Title

Antoine Merel, Elliot Jacquet-Francillon

***Abstract*—This project investigates whether lead-lag relationships between cryptocurrencies exist and can be leveraged to design a trading strategy. Using second-by-second price data for the 10 largest cryptocurrencies by market cap provided by Binance, lead-lag dynamics were computed using Optimal Thermal Causal Path (OTCP). A basic rules-based strategy was implemented to exploit significant price movements in leading assets not yet mirrored by lagging counterparts. The strategy achieved average returns of 20–50% per month, taking between 3,000 and 7,000 long or short positions monthly. However, when accounting for trading fees, net returns would have been negative. These findings suggest that lead-lag relationships do provide valuable information about market dynamics, and that it may be possible to design a more sophisticated strategy incorporating mutual information on pairwise lead-lags that would be profitable.**

I. Introduction

Cryptocurrency market is among the most dynamic and volatile in the financial world, operating all the time and exhibiting significant price fluctuations. While Bitcoin (BTC) has seen an increasing institutional adoption, most of the other cryptocurrencies remain heavily influenced by retail investors. The fast-moving nature of this market and the low financial knowledge of its investors create fertile ground for exploring innovative trading strategies.

The inherent complexity of cryptocurrency price movements suggests the existence of exploitable patterns, particularly among highly correlated assets. Identifying and leveraging such patterns requires advanced analytical techniques capable of handling large-scale, high-frequency data. One promising area of exploration lies in uncovering relationships between asset price movements, where signals from one cryptocurrency might lead or lag those of another. These lead-lag relationships, if properly identified and understood, could provide valuable insights for designing actionable trading strategies.

This project focuses on investigating such opportunities through the lens of lead-lag relationships, employing specialized tools tailored to the unique challenges of cryptocurrency data.

Lead-lag relationships, where one asset's price movement precedes or lags behind another, have long been studied in traditional financial markets. However, it is unclear whether these relationships can be used effectively in a trading strategy. In fact, while lead-lag analysis can identify relationships between assets in the past, it is not guaranteed that these relationships are predictive or provide actionable signals for trading.

In this project, we aim to address these uncertainties by testing whether lead-lag relationships are a valuable tool for cryptocurrency trading. Specifically, we analyze historical price data that has already occurred to measure the existence and strength of lead-lag relationships. Furthermore, we test whether these relationships can be used to inform a trading strategy, generating predictive signals that lead to profitable outcomes. By focusing on real historical data, our goal is to evaluate whether lead-lag relationships have practical value or are simply artifacts of past price behavior.

We analyze historical price data from the 10 largest cryptocurrencies by market capitalization. Using Optimal Thermal Causal Path (OTCP), we compute the lead-lag dynamics between asset pairs, aiming to identify situations where the price movement of one cryptocurrency leads or lags behind another. The ultimate goal is to determine whether these relationships provide meaningful information that can be used for short-term trading decisions.

We then implement a basic trading strategy designed to exploit these lead-lag dynamics. By defining clear rules for entering long or short positions based on significant price movements and lagging signals, we test whether the identified relationships can generate profitable outcomes.

II. Related Work

A. Lead-Lag Analysis in Financial Markets

Lead-lag analysis has long been a critical tool in financial markets for uncovering relationships between assets, where the price movement of one asset precedes or lags behind another. These relationships often arise due to shared market drivers, such as macroeconomic factors, correlated investor behavior, or liquidity flows, and can provide valuable insights for trading strategies. Understanding and quantifying these lead-lag dynamics is particularly important in high-frequency trading, where actionable signals must be identified and acted upon in fractions of a second.

Traditional methods for lead-lag analysis include cross-correlation and Granger causality tests. Cross-correlation measures the statistical relationship between two time series at various lags, while Granger causality tests whether one time series can predict another by examining lagged regressions. Despite their widespread use, these methods have notable limitations. Cross-correlation assumes stationarity and linearity in the relationships, which are often violated in real-world financial time series. Similarly, Granger causality requires fixed lag assumptions and parametric modeling, making it less effective in capturing non-linear or time-varying dependencies.

B. The Optimal Thermal Causal Path (OTCP) Method

The OTCP method, introduced by Didier Sornette and Wei-Xing Zhou, provides a non-parametric framework for detecting and quantifying dynamic lag structures between two time series. Unlike traditional methods such as cross-correlation or Granger causality, OTCP does not assume linearity, stationarity, or fixed lags. Instead, it allows for a time-evolving, real-time determination of lead-lag relationships, making it particularly well-suited for noisy, non-linear datasets such as financial time series.

