

MENG INDIVIDUAL PROJECT

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

An Empirical Analysis of Mean Reversion Strategies for Pairs Trading in Decentralised Exchanges

Author:
Devam Savjani

Supervisor:
Dr. Thomas E. Lancaster

Second Marker:
Mr. Ivan Procaccini

June 19, 2023

Abstract

Given the rise in popularity of blockchain technology and decentralised finance, trading cryptocurrencies has grown massively, leading to numerous markets that have exposure to countless arbitrage opportunities. Further increasing this, the introduction of decentralised exchanges have introduced different dynamics into market prices. Methods such as cyclic and triangular arbitrage on decentralised exchanges have been researched and found to be more costly than profitable due to transaction fees. This raises the question of whether more advanced arbitrage strategies can result in a significant profit.

This project provides novel research into the application of quantitative trading strategies, based on mean reversion, on decentralised exchanges. The mean reversion strategies implemented include a constant hedge ratio strategy using ordinary least squares (OLS) as a benchmark, sliding window using OLS, lagged using OLS, Granger Causality test-based OLS model, and Kalman Filter-based hedge ratio estimation. A sophisticated live trading system was also implemented to enable the strategies to trade live on the Ethereum blockchain. Furthermore, to evaluate the performance of the strategies, a backtesting system was developed to simulate the market conditions accurately.

We found that three of the five strategies implemented surpassed current state-of-the-art research, using similar methods on different financial products, accumulating annual percentage returns of 81.14%, 48.85% and 38.70% by the Kalman Filter, Granger Causality and Lagged strategies. Furthermore, each strategy's β and Sharpe ratio display an attractive investment opportunity for investors, as it displays a low correlation to the market portfolio and favourable risk-adjusted returns.

Acknowledgements

My sincere gratitude goes to my supervisor Dr. Thomas E. Lancaster and second marker Mr. Ivan Procaccini for their advice and guidance throughout the project.

I would also send my gratitude to my family and friends whose motivation helped me throughout this degree.

Contents

1	Introduction	3
1.1	Motivation	3
1.2	Objectives	3
1.3	Contributions	4
1.4	Ethical Issues	4
2	Background	5
2.1	Cryptocurrencies	5
2.1.1	Blockchain	5
2.1.2	Ethereum	7
2.1.3	Decentralised Finance	8
2.2	Arbitrage	9
2.3	Pure Arbitrage Techniques	9
2.4	Statistical Arbitrage Techniques	9
2.4.1	Basic Mean Reversion	10
2.4.2	Statistical Arbitrage using the Kalman Filter	10
2.4.3	Analysis on Cryptocurrency Arbitrage on Centralized Exchanges	13
3	Design Decisions	16
3.1	Mean Reversion Strategies	16
3.2	Buying and Short Selling	16
3.2.1	Buying	16
3.2.2	Short Selling	17
3.3	Protocols of Interest	17
3.3.1	Uniswap	17
3.3.2	Aave	19
3.4	Fees that are incurred when trading	20
3.4.1	Gas Fees	20
3.4.2	Uniswap Fees	21
3.4.3	Aave Fees & Collateral	21
3.5	Liquidity Pool Pairs of Interest	22
3.5.1	Correlated and Cointegrated Liquidity Pools	22
4	Implementation of the Trading Systems	24
4.1	Backtesting System	24
4.1.1	Data Collection and Storage	26
4.1.2	Types of Orders and Execution	27
4.1.3	Gas Fees	28
4.1.4	Validating Balance Health	30
4.2	Live Trading	30
4.2.1	Smart Contracts	31
4.2.2	State	34
4.2.3	Retrieval of Data and Signal Generation	34

4.2.4	Trade Execution	34
5	Trading Strategies	35
5.1	Abstract Strategy	35
5.1.1	Volume of Trades	36
5.2	Constant Hedge Ratio Strategy	37
5.3	Sliding Window OLS Strategy	37
5.4	Lagged OLS Strategy	38
5.5	Unrestricted Model from the Granger Causality Test	38
5.6	Kalman Filter Strategy	39
6	Evaluation	41
6.1	Trading System	41
6.2	Strategy Parameters	41
6.2.1	Number of Standard Deviations away from the mean	42
6.2.2	Gas Fees	42
6.2.3	Initial Investment Volume	46
6.2.4	Window Size	46
6.3	Liquidity Pool Pairs Selection	47
6.4	Evaluation of Strategies	48
6.4.1	Beta	50
6.4.2	Volatility & Sharpe Ratio	51
6.4.3	Comparison with Previous results	52
7	Conclusion	57
7.1	Future Work	58
7.1.1	Running the Strategies Live	58
7.1.2	Different Decentralised Exchanges	58
7.1.3	Different Tokens	58
7.1.4	Different Blockchains	58
A	Additional Background	59
A.1	Pure Arbitrage	59
A.2	Optimal Portfolio Design for Mean Reversion	62
B	Supporting Tables	65
B.1	Exerpt of the liquidity_pools Table	65
C	Supporting Algorithms	66
C.1	Gas Price Collection	66
D	Supporting Code Snippets	67
D.1	Data Collection GraphQL Queries	67
D.1.1	Uniswap Pool Historical Price Data Collection	67
D.1.2	Uniswap Pool Live Price Data Collection	67
D.1.3	Uniswap Historical Gas Price Data Collection	67
D.1.4	Uniswap Live Gas Price Data Collection	68
D.1.5	Aave Historical Lending Pool Data Collection	68
D.2	Backtesting Code Snippets	68
D.2.1	Order Execution Code	68
D.2.2	Strategy Hedge Ratio Calculation Code	70
D.2.3	Live Trading Code Snippets	71
	References	76

Chapter 1

Introduction

1.1 Motivation

Since the advent of Bitcoin, the groundbreaking peer-to-peer payment network and cryptocurrency, the blockchain technology ecosystem has witnessed an exponential surge, propelling the emergence of numerous cryptocurrencies that are now widely utilised and traded. The soaring volatility of cryptocurrencies has not only captured the attention of retail investors, with some reaping substantial returns while others suffered significant losses [1]. Moreover, prominent financial institutions have joined the fray, aiming to capitalise on this novel tradable asset [2]. Consequently, this surge of interest has spurred the development of more sophisticated approaches to cryptocurrency trading.

Initially, trading cryptocurrencies posed a significant challenge for individuals lacking technical expertise, with the first recorded transaction occurring on 12th October 2009 via a PayPal transaction [3]. However, the landscape has since evolved, giving rise to various centralised and decentralised cryptocurrency exchanges. Centralised exchanges have gained popularity, offering a user-friendly interface allowing investors with limited technical knowledge to access the ecosystem. In contrast, decentralised exchanges (DEXes) operate on blockchain networks and facilitate direct peer-to-peer transactions using smart contracts. The prevalence of centralised exchanges is evident, as depicted in Figure 1.1, where the proportion of trades on DEXes remained minimal until the summer of 2020 but has steadily increased.

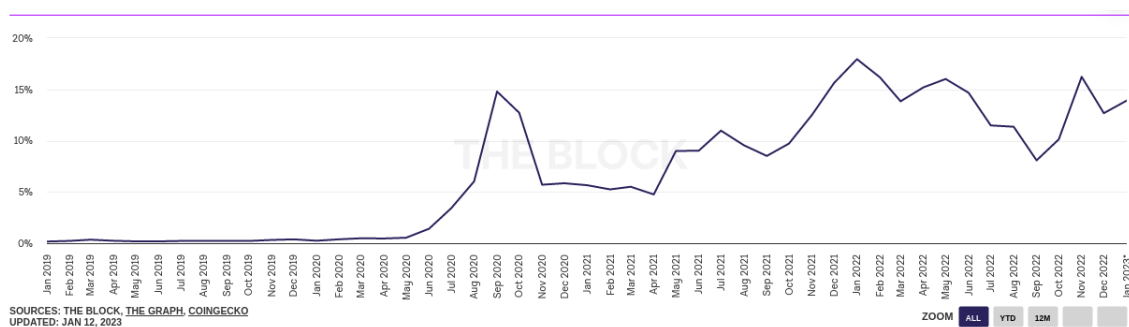


Figure 1.1: DEX to CEX Spot Trade Volume [4]

1.2 Objectives

This poses a question about whether advanced trading strategies can exploit arbitrage opportunities on decentralised exchanges. Surprisingly, existing research on cryptocurrency arbitrage has solely focused on triangular and cyclic arbitrage. Therefore, this project aims to be the first to explore the potential profitability of employing mean reversion techniques in cryptocurrency pairs trading on decentralised exchanges. By implementing diverse variations of the mean reversion strategy, the project aims to unravel the most effective methods for achieving remarkable profits.

1.3 Contributions

The work presented in the report is motivated by the rise in trading activity on decentralised exchanges. Exploring statistical arbitrage techniques on DEXes provides an opportunity to gain insights into decentralised markets' unique dynamics and potential profitability. The contributions are as follows:

1. **Analysis of Liquidity Pool Pairs** - The analysis of liquidity pool pairs involves examining the correlation and cointegration between different liquidity pools. This analysis provides insights into the pricing dynamics and long-term inter-pool relationships, exposing potential arbitrage opportunities.

2. **Backtesting System** - A meticulous backtesting system was developed to replicate the execution of trades by utilising comprehensive historical data directly sourced from the Ethereum blockchain. This system enabled the evaluation of trading strategies under realistic and reliable market conditions, allowing for a thorough assessment of their performance.

3. **Live Trading System** - To facilitate real-time trading and interaction with the Ethereum blockchain, a dynamic live trading system was designed and implemented. This system enabled the seamless execution of trading strategies in a live market environment.

4. **Trading Strategies** - The mean reversion strategy relies on the hedge ratio, a key parameter that impacts the strategy's risk exposure and trading volumes. To assess the performance of different hedge ratio estimation methods, multiple strategies were employed: constant hedge ratio using ordinary least squares (OLS) as a benchmark, sliding window using OLS, lagged using OLS, Granger Causality test-based OLS model, and Kalman Filter-based hedge ratio estimation.

5. **Novel Analysis of Applying Mean Reversion on Decentralised Exchanges** - The analysis showed that the Kalman Filter strategy outperformed other mean reversion strategies and current research, achieving an Annual Percentage Return of 81.14%. Equally noteworthy, the Granger Causality and Lagged strategies delivered substantial returns of 48.85% and 38.70%, respectively, surpassing returns observed in prior research on mean reversion and pure arbitrage across diverse asset classes.

1.4 Ethical Issues

The ethics of cryptocurrencies are widely debated for reasons such as anonymity, leading it to be the choice of currency used by criminals and illegal institutions, volatility and lack of regulation. The high volatility makes cryptocurrencies and decentralised finance very risky for retail investors that do not have the technical or financial know-how making investing in cryptocurrencies.

Another aspect of cryptocurrencies that has raised ethical questions is the energy consumption and carbon dioxide emission from the mining of cryptocurrencies. Formal research has also found that 'approximately 69 million metric tons of CO₂ (Carbon dioxide) emission due to bitcoin mining' [5]. Thus, this is an ethical concern that was considered when designing the strategies so that the number of transactions that do not result in a profit, i.e. do not add value to the project, is limited.

In addition to the concerns above, although this project aims to find riskless profits, *'free lunches'*, it is not, in any form, of financial advice. Those who use the research or software used in the development and research process to attempt to get favourable results, are liable for the losses or gains.

Chapter 2

Background

2.1 Cryptocurrencies

Before delving into the financial side of the project, it is important to understand the underlying assets and the technology that drives them.

2.1.1 Blockchain

The building blocks of cryptocurrencies come from blockchain technology. Blockchain is a distributed ledger that stores data, in blocks, in a chain, comprising the data itself and a full transaction history [6]. They can be thought of as a State Transition System where the state stores data, i.e. the owners of each token, and the state-transition function is a function of state and a transaction [7]. The transition function defines how the transaction, T , should affect the state, σ .

$$\sigma' = transition_func(\sigma, T)$$

Thus, the state, σ , and transition function $transition_func$ are defined by the blockchain implementation, which tends to be more complex.

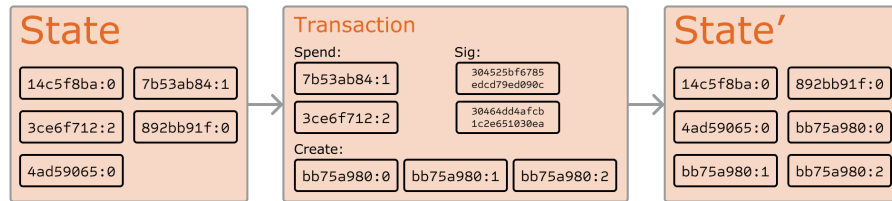


Figure 2.1: State Transition Diagram [7]

In reality, a collection of transactions are batched together into a block which is then committed to the network. The components of a block can be seen below.

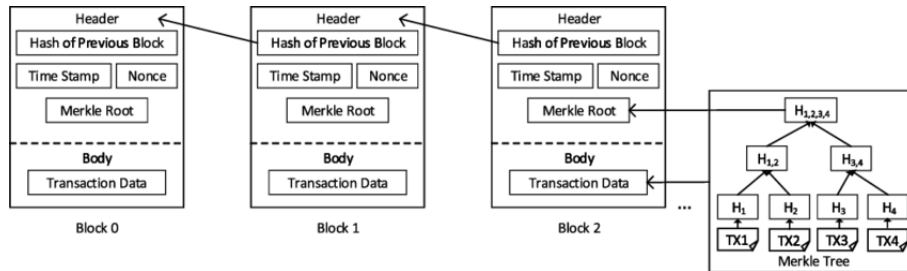


Figure 2.2: Blockchain Diagram [8]

A block has two significant components, the header and the body. The block's body contains the list of transactions to be applied. Whereas the header components may depend on the blockchain's implementation, they typically have the following key components.

- **Block number** - A unique identifier assigned to each block in the chain. It serves as a chronological ordering mechanism, indicating the block's position within the blockchain.
- **Timestamp** - Timestamp of the block's creation time.
- **Merkle Root** - A Merkle root is also stored in each block to validate transactions in the body of the block efficiently, in terms of storage and searching [9]. A Merkle tree is a tree of hashes where each leaf node is its data hash, and its parent node is the hash of their children's hashes. In storing the Merkle root, we do not need to store each transaction in each block directly, and it also allows a quick search for any malicious alterations in differing blocks.
- **Hash of Previous Block** - Acts as a pointer to the preceding block. Hence, the term "Blockchain" is coined as one can track back to the first block from the latest block in a chain-like manner.

In addition to these components, a major feature of blockchain technology is its decentralised property. The state of the blockchain is managed by a distributed network of nodes maintained by voluntary node operators. Node operators play a crucial role in maintaining the blockchain. They validate and propagate transactions, participate in consensus mechanisms, and ensure unanimous agreement on the current state of the blockchain before it is updated. To facilitate these operations, blockchain networks commonly adopt a peer-to-peer (P2P) architecture. Each participant in the network, or node, has equal status and communicates directly with other nodes. Transactions and data are shared among participants without intermediaries or central servers. When transitioning to the next state, the node broadcasts the transaction, T , to the blockchain network, and each node applies the transition function to obtain the new and same state, $\sigma' = transition_func(\sigma, T)$. All nodes arrive at the same state as the transition function is deterministic. Once each node has applied the transition function, the network uses a consensus mechanism to the globally agreed σ' , and the transactions become actualised.

Consensus mechanisms ensure agreement among participants regarding the validity and order of transactions added to the blockchain. Popular examples of consensus mechanisms include Proof of Work (PoW), Proof of Stake (PoS), and Delegated Proof of Stake (DPoS). The most widely adopted mechanisms in current blockchains are Proof of Work and Proof of Stake [10]. In addition to reaching a consensus, the consensus algorithms also determine which node should create a new block and which node should be granted a monetary reward.

The Proof of Work (PoW) consensus algorithm was first used in Bitcoin. The nodes compete to solve complex mathematical problems to verify that the transactions from a new block are valid to add them to the blockchain. These nodes are called miners. Once the first miner obtains the solution to the mathematical puzzle, it broadcasts the new block on the network so that the other nodes verify the solution and update their local replica. Consensus is then reached when 51% of the nodes agree on the new state of the blockchain. The main drawback of this algorithm is the expensive computational calculations required by miners to repeatedly find the solution to the mathematical problem. Once a miner discovers the solution, it is broadcasted to other nodes for easy and cheap verification.

To overcome this high cost, some blockchains have elected to use the Proof of Stake (PoS) consensus algorithm, which is more energy efficient and scalable. In PoS, validators are

responsible for creating new blocks by staking a certain amount of cryptocurrency as collateral. The block creator is randomly selected from the validators, with the probability of selection determined by the amount they have staked. Due to the advantages that PoS has over PoW, Ethereum has transitioned from PoW to PoS in Ethereum 2.0 in 2022 [11].

2.1.2 Ethereum

One of the most prominent blockchains is Ethereum which was proposed in a whitepaper by Vitalik Buterin. It is centred around using smart contracts to create financial products and services that could operate independently of traditional financial institutions. Hence decentralised finance was birthed [12].

Smart Contracts

Smart contracts are self-executing contracts between buyers and sellers that deploy on a blockchain. The programs are reactive because they are only executed on the blockchain, providing that certain conditions are met [13]. These conditions are set in the lines of code in the smart contract. To possess the reactive property, smart contracts are stored and executed on every participating node on the network as part of the state σ . They are also immutable and decentralised, which makes smart contracts advantageous as they can facilitate, verify, and enforce the negotiation or performance of a contract [14, 15].

Ethereum's Architecture

Ethereum's architecture is similar to Bitcoin's but has a few differences; one is that rather than managing a distributed ledger, it uses a distributed state machine. The Ethereum Virtual Machine (EVM) defines the rules of changing states from block to block. Each node on the Ethereum blockchain contains an immutable instance of the EVM [16].

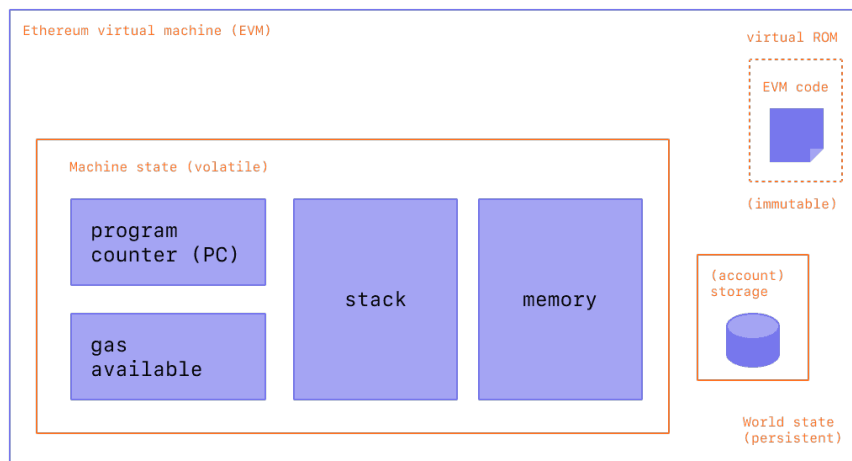


Figure 2.3: EVM components [16]

Ethereum's state σ consists of the states of all accounts where the state of an account is represented by four components: balance of ETH (the native cryptocurrency on Ethereum), a nonce, hash of the smart contract code (if the account is a smart contract) and the Merkle root of smart contract storage which maintains the contracts variables (again, if the account is a smart contract) [17]. A nonce, 'Number only used once', is a number that is added to a hashed block to make the transaction more secure. It is randomly generated and used to validate a transaction. A miner first guesses a nonce and appends the guess to the hash of the current header. The miner then rehashes the value and compares this to the target hash. If the guess is correct, the miner is granted the block [18].

