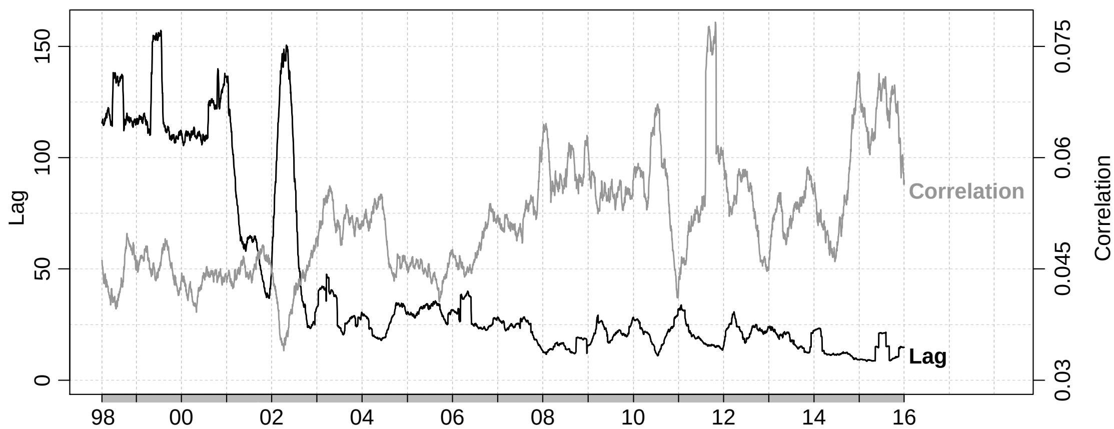


Figure 7: Average lag (left axis) and average correlation (right axis) for the top 10 pairs of OCP in 60-day moving windows from January 1998 until December 2015.

Last but not least, figure 8 portrays the portfolio constituency for the top 10 pairs of OCP over time (daily data is clustered quarterly). According to the Global Industry Classification Standard, all companies of the top pairs are categorized into the following 10 economic sectors (valuation date: 2015/12/31): Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunications Services, and Utilities. Notably, the strategy possesses an anti-cyclical constituent portfolio, i.e., sectors are avoided in times of bull markets and vice versa. As such, stocks from the IT sector are completely taken out of the portfolio during the dot-com bubble at the turn of the millennium. In contrast, the portfolio consists of a large number of technology companies in the years after the crash – top value of approximately 50 percent is achieved in 2006. On the same note, the percentage of financial stocks is close to zero in the years 2006 and 2007, the height of subprime lending and fraudulent underwriting practices. In times of the global financial crisis and its aftermath, the share rises up to 25 percent during the phase of high market turmoil.

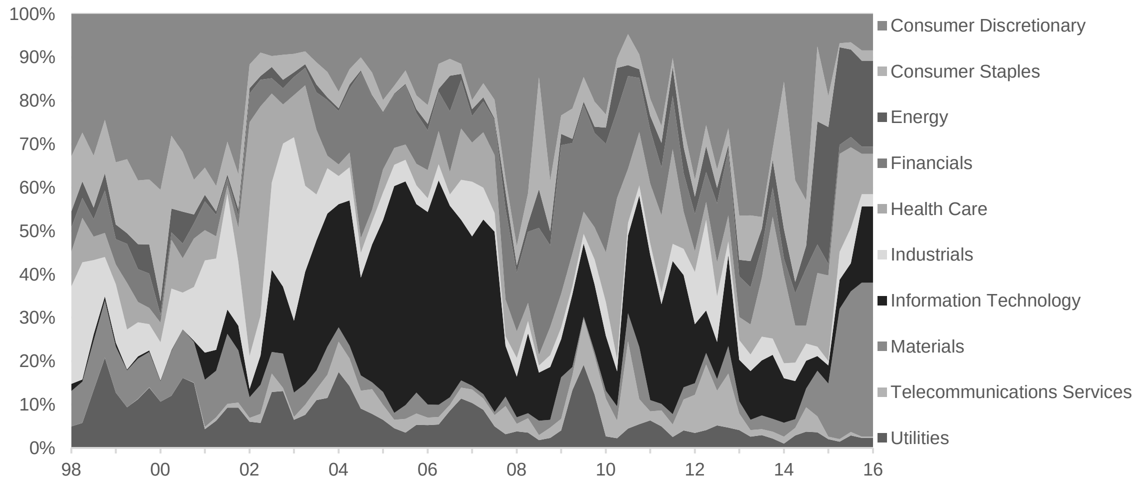


Figure 8: Constituent portfolio for the top 10 pairs of OCP from January 1998 until December 2015.