The core of the OTCP method lies in mapping one time series onto another by identifying the "optimal path" in a lag-time plane. This path represents the temporal offsets that minimize the total cost between the two series, subject to certain constraints. The cost function used in OTCP accounts for the similarity of values between the two series as well as the lag structure, ensuring that the method adapts dynamically to changes in the lead-lag relationship over time.

A unique feature of OTCP is its incorporation of a "thermal" parameter, which introduces stochasticity into the optimization process. This thermal component allows the algorithm to probabilistically explore suboptimal paths, helping it avoid being trapped in local minima. As a result, OTCP is able to identify more robust and accurate lead-lag relationships, even in the presence of significant noise or non-linear dependencies in the data.

Another strength of OTCP is its ability to provide a lag value at every point in time, reflecting the dynamic nature of the lead-lag relationship. This contrasts with methods like Granger causality, which typically assume a fixed lag, and cross-correlation, which requires predefined lag intervals for analysis. OTCP's dynamic lag computation is critical for applications in high-frequency trading, where lead-lag relationships can evolve rapidly.

C. Intuition behind OTCP

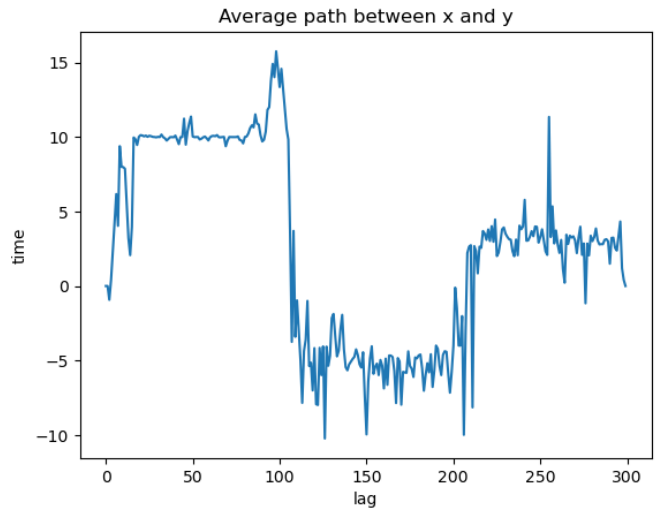
The OTCP method dynamically detects the lag between two time series, adapting to changes in their relationship over time. To build intuition, consider the following constructed example where two time series, x and y, are related by a shifting lag structure. Let x be a time series, and let's define y as:

$$y = \begin{cases} \varepsilon_y, & \text{for } t \in [0, 10) \\ x[t-10] + \varepsilon_y, & \text{for } t \in [10, 100) \\ x[t+5] + \varepsilon_y, & \text{for } t \in [100, 200) \\ x[t-3] + \varepsilon_y, & \text{for } t \in [200, 300] \end{cases}$$

, where ε_y represents noise added to y to mimic real-world imperfections.

When we apply OTCP to these two time series, we obtain the result in figure 1. As we can see, in this simple example, the method is able to identify the lag k at every point in time.