Ethereum supports two types of transactions: contract creation and message calls. Message calls transfer ETH from account to account and invoke smart contracts. The process of how transactions are validated is described below:

1. Validate the parent block
2. Validate that the current timestamp is greater than the previous timestamp
3. Check that the Ethereum concepts are valid
4. Perform Proof of Stake on the block
5. Check for errors and gas
6. Validate the final state

Gas Fees

Due to the capability of expressing loops in EVM bytecode, there is a possibility that smart contract execution could continue indefinitely. This presents a challenge as invoking a contract with an infinite loop would cause all Ethereum network nodes to become stuck executing it, resulting in a denial of service. To mitigate this issue, Ethereum has implemented a pricing mechanism. Every computation performed by a smart contract requires the payment of a fee called gas, denominated in Ether.

To measure how much computational effort is required to execute operations on the Ethereum network, gas is used as a unit [19]. Every block has a base fee derived from the demand for the block space, which is burnt. Therefore, network users are expected to set a tip (priority fee) to reimburse miners for adding their transactions in blocks; thus, the higher the tip, the greater the incentive for miners to validate the transaction. Using gas means that the Ethereum network is tolerant to spam and has a maximum gas fee to make Ethereum tolerant to malicious code that would be used to waste resources.

2.1.3 Decentralised Finance

One of Ethereum and smart contract applications is Decentralised Exchanges (DEXes). Before delving into DEXes, it is important to understand the mechanisms of centralised exchanges.

Centralized Exchanges

Centralised exchanges (CEXes) allow agents to discover and trade assets. CEXes facilitate trading between buyers and sellers by providing an online platform that manages and maintains an order book. An order book aggregates buy and sell orders and execute matching buy and sell orders. The order book and transactions are typically managed on a central database. When trading, exchanges charge trading fees for the maker and the taker to operate the exchange and do not charge any gas fees as there is no interaction with the blockchain.

Decentralised Exchanges

In contrast, DEXes utilise blockchain technology and smart contracts to execute trades, thus providing a high level of determinism by nature of the technology. These trades are executed on the blockchain via smart contracts and on-chain transactions. There are two types of DEX: order book DEXes and Automated Market Makers (AMMs). An order book DEX is less common and is akin to a CEX; an order book is stored on the blockchain rather than on a central database. This means each order placed requires the order book to be posted on the blockchain at each transaction. Automated Market Makers are more common and provide instant liquidity by using liquidity pools so that users can swap

their tokens for a price determined by the portions within the liquidity pool [20]. DEXes have multiple advantages, including lower transaction fees, privacy, diversity and trustless transactions, but they also have drawbacks, such as scalability and the potential of having poor liquidity [21].

2.2 Arbitrage

Arbitrage is when a trader simultaneously buys and sells an asset to exploit a market inefficiency [22]. Arbitrage is also possible in other types of securities by finding price inefficiencies in the prices of options, forward contracts and other exotics.

The fundamental theory for arbitrage is the “Law of One Price”. The “Law of One Price” states that every identical commodity or asset should have the same price regardless of exchange or location, given there are no transaction costs, no transportation costs, no legal restrictions, the exchange rates are the same and no market manipulation occurs [23]. This is because if this were not the case, an arbitrage opportunity would arise, and someone would take advantage of the scenario causing the prices on both markets to converge due to the market forces. In the real world, arbitrage opportunities are tremendously common, thus allowing a risk-free investment [24, 25].

There are countless types of arbitrage, such as spatial arbitrage, which profits off of different prices on exchanges in different locations, temporal arbitrage, which takes advantage of price differences at different times, risk arbitrage, which profits from perceived discrepancies in their risk-return profiles and finally market arbitrage which takes advantages of different prices on different exchanges/markets. Statistical methods include pairs trading, which involves buying and selling assets that are believed to be mispriced relative to one another; momentum trading, which identifies if assets have a strong momentum (either up or down) and profiting off of that; and finally, algorithmic trading which uses algorithms to analyse data and trades based on statistical analysis.

2.3 Pure Arbitrage Techniques

Research into cryptocurrency arbitrage is still in its infancy, and previous research has mainly focussed on the economics of cryptocurrencies, i.e. miner/trader behaviour and influence of cryptocurrency trading [26, 27, 28, 29, 30, 31, 32]. Furthermore, very limited research has compared statistical strategies and pure methods of arbitrage of cryptocurrencies. However, it is still important to understand the basics of pure arbitrage; this form of arbitrage takes advantage of price discrepancies between three or more different currencies in the foreign exchange market. It involves executing a series of trades to profit from the imbalance in exchange rates between the currencies involved. An example of this can be seen in Figure 2.4, given the exchange rates are $\$1 = \text{€}0.85$, $\text{€}1 = \text{£}0.75$, $\text{£}1 = \$1.20$, we can make a series of exchanges (trades) such that by starting with \$100, the result of this cyclic arbitrage I am left with \$130.72, hence a \$30.72 risk-free profit. Although the example uses fiat currencies (traditional currencies that are issued by governments, such as the US Dollar (USD) and the Great British Pound Sterling (GBP)), the same also holds for cryptocurrencies. Additional research into pure cryptocurrency arbitrage can be found in Appendix A.1.

2.4 Statistical Arbitrage Techniques

As mentioned previously, the project’s subject is to analyse the profitability of statistical arbitrage methods in decentralised exchanges. The elected technique of statistical arbitrage for this project is mean reversion in pairs trading. Research within pair trading has

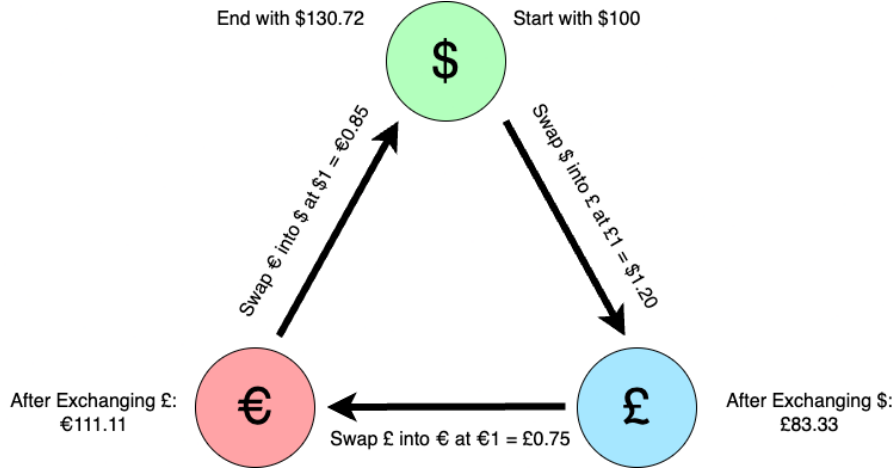


Figure 2.4: Triangular Arbitrage Diagram

been vast, with many streams of approaches emerging; distance approach, cointegration approach, time-series approach, stochastic approach and some others, including using machine learning [33]. However, this project will only look at cointegration/co-correlation approaches.

2.4.1 Basic Mean Reversion

The cointegration approach follows three key steps. The first is the selection of pairs based on similarity measures; the next is assessing the credibility; finally, thresholds are set for trading. The spread is defined as

$$\varepsilon_{ij,t} = P_{i,t} + \gamma P_{j,t}$$

where $P_{i,t}$ and $P_{j,t}$ denote the $I(1)$ non-stationary price processes of the assets i and j , γ is the cointegration coefficient, also referred to in literature as the hedge ratio. $\varepsilon_{ij,t}$ is the linear combination of the non-stationary prices and is $I(0)$ stationary and hence mean-reverting; note that stationary processes are those of which have a constant mean. Rad's implementation of this approach on stocks results in a 0.83% return before considering transaction costs [34]. Another paper, [35], looked into setting the thresholds and setting a minimum profit, MP_{ij,t_c} :

$$MP_{ij,t_c} = \frac{n(\varepsilon_{ij,t_0} - \varepsilon_{ij,t_c})}{|\gamma|}$$

Where t_0 and t_c are the opening and closing times, n is the volume longed of asset j . Additional research on the optimal portfolio designed for mean reversion can be found in Appendix A.2.

2.4.2 Statistical Arbitrage using the Kalman Filter

Another method that is used in statistical arbitrage is using the Kalman Filter. Recall the equation for spread:

$$\varepsilon_{ij,t} = P_{i,t} + \gamma P_{j,t}$$

The Kalman filter is a recursive algorithm for estimating the state of noisy data that needs to be filtered to be able to estimate the state of a system based on a sequence of observations, taking into account both current measurements and the system dynamics [36]. This makes it very useful to estimate the hedge ratio γ . Initially, a book by Vidyamurthy discusses best practices for choosing cointegrated equities and found that the Kalman filter was found optimal when the state-space and observation equations are linear, and the noise is Gaussian [37]. Since then, there have been many extensions of the filter, such as the

Extended Kalman Filter (EKF) and Unscented KF, aimed at handling when the state-space and observation equations are non-linear and the noise is not Gaussian.

The Kalman Filter works in 2 phases, prediction and update. The prediction phase is as follows

$$\begin{aligned}\hat{\mathbf{x}}_k &= \mathbf{F}_k \hat{\mathbf{x}}_{k-1} + \mathbf{B}_k \vec{\mathbf{u}}_k + \mathbf{w}_k \\ \mathbf{P}_k &= \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k\end{aligned}$$

Where $\hat{\mathbf{x}}_k$ is the new best estimate (prediction) that is derived from $\hat{\mathbf{x}}_{k-1}$, the previous estimate and the prediction function \mathbf{F}_k . $\vec{\mathbf{u}}_k$ is the correction term, called the control vector, that is used when it is known that there are external influences in combination with \mathbf{B}_k which is called the control matrix. In addition to this, the new uncertainty (covariance matrix), \mathbf{P}_k , is calculated using the previous uncertainty and additional uncertainty from the environment, \mathbf{Q}_k , \mathbf{w}_k is called the state noise. The update is as follows

$$\begin{aligned}\hat{\mathbf{x}}'_k &= \hat{\mathbf{x}}_k + \mathbf{K}'(\vec{\mathbf{z}}_k - \mathbf{H}_k \hat{\mathbf{x}}_k) \\ \mathbf{P}'_k &= \mathbf{P}_k - \mathbf{K}' \mathbf{H}_k \mathbf{P}_k \\ \mathbf{K}' &= \mathbf{P}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \mathbf{R}_k)^{-1}\end{aligned}$$

Where \mathbf{K}' is defined as the Kalman gain, \mathbf{H}_k is the measurement matrix, $\vec{\mathbf{z}}_k$ is mean of the observed values, which is also calculated by $\vec{\mathbf{z}}_k = \mathbf{H}_k \hat{\mathbf{x}}_k + \mathbf{v}_k$ where \mathbf{v}_k is the measurement noise, and \mathbf{R}_k is the covariance of the uncertainty of the observed values [38].

A paper investigating the use of the Kalman Filter on ETFs found that its employed strategy worked well for in-sample data points and worse, but still profitable, results for out-of-sample data [39]. The paper adapted the Kalman Filter to be able to use it for pairs trading to the following:

$$\begin{aligned}\mathbf{y}_t &= \mathbf{x}_t \beta_t + \epsilon_t \\ \beta_t &= \mathbf{I} \beta_{t-1} + \omega_t\end{aligned}$$

Then calculating the Kalman Gain:

$$\text{Kalman Gain} = \frac{\text{Error in the estimate}}{\text{Error in the estimate} + \text{Error in the measurement}}$$

Then to calculate the estimate:

$$\text{Estimate}_t = \text{Estimate}_{t-1} + \text{Kalman Gain} \times (\text{Measurement} - \text{Estimate}_{t-1})$$

And finally, calculating the new error:

$$\begin{aligned}E_{\text{estimate}_t} &= \frac{E_{\text{measurement}} \times E_{\text{estimate}_{t-1}}}{E_{\text{measurement}} + E_{\text{estimate}_{t-1}}} \\ E_{\text{estimate}_t} &= E_{\text{estimate}_{t-1}} \times (1 - \text{Kalman Gain})\end{aligned}$$

The author states, “pairs trading strategies have gained widespread acceptance, thus making profitability much more elusive” to justify the disappointing results. However, the author fails to find evidence or provide sufficient evidence to justify the claim [39].

Another paper used the combination of the Kalman Filter and Machine Learning, more specifically, Extreme Learning Machine and Support Vector Regression (SVR), to build a statistical arbitrage strategy on the Brazilian Stock Exchange [40]. The strategies can simply be explained as using SVR and ELM to forecast returns and the Kalman Filter to improve the forecast.

The paper also compares methods, such as LASSO, BMA, and GRR, to benchmark the

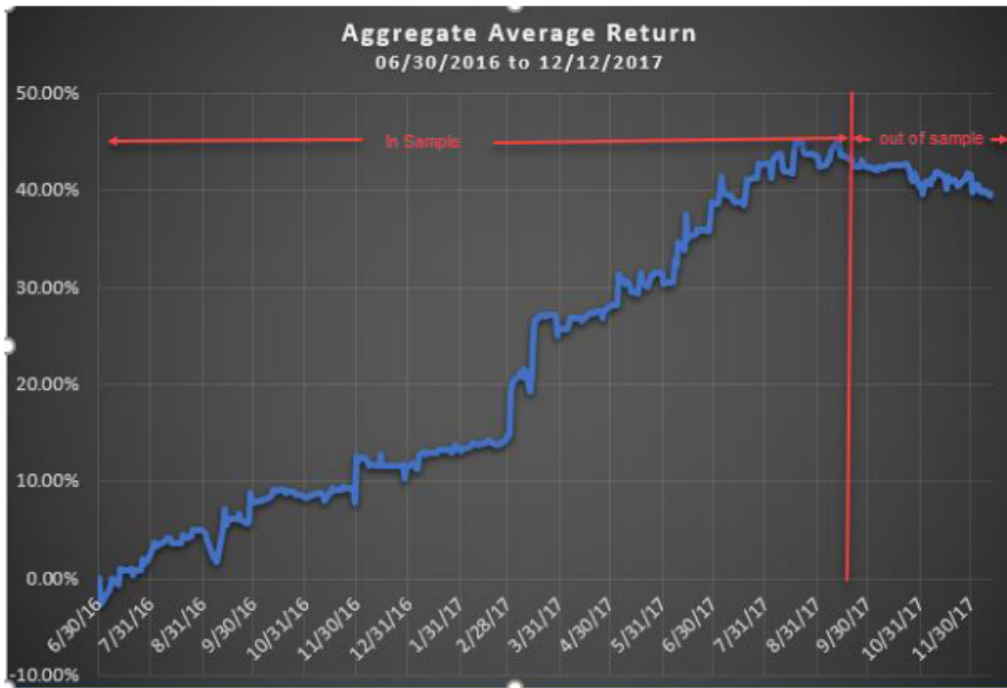


Figure 2.5: Aggregate average return of using the Kalman filter for pairs trading on ETFs [39]

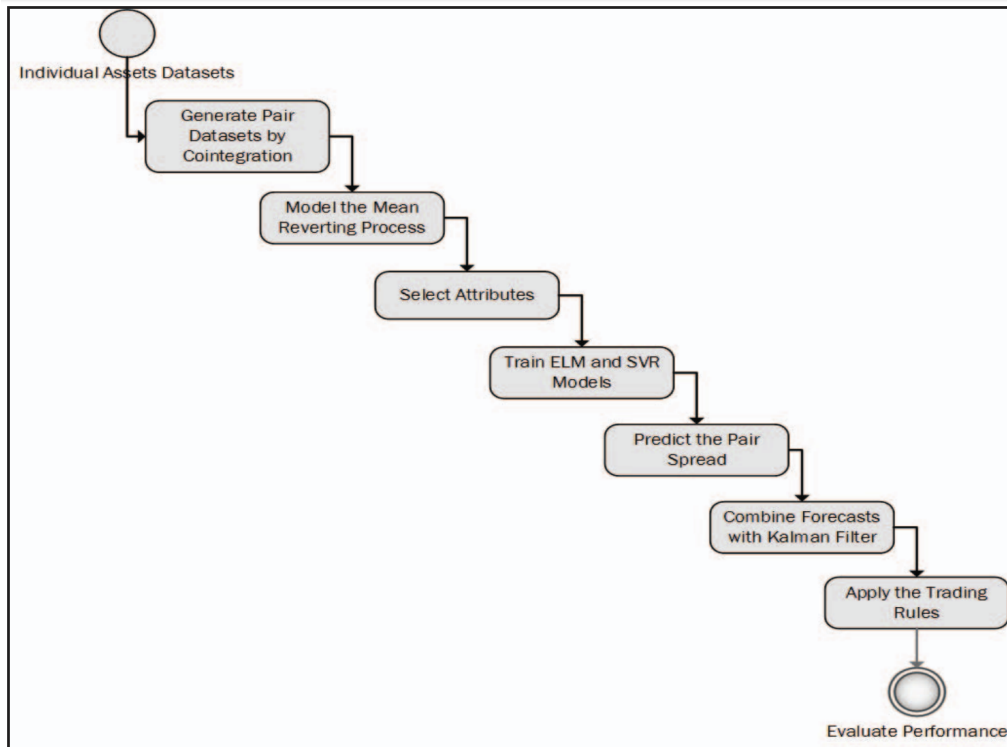


Figure 2.6: Visualisation of the trading strategy used in [40]

performance of the Kalman Filter. The research found that using the ELM and SVR forecasts results in a return of 20.19% and 21.32%, respectively, for out-of-sample data points, and using a combination with the Kalman Filter gives a return of 26.13% for out-of-sample data points. Figure 2.7 shows the full results below. In addition, it can be seen that the return volatility also decreases, which is ideal for investment managers.

TABLE IV. ECONOMETRIC PERFORMANCE – ELM AND SVR – IN-SAMPLE

MODEL	MAX. DD	SHARPE	VOLATILITY	RETURN
ELM	-2.31%	1.80	3.05%	5.83%
SVR	-2.18%	1.73	3.20%	5.39%

TABLE V. ECONOMETRIC PERFORMANCE ELM AND SVR – OUT-OF-SAMPLE

MODEL	MAX. DD	SHARPE	VOLATILITY	RETURN
ELM	-2.80%	3.83	5.64%	20.18%
SVR	-2.72%	4.31	5.29%	21.32%

TABLE VI. ECONOMETRIC PERFORMANCE – COMBINATION MODELS – IN-SAMPLE

MODEL	MAX. DD	SHARPE	VOLATILITY	RETURN
BMA	-2.24%	1.38	3.05%	4.94%
GRR	-2.34%	2.07	3.20%	6.37%
KALMAN	-2.30%	1.97	3.57%	6.89%
LASSO	-2.43%	2.06	3.39%	6.30%

TABLE VII. ECONOMETRIC PERFORMANCE – COMBINATION MODELS – OUT-OF-SAMPLE

MODEL	MAX. DD	SHARPE	VOLATILITY	RETURN
BMA	-2.97%	3.83	5.12%	19.33%
GRR	-2.53%	4.76	5.49%	23.69%
KALMAN	-2.64%	5.29	5.17%	26.13%
LASSO	-2.64%	4.76	5.49%	23.79%

Figure 2.7: Econometric results [40]

Other papers/articles such as [41, 42, 43] have designed, compared and analysed other statistical arbitrage techniques using Machine Learning algorithms and revealed that some algorithms are profitable. Most research on machine learning trading strategies has been on assets like stocks on centralised exchanges. The little research that has been done on statistical arbitrage on cryptocurrencies has all been on analysing arbitrage on centralised exchanges and not decentralised exchanges. One of the research projects that analysed machine learning methods of statistical arbitrage on cryptocurrencies on a centralised exchange compared a logistic regression approach with a random forest approach [43].