6. Conclusion

This paper presents an integrated statistical arbitrage trading framework relying on the novel introduced optimal causal path algorithm and deploys it on minute-by-minute data of the S&P 500 constituents from January 1998 to December 2015. In this respect, the manuscript makes three main contributions to the existing literature.

The first contribution refers to the developed optimal causal path algorithm and its use for identifying promising stock pairs and for generating buy and sell signals. Essentially, the flexible algorithm efficiently identifies the optimal non-linear mapping given two time series and estimates its corresponding lead-lag structure. Therefore, the established trading strategy is in a position to predict the future returns of the following stock by exploiting information about the leading stock.

The second contribution focuses on the performance of the proposed strategy and its value-add compared to well established frameworks in this area of research. In the empirical back-testing study, the trading algorithm achieves statistically and economically significant returns of 54.98 percent p.a. after transaction costs – Fama-French models do not indicate any loading on common sources of systematic risk. Results are well superior to the benchmark approaches ranging between 2.19 percent for a naive buy-and-hold strategy of the S&P 500 index to 33.72 percent for the variant based on lagged cross-correlation. A series of robustness checks confirms the necessity of regarding a model that permits an elastic adjustment of the time axis.

The third contribution bears on the fact that the strategy outperformances in the context of cryptocurrencies even in the sample period from 2012 to 2015. Interestingly, a more granular analysis shows that stock returns contain substantial information about the future bitcoin returns. This finding poses a severe challenge to the semi-strong form of market efficiency.

For further research in this field, hidden Markov models may be explored in order to receive probability distributions. Second, a multivariate algorithm and arbitrage framework that accounts for common interactions could be implemented. Finally, the presented methodology might be a promising tool for efficiently coping with time deformations in other areas of application, such as human action recognition or robot programming.

Bibliography

Al-Naymat, G., Chawla, S., Taheri, J., 2009. SparseDTW: A novel approach to speed up dynamic time warping. In: Kennedy, P. J., Ong, K., Christen, P. (Eds.), Proceedings of the Eighth Australasian Data Mining Conference. Australian Computer Society, Melbourne, Australia, pp. 117–127.

Alexander, C., 2001. Market models: A guide to financial data analysis. John Wiley & Sons, Chichester, UK and New York, NY, USA.

Arici, T., Celebi, S., Aydin, A. S., Temiz, T. T., 2014. Robust gesture recognition using feature pre-processing and weighted dynamic time warping. Multimedia Tools and Applications 72 (3), 3045–3062.

Avellaneda, M., Lee, J.-H., 2010. Statistical arbitrage in the US equities market. Quantitative Finance 10 (7), 761–782.

Baek, C., Elbeck, M., 2015. Bitcoins as an investment or speculative vehicle? A first look. Applied Economics Letters 22 (1), 30–34.

Berndt, D. J., Clifford, J., 1994. Using dynamic time warping to finder patterns in time series. In: Fayyad, U. M., Uthurusamy, R. (Eds.), Knowledge Discovery in Databases: Papers from the 1994 AAAI Workshop,. AAAI Press, Menlo Park, CA, USA, pp. 359–370.

Bollinger, J., 1992. Using Bollinger bands. Stocks & Commodities 10 (2), 47–51.

Bollinger, J., 2001. Bollinger on Bollinger bands. McGraw-Hill, New York, NY, USA.

Bouoiyour, J., Selmi, R., Tiwari, A. K., Olayeni, O. R., 2016. What drives bitcoin price. Economics Bulletin 36 (2), 843–850.

Bowen, D. A., Hutchinson, M. C., 2016. Pairs trading in the UK equity market: Risk and return. The European Journal of Finance 22 (14), 1363–1387.

Chen, H., Chen, S. J., Li, F., 2012. Empirical investigation of an equity pairs trading strategy. Working paper, Columbia University.

Cheng, H., Dai, Z., Liu, Z., Zhao, Y., 2016. An image-to-class dynamic time warping approach for both 3D static and trajectory hand gesture recognition. Pattern recognition 55, 137–147.

Chinthalapati, V. L., 2012. High frequency statistical arbitrage via the optimal thermal causal path. Working paper, University of Greenwich.

Chohan, U. W., 2017. Cryptocurrencies: A Brief Thematic Review. Working paper, University of New South Wales.