Figure 1. Lead-lag relationship between x and y

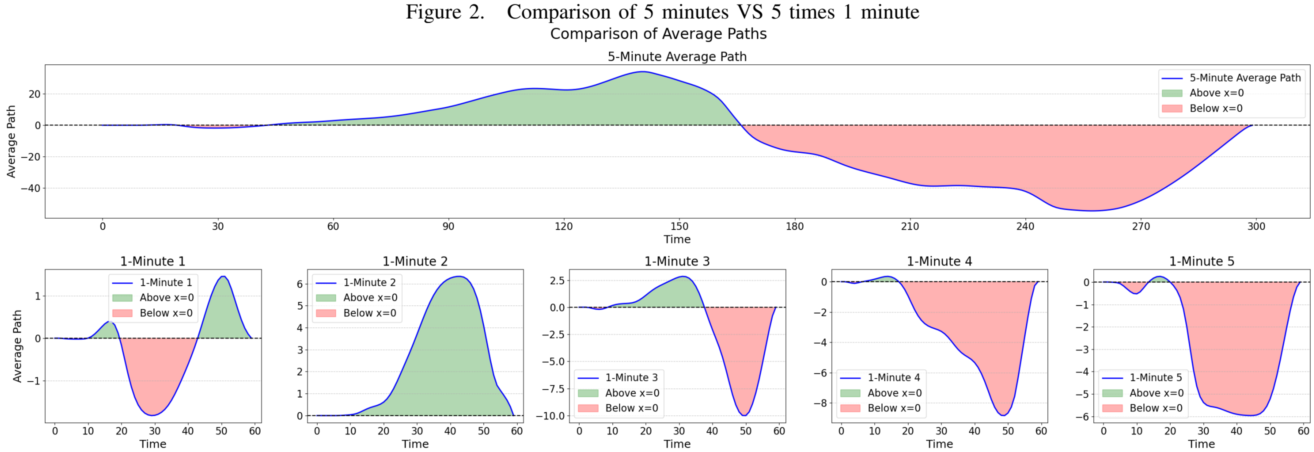


One characteristic of the algorithm is that it always starts and ends at a lag of 0 by design. This can introduce biases if the window size is too small, as the algorithm tends to "pull" toward 0 at the edges of the window. Ideally, we would compute OTCP over a window as long as possible to counter this effect and obtain more accurate lag estimates. However, since the computation time for OTCP scales quadratically with the window size, increasing the window length significantly can quickly become computationally infeasible.

To explore the impact of using smaller windows, we applied OTCP to a random 5-minute window of BTC and ETH prices, and then divided the same 5-minute period into five separate 1-minute windows. We computed the lead-lag relationship across the entire 5-minute window and compared it with the results obtained from the five 1-minute windows. We can observe the results in figure 2. As expected, the amplitude of the detected lags was smaller in the shorter 1-minute windows due to the "pull" toward 0 at the boundaries. However, the sign of the lags remained consistent between the 1-minute and 5-minute computations. This consistency suggests that while shorter windows reduce the magnitude of the detected lags, the key directional information (lead or lag) is preserved.

Given this observation, we empirically determined that a window size of 40 seconds is optimal for our analysis. This size strikes a balance between computational efficiency and the ability to detect meaningful lag relationships. It is small enough to handle the large dataset of second-by-second cryptocurrency prices efficiently, yet large enough to preserve significant lead-lag structures. Furthermore, lead-lag relationships with lags exceeding 40 seconds are rare

Figure 2. Comparison of 5 minutes VS 5 times 1 minute



in our dataset, making this window size both practical and sufficient for the purpose of our study.

III. Data Processing

A. Data Collection

The data used in this study consists of second-by-second price information for the 10 largest cryptocurrencies by market capitalization excluding stablecoins as of January 10, 2025. The cryptocurrencies included are BTC, Ethereum (ETH), Binance Coin (BNB), Ripple (XRP), Cardano (ADA), Solana (SOL), Dogecoin (DOGE), TRON (TRX), Avalanche (AVAX), and Chainlink (LINK).

Cryptocurrency price data was downloaded from Binance, which is one of the largest cryptocurrency exchanges by trading volume. Since Binance only provides price data in terms of cryptocurrency pairs rather than prices in US dollars, we collected prices for all assets against Tether (USDT), because it is a stable coin whose price is always fixed to one US dollar.

B. Data Storage and Formatting

We are looking at lead-lags of 45 pairs of cryptocurrencies, which makes it challenging in size and computations, so we needed to undertake different actions to compensate from that.

Given the size and frequency of the dataset, efficient storage and formatting were essential to facilitate analysis. The raw data collected from Binance was initially stored in CSV format, which, while accessible, proved inefficient for large-scale storage and computational tasks due to its size and slower read/write performance. To address this, the data was converted into parquet files, which offer better compression and faster access speeds, making them more suitable for high-frequency financial data.

Each cryptocurrency's data was stored in separate parquet files, organized by day. For the 10 selected cryptocurrencies, this resulted in one parquet file per day per asset, ensuring that the data could be accessed in manageable chunks. This daily partitioning of the data further optimized loading times, as only the required files for a specific period needed to be accessed during analysis.

Moreover, to reduce unnecessary storage and improve processing efficiency, only the relevant columns for our analysis were retained: the timestamp and the price at the opening of each one-second candlestick.