2.4.3 Analysis on Cryptocurrency Arbitrage on Centralized Exchanges

Although the research in the previously mentioned papers does not investigate the cointegration approach to cryptocurrencies, the takeaways are the mathematical fundamentals used in statistical arbitrage. Kristoufek and Bourri researched the sources of statistical arbitrage of bitcoin in multiple centralised exchanges. The Grey correlation is built on top of the Grey system theory [44] and can capture non-linear correlations without assuming a Gaussian distribution; thus, using the Grey correlation provides a more robust metric to understand correlations between both series. The Grey correlation $\gamma(X_0, X_i)$ is defined with two steps:

1. $\gamma(x_0(k), x_i(k)) = \frac{\min_k \min_k |x_0(k) - x_i(k)| + \varepsilon \max_k \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \varepsilon \max_k \max_k |x_0(k) - x_i(k)|}$
2. $\gamma(X_0, X_i) = \frac{1}{n} \sum_{i=1}^n \gamma(x_0(k), x_i(k))$

With $\varepsilon \in [0, 1]$, the standard is set to $\varepsilon = 0.5$.

The DCC-GARCH(1,1) model is also used to obtain conditional correlations for Bitcoin exchanges [45]. The model was designed to use a combination of parameters such as the standard deviation of Bitcoin returns, traded volume, the volume of on-chain transactions, fees paid to miners, the ratio of current price and recent price history and internet hype/trends.

Upon analysis of Grey and DCC-GARCH(1,1) correlations, it is found that the DCC correlations show little variability. In contrast, the Grey correlations are a lot more variable, ranging from 0.29 to 1. In addition, the paper further investigates these sources and finds that these opportunities are introduced when there is a large number of inter-exchange transfer requests, i.e., the network is congested and has high price volatility. In contrast,

the high volume of exchanges and on-chain activity cause the arbitrage opportunities to decrease. This paper finds and explains these sources of statistical arbitrage however does not implement or devise an algorithm that uses statistical arbitrage to generate a profit from price discrepancies of Bitcoin on different exchanges.

A paper investigating statistical arbitrage on multiple cryptocurrencies is [46]. The authors of this paper analysed co-movements and cointegration of different cryptocurrencies on a centralised exchange using Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS), Ljung-Box autocorrelation tests on both stationary forms ($I(0)$) and the original form ($I(1)$). The paper then develops a dynamic factor model based on the assumption that the price dynamics of cryptocurrencies are driven by Bitcoin [47]; this is then evidenced by similar paths found in cryptocurrencies shown in Figure 2.8.

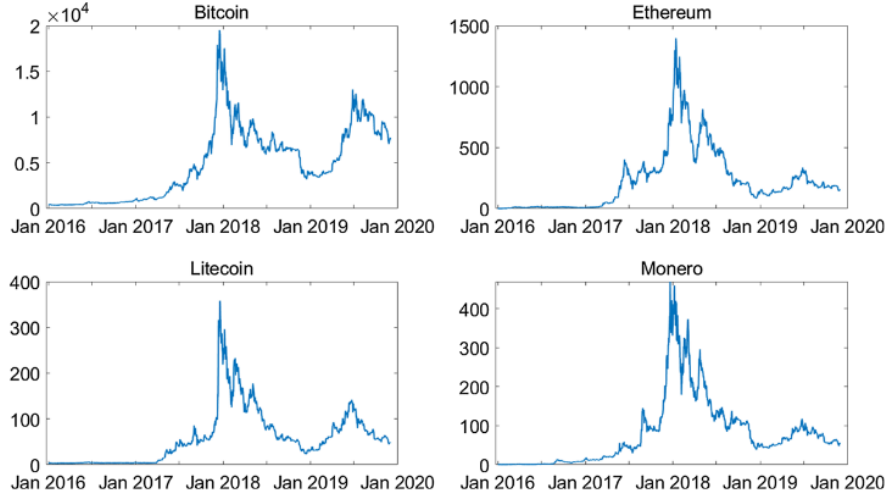


Figure 2.8: Price behaviour of Bitcoin, Ethereum, Litecoin, Monero [46]

For simplicity, the authors set the number of hidden factors to 2, and upon analysis, f_1 is a $I(1)$ process, and the second factor f_2 is a stationary process that is independent of f_1 . After overlaying f_1 with the price of Bitcoin, it is also found that the first factor strongly correlates with the price of Bitcoin.

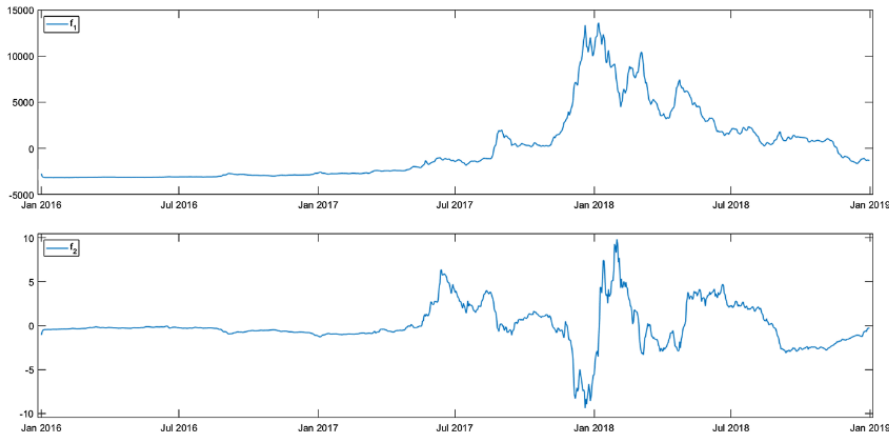


Figure 2.9: Hidden factors f_1 and f_2 from Jan 2016 to Dec 2018 [46]

The paper then uses this model to build an investment strategy, using forecasting using the estimated parameters:

$$\hat{p}_{i,\tau+1} = \mathbb{E}_\tau(p_{i,\tau+1}) = \hat{\alpha}_i + \hat{\beta}_{i1}\mathbb{E}_\tau(f_{1,\tau+1}) + \hat{\beta}_{i2}\mathbb{E}_\tau(f_{2,\tau+1})$$

Where

$$f_{1,t} = \lambda_1 f_{1,t-1} + \eta_{1,t}$$

$$f_{2,t} = \lambda_2 f_{2,t-1} + \eta_{2,t}$$

The expected gains one day ahead are given by:

$$g_{\tau+1} = \mathbb{E}_{\tau}[v_{\tau+1}] = \sum_{i=1}^{\lfloor I/2 \rfloor} \hat{p}_{\tau+1}^{(i)} - \sum_{i=\lfloor I/2 \rfloor + 1}^I \hat{p}_{\tau+1}^{(i)}$$

Using this and a threshold which is calculated by the combination of the current price and standard deviation of the trading position value:

- if $g_{\tau+1} > v_{\tau} + c\sigma_{\tau}^v$, go long
- if $g_{\tau+1} < v_{\tau} - c\sigma_{\tau}^v$, go short
- if $v_{\tau} - c\sigma_{\tau}^v \leq g_{\tau+1} \leq v_{\tau} + c\sigma_{\tau}^v$, no trade

The paper's researchers evaluated their trading strategy for 334 days and a moving window of 3 years, 1096 observations, every day to estimate the parameters for the dynamic factor model. We can see in Figure 2.10 that the strategy consistently generated a profit even when considering transaction costs.

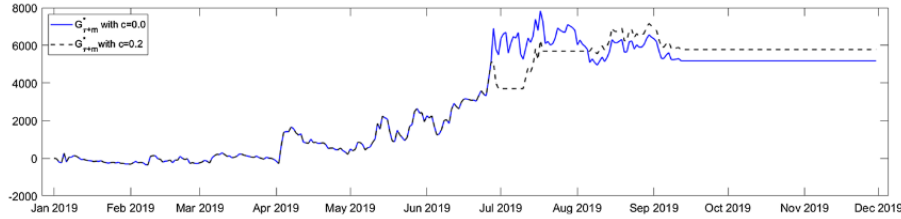


Figure 2.10: Net gains taking transaction fees into account [46]

Overall, although countless studies have explored statistical analysis of arbitrage opportunities in centralised exchanges, little attention has been given to investigating such opportunities in decentralised exchanges. Profitable methods like a mean reversion, mean reversion utilising the Kalman filter, and analysing cointegrated cryptocurrencies have been identified in centralised exchange settings. Consequently, it would be intriguing to explore whether these findings hold when applied to decentralised exchanges.

Chapter 3

Design Decisions

3.1 Mean Reversion Strategies

Considering the current state of the art and the limited exploration of statistical arbitrage in decentralised exchanges, the objective of this project is to examine the profitability of mean reversion techniques in pairs trading and determine if they offer a viable investment opportunity. Additionally, by employing various methods to estimate the hedge ratio and effectively manage risk, we can further analyse the overall performance and appeal of the trading strategy to potential traders.

In the mean reversion strategies, the hedge ratio refers to the ratio or proportion between the positions of two assets involved in a pairs trading strategy. It determines the optimal allocation of capital between the assets to minimise risk and create a market-neutral position. The hedge ratio represents the number of units of one asset that should be bought for each unit of the other asset to be shorted in order to create a balanced or hedged position. It is derived through statistical techniques such as regression analysis. The methods that are explored in this paper are the use of a Constant Hedge Ratio, Sliding Window Ordinary Least Squares, Lagged Ordinary Least Squares, the unrestricted OLS model returned by the Granger Causality Test and also the Kalman Filter. These methods are selected due to their ability to test whether underlying dynamics affect the relationships between liquidity pools, hence affecting the hedge ratio. To further see the details and implementation of each strategy, refer to [Section 5](#).

3.2 Buying and Short Selling

The process of buying and short selling (often called selling) in traditional markets is relatively straightforward, as brokers and exchanges play a crucial role in executing orders on behalf of individuals. However, when it comes to trading cryptocurrencies, the responsibility falls directly on the trader. Therefore, it becomes essential to clearly define what buying and selling entail in the context of cryptocurrency trading.

3.2.1 Buying

Prices on decentralised exchanges are represented as ratios, for example, $1 \text{ USDC} = P \text{ WETH}$. With this understanding, let us delve into the process of buying one unit of USDC/WETH on a decentralised exchange:

- Opening a Buy position:
 - Starts with P_0 WETH
 - Swaps the WETH for USDC, hence ends with 1 USDC
- Closing a Buy position:

- Starts with 1 USDC
- Swaps the USDC for WETH, hence ends with P_1 WETH

Consequently, if the price of USDC/WETH increases from the moment the buy position is opened to the time it is closed, the trader realises a profit. On the other hand, if the price declines during this period, the trader incurs a loss.

3.2.2 Short Selling

Short selling an asset is a more complex process compared to buying because it involves trading an asset the trader does not initially possess. In traditional markets, this consists of the trader borrowing the desired asset from a broker or another party and selling it. When the trader decides to close the position, they repurchase the same amount of the borrowed asset and return it to the lender, hoping that its value has declined. In traditional settings, brokers and exchanges typically automate the borrowing and returning processes. However, in the context of decentralised exchanges (DEXes), such mechanisms are absent or not readily available.

To carry out the short-selling process, the utilisation of a lending protocol to borrow any required assets is employed. By leveraging a lending protocol, short-selling becomes easily accessible. The process of selling is as follows:

- Opening a Sell position:
 - Borrow 1 USDC
 - Swaps the USDC for WETH, hence ends with P_0 WETH
- Closing a Sell position:
 - Starts with P_0 WETH
 - Swaps the WETH for USDC, hence ends with $\frac{P_0}{P_1}$ USDC
 - Return the borrowed USDC, leaving $\frac{P_0}{P_1} - 1$ USDC

Consequently, if the price of USDC/WETH decreases, i.e. $P_1 < P_0$ from the moment the sell position is opened to the time it is closed, the trader realises a profit. On the other hand, if the price increases during this period, the trader incurs a loss.

Note that this is an illustrative example and does not account for any potential fees that the trader might incur.

3.3 Protocols of Interest

For this project, Uniswap and Aave protocols were selected as the DEX and lending protocol of choice. Uniswap is selected as it possesses the highest share of volume being swapped, around 60%, among all of the DEXes [48]. Also, Aave was used as it is one of the most popular DeFi lending protocols for the wide range of supported tokens [49].

3.3.1 Uniswap

Uniswap is a decentralised exchange protocol built on the Ethereum blockchain. It allows users to trade ERC-20 tokens directly from their wallets without the need for intermediaries or traditional order books. Uniswap utilises automated market-making, where liquidity providers contribute funds to liquidity pools, earning fees on trades made in the pool. The protocol employs a mathematical formula called the constant product formula to maintain balanced token ratios in the pool. When users want to trade, Uniswap calculates the conversion based on the pool's token ratios and executes the trade through a smart contract.

Automated Market Maker (AMM) Model and Liquidity Pools

The AMM model is a system that replaces traditional order books with liquidity pools to facilitate trading between different tokens. Automated Market Makers do this with the aid of liquidity pools. They are pools of tokens contributed by liquidity providers (LPs) to facilitate trading between tokens within the exchange. One paper describes pools as “a smart contract that holds at least two cryptoassets and allows trading through depositing a token of one type and thereby withdrawing tokens of the other type” [50]. These liquidity pools are maintained by smart contracts on Ethereum and Liquidity Providers (LPs). Liquidity providers are individuals that voluntarily contribute an equal amount of cryptocurrencies to liquidity pools; for example, in a pool for trading ETH and DAI, LPs would contribute an equal value of ETH and DAI tokens. LPs are incentivised to provide liquidity in exchange for a share of the trading fees in the liquidity pool. LPs can later withdraw their shares along with the accumulated fees. On Uniswap V3, the fee varies depending on the set-up of the liquidity pool; currently, the fees can be any of the following: 0.01%, 0.05%, 0.3%, or 1%.

Constant Product Formula

The AMM model relies on a mathematical formula called the constant product formula ($xy = k$) to maintain a balanced ratio between the tokens in the pool and calculate the exchange rate. As trades occur, the product of the token balances remains constant. When one token’s proportion in the pool decreases, its value increases, ensuring an automatic price adjustment. When a trade is executed, the price change can be described in this formula $(x + \Delta x)(y + \Delta y) = k$. Hence, after rearranging: $\Delta y = \frac{k}{x + \Delta x} - y$. Under this model, the balance of the tokens in a liquidity pool can never be depleted as the token will get infinitely more expensive as the reserves approach 0.

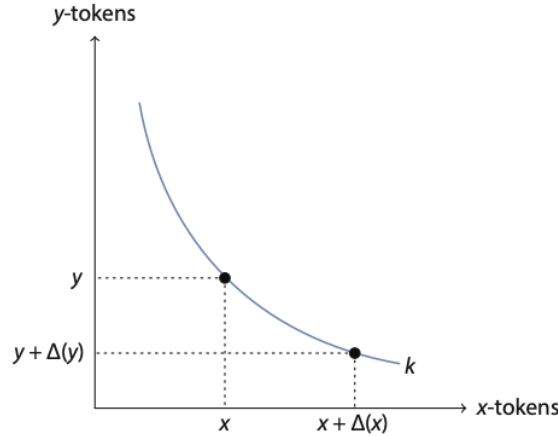


Figure 3.1: Constant Product Formula [50]

Slippage

As seen in Figure 3.2, we can see how the constant product formula is used in Uniswap. Uniswap allows users to set slippage tolerance levels, determining the maximum acceptable difference between the expected and executed prices. Slippage refers to the difference between the expected price of a trade and the executed price due to market volatility and liquidity conditions. Slippage happens because the constant product formula adjusts prices based on the ratio of tokens in the pool. As trades are executed, token balances and prices change accordingly. Thus, larger trades can cause a more significant price impact, resulting in slippage.



Figure 3.2: How Uniswap works [51]

3.3.2 Aave

Aave is a decentralised lending and borrowing protocol built on the Ethereum blockchain. It enables users to lend and borrow a wide range of cryptocurrencies directly without the need for intermediaries such as banks. Aave operates through liquidity pools and smart contracts, providing a secure, transparent, and efficient platform for lending and borrowing.

Users can deposit their cryptocurrency assets into Aave's liquidity pools and earn interest on their deposits. These funds contribute to the available liquidity for borrowers. On the other hand, borrowers can use their deposited assets as collateral to borrow other assets from the pool. The amount they can borrow is determined by the value of their collateral and specific borrowing parameters set by the protocol.

Liquidity Pools

The fundamental mechanism enabling Aave's lending and borrowing functionality is liquidity pools. Users deposit their cryptocurrency assets into Aave's liquidity pools. These assets serve as collateral and contribute to the overall liquidity of the protocol. The deposited assets in the liquidity pools create reserves of available liquidity. These reserves are utilised to fulfil borrowing demands from other users within the Aave ecosystem.

Lending and Borrowing

Lending works by lenders depositing their cryptocurrency assets into Aave's liquidity pools. These assets act as collateral and are represented by interest-bearing tokens called aTokens. The aTokens represent the user's share of the deposited assets. Interest is earned immediately and is accrued in real-time.

The borrowing process is enabled by borrowers using their deposited assets as collateral to borrow other assets from the liquidity pools. The value of the collateral determines the borrowing capacity. The borrowing capacity is calculated by parameters set by Aave, one of which is the maximum loan-to-value (LTV) ratio. Once the borrower requests a loan, if the borrower's collateral meets the necessary requirements, they can proceed with the loan and transfer the borrowed funds to the borrower's wallet. In addition, borrowers pay interest on the borrowed amount based on the prevailing interest rates. Aave offers both variable and stable interest rates for borrowers. Variable rates fluctuate based on market dynamics, while stable rates remain fixed. This flexibility allows users to choose the borrowing option that best suits their needs. The interest on a loan is accrued in real-time, second by second, and the borrower decides when to repay it, as long as the loan is safe from liquidation. Liquidation happens if the value of a borrower's collateral falls

below the liquidation threshold due to market volatility; the collateral can be liquidated. Liquidators can purchase the collateral at a discounted price to ensure the solvency of the liquidity pool.

3.4 Fees that are incurred when trading

As previously mentioned, when engaging in trading activities, there are several fees that are deducted. These fees serve multiple purposes, including compensating the liquidity providers and addressing the potential lack of access to assets from the lender. Therefore, in this section, we will delve into the various costs associated with implementing and executing the strategy.