Clegg, M., Krauss, C., 2017. Pairs trading with partial cointegration. Quantitative Finance 18 (1), 121–138.

Coelho, M. S., 2012. Patterns in financial markets: Dynamic time warping. Working paper, NOVA School of Business and Economics.

Cont, R., 2001. Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance 1 (2), 223–236.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E. J., 2008. Querying and mining of time series data: Experimental comparison of representations and distance measures. In: Jagadish, H. V. (Ed.), Proceedings of the VLDB Endowment. ACM, New York, NY, USA, pp. 1542–1552.

Do, B., Faff, R., 2010. Does simple pairs trading still work? Financial Analysts Journal 66 (4), 83–95.

Do, B., Faff, R., 2012. Are pairs trading profits robust to trading costs? Journal of Financial Research 35 (2), 261–287.

Dupas, R., Tavenard, R., Fovet, O., Gilliet, N., Grimaldi, C., Gascuel-Odoux, C., 2015. Identifying seasonal patterns of phosphorus storm dynamics with dynamic time warping. Water Resources Research 51 (11), 8868–8882.

Fama, E. F., French, K. R., 1996. Multifactor explanations of asset pricing anomalies. The Journal of Finance 51 (1), 55–84.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116 (1), 1–22.

Fu, C., Zhang, P., Jiang, J., Yang, K., Lv, Z., 2017. A Bayesian approach for sleep and wake classification based on dynamic time warping method. Multimedia Tools and Applications 76 (17), 17765–17784.

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.

Huck, N., Afawubo, K., 2015. Pairs trading and selection methods: Is cointegration superior? Applied Economics 47 (6), 599–613.

Ilzetzki, E., Mendoza, E. G., Végh, C. A., 2013. How big (small?) are fiscal multipliers? Journal of Monetary Economics 60 (2), 239–254.

Itakura, F., 1975. Minimum prediction residual principle applied to speech recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing 23 (1), 67–72.

Jiao, L., Wang, X., Bing, S., Wang, L., Li, H., 2014. The application of dynamic time warping to the quality evaluation of Radix Puerariae thomsonii: Correcting retention time shift in the chromatographic fingerprints. Journal of Chromatographic Science 53 (6), 968–973.

Juang, B.-H., 1984. On the hidden Markov model and dynamic time warping for speech recognition – A unified view. Bell Labs Technical Journal 63 (7), 1213–1243.

Keogh, E. J., Pazzani, M. J., 2000. Scaling up dynamic time warping for datamining applications. In: Ramakrishnan, R., Stolfo, S., Bayardo, R., Parsa, I. (Eds.), Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 285–289.

Keogh, E. J., Ratanamahatana, C. A., 2005. Exact indexing of dynamic time warping. Knowledge and Information Systems 7 (3), 358–386.

Kim, J. H., Shamsuddin, A., Lim, K.-P., 2011. Stock return predictability and the adaptive markets hypothesis: Evidence from century-long US data. Journal of Empirical Finance 18 (5), 868–879.

Kim, S., Baginski, M. E., 2016. A cross correlation-based stock forecasting model. Working paper, Auburn University Journal.

Kim, S., Heo, J., 2017. Time series regression-based pairs trading in the Korean equities market. Journal of Experimental & Theoretical Artificial Intelligence 29 (4), 755–768.

Knoll, J., Stübinger, J., Grottke, M., 2017. Exploiting social media with higher-order factorization machines: Statistical arbitrage on high-frequency data of the S&P 500. FAU Discussion Papers in Economics (13), University of Erlangen-Nürnberg.

Kristoufek, L., 2015. What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. PLOS One 10 (4), 1–15.

Létourneau, P., Stentoft, L., 2014. Refining the least squares Monte Carlo method by imposing structure. Quantitative Finance 14 (3), 495–507.

Li, Q., Clifford, G. D., 2012. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. Physiological measurement 33 (9), 1491–1502.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics, 447–470.

Meng, H., Xu, H.-C., Zhou, W.-X., Sornette, D., 2017. Symmetric thermal optimal path and time-dependent lead-lag relationship: Novel statistical tests and application to UK and US real-estate and monetary policies. Quantitative Finance 17 (6), 959–977.

Miao, G. J., 2014. High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach. International Journal of Economics and Finance 6 (3), 96–110.

Muda, L., Begam, M., Elamvazuthi, I., 2010. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. Journal of Computing 2 (3), 138–143.