C. Precomputations

To optimize the trading strategy and reduce the computational load during backtesting, several key metrics were precomputed and stored. For each cryptocurrency, we calculated the returns, defined as the percentage change in price between consecutive seconds, as well as the 5-second returns, which represent the percentage price change over rolling 5-second windows. Additionally, rolling percentiles of the 5-second returns over the past hour were computed, specifically the 5th, 25th, 75th, and 95th percentiles. These metrics were stored alongside the original price data for each cryptocurrency in the same parquet files, organized by day. This structure ensured that all relevant metrics for a given cryptocurrency and day were easily accessible in a single file, improving computational efficiency.

In addition to the individual cryptocurrency metrics, lead-lag values were precomputed for all 45 unique cryptocurrency pairs formed from the 10 selected cryptocurrencies. For each pair, rolling weighted averages of the lead-lag values were calculated over 40-second windows. We will explain why we needed to do this later on, when explaining the trading strategy that we tested. To maintain efficient storage and organization, the lead-lag values were stored separately in their own parquet files, with one file per pair per day.

These precomputations were critical for handling the high-frequency, large-scale dataset. By performing these calculations in advance, the strategy was able to operate efficiently, accessing key metrics and relationships without the need for expensive real-time computation.

In the end, even with our file compressions and optimizations, our combined dataset weighted 50 GB.

D. Computational Optimizations

The computational demands of this study were driven primarily by the need to compute lead-lag relationships for all 45 unique cryptocurrency pairs at second-level frequency over an entire year. For each pair, lead-lag values were calculated over rolling 40-second windows, with the computation for each second involving operations proportional to n^2, where n is the length of the time series within the rolling window. This quadratic complexity, coupled with the high-frequency nature of the data, made lead-lag computations particularly demanding.

To address these challenges, we employed Just-In-Time (JIT) compilation using the numba library, which dramatically improved the performance of these computations. A naive implementation of the lead-lag computation, without JIT optimization, took approximately 3.38 seconds for 7 computations, while by applying JIT compilation, the computation time was reduced to approximately 2.12 milliseconds. This optimization made it feasible to precompute lead-lag values for the entire dataset, significantly reducing the time required for this critical step.

Additionally, we utilized parallelization to distribute the workload across multiple CPU cores. Since the lead-lag computations for different cryptocurrency pairs and days are independent, parallelization allowed us to compute values for multiple pairs or time intervals simultaneously. By leveraging all available hardware resources, we were able to process the full dataset in a reasonable amount of time, despite its size and computational complexity.

IV. Trading Strategies

A. Inspiration

The motivation for this project stems from the observable patterns of price movements across cryptocurrencies, particularly their tendency to follow similar trends over time. Cryptocurrencies are highly interconnected due to shared market drivers such as investor sentiment, macroeconomic factors, and liquidity flows. This interconnectedness often results in correlated price movements between assets, even at high frequencies.

For example, an analysis of BTC and ETH price variations over a randomly selected one-hour period shown in figure 3 reveals a strong similarity in their overall trends. When zooming in on a smaller window of 200 seconds within the same period, this pattern persists, as we can see in figure 4, with the prices of the two assets following one another closely, although with slight temporal offsets. These observations hint at the potential for identifying leading and lagging behavior between assets in high-frequency timeframes.

Figure 3. Comparison of price between BTC and ETH over one hour

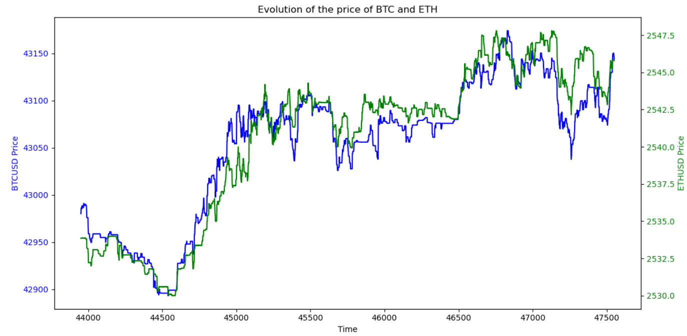
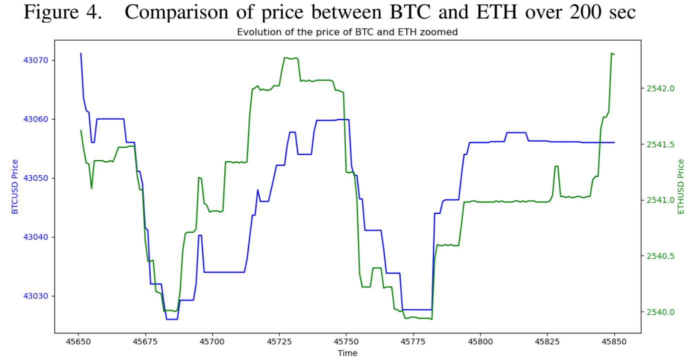