3.4.1 Gas Fees

Gas fees in Ethereum are a crucial component of the network's operation. They serve two primary purposes: to prevent spam and denial-of-service attacks and to incentivise miners to include transactions in the blockchain. Each operation, such as sending tokens, executing a smart contract function, or interacting with decentralised applications, consumes a certain amount of gas.

The fee is calculated by multiplying the gas price (expressed in Gwei, where 1 Gwei is equal to 0.000000001 ETH) by the gas used [19]. As mentioned in previous sections, when a user initiates a transaction, the user specifies the gas limit, which defines the maximum amount of gas that a user is willing to pay for a transaction; if the gas limit is too low and the gas used is more significant than its limit, the transaction will fail and be reverted. This means that the state changes made by the transaction will not be recorded on Ethereum, and the gas fees paid for that transaction will be lost.

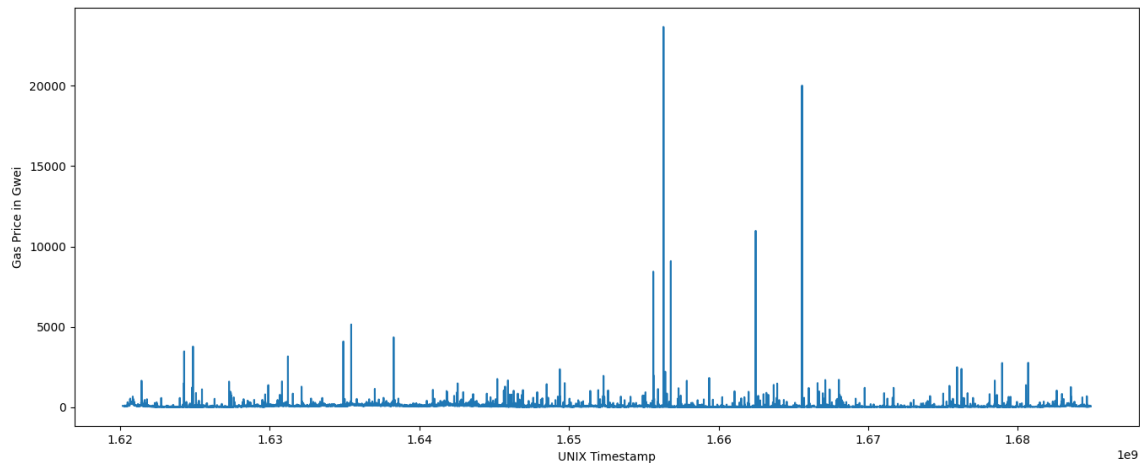


Figure 3.3: The gas price over time

Figure 3.3 displays the gas price history over time, showing the price surges and drops. The market forces of supply and demand determine the reason for these fluctuations. It is the amount of Ether a user is willing to pay for each unit of gas consumed in a transaction. Miners can prioritise transactions with higher gas prices and thus are incentivised to include transactions with higher gas prices because they receive the gas fees as a reward for their mining efforts. Therefore, setting a higher gas price increases the chances of a transaction being executed.

The second part of the gas fee is how much gas is used; it is calculated by adding the gas used for each operation in the transaction. Each operation has a predefined gas cost

associated with it, which reflects the complexity and resource requirements of that operation. When a transaction or smart contract execution is initiated, the Ethereum Virtual Machine (EVM) processes each operation and deducts the corresponding gas cost from the gas limit. In addition to this, Ethereum also has a base fee with all of their transactions; thus, having multiple small transactions becomes more costly than using a large combined transaction [19].

3.4.2 Uniswap Fees

In addition to transaction fees, there is another fee mechanism in place within the system when swapping on Uniswap. This fee is designed to incentivise and reward liquidity providers on the platform. When a swap occurs, a fee is deducted from the tokens expected to be returned to the user.

The fee is calculated as a percentage of the swap volume, meaning that the larger the swap, the higher the fee. For example, let us consider a scenario where the exchange price is 1 TKNA = 100 TKNB, and the fee is set at 0.3%. If someone exchanges 1 TKNA for TKNB, instead of receiving the full 100 TKNB based on the exchange rate, they will receive $100 \times (1 - 0.3\%)$ TKNB = 99.7 TKNB. The fee reduces the total amount of tokens received in the swap.

It is important to note that the specific fee percentage may vary between different liquidity pools on Uniswap. Each liquidity pool can have its own fee tier, and the fees associated with each interested liquidity pool can be seen in Table 3.1.

Pool Address	Token0	Token1	Fee Tier as a %
0x88e6a0c2ddd26feeb64f039a2c41296fcb3f5640	USDC	WETH	0.05
0x8ad599c3a0ff1de082011efddc58f1908eb6e6d8	USDC	WETH	0.30
0x11b815efb8f581194ae79006d24e0d814b7697f6	WETH	USDT	0.05
0x4e68ccd3e89f51c3074ca5072bbac773960dfa36	WETH	USDT	0.30
0x60594a405d53811d3bc4766596efd80fd545a270	DAI	WETH	0.05
0xc2e9f25be6257c210d7adf0d4cd6e3e881ba25f8	DAI	WETH	0.30
0xe0554a476a092703abdb3ef35c80e0d76d32939f	USDC	WETH	0.01
0xc5af84701f98fa483ece78af83f11b6c38aca71d	WETH	USDT	0.1

Table 3.1: Uniswap fees for each liquidity pool of interest

3.4.3 Aave Fees & Collateral

Aave facilitates lending and borrowing transactions among users, and as part of its operations and incentive mechanisms, it imposes certain fees. In the context of this trading strategy, the only fees applicable would be the interest rates on the loan. The interest rate is typically expressed as an Annual Percentage Yield (APY) and is accrued continuously. Currently, Aave supports variable interest rates, which fluctuate based on market conditions, the borrowed asset, and supply and demand dynamics within Aave.

Another aspect of Aave for the trading strategy is collateralisation and the avoidance of liquidation. Liquidation occurs when a borrower's position is forcibly closed, and their collateral assets are sold off in cases of default or insufficient collateral. When borrowing from Aave, borrowers are required to allocate a certain percentage of the loan value as collateral, known as the Loan-to-Value (LTV) ratio [52]. For instance, if a token has a 75% LTV, borrowers can borrow 0.75 ETH worth of the corresponding token for every 1 ETH worth of collateral provided. However, as token prices fluctuate, the ratio between the borrowed token value and the collateral value changes, posing risks for both lenders and Aave. To safeguard lenders and maintain protocol solvency, a liquidation threshold is established. If

the collateral value falls below this threshold, the borrower’s position becomes vulnerable to liquidation. In such cases, the collateral is auctioned off to repay the outstanding debt to the lenders. Typically, the liquidation threshold is set 5-10% higher than the asset’s LTV. For WETH, the Loan-to-Value ratio is 82.5%, and the liquidation threshold is 86%.

3.5 Liquidity Pool Pairs of Interest

One of the remarkable features of Uniswap is its support for even the smallest cryptocurrencies, currently possessing 12,182 liquidity pools.

Upon inspection, it can be seen that some pools exhibit high trading volumes, indicating significant activity and demand for those specific pairs. On the other hand, there are also pools with zero liquidity, which could be attributed to various reasons. It could be due to a lack of demand for one of the tokens offered in the pool, or it may indicate that the pool is relatively new and has yet to attract significant participation. Table B.1 in the Appendix showcases the diversity of these pools.

The mean reversion strategies are centred on certain assertions of the liquidity pools that are being traded on in order to enhance the chances of successful swaps and maximise potential profitability; only liquidity pools with a trading volume surpassing \$10 million and pools that include tokens supported by Aave are used. The tokens Aave supports are DAI, EURS, USDC, USDT, AAVE, LINK, and WBTC. Additionally, any liquidity pools not involving WETH, a tokenised form of ETH (Ethereum’s native cryptocurrency), are excluded. WETH’s key advantage is that it can easily be used to convert back into ETH in case the balance of ETH falls too low.

3.5.1 Correlated and Cointegrated Liquidity Pools

Once the liquidity pools of interest have been identified, the pool pairs are filtered based on their correlation, as correlated pairs tend to exhibit similar pricing movements; thus would be possible to apply the mean reversion trading strategies. However, it is important to strike a balance between a highly correlated pair and a low correlated pair, as an excessively high correlation can result in minimal price deviations, leading to fees that exceed the deviations and resulting in overall losses instead of profits. To address this, a condition is set such that the correlation coefficient (ρ) should fall within the range of $0.992 \leq \rho \leq 0.997$. Figure 3.4 shows the correlation matrix of the liquidity pools that meet the initial requirement of having a volume greater than \$10 million volume ratio, include WETH and other Aave-supported cryptocurrencies. To determine which liquidity pools are cointegrated, the Engle-Granger approach is employed. The Engle-Granger test for cointegration involves several steps. Firstly, a unit root test is performed individually on each time series using methods like the Augmented Dickey-Fuller (ADF) test. These tests assess whether the time series is stationary or exhibits unit roots (non-stationary) individually. For cointegration to hold, both time series must be non-stationary.

Once it is established that both time series are non-stationary, the cointegration equation is estimated. Typically, a linear regression model captures the long-term relationship between the time series variables. This equation explains how the variables are linked over a long period.

Subsequently, a unit root test is conducted on the residuals obtained from the regression analysis. If the residuals are stationary, it is sufficient to conclude that the time series are cointegrated, implying that the variables move together in the long run, even if short-term deviations occur.

Following the sequence mentioned earlier in the Engle-Granger test enables us to determine

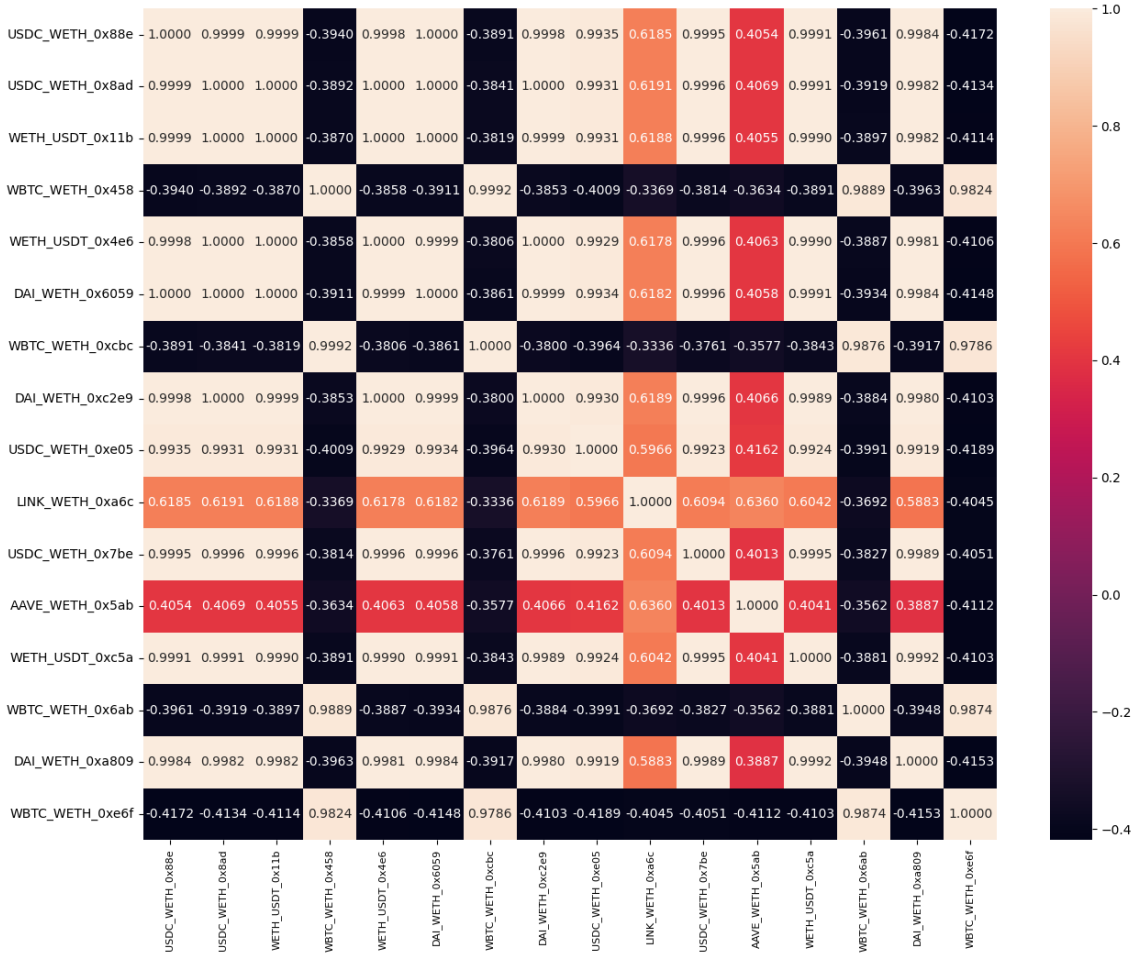


Figure 3.4: Correlation Matrix of the Liquidity Pools that meet the \$10 million volume and include WETH

which liquidity pools demonstrate cointegration, emphasising their interconnectedness and mutual relationship over time. The outcome of the cointegration tests for all possible combinations of liquidity pools is presented in Table 3.2. This table provides a comprehensive list of the liquidity pool pairs, the results of their respective cointegration tests, and the correlation coefficient of each combination. By concentrating on the cointegrated pairs,

pool1	pool2	t-statistic of unit-root test	Critical Values			Corr Coeff
			1%	5%	10%	
USDC_WETH_0x88e6...	USDC_WETH_0xe055...	-11.306176	-3.898069	-3.337038	-3.045081	0.99347
USDC_WETH_0x8ad...	USDC_WETH_0xe055...	-11.210146	-3.898082	-3.337046	-3.045086	0.99307
WETH_USDT_0x11b...	USDC_WETH_0xe055...	-10.010746	-3.898069	-3.337038	-3.045081	0.99315
WETH_USDT_0x4e68...	USDC_WETH_0xe055...	-9.913205	-3.898081	-3.337045	-3.045085	0.99291
DAI_WETH_0x60594...	USDC_WETH_0xe055...	-11.395276	-3.898069	-3.337038	-3.045081	0.99338
DAI_WETH_0xc2e9f2...	USDC_WETH_0xe055...	-10.418722	-3.89831	-3.337173	-3.045174	0.99298
USDC_WETH_0xe055...	WETH_USDT_0xc5af...	-9.7789	-3.901414	-3.338902	-3.046374	0.99243

Table 3.2: Cointegration test on Liquidity Pool pairs

we can delve deeper into exploring their dynamics, assessing their behaviour, and leveraging the interconnectedness between these liquidity pools. This approach enables us to prioritise our analysis and concentrate on the most relevant liquidity pool combinations that offer potential opportunities for profitable trading or other related activities.

Chapter 4

Implementation of the Trading Systems

This section provides a comprehensive overview of the two trading system implementations. The first focuses on the backtesting system, which simulates and evaluates strategies in a controlled environment. In this system, historical market data is utilised to test and analyse the performance of various strategies. The second implementation is the live system, designed to execute trades on the Ethereum blockchain using smart contracts. This system operates in real time and interacts with the live market, allowing actual trades to be executed based on predefined strategies. Together, these two implementations provide a comprehensive framework for testing, evaluating, and deploying trading strategies in simulated environments and on the live blockchain.

4.1 Backtesting System

To develop a resilient backtesting system, a dedicated class was constructed to streamline the process of testing trading strategies; the system is entirely written in Python as it possesses countless libraries, including Pandas and Numpy, which are extensively used for data manipulation and analysis, it also is good with integrating with many data sources such as GraphQL which is required for this project. To evaluate a particular strategy, the `backtest` function requires `cointegrated_pair`, a tuple of the names of the liquidity pools that the strategy is to be evaluated on, `strategy`, the strategy, and finally, the `initial_investment` in ETH. The first argument is required and used to retrieve the relevant historical prices. The second parameter is self-explanatory, the strategy to test. Finally, including the initial investment amount as an input parameter enables users to simulate the performance of their strategies with a specific starting capital; it also helps analyse how the various transaction costs affect the ability to trade if any trades result in a loss. The `backtest` function iterates through historical data calling the strategy's `generate_signal` and executing the trade orders it receives at each timestep. However, historical data is required; hence, the first step is to collect the historical data. The logical process of backtesting a pair can be seen in Figure [4.1](#).

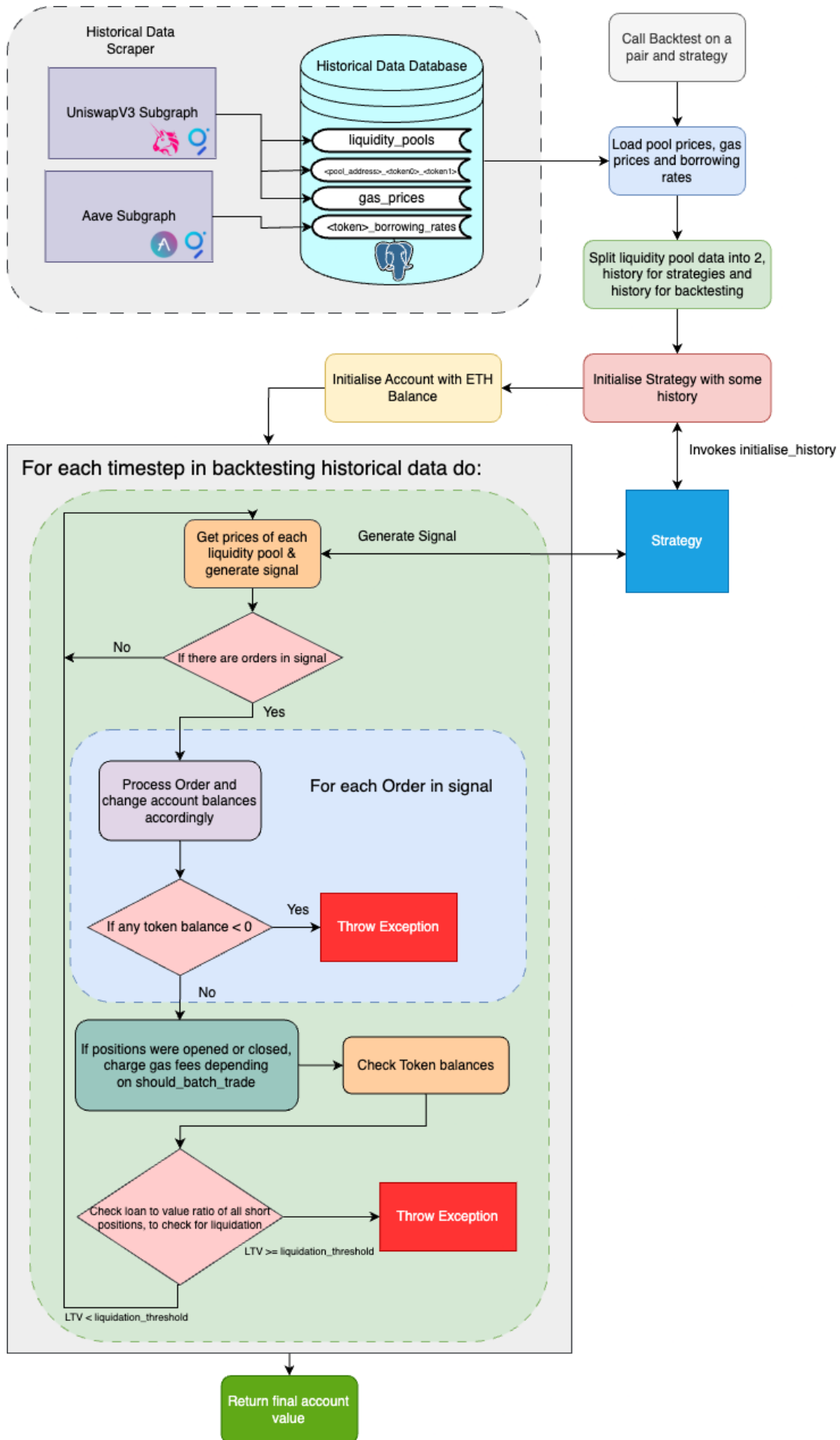


Figure 4.1: Flowchart of backtesting a strategy on a liquidity pool pair

4.1.1 Data Collection and Storage

In order to simulate the market as accurately as possible, the system should have access to reliable and accurate historical market data, including price, volume, and other relevant indicators. Therefore, all of the data is retrieved from the Uniswap and Aave protocols' subgraph using the Graph [53]. The Graph is a decentralised protocol for indexing and querying blockchain data, making the data provided 100% reliable as it indexes directly on the Ethereum blockchain. Subgraphs are GraphQL APIs designed to facilitate querying and extracting data from the blockchain. These APIs adhere to a specific schema outlined by the protocol, enabling seamless communication between the protocol and the underlying blockchain. Therefore, the Uniswap V3, Aave V2, and Aave V3 subgraphs are used to obtain the data required for the backtesting system.