Müller, M., 2007. Information retrieval for music and motion. Springer, Berlin, Germany and Heidelberg, Germany.

Müller, M., Mattes, H., Kurth, F., 2006. An efficient multiscale approach to audio synchronization. In: Tzanetakis, G., Hoos, H. (Eds.), Proceedings of the 7th International Conference on Music Information Retrieval. University of Victoria, Victoria, Canada, pp. 192–197.

Myers, C., Rabiner, L., 1981. A level building dynamic time warping algorithm for connected word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing 29 (2), 284–297.

Myers, C., Rabiner, L., Rosenberg, A., 1980. Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing 28 (6), 623–635.

Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org.

Narayanan, A., Bonneau, J., Felten, E., Miller, A., Goldfeder, S., 2016. Bitcoin and cryptocurrency technologies: A comprehensive introduction. Princeton University Press, Princeton, NJ, USA.

Pearson, K., 1895. Note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London 58, 240–242.

Prager, R., Vedbrat, S., Vogel, C., Watt, E. C., 2012. Got liquidity? BlackRock Investment Institute.

Prätzlich, T., Driedger, J., Müller, M., 2016. Memory-restricted multiscale dynamic time warping. In: Ding, Z., Luo, Z.-Q., Zhang, W. (Eds.), Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Danvers, MA, USA, pp. 569–573.

QuantQuote, 2016. QuantQuote market data and software. https://quantquote.com.

R Core Team, 2017. stats: A language and environment for statistical computing. R package.

Rabiner, L., Juang, B. H., 1993. Fundamentals of speech recognition. PTR Prentice Hall, Upper Saddle River, NJ, USA.

Rad, H., Low, R. K. Y., Faff, R., 2016. The profitability of pairs trading strategies: Distance, cointegration and copula methods. Quantitative Finance 16 (10), 1541–1558.

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., Keogh, E. J., 2012. Searching and mining trillions of time series subsequences under dynamic time warping. In: Yang, Q. (Ed.), Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 262–270.

Rath, T. M., Manmatha, R., 2003. Word image matching using dynamic time warping. In: Dyer Chuck, Perona, P. (Eds.), Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, Danvers, MA, USA, pp. 521–527.

Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing 26 (1), 43–49.

Salvador, S., Chan, P., 2007. Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis 11 (5), 561–580.

Senin, P., 2008. Dynamic time warping algorithm review. Working paper, University of Hawaii at Manoa.

Silva, D., Batista, G., 2016. Speeding up all-pairwise dynamic time warping matrix calculation. In: Venkatasubramanian, S. C., Wagner, M. (Eds.), Proceedings of the Sixteenth SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp. 837–845.

Sornette, D., Zhou, W.-X., 2005. Non-parametric determination of real-time lag structure between two time series: The "optimal thermal causal path" method. Quantitative Finance 5 (6), 577–591.

S&P Dow Jones Indices, 2015. S&P Global – Equity S&P 500 index. https://us.spindices.com/indices/equity/sp-500.

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., 2018. Pairs trading with a mean-reverting jump-diffusion model on high-frequency data. Quantitative Finance, Forthcoming.

Stübinger, J., Mangold, B., Krauss, C., 2016. Statistical arbitrage with vine copulas. FAU Discussion Papers in Economics (11), University of Erlangen-Nürnberg.

Vidyamurthy, G., 2004. Pairs trading: Quantitative methods and analysis. John Wiley & Sons, Hoboken, NJ, USA.

Vlachos, M., Kollios, G., Gunopulos, D., 2002. Discovering similar multidimensional trajectories. In: Agrawal, R., Dittrich, K. (Eds.), Proceedings of the 18th International Conference on Data Engineering. IEEE, Danvers, MA, USA, pp. 673–684.

Wang, G.-J., Xie, C., Han, F., Sun, B., 2012. Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree. Physica A: Statistical Mechanics and its Applications 391 (16), 4136–4146.

Zhang, Y., Adl, K., Glass, J., 2012. Fast spoken query detection using lower-bound dynamic time warping on graphical processing units. In: Sakai, H., Nishitani, T. (Eds.), Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing. IEEE, Danvers, MA, USA, pp. 5173–5176.

Zhou, W.-X., Sornette, D., 2006. Non-parametric determination of real-time lag structure between two time series: The "optimal thermal causal path" method with applications to economic data. Journal of Macroeconomics 28 (1), 195–224.