Figure 4. Comparison of price between BTC and ETH over 200 sec



This visual evidence of correlated movements led us to hypothesize that lead-lag relationships could be systematically analyzed and exploited for trading. If one cryptocurrency exhibits a price movement that the other has yet to replicate, this could serve as a signal for entering long or short positions on the lagging asset. This motivated us to explore the use of OTCP to quantitatively identify and measure these relationships, forming the basis of our trading strategy.

B. Lead-Lag Computation

To identify exploitable relationships between cryptocurrency pairs, we computed lead-lag dynamics over short time intervals, ensuring that the computation aligns with a realistic real-time trading setup. Specifically, lead-lags were calculated using 40-second rolling windows, with a new computation performed every second, with the prices of the last 40 seconds. Within each window, a weighted average of the lead-lag values was applied, assigning greater weight to the most recent seconds. This approach prioritizes recent price movements, which are more reflective of the current market dynamics, while smoothing the effects of older data.

Importantly, our computation method does not rely on the current price or future prices to determine the lead-lag value for the current second. This ensures that the methodology remains consistent with a real-time trading context, where future prices are unknown. By focusing only on historical data up to the current second, the computed lead-lag value represents an actionable metric that could be used in live trading decisions.

The result of this process is a dynamic, time-varying measure of lead-lag relationships that evolves second-by-second. These values serve as the foundation for identifying potential trading opportunities, as they provide insights into which asset is leading or lagging within the most recent 40-second window.

Interestingly, the distribution of computed lead-lag values was nearly identical across all cryptocurrency pairs, with the exception of one outlier pair. This consistency suggests that the dynamics governing lead-lag relationships in the cryptocurrency market are shared broadly across assets. Even more notably, the distributions were centered around 0, indicating that no single asset consistently leads or lags another over time. This symmetry around 0 reflects the dynamic nature of cryptocurrency markets, where leadership among assets tends to shift frequently.

When analyzing the computed lead-lags, values exceeding one standard deviation were treated as significant signals for potential trading opportunities. In this context, a lead-lag value greater than one standard deviation corresponds to approximately the 84th percentile. These thresholds indicate that the observed lead-lag is relatively rare and may represent a meaningful dive). rgence from typical behavior, making it a potential candidate for actionable trading decisions.

C. Decision Rules for Trading

We designed a basic strategy based on that to see whether we could use the lead-lags to extract returns in a high-frequency trading strategy. The strategy follows a rules-based approach designed to exploit significant lead-lag relationships between cryptocurrency pairs, combined with extreme short-term price movements in the leading asset. Below is the pseudocode describing the algorithm in detail:

Algorithm 1 - Simple Trading Strategy Based

Step 1: Identify Trading Signals

- Each second:
  - Identify if there is a significant lead-lag relationship between two assets
  - Check if the leading asset has a significant 5-second return.
  - Check if the lagging asset has flat returns over the same 5-second period.

Step 2: Execute Trades

- If the conditions above are satisfied:
  - Take a long position on the lagging asset if the leading one exhibits positive movement.
  - Take a short position otherwise

Step 3: Close Trades

- Close the position on the lagging asset when:
  - The lagging asset moves by a significant amount from the entry price, or
  - 40 seconds have passed since the trade was initiated.

Return: Profit/loss results for all executed trades.