In order to store the necessary data for backtesting and trading, a PostgreSQL database is used to provide consistency and flexibility, supporting numerous data types and a large set of SQL functions that allow for advanced querying. The database adopts the following schema:

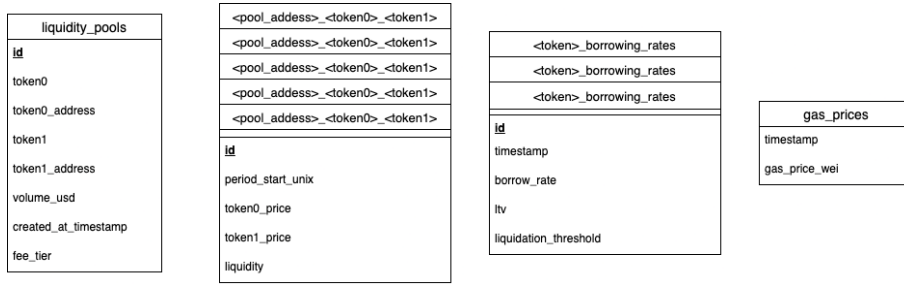


Figure 4.2: Tables contained in the database

The `liquidity_pools` contains data about the liquidity pools on Uniswap V3. After obtaining all of the data, it is found that it possesses 12,182 liquidity pools. However, most of these pools exhibit minimal or negligible trading volume. As a result, a criterion is established to selectively include only those liquidity pools that involve tokens supported by Aave and possess a trading volume exceeding \$10,000,000 (or \$10 million). This filtering condition ensures that the collected and stored data holds significance and relevance for research purposes, as these pools would allow for short selling.

Once these pools have been identified, pricing data about each pool that meets this condition is collected again using the Uniswap V3 subgraph. The query returns an array of dictionaries containing the pre-specified pricing data points at a frequency of every hour. Due to the limitations imposed by the subgraph, the results are constrained to a maximum size of 1000 entries. Hence, the query is executed iteratively to overcome this limitation and obtain the complete dataset. The previous maximum time, referred to as `prev_max_time`, is passed as an argument in subsequent queries to fetch the remaining data, i.e. `prev_max_time = hourlyData[-1]['periodStartUnix'] if len(hourlyData) > 0 else prev_max_time`. This data is stored in tables of the form `<pool_address>_<token0>_<token1>`.

Similarly, obtaining the interest rates for borrowing necessitates the utilisation of two of Aave's subgraphs. This requirement arises due to Aave's migration to Version 3 in March 2022, resulting in a transitional period where Uniswap V3 and Aave V2 were concurrently utilised. The schemas for the two are different; however, for the data we require, the borrowing rate, loan-to-value and liquidity threshold, the schema is consistent, and the same query can be used; this can be seen in Appendix D.1.5. During the table initialisation process, requests are made to the V2 and V3 GraphQL endpoints. However, only the V3 endpoint is utilised for sending requests when updating the table. The data is stored in

tables of the form `<symbol>_borrowing_rates`, where `<symbol>` are all of the tokens that are present in the liquidity pools that are of interest.

The transaction history is queried at each hour to collect the gas price history since Uniswap migrated to V3. This is because querying in the same as the pricing data and borrowing rate history is too exhaustive as transactions occur every second. Therefore, it is more efficient to query at each hour with a window to retrieve the gas price at the closest hour as follows. Algorithm 5 in the Appendix shows the pseudocode gas price at each hour is queried.

4.1.2 Types of Orders and Execution

There are numerous order types that the backtesting supports due to measures to ensure a positive balance and avoid liquidation. The types of orders are; **BUY ETH**, **WITHDRAW**, **DEPOSIT**, **SWAP**, **OPEN BUY**, **CLOSE BUY**, **OPEN SELL**, **CLOSE SELL**. The procedures for the execution of each order type are outlined below. The code of the execution, excluding gas fees, of the order is outlined in Appendix D.2.1.

BUY ETH The **BUY ETH** order is used to swap a percentage of the accounts balance of WETH to ETH. The argument is a float between 0 and 1.

SWAP The **SWAP** order is to execute a simple swap for either token0 or token1 of the liquidity pool. This order is only used as a precautionary order if the balance of the WETH balance gets too low. It takes two arguments, the first indicating whether it is swapping for token0 or token1, and the second parameter is a list of swaps, tuples containing with pool and the quantity that would like to be swapped. The amount received in the account is multiplied by $(1 - \text{swap_fees}[\text{swap_token}])$, this is because Uniswap charges a percentage of the swap specified by `swap_fees[swap_token]`.

OPEN BUY Similar to the **SWAP** order, **OPEN BUY** opens a buy position. As mentioned above, opening a buy order is simply swapping token1 for token0. The parameters of buying are the target token and the volume.

CLOSE BUY Closing a buy position is similar to opening a buy position; however, the swap is in the other direction, i.e. from token to WETH. The only argument is the id of the buy order. To account for Uniswap's fees, the volume received after the initial swap is calculated and then swapped back to WETH.

OPEN SELL Opening a sell position involves borrowing a token and then swapping the token. The parameters of selling are the target token and the volume. It also consists of depositing the required amount to borrow the volume ordered, calculated by the value-to-loan ratio.

CLOSE SELL Closing a sell position is more complicated; it requires swapping back from the token to WETH, repaying the loan with interest and finally returning collateral. Swapping back to the required amount of tokens is the first step. For this, the variable Annual Yield Rates between the timestamp of the opening and the closing position are used to calculate the amount of interest required to be paid, as this is cumulated every second. Once the volume of tokens required to be returned is calculated, the equivalent amount of WETH plus accounting for Uniswap fees is swapped to obtain these tokens. Finally, the tokens are repaid, and the collateral is returned.

WITHDRAW The **WITHDRAW** order's function is to withdraw some collateral stored in Aave. It is not used in the strategies however is implemented in case of developing further strategies that may require this functionality. The only parameter is the amount that would like to be withdrawn.

DEPOSIT The **DEPOSIT** order's function is to deposit some additional collateral stored in Aave. This order is used when the collateral value not properly covering the loan value, hence avoiding liquidation. The only parameter is the amount that would like to be deposited.

4.1.3 Gas Fees

To best approximate the gas fee, it is paramount to have an approximation of how much gas is used in each. Therefore, the smart contract for trading was written and run on a forked Ethereum mainnet to accurately simulate the contracts behaviour; see Subsection 4.2.1 to see the implementation of the smart contract. Using Hardhat, the gas usage of each function was tested, and Figure 4.3 shows the results. It can be seen that some functions have a tighter spread, between the maximum and minimum, compared to others; therefore, to better understand this, plots using various parameters on the different functions provided a deeper insight into the cause of this, Figures 4.4 and 4.5.

Solc version: 0.7.6		Optimizer enabled: true		Runs: 200	Block limit: 30000000 gas	
Methods						
Contract	Method	Min	Max	Avg	# calls	eur (avg)
Swaps	borrowToken	423221	429540	423777	92	-
Swaps	closeBuySellPositions	501246	511946	502811	182	-
Swaps	depositCollateral	211596	251105	214353	54	-
Swaps	openBuySellPositions	523792	566537	526749	184	-
Swaps	repayBorrowedToken	400754	401069	401056	90	-
Swaps	swapEthForWeth	50776	64456	58136	342	-
Swaps	swapExactUsingPool	93702	181561	126108	54	-
Swaps	swapExactUsingRouter	128093	178762	152002	54	-
Swaps	swapWethForEth	61476	61485	61482	54	-
Swaps	withdrawCollateral	181067	198163	181733	52	-
Deployments					% of limit	
Swaps		-	-	1581910	5.3 %	-

Figure 4.3: Results from running smart contract functions

Upon analysing these plots, it becomes apparent that the majority of the functions exhibit a high level of consistency. However, it is noteworthy that for the smallest quantity tested, the gas consumption is significantly higher than that of larger quantities. Nevertheless, once this threshold is surpassed, gas usage remains consistent. Therefore the average gas usage of each operation is used in the backtesting system as it provides a representative estimate of the gas consumption across different quantities.

Furthermore, another interesting feature is the gas usage by swaps and the different methods. The initial method utilises Uniswap's Router, which scans all available liquidity pools to identify the optimal price, whereas the second method involves direct swapping within a specific liquidity pool. Figure 4.4 illustrates that, on average, the router-based approach exhibits higher gas usage than direct swapping using a liquidity pool. However, directly swapping using a liquidity pool exhibits greater volatility in gas consumption but also consistently returns to values above 100,000; because of this, the average gas usage of swapping using a liquidity pool is used.

It can also be seen that combining the operations to open/close buy and sell positions consumes less gas compared to if they were to be executed separately. This also confirms the claims in Wang's paper that batching the operations together is more effective than executing them separately [54]. Thus incorporating this into the backtesting system, each order deducts some ETH, which is calculated by how much gas is used by the order type

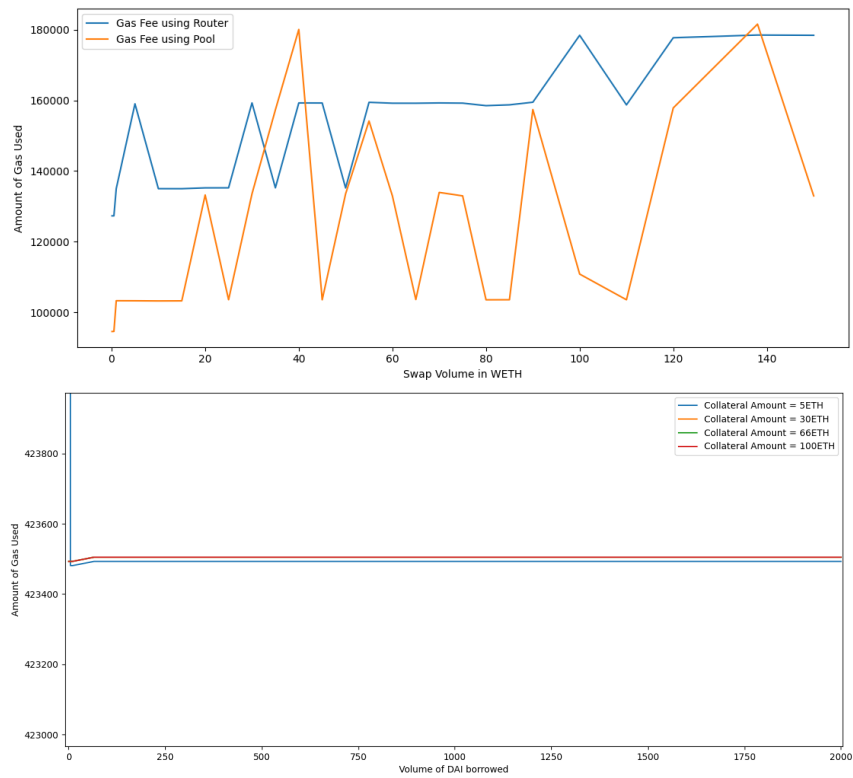


Figure 4.4: Gas Usage Plots of Gas Used by Swapping (Top) and Borrowing (Bottom)

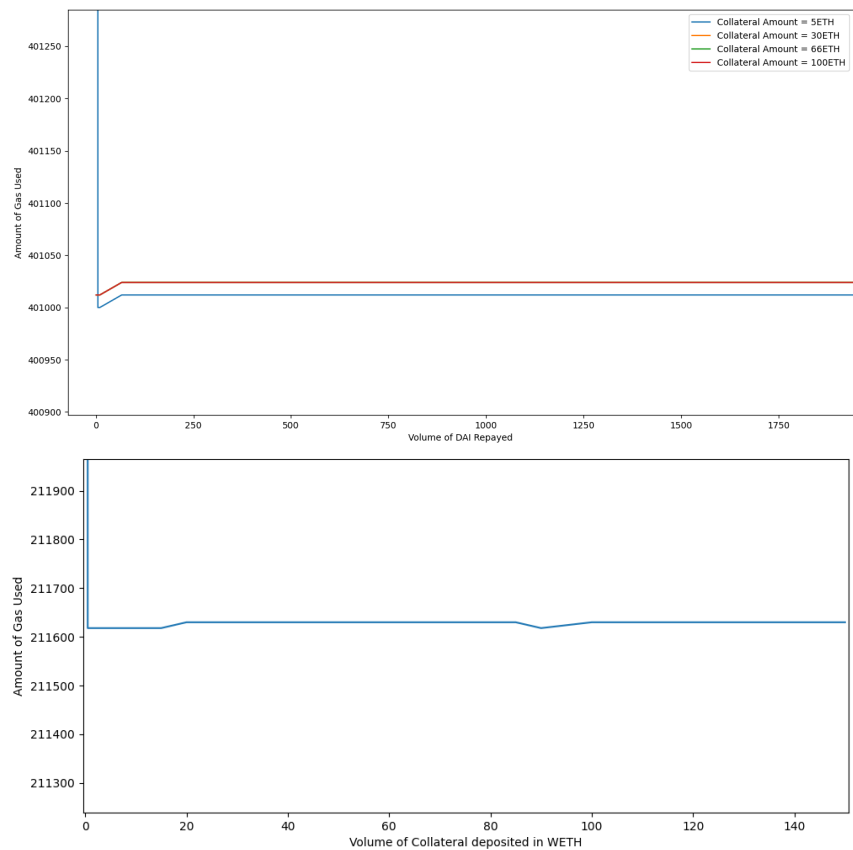


Figure 4.5: Gas Usage Plots of Gas Used by Repaying Borrowed Tokens (Top) and Depositing additional tokens (Bottom)

multiplied by the gas price. For example, in the case of buying more ETH:

```
1 amount_to_swap = self.account['WETH'] * order[1]
```



```

2 self.account['WETH'] = self.account['WETH'] - amount_to_swap
3 self.account['ETH'] = self.account['ETH'] + amount_to_swap - (
    GAS_USED_BY_BUYING_ETH * gas_price_in_eth)

```

However, in order to analyse the impact of combining operations on strategy returns, the deduction of ETH for gas fees for both OPEN and CLOSE orders is handled differently. Instead of deducting the fees immediately upon each order, the deduction is performed after iterating through the trading signal. If the signal contains such orders, the necessary amount of gas fee is deducted based on whether the orders are specified, by the parameter `should_batch_trade`, to be batched or executed separately. The code for handling this can be seen in Appendix D.2.1.

4.1.4 Validating Balance Health

To maintain the integrity and effectiveness of the trading strategy, it is crucial to incorporate various checks and safeguards after each trade and at each timestep. One key aspect involves monitoring the balance of each token to ensure that it remains positive. After every order execution, the tokens' balances are checked, and if any of them fall below zero, an exception is triggered.

```

1 if self.account['T1'] < negative_threshold:
2     raise Exception('Account balace goes below 0 - T1')
3 if self.account['T2'] < negative_threshold:
4     raise Exception('Account balace goes below 0 - T2')
5 if self.account['WETH'] < negative_threshold:
6     raise Exception('Account balace goes below 0 - WETH')
7 if self.account['ETH'] < negative_threshold:
8     raise Exception('Account balace goes below 0 - ETH')

```

Furthermore, the loan-to-value ratio is calculated to simulate the potential liquidation of a sell position's loan. This ratio serves as an indicator of the position's health and risk level. If the loan-to-value ratio breaches the predefined liquidation threshold, an exception is thrown, indicating that the position is approaching an unsustainable state.

```

1 sell_token, sold_price, sell_volume, _ = sell_trade
2 current_token_price = prices[f'P{sell_token[1]}']
3 curr_value_of_loan_pct = (sell_volume * current_token_price) / self.account
    ['collateral_WETH']
4 if round(curr_value_of_loan_pct, 4) > liquidation_threshold:
5     raise Exception(f'Short position liquidated')

```

4.2 Live Trading

A similar approach is required to run the strategies live; however, the execution system and balance tracking are altered. For the system, multiple components are required for live trading to occur. The first is the ability to interact with the blockchain to execute the orders; handled by smart contracts; the second is to manage the state of the strategy, accounts and open positions for the strategies to use to generate signals, and finally, a method to convert the signal into smart contract invocations to execute the orders—the workflow of trading on the live execution system.