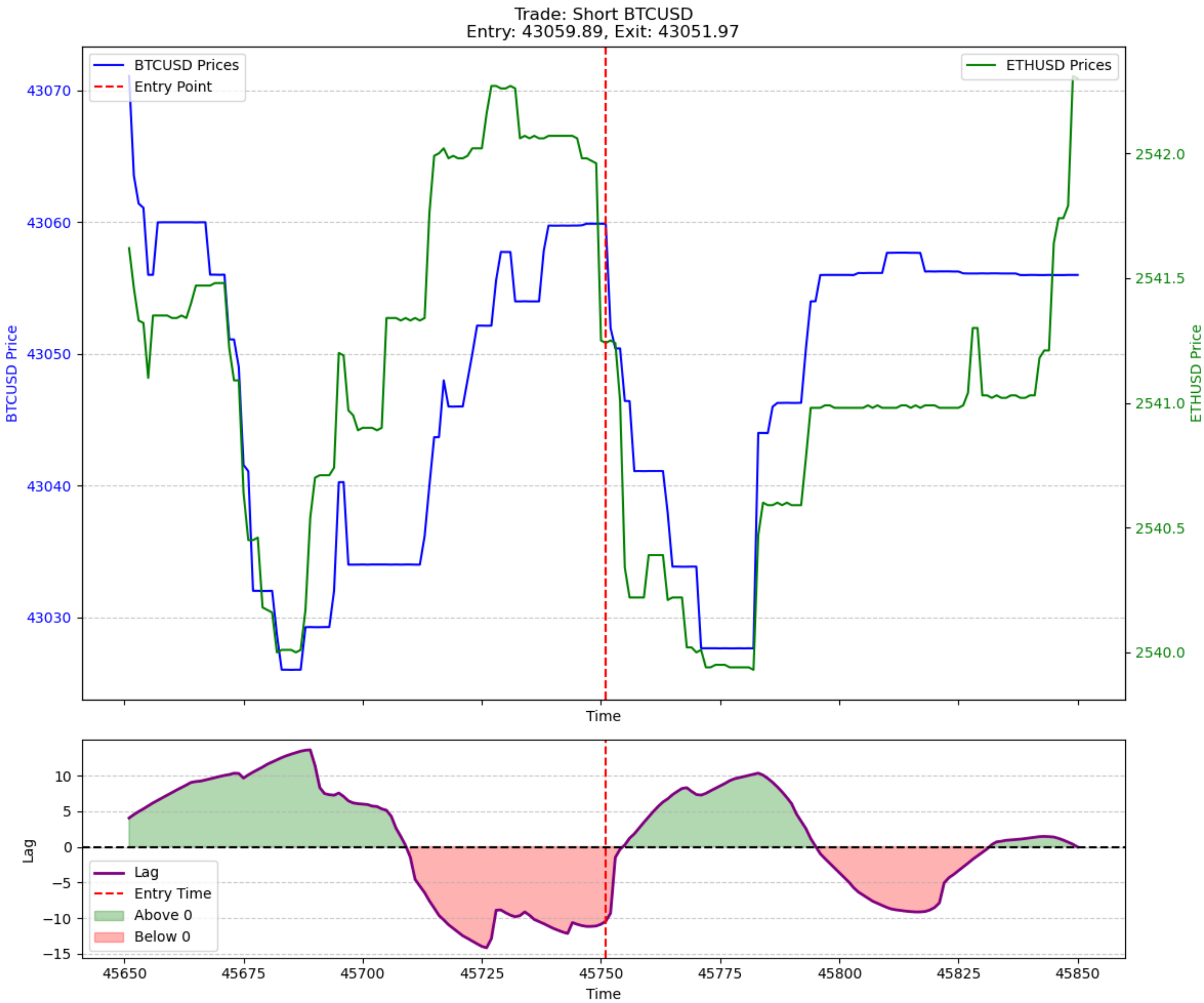
For this strategy, we define a significant 5-second return as a return exceeding the 95th percentile of 5-second returns calculated over a rolling window of the past hour. We chose a one-hour window because it strikes a balance between being long enough to capture meaningful high-volatility events while remaining short enough to reflect the current state of the market. Using a longer period, such as a day, would include returns from periods of lower activity, like nighttime trading, which could distort the percentile thresholds. Moreover, returns from 24 hours ago are unlikely to correlate strongly with current price dynamics. We also consider a cryptocurrency's 5-second return to be flat if it lies between the 25th and 75th percentiles of the rolling 5-second return distribution over the past hour.

D. Example of a trade

To illustrate the practical implementation of our algorithm, we provide an example of a trade where the strategy successfully exploited a lead-lag relationship between BTC and ETH. In this scenario, depicted in Figure 5, we observe a significant lag between BTC (blue) and ETH (green), with ETH acting as the leading asset, and our algorithm entered a short position on BTC at the time marked with the red line.

Just before the red line, we can see that ETH exhibits a significant price drop, while BTC remains flat. This situation satisfies the algorithm's criteria for identifying a trading signal: a significant lead-lag relationship combined with extreme movement in the leading asset and flat movement in the lagging asset.