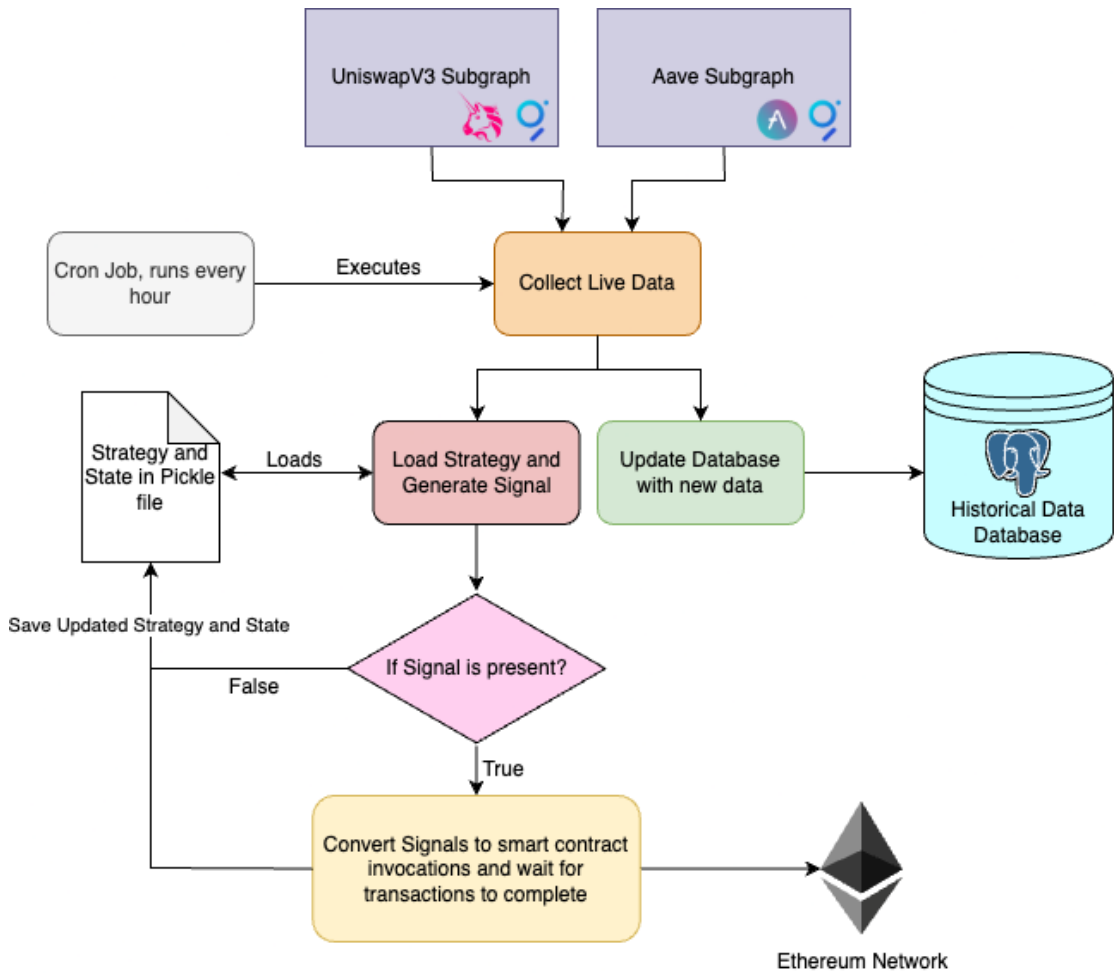


Figure 4.6: Flowchart of trading a strategy live

4.2.1 Smart Contracts

In order to be able to interact with Uniswap and Aave over Ethereum, smart contracts are written, which are then deployed on the blockchain network. Functions that the smart contract possesses can then be executed by anyone who knows the contract's address on the network. Smart contracts are programmable protocols written in Solidity, a high-level programming language designed for creating and executing smart contracts on Ethereum. For trading using the mentioned strategies, a few the following functionalities are required to be implemented; wrapping ETH for WETH, unwrapping WETH, swapping token A for token B using a given liquidity pool and finally, borrowing and repaying some of token A using WETH as collateral. The first two functions are required if the trader has only ETH in their account; thus, by wrapping the ETH, they can trade ETH as an ERC-20 token. ERC-20 is a widely adopted technical standard to create interchangeable Ethereum blockchain tokens. It provides developers with the framework to design tokens compatible with various applications and services operating within the Ethereum ecosystem. These tokens are designed to represent a variety of assets, most importantly cryptocurrencies.

Wrapping of ETH and Unwrapping WETH

This functionality is required if the trader has only ETH in their account; thus, by wrapping the ETH, they could trade ETH in its ERC-20 format, WETH. The functions of wrapping ETH and unwrapping WETH can be seen in Figure 4.7, where `swapEthForWeth` wraps ETH and `swapWethForEth` unwraps WETH. The `swapEthForWeth` function first uses the `IWETH` interface, which is itself an ERC20 interface, to allow functions such as transfer of

ownership of the token, to deposit an amount of ETH in exchange for WETH, which is then transferred back to the address of the caller. The `swapWethForEth` function unwraps the WETH by first sending the specified volume of WETH from the caller to the smart contract, withdrawing the specified amount from the IERC20, which calls the `receive` function to enable the receiving of ETH and once the ETH is withdrawn into the smart contract's account, it is transferred back to the caller.

```

1 interface IWETH is IERC20 {
2     function deposit() external payable;
3     function withdraw(uint amount) external;
4 }
5
6 address public immutable wethAddress = 0xC02aaA39b223FE8D0A0e5C4F27eAD9083C756Cc2;
7
8 function swapEthForWeth() external payable {
9     IWETH weth = IWETH(wethAddress);
10    weth.deposit{ value: msg.value }();
11    weth.transfer(msg.sender, msg.value);
12 }
13
14 function swapWethForEth(uint256 amount) external payable {
15     IWETH(wethAddress).transferFrom(msg.sender, address(this), amount);
16     IWETH(wethAddress).withdraw(amount);
17     msg.sender.transfer(address(this).balance);
18 }
19
20 receive() external payable {}

```

Figure 4.7: Solidity Code for Interchanging between ETH and WETH

Swapping using a Uniswap Liquidity Pool

Swapping using a Uniswap liquidity pool is also another critical functionality. This is done by using the `IUniswapV3Pool` interface to expose the liquidity pool's functions and calling `swap` on it using the necessary parameters. The parameters, in their respective order, are the address of the recipient of the swapped tokens, a boolean on whether the swap direction is from token0 to token1 or vice versa, the volume of the source token that would like to be swapped, an integer (`sqrtPriceLimitX96`) that manages how much slippage is acceptable, and finally the ABI encoded data that is required in the callback function. When calling `swap` function, the `sqrtPriceLimitX96` parameter is set to ignore the effects of slippage and proceed with the swap. Finally, once the swap has taken place, Uniswap calls the callback function `uniswapV3SwapCallback`, with transfers the swapped tokens to the caller of the function. See Figure 4.8.

Borrowing and Repaying Tokens using Aave

To allow for short selling, borrowing and repaying tokens are required; therefore, interacting with Aave in the smart contract is essential. The first step is to instantiate the lending pool using Aave's `PoolAddressesProvider`. When borrowing, some collateral is set aside to be able to borrow any number of tokens; thus, the first step is to deposit this collateral into the lending pool and allow it to be used as collateral (Lines 1-8). Secondly, the specified token is borrowed from the lending pool using the variable interest rate specified by the third argument in the `borrow` function; once received, the tokens are sent to the caller's address. The opposite methodology is true for loan repayment; the tokens are transferred from the caller's address to the smart contract's address, which is then sent back to the lending pool by the `repay` function. The collateral is withdrawn and automatically sent to the caller's address. See Figure 4.9.

```

1 function swapExactUsingPool(address poolAddress, bool zeroForOne, int256 amountIn)
  public returns (int256, int256) {
2     IUniswapV3Pool pool = IUniswapV3Pool(poolAddress);
3     return
4         pool.swap(msg.sender, zeroForOne, amountIn, zeroForOne ? TickMath.MIN_SQRT_RATIO
          + 1 : TickMath.MAX_SQRT_RATIO - 1, abi.encode(poolAddress, pool.token0(), pool.
            token1(), msg.sender));
5 }
6
7 function uniswapV3SwapCallback(int256 amount0Delta, int256 amount1Delta, bytes calldata
  data) external override {
8     (address poolAddress, address token0, address token1, address userAddress) = abi.
      decode(data, (address, address, address, address));
9     require(msg.sender == address(poolAddress));
10    if (amount0Delta > 0) {
11        IERC20(token0).transferFrom(
12            userAddress,
13            msg.sender,
14            uint256(amount0Delta)
15        );
16    }
17    if (amount1Delta > 0) {
18        IERC20(token1).transferFrom(
19            userAddress,
20            msg.sender,
21            uint256(amount1Delta)
22        );
23    }
24 }

```

Figure 4.8: Solidity Code for Swapping using Uniswap

```

1 IPool public immutable lendingPool = IPool(IPoolAddressesProvider(0
  x2f39d218133AFaB8F2B819B1066c7E434Ad94E9e).getPool());
2
3 function borrowToken(address tokenAddress, uint256 borrowAmount, uint256
  collateralAmount) public {
4     // Deposit Collateral
5     IERC20(wethAddress).transferFrom(msg.sender, address(this), collateralAmount);
6     IERC20(wethAddress).approve(address(lendingPool), collateralAmount);
7     lendingPool.deposit(wethAddress, collateralAmount, address(this), 0);
8     lendingPool.setUserUseReserveAsCollateral(wethAddress, true);
9
10    // Borrow token
11    lendingPool.borrow(tokenAddress, borrowAmount, 2, 0, address(this));
12    IERC20(tokenAddress).transferFrom(address(this), msg.sender, borrowAmount);
13 }
14
15 function repayBorrowedToken(address tokenAddress, uint256 repayAmount, uint256
  collateralWithdrawAmount) public {
16    IERC20(tokenAddress).transferFrom(msg.sender, address(this), repayAmount);
17    IERC20(tokenAddress).approve(address(lendingPool), repayAmount);
18    lendingPool.repay(tokenAddress, repayAmount, 2, address(this));
19    lendingPool.withdraw(wethAddress, collateralWithdrawAmount, msg.sender);
20 }

```

Figure 4.9: Solidity Code for Borrowing and Paying on Aave

Opening and Closing of Trading Positions

In addition to these functions, it is known that batching transactions together is more efficient than having to execute them separately, therefore as the strategies being ex-

plored for this project require both a buying and selling trade each time, the functions `openBuySellPositions` and `closeBuySellPositions` have been implemented such that it follows the method mentioned in Subsection 3.2, combining the internal functions `swapExactUsingPool`, `borrowToken` and `repayBorrowedToken`. The full implementation can be found in Appendix D.2.3.

4.2.2 State

Storing the state of the strategy is essential as the strategies need to store and maintain its history and variables to ensure consistency when trading; therefore, the state is stored as a dictionary in a pickle file. The state contains the liquidity pool pair being traded, the strategy instance, a dictionary of the open positions (initially empty) and the account balance. The state is updated after every execution and stored in the pickle file.

4.2.3 Retrieval of Data and Signal Generation

The frequency at which signals and data are collected is easily variable; however, the current design has a frequency of every hour to ensure consistency. This frequency is set in a Cron job. The Cron job runs a script to collect the current price data and generate a signal. Using Uniswap's subgraph, two queries are sent to retrieve the necessary data. The first query queries the liquidity pool returning its current prices, and the second, retrieves the most recent transaction, regardless of the type of transaction, to collect the cost of gas at the time. These data points are inserted into their corresponding tables in the database. They are also sent to the strategy to generate a signal, which is sent to the execution system for execution. The GraphQL queries are in Appendix D.1.2 and D.1.4.

4.2.4 Trade Execution

Once the signal has been received, it is sent to the trade execution system. Once the signal is required to be actualised, the execution system breaks the signal into different types of orders and executes them. For each type of order in signal, its parameters are extracted, e.g. volume to buy, volume to sell, the target token and the prices when opening a position. These parameters are then used to calculate the remaining parameters required to call the corresponding smart contract function, e.g. the swap direction's `zeroForOne`. Finally, once all the necessary function parameters have been defined, the order is executed on the blockchain by performing four steps. The first is calling the smart contract function, signing the transaction, sending the transaction and waiting for the transaction to complete. After all of the orders have been executed, the new balances of each token, along with the updated `open_positions`, are returned to then be stored as the new state.

```

1 # Call the order's corresponding function
2 call_function = contract.functions.<Smart Contract Function Name>(<Function Paramters>).
   buildTransaction({"chainId": Chain_id, "from": caller, "nonce": nonce})
3 # Sign transaction
4 signed_tx = web3.eth.account.sign_transaction(call_function, private_key=private_key)
5 # Send transaction
6 send_tx = web3.eth.send_raw_transaction(signed_tx.rawTransaction)
7 # Wait for the transaction receipt
8 tx_receipt = web3.eth.wait_for_transaction_receipt(send_tx)

```

Chapter 5

Trading Strategies

As previously mentioned, the strategies employed are variations of the mean reversion trading strategy. The basic concept of a mean reversion strategy identifies two closely related assets, typically referred to as a pair, and takes advantage of deviations from their historical price relationship. This works by identifying a pair of assets with a historically stable relationship, as seen in Section 3.5. Then as prices change, the spread between the prices is calculated, and if the spread has deviated from the historical mean spread, the strategy should buy the undervalued asset and short-sell the other. The positions are then closed once the spread reverts to its historical mean.

5.1 Abstract Strategy

The implementation of the generalization of these strategies is intricate and designed to be scalable and highly customizable with various parameters. Additionally, since the trading strategies being investigated are based on mean reversion, an abstract strategy is created to allow for quick exploration and research into different approaches and the impact of hedge ratio calculations on returns. This abstract strategy encompasses the core logic of generating trading signals, determining optimal trade timing and volumes, and transmitting them to the live or backtesting system. The strategy-specific classes contain operations that calculate the hedge ratio and establish the thresholds that trigger the generation of trading signals. Below are the functions that the abstract strategy possesses.

```
1 class Abstract_Strategy():
2     def __init__(self, number_of_sds_from_mean, window_size_in_seconds,
3         percent_to_invest, strategy_name, gas_price_threshold,
4         rebalance_threshold_as_percent_of_initial_investment):
5         ...
6
7     def initialise_historical_data(self, history_p1, history_p2):
8         ...
9
10    def recalculate_thresholds(self, has_trade=False):
11        raise NotImplementedError("recalculate_thresholds not implemented")
12
13    def new_tick(self, price_of_pair1, price_of_pair2, has_trade):
14        ...
15
16    def generate_signal(self, ctx, prices):
17        ...
```

The functions `__init__`, `initialise_historical_data`, `new_tick` and `generate_signal` are all inherited by each strategy and `recalculate_thresholds` is implemented in each instance of the strategy. `__init__`, `initialise_historical_data` are self explanatory. `new_tick` is called in `generate_signal` to update the strategy's knowledge of historical prices; this function also triggers a call to `recalculate_thresholds` which re-calculates the hedge ratio and determines the thresholds at which an arbitrage opportunity becomes

apparent. `generate_signal` is the function that the trading system calls at each price update. It first invokes `new_tick`, which in turn updates the thresholds; finally, the updated hedge ratio and thresholds are used in generating a signal. The steps of `generate_signal` are outlined below:

1. The first step is to call `new_tick` which updates the thresholds
2. Second is to check whether a position is already held;
 - (a) If a position IS held, the first is to check if the spread is between the thresholds, and if the gas price is below the specified limit, the positions are closed. However, before this, a precautionary check is made to ensure the balance of any token does not go below 0; if it does, a swap order is placed before the closing of the positions such that the balance of each token remains positive. Otherwise, the loan-to-value health factor is calculated; if it may cause liquidation, additional collateral is deposited in Aave.
 - (b) However, if a position is NOT held and the gas price is below the specified limit, there are 3 cases.
 - i. $spread > upper_threshold$ - Buy pair 2 and short sell pair 1
 - ii. $spread > lower_threshold$ - Buy pair 1 and short sell pair 2
 - iii. No trade

Furthermore, if a trade is ordered, it is also checked; if the WETH balance falls below the rebalance threshold, the additional tokens are converted back to WETH.

In addition, prior to opening or closing a position, a preliminary check is conducted to ensure that the ETH balance remains positive. If the anticipated balance after executing the orders indicates a potential shortfall, an order to BUY ETH is automatically initiated. This precautionary measure ensures that the strategy maintains a positive ETH balance and avoids entering into positions that could lead to negative balances.

Refer to Appendix [D.2.2](#) to see the implementations of calculating the hedge ratios of each strategy.

5.1.1 Volume of Trades

A major part of the strategy is at which volumes to trade; therefore, in order to be able to trade the maximum amount, a minimization problem must be solved. However, before delving into this problem, it is important to highlight the influence of the hedge ratio on the trading volumes. If the hedge ratio is positive, for every unit of pair0 traded, `hedge_ratio` units of pair1 should be traded. However, the inverse is true if the hedge ratio is negative; for every $-\text{hedge_ratio}$ units of pair0 traded, 1 unit of pair1 should be traded.

Given that Z is the amount of WETH in the account, p_0, p_1 are the prices of pairs 0 and 1 respectively, LTV is the loan-to-value ratio and x, y are the volumes to be traded on pair0 and pair1 respectively. Additionally, to generalize the problem for positive and negative hedge ratios, let $n : k$ represent the ratio of volume traded for pair0:pair1. Assuming we are buying pair0 and selling pair1. The solution to the following minimization problem enables the strategy to find the maximum volumes that can be traded to leave the least

amount of WETH remaining.

$$\begin{aligned}
\min_{x,y} \quad & Z - xp_0 - \frac{yp_1}{LTV} \\
\text{s.t.} \quad & x = \frac{k}{n}y \\
& x, y \geq 0 \\
& p_0, p_1 \geq 0 \\
& Z \geq 0 \\
& 0 \leq LTV \leq 1
\end{aligned} \tag{5.1}$$

Therefore, as all terms in the objective function are positive, the minimal objective value, in theory, must be 0:

$$\begin{aligned}
Z - xp_0 - \frac{yp_1}{LTV} &= 0 \\
Z - \frac{k}{n}yp_0 - \frac{yp_1}{LTV} &= 0 \\
Z - (\frac{k}{n}p_0 + \frac{p_1}{LTV})y &= 0 \\
y &= \frac{Z}{\frac{k}{n}p_0 + \frac{p_1}{LTV}} \\
x &= \frac{k}{n} \frac{Z}{\frac{k}{n}p_0 + \frac{p_1}{LTV}}
\end{aligned}$$

Thus, the resulting values of x and y represent the solution that minimizes the problem while maximizing the tradable volume. The case above shows that x units of pair0 can be bought, and thus y units of pair1 must be sold. Similarly, if pair0 is sold and pair1 is bought, the maximum volumes are determined by $x = \frac{Z}{\frac{p_0}{LTV} + \frac{n}{k}p_1}$ and $y = \frac{n}{k} \frac{Z}{\frac{p_0}{LTV} + \frac{n}{k}p_1}$.

5.2 Constant Hedge Ratio Strategy

In the most simplistic mean reversion strategy, the approach relies on a given historical dataset, assuming the hedge ratio between the paired assets remains consistent over the long term. The Ordinary Least Squares (OLS) regression method is employed to determine this hedge ratio.

In Figure 5.1, the observed trend reveals that the gradient, representing the rate of change, exhibits proximity to the value of 1. This indicates a relatively balanced relationship between the prices of each pair. However, it is important to note that as time progresses, any fluctuations or deviations from the exact value of 1 are minimal and have a negligible impact on the overall trend. The stability of the gradient over time suggests a consistent and relatively stable relationship between the liquidity pools, reinforcing the notion that their interdependence remains relatively constant.

5.3 Sliding Window OLS Strategy

As mentioned earlier, the hedge ratio is crucial in minimizing risk and achieving a market-neutral position in trading strategies. However, a static hedge ratio may prove suboptimal since market conditions and the correlation between paired assets can evolve. A dynamic approach is adopted to tackle this challenge using a sliding window approach. This approach continuously updates the hedge ratio at each tick by considering the most recent data. The length of the data window used for the calculation is determined by the value specified in the `window_size` argument. By incorporating the sliding window mechanism, the hedge ratio remains adaptive to the changing market dynamics, providing a more accurate and responsive estimation for effective risk management and trading decisions.

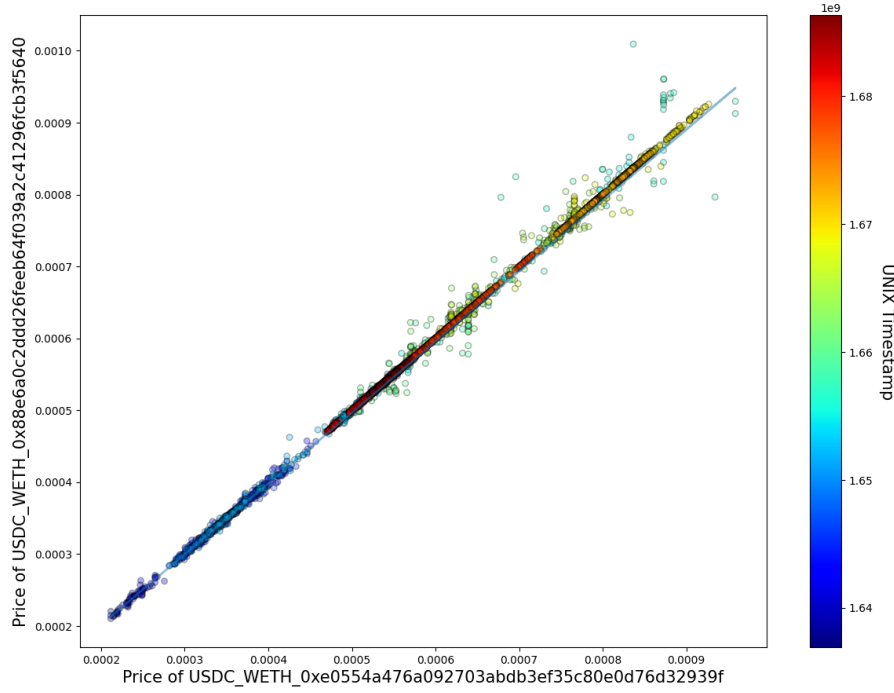


Figure 5.1: Conducting OLS to obtain the Hedge Ratio

5.4 Lagged OLS Strategy

A lagged strategy was also implemented as an extension of the sliding window strategy. In the sliding window strategy context, the hedge ratio is updated using the most recent data within a specified window size. However, due to potential delays or lags in the synchronization of liquidity pools, the instantaneous prices of the assets may not fully reflect their actual relationship. Therefore, this approach aims to account for potential lags in the synchronization of liquidity pools. The lagged strategy extends the existing sliding window approach using `sm.tsa.lagmat` function to perform the lag, which is then used to perform OLS.

5.5 Unrestricted Model from the Granger Causality Test

Similarly, to account for lag and causality, a strategy that was employed was using the Granger Causality Test. The Granger Causality Test is a statistical hypothesis test that assesses the predictive value of one time series in forecasting another [55]. The Granger causality test is based on the idea that if variable X is causing variable Y, then the past values of X should contain information that helps predict the future values of Y beyond what can be predicted by using only the past values of Y itself. The test involves fitting two regression models: the restricted and unrestricted models. In the restricted model, only the past values of the dependent variable are used as predictors. In the unrestricted model, the past values of the dependent variable and the potentially causal variable are included as predictors. By comparing the performance of these two models, the test determines whether including the potentially causal variable improves the predictive power.

In order to estimate the hedge ratio, the unrestricted model's regression parameters are used as it captures the potential influence of the potentially causal variable on the dependent variable beyond what can be explained by the lagged values of the dependent variable alone.

5.6 Kalman Filter Strategy

Another method often used to ensure market-neutral positions is using Kalman Filters as a dynamic tool to update and adapt the hedge ratio, taking into account the changing market conditions and the evolving relationship between the paired assets. This allows for a more responsive and adaptive approach to maintaining a market-neutral position. The Kalman Filter considers not only the current price data but also the historical information and the underlying dynamics of the assets. It provides a more robust and accurate estimation of the optimal hedge ratio, which aligns with the mean reversion strategy's objective of capturing potential price divergences and profiting from their eventual convergence.

The Kalman Filter works by iteratively updating and refining its estimate of the system state using a prediction step and an update step. An Ordinary Least Squares regression is used as an initial 'guess' of the parameters using the historical information provided. The remainder of the parameters that are selected can be seen below:

```
1 model = sm.OLS(price_history_2, sm.add_constant(price_history_1))
2 initial_state = model.fit().params[:-1]
3
4 kf = KalmanFilter(
5     n_dim_state=2,
6     initial_state_mean=initial_state,
7     transition_matrices=np.eye(2),
8     observation_matrices=obs_mat,
9     transition_covariance=1e-5 * np.eye(2)
10 )
```

Below describes the reasoning behind the choices of parameters:

- **n_dim_state** - This parameter to the number of elements in the state. In our scenario, the state consists of the y-intercept and the gradient.
- **initial_state_mean** - As mentioned earlier, the Kalman Filter utilizes past estimates to iteratively estimate the true states. Therefore, the OLS method calculates the initial state of the price history's regression, which returns both the y-intercept and gradient.
- **observation_matrices** - The observation matrix defines the relationship between the observed measurements and the hidden state variables. This takes the provided historical data zipped along with 1 as the observed measurements directly correspond

to the hedge ratio. This can be represented as
$$\begin{bmatrix} y_1 & 1 \\ y_1 & 1 \\ \vdots & 1 \\ y_n & 1 \end{bmatrix}.$$

- **transition_matrices** - In the context of our strategy, we anticipate a consistent long-term hedge ratio. Therefore, we set the **transition_matrices** parameter to
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$
- **transition_covariance** - Finally the **transition_covariance** specifies the covariance matrix of the process noise. It represents the uncertainty or variability in the state transition; hence it is set to be very small,
$$\begin{bmatrix} 10^{-5} & 0 \\ 0 & 10^{-5} \end{bmatrix}.$$

Figures 5.2 and 5.3 provide visual representations of the hedge ratio's evolution over time. These figures illustrate that the hedge ratio, while exhibiting subtle variations, changes over the observed period. One notable event is the significant rise at the timestamp 1.655e9 or June 2022. This shift in the hedge ratio is likely attributed to the migration from the Proof of Work (PoW) consensus mechanism to the Proof of Stake (PoS) consensus mechanism

on the Ethereum network.

During the transition from PoW to PoS, there are fundamental changes in how the Ethereum network reaches consensus and validates transactions. This change in the consensus mechanism can impact various aspects of the network, including the dynamics of the hedge ratio. It is plausible that the observed drop in the hedge ratio during the migration period reflects the adjustments and reconfiguration of the underlying mechanisms in response to the shift to PoS.

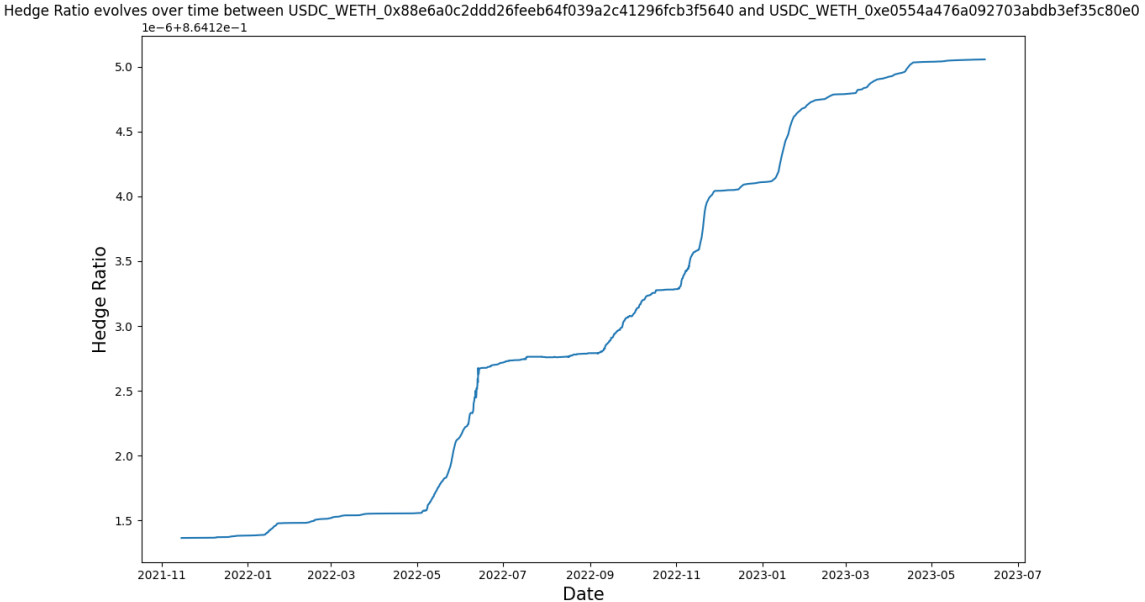


Figure 5.2: Plot of how the Hedge Ratio evolves using the Kalman Filter

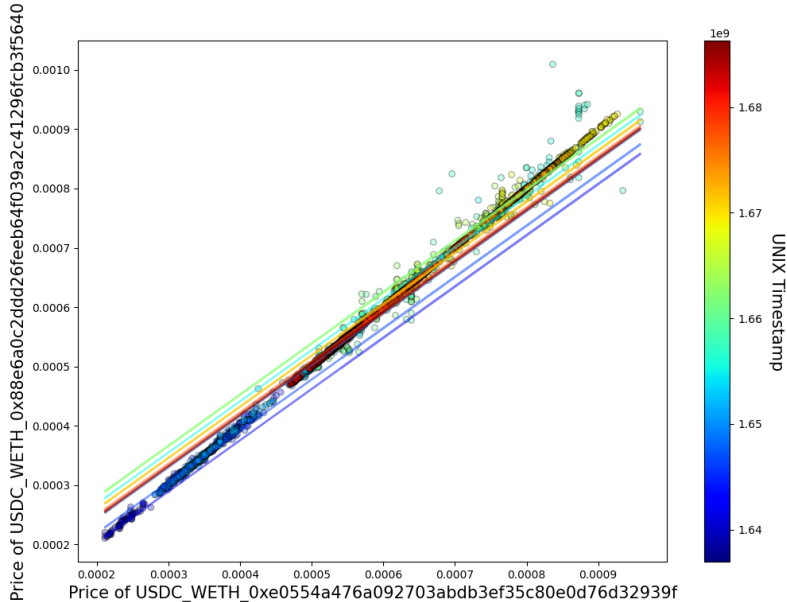


Figure 5.3: The plot shows the Kalman Filter estimating the regression line over time

Chapter 6

Evaluation

6.1 Trading System

The implemented backtesting system is designed to be robust and comprehensive, considering all the necessary costs associated with trading, as discussed in Section 4.1. However, factors such as data granularity and accounting for slippage are some areas that could benefit from further improvement.

One aspect that could be enhanced is the data granularity of the pricing data used in the backtesting system. Currently, the system uses hourly data points, which may not capture all the price movements and volatility. By incorporating more granular data, such as minute-by-minute or even tick-level data, the system could provide a more accurate representation of market dynamics and potentially improve the accuracy of the backtest results.

Additionally, when executing swaps or trades, it is important to consider the impact of slippage. Slippage refers to the price difference that occurs when executing a trade due to changes in the market conditions, such as supply differences between trading pairs. To account for slippage, the system should incorporate realistic price adjustments to accurately simulate the trading environment. This can help ensure that the backtest result more accurately reflects the actual performance in real trading scenarios. However, this is somewhat mitigated as only larger liquidity pools, with a volume of \geq \$10 million, are used and with only a maximum of 100ETH (\$394,626.00 at the beginning of the backtesting period) being traded in each pool, the effects of slippage become minimal in such scenarios.

By addressing these aspects, refining the data granularity, and furthering the slippage considerations, the backtesting system can provide more precise and reliable results, leading to improved insights into the profitability and viability of the trading strategy under evaluation.

6.2 Strategy Parameters

The design of each strategy allows for extensive parameterisation, providing multiple axes for analysis. Parameters that can be adjusted are at what number of standard deviations away from the mean spread a threshold should be set and a signal to be generated, how the trades are executed, limits on how much gas price should be when executing a trade, and also the volume of the initial investment. These are all factors that can affect the returns of the trading strategy. In the following sections, we will examine the strategy's performance by varying these parameters, providing valuable insights into their influence on trading outcomes. For data visualisation, we denote the liquidity pool pairs by their index; see Table 6.1. Note that -100% in the tables below indicate one of the token's balance in the account falls below 0, resulting in bankruptcy.

Pool1	Pool2	Label
USDC_WETH_0x88e6a0c2ddd26feeb64f...	USDC_WETH_0xe0554a476a092703abd...	0
USDC_WETH_0x8ad599c3a0ff1de08201...	USDC_WETH_0xe0554a476a092703abd...	1
WETH_USDT_0x11b815efb8f581194ae7...	USDC_WETH_0xe0554a476a092703abd...	2
WETH_USDT_0x4e68ccd3e89f51c3074c...	USDC_WETH_0xe0554a476a092703abd...	3
DAI_WETH_0x60594a405d53811d3bc47...	USDC_WETH_0xe0554a476a092703abd...	4
DAI_WETH_0xc2e9f25be6257c210d7adf...	USDC_WETH_0xe0554a476a092703abd...	5
USDC_WETH_0xe0554a476a092703abd...	WETH_USDT_0xc5af84701f98fa483ece7...	6

Table 6.1: Liquidity pool pairs and their corresponding labels

6.2.1 Number of Standard Deviations away from the mean

Varying the standard deviation in a mean reversion pairs trading strategy can significantly impact the strategy's returns. When the number of standard deviations is too high, the spread between the pair's prices must deviate significantly from the mean before generating a signal. In this case, the strategy may generate fewer trades, resulting in longer holding periods for positions; however, this tends to have a greater return per trade. On the other hand, setting a lower number of standard deviations makes the strategy more sensitive to smaller deviations from the mean. This increases the frequency of trade signals and potentially leads to more active trading. However, it also exposes the strategy to a higher risk of false signals or trading on noise in the price data. It can result in higher transaction costs and potentially lower returns due to increased trading activity, potentially resulting in losses. For simplicity, $\#_\sigma$ denotes this parameter.

Table 6.2 displays the returns as the standard deviations away from the mean the threshold is set. For this experiment, the other parameters were; 100 ETH worth of initial investment, the window size set to 30 days, a gas price threshold of 1 ETH (i.e. not having a threshold), the interactions with the blockchain to be separately executed and finally the lag that the lagged strategy uses are set to 1 hour. The period that trading occurs is from 18th December 2021 to 9th June 2023.

We can see that as $\#_\sigma \rightarrow 0$, the number of trades increases; however, this causes the returns to be more negative than positive, likely due to trades being opened when gas fees are substantially larger than the profit. However, if $\#_\sigma$ is larger, although the profits from each trade increase, the number of actionable opportunities is limited. Another interesting feature that can easily be spotted is that the Constant Hedge Ratio Strategy makes little to no profit in all cases. In contrast, the other strategies perform better as $\#_\sigma \rightarrow \sim 1.75$. Furthermore, all strategies return a loss on pool pair 6 except the Kalman Filter Strategy, albeit with a higher threshold.

Consistent profit among all pool pairs, excluding pool pair 6, is also achieved at different numbers of standard deviations. We can see in Figure 6.1 that the average return across all liquidity pool pairs that each strategy generates at each standard deviation. The Kalman Filter becomes largely profitable at $\#_\sigma = 0.5$, whereas the lagged and sliding window strategies become profitable at $\#_\sigma = 2$. This is because the Kalman Filter and the Granger Causality tests can better find underlying trends, thus better selecting the hedge ratio, affecting the volumes being traded and hence the effect of the returns.

6.2.2 Gas Fees

Another aspect that could be changed is fees; the first is how the transactions are executed on the blockchain and the maximum gas price if we want to open a position. As previously mentioned, the gas fee is calculated by combining the gas price and the gas used. Therefore, these two parameters allow the strategy to control the gas used and the gas price.

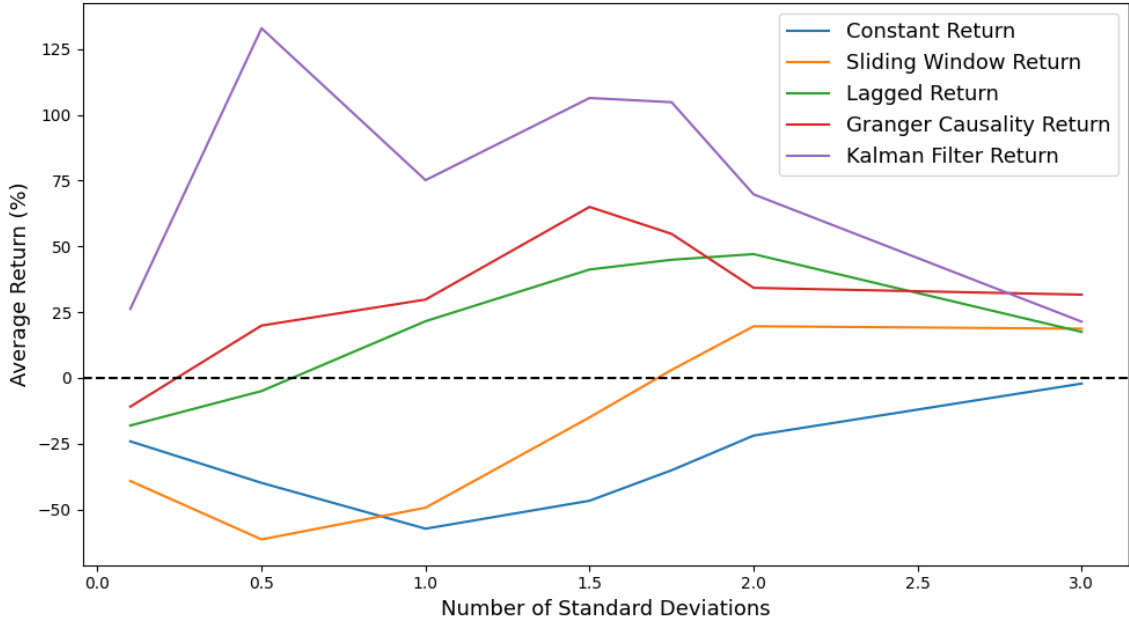


Figure 6.1: Average Return across all pool pairs using various standard deviations

Batched Transactions vs Separately Executed Transactions

The method in which the transactions are executed affects how much gas is used; when transactions are executed separately, each transaction incurs its own gas cost. This means that gas will be consumed for various operations such as contract interaction, data storage, and computational tasks for every individual transaction. As a result, executing many separate transactions can lead to a high cumulative gas cost. On the other hand, when transactions are batched together, they are consolidated into a single transaction. This consolidation reduces the overhead associated with executing multiple transactions individually. By combining the operations of multiple transactions into a single execution, redundant computations and storage operations can be eliminated, resulting in more efficient gas usage. It can immediately be seen in Figure 4.3 that the process of batching the transactions consumes less gas than the combined gas of the transactions required to execute the opening and closing of the positions.

Furthermore, the effects of separate versus batched transactions can be seen in Table 6.3. The parameters used for this investigation were the same as above with the selection, e.g. $\#_{\sigma} = 1.75$, 100 ETH worth of initial investment, the window size set to 30 days, a gas price threshold of 1 ETH (i.e. not having a threshold) and the testing period remained from 18th December 2021 to 9th June 2023.