Figure 5. Short position where it worked



Upon detecting this signal, the algorithm executes a short position on BTC, anticipating that BTC's price will follow ETH's downward trend. As the figure shows, BTC's price begins to drop shortly afterward, validating the algorithm's prediction. The position is closed when BTC's price movement reaches the 95th percentile of its historical returns over the past hour, generating a profit from the trade.

It is worth noting that while the algorithm successfully captured the price drop in BTC, it exited the position earlier than optimal, as BTC continued to decline after the trade was closed. This highlights a potential area for improvement in the selling strategy, such as incorporating dynamic exit conditions, allowing for higher returns when we are in such situations. Nevertheless, this example clearly demonstrates the fundamental idea behind the strategy.

V. Results

VI. Limitations and Future Work

A. Limitations

This study operates under several limitations, reflecting its primary objective of evaluating the feasibility of using lead-lag relationships to inform trading strategies rather than developing a fully operational system. The most significant limitation is the exclusion of trading fees, slippage, and other transaction costs. High-frequency trading strategies often involve thousands of trades per month, and these costs can accumulate rapidly, significantly reducing or even eliminating profitability. While this study focuses on gross

Pair	Monthly Return (%)
BNB-XRP-yearly-results.parquet	39.04
SOL-LINK-yearly-results.parquet	38.46
BNB-AVAX-yearly-results.parquet	36.61
BNB-LINK-yearly-results.parquet	35.72
DOGE-LINK-yearly-results.parquet	33.04
XRP-AVAX-yearly-results.parquet	32.05
TRX-LINK-yearly-results.parquet	29.02
ADA-LINK-yearly-results.parquet	26.98
ETH-BNB-yearly-results.parquet	26.09
ADA-AVAX-yearly-results.parquet	25.32

Table I
TOP 10 PAIRS WITH HIGHEST MONTHLY RETURNS (31 DAYS)

returns to explore the potential of lead-lag relationships, incorporating realistic transaction costs would be essential for a complete assessment of the strategy's viability.

Another limitation is the simplicity of the trading strategy itself. The decision rules were intentionally designed to be straightforward, without advanced features such as dynamic position sizing, stop-loss mechanisms, or risk-adjusted trade sizing. This simplicity limits the strategy's results, and make the returns appear really small against the trading fees, which indicates that it couldn't be applied in a real-time context.

Finally, the study assumes ideal market conditions, including perfect liquidity and instantaneous trade execution at the prices provided by Binance. In reality, market liquidity can vary significantly, particularly for less popular cryptocurrencies, and large trades can influence market prices, resulting in slippage. Execution delays and latency, which are not considered in this study, could also impact real-world performance, especially in high-frequency trading scenarios. These assumptions simplify the testing environment but do not fully reflect the challenges of deploying the strategy in live trading.

B. Future Work

Several directions for future research could address its limitations and improve the strategy's practicality and effectiveness. The most immediate next step is to incorporate trading fees, slippage, and other transaction costs into the analysis. These factors are critical in high-frequency trading, where the cumulative effect of thousands of trades can substantially impact net returns. A more realistic evaluation of the strategy's profitability would involve optimizing trade frequency to minimize costs while retaining the core insights provided by lead-lag relationships.

The decision rules underpinning the strategy could also be refined to enhance robustness and adaptability. Incorporating dynamic exit conditions, such as trailing stop-loss mechanisms, could reduce premature trade closures, allowing the strategy to capture larger price movements when market trends persist. Additionally, introducing position sizing rules based on the magnitude of the lead-lag signals or the strength of recent price movements could improve the strategy's risk management and responsiveness to market conditions. These enhancements would enable the strategy to perform better in more diverse and volatile market environments.

Future research should also focus on implementing the strategy in a live trading environment to test its real-world applicability. This would allow for the evaluation of factors such as latency, execution delays, and data discrepancies, which are not accounted for in backtesting. Real-time testing could reveal practical challenges and provide opportunities to fine-tune the algorithm for deployment in live markets.

Another promising avenue for future work involves incorporating multiple lead-lag signals across cryptocurrency pairs. For instance, if one cryptocurrency, such as crypto A, leads crypto B with a positive price movement, while another leading cryptocurrency, crypto C, shows a conflicting downward movement, the strategy could abstain from trading crypto B to avoid noise from contradictory signals. Conversely, if both crypto A and crypto C exhibit upward movements, this alignment could justify a larger position on crypto B, amplifying the trade signal. By accounting for the combined dynamics of multiple assets, this extension could improve decision-making and reduce false signals.

REFERENCES