	Pool Pair	Strategy's Annual Percentage Rate (APR) - Trading from 18th December 2021 to 9th June 2023									
		Constant		Sliding Window		Lagged		Granger Causality		Kalman Filter	
		Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades
Seperate	0	-7.14	288	24.57	342	81.97	353	61.58	192	98.76	153
	1	-52.98	358	-20.9	293	-5.29	323	12.99	213	45.66	178
	2	-1.71	225	21.43	277	68.39	270	68.62	180	95.11	133
	3	-45.93	309	-14.98	268	-1.01	294	22.19	211	53.56	145
	4	-2.54	241	26.46	294	55.77	291	65.92	201	84.62	152
	5	-41.71	287	-9.73	225	9.79	250	20.48	200	61.97	152
	6	-35.28	76	-19.16	67	-29.96	76	-23.36	37	-16.02	74
Batched	0	-2.2	288	30.8	342	87.79	353	66.76	192	100.13	153
	1	-47.23	358	-16.28	293	0.91	323	16.58	213	46.78	178
	2	2.63	225	28.29	277	74.07	270	75.52	180	97.43	133
	3	-42.06	309	-9.95	268	4.71	294	25.33	211	52.93	145
	4	1.72	241	29.97	294	62.84	291	69.21	201	86.18	152
	5	-37.04	287	-4.83	225	14.37	250	23.51	200	62.31	152
	6	-33.66	76	-17.93	67	-27.78	76	-21.93	37	-14.77	74

Table 6.3: Returns of each strategy when using seperate and batched transactions

When comparing the use of batched transactions to executing transactions separately, it is evident that utilising batched transactions yields a higher return, as expected. The observed increase in return ranges from approximately 1% to 8% compared to executing the transactions individually. The higher return achieved through batched transactions can be attributed to several factors. First, as mentioned earlier, batching transactions reduces the gas costs of executing multiple transactions. By consolidating operations into a single transaction, redundant computations and storage operations are eliminated, resulting in more efficient gas usage. The reduction in gas costs directly contributes to the overall profitability of the trading strategy.

Gas Price Threshold

A threshold on the gas price can also be set to ensure positions with a high transaction cost are avoided. As the gas price is variable, we can see in Figure 6.2 the distribution of the gas prices over time is similar to a geometric distribution; therefore, by thresholding high gas prices, the strategies would avoid making a trade where the cost of trading would overshadow the trade's profit.

Therefore, our investigation consists of varying this threshold to see its effects on the returns of each strategy, using batched transactions and the same parameters used in the above experiment, the thresholds are set to the 60th, 70th, 80th, 90th, 100th quantiles. Table 6.4 displays the results of varying these thresholds. It can be seen that filtering more costly opportunities, expectedly, reduces the number of actionable; however, the higher thresholds also show that these opportunities are still just as if not more profitable. In addition, to this, Figure 6.3 shows that a low threshold benefits the constant hedge ratio strategy most as with the low threshold, it has a slight loss, and as the threshold increases, the return using the strategy decreases. Whereas the returns of the Kalman Filter, Granger Causality and Lagged strategies are hindered with a lower threshold as many profitable opportunities are missed due to the overly low threshold, resulting in a reduction in the overall profitability of the trading strategy. The figure also shows how setting the threshold after the 90th percentile does little to no effect on the return, giving a slightly positive im-

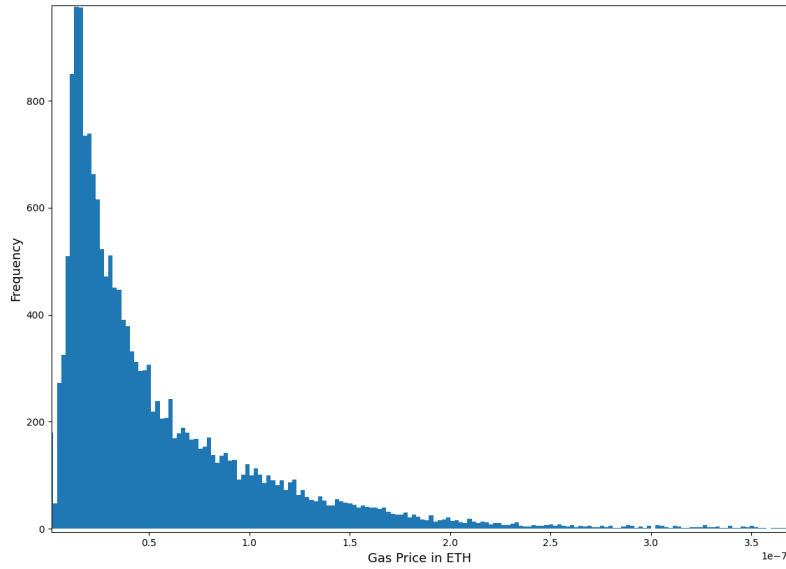


Figure 6.2: Histogram Plot of Gas Prices in ETH

impact on the Lagged strategy and a slight negative impact on the remaining. This is likely because there are only a few arbitrage opportunities to be present in such times when the gas price is high, thus having little impact. It is also interesting to see that the Lagged and the Kalman Filter strategies perform better by setting the threshold to the 90th percentile compared to the 80th percentile, whereas the other strategies perform worse.

Overall, the analysis of varying gas price thresholds reveals that adjusting the threshold level directly impacts the number of actionable opportunities and the resulting returns. Higher thresholds filter out less favourable opportunities while still maintaining profitability. However, it is crucial to strike a balance and avoid setting the threshold too low, which may lead to missed profitable opportunities. Hence, the threshold is set to the 90th percentile to get the best balance for the following experiments.

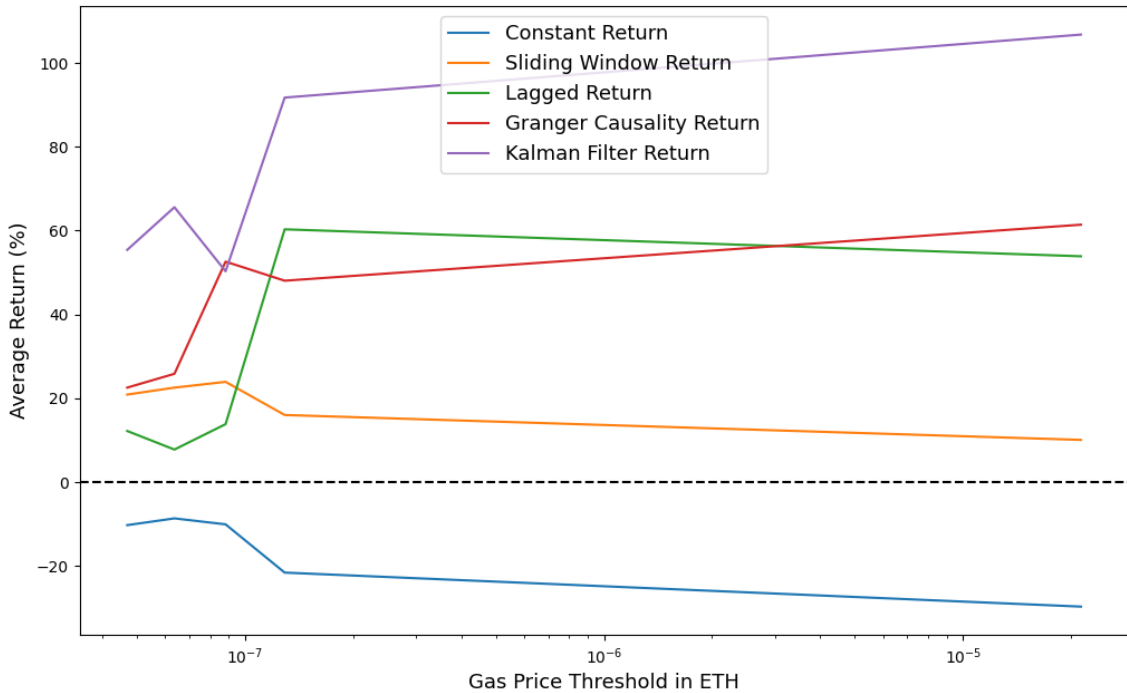


Figure 6.3: Average Returns of each strategy when using different gas price thresholds

6.2.3 Initial Investment Volume

Another important factor contributing to the profit is the initial investment; therefore, Table 6.5 and Figure 6.4 display the returns of various initial investments. It can immediately be seen that the higher the investment, the greater the return until the returns plateau. For the smallest investments result in a loss for all strategies because the fee incurred by gas fees is fixed regardless of the volume. In contrast, the costs incurred by Aave and Uniswap are proportional to the volume being traded. This fixed fee means that a greater initial investment is required to cover the cost of the gas fee in order to generate profit, which is, of course, proportional to the volume being traded. Hence, as the investment increases, the return increases; however, it begins to plateau after 25ETH as the fixed fee becomes negligible compared to the investment.

Upon comparing the different strategies, the Kalman Filter can generate a profit with the least investment, followed by the Granger Causality strategy, then the Lagged and Sliding Window strategies and finally, the Constant strategy fails to make any money with any investment. Interestingly, although the Granger Causality test strategy can churn a better return at lower initial investments, when initial investment ≤ 20 , the Lagged strategy overtakes the returns of the Granger Causality strategy once returns hit 0%. The Lagged strategy plateaus to around 60% whereas the Granger Causality strategy plateaus to 50%, indicating that although both of these methods assess the causality of the price dynamic between the two liquidity pairs, the lagged strategy estimates the hedge ratio better than the unrestricted model generated from the Granger Causality test.

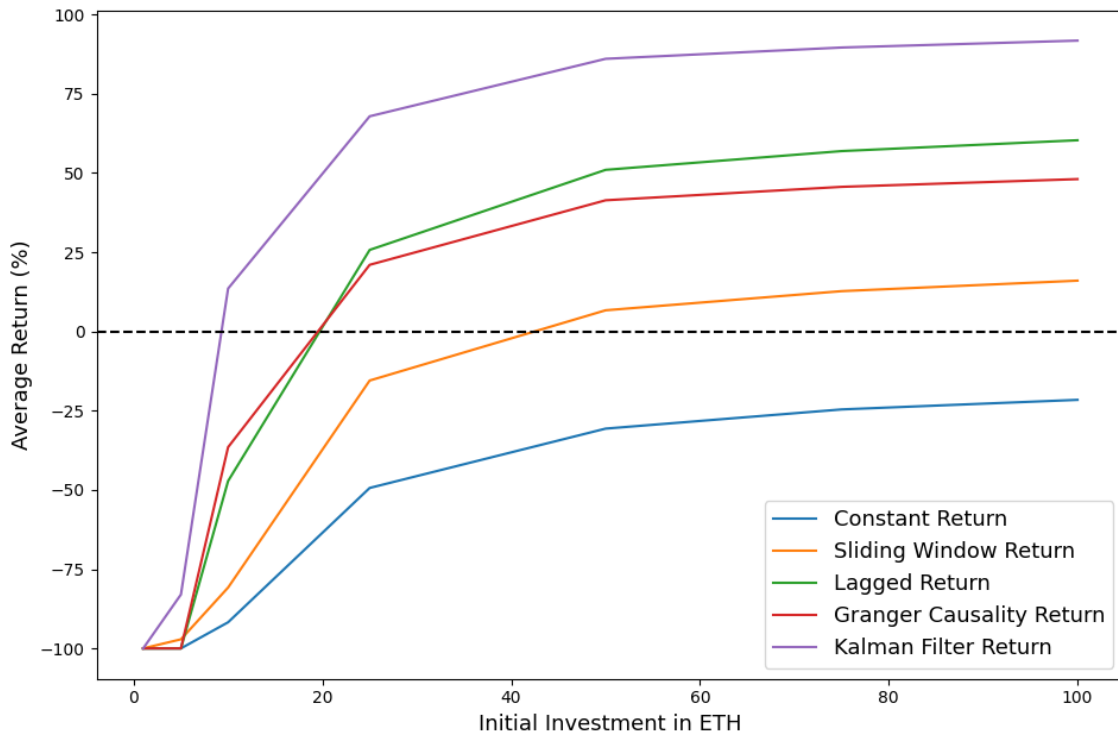


Figure 6.4: Average Returns of each strategy using different initial investment volumes

6.2.4 Window Size

The final parameter that also affects the returns is the window size. This window size is used when calculating the mean and standard deviation to calculate the thresholds. As we can see in the excerpt below, the greater the window size, the more data the strategy would use to calculate the thresholds.

```

1 spread = self.history_p1[-self.window_size_in_hours:] - self.hedge_ratio *
  self.history_p2[-self.window_size_in_hours:]
2 spread_mean = spread.mean()
3 spread_std = spread.std()
4 self.upper_threshold = spread_mean + self.number_of_sds_from_mean *
  spread_std
5 self.lower_threshold = spread_mean - self.number_of_sds_from_mean *
  spread_std

```

To evaluate how the window size affects the returns, backtesting is conducted from 9th June 2022 to the 9th June 2023, exactly 1 year. The results can be found in Table 6.6 and Figure 6.5 where we can see that, as the window sizes increase, the returns increase; however, for the smaller window sizes, the return are more erratic and volatile. This volatility decreases as the window size increases, plateauing to each strategy's returns.

It is also interesting to see that in this period of trading, the Kalman Filter performs the worst, which is surprising as it has outperformed each of the strategies in the experiment that have been previously mentioned.

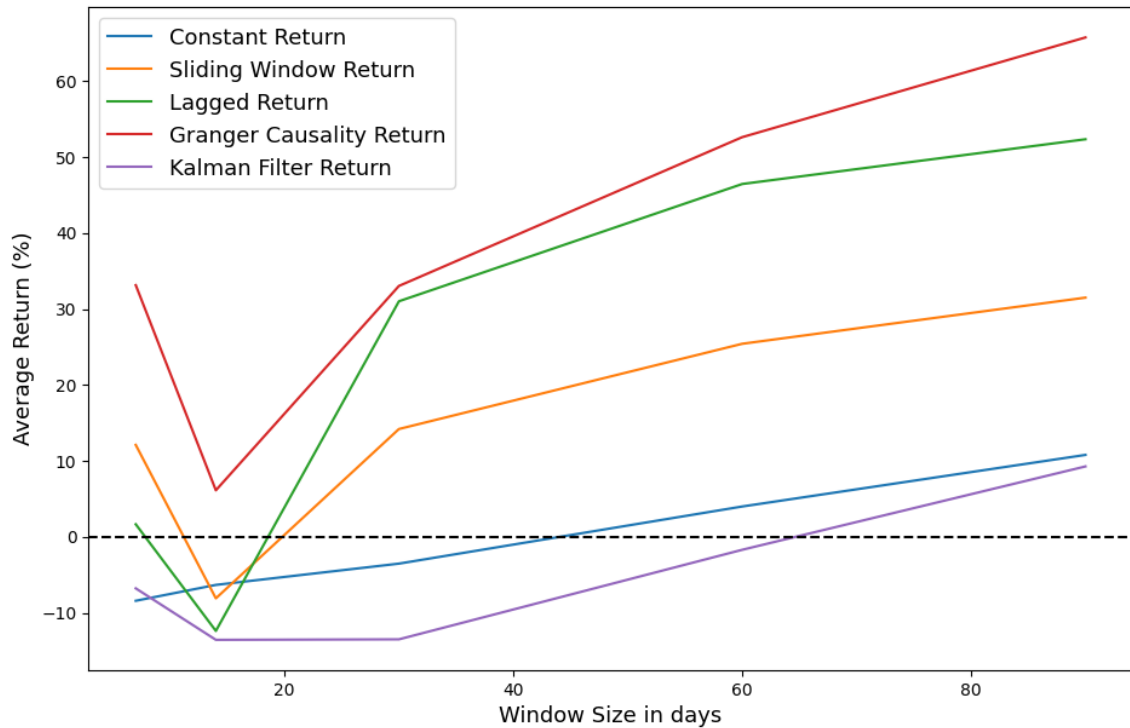


Figure 6.5: Average Returns of each strategy using different Window Sizes

6.3 Liquidity Pool Pairs Selection

The choice of liquidity pools has a direct impact on the returns. Immediately, by looking at Section 6.2, the liquidity selection has proven successful. However, we could investigate why these pairs performed particularly well. The main limitation of the pools that could be investigated was that Aave must support both tokens in the liquidity pools, and a majority of the tokens that Aave supports are stablecoins. Therefore, looking at Table 3.2, it can quickly be seen that all liquidity pairs have some form of a stablecoin aimed to be pegged to the US Dollar.

Furthermore, looking at the correlation matrix, Figure 3.4, we can immediately see that the pairs associated with tokens pegged to the US Dollar, such as USDT, USDC, and DAI,

exhibit the expected high correlation with each other, as expected. However, an exception is observed with the liquidity pool 0xe0554a476a092703abdb3ef35c80e0d76d32939f, which exhibits a lower correlation with other liquidity pools containing these stablecoins. This lower correlation is highly likely due to Uniswap routing swaps via different liquidity pools, causing the price correlation to be lower. Consequently, this lower correlation opens up additional avenues for potential arbitrage opportunities within the pool.

Upon further investigation, it can be seen that the liquidity pool 0xe0554a476a09... had a lot of volume being traded from around November 2022 through to March 2023, however since then, the number of transactions has been limited, likely increasing the probability for misprices, hence a greater chance of profit.

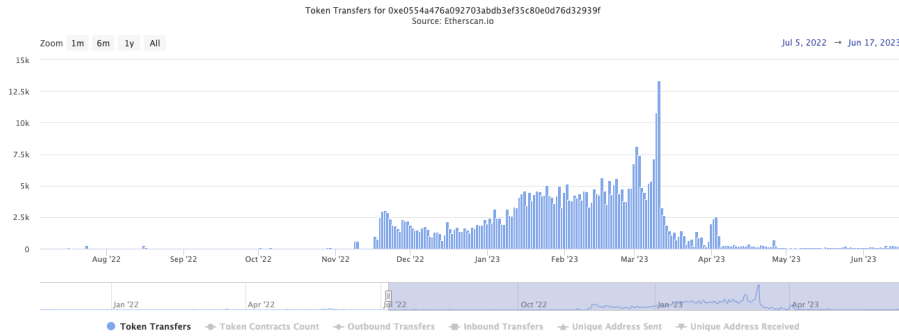


Figure 6.6: Number of Token Transfer over time in 0xe0554a476a092703abdb3ef35c80e0d76d32939f [56]

6.4 Evaluation of Strategies

Overall, given the experiments conducted, the optimal parameters for each strategy are different. Thus, to evaluate each strategy, Table 6.7 displays the optimal parameters used for each strategy.

Strategy	Number of Standard Deviations from the mean, $\#\sigma$	Gas Price Threshold in ETH	Window Size in Days	Type of Execution, Separate or Batched
Constant	3	6.36e-08	60	Batched
Sliding Window	2	8.83e-08	60	Batched
Lagged	2	1.29e-07	60	Batched
Granger Causality	1.5	2.13e-05	60	Batched
Kalman Filter	0.5	2.13e-05	60	Batched

Table 6.7: Optimal Parameters for each strategy

By running the backtesting system from 14th January 2022 to 9th June 2023 with an initial investment of 100ETH, we can see in Table 6.8 the Kalman Filter outperforms all of the other strategies in all liquidity pool pairs except pair 6. The Granger Causality strategy has also performed well. In contrast, the Lagged strategy returned mediocre return, with the Constant and Sliding Window strategies performing similarly, resulting in around 20% APY.

Pool Pair	Strategy's Annual Percentage Rate (APR) - Trading from 14th January 2022 to 9th June 2023									
	Constant		Sliding Window		Lagged		Granger Causality		Kalman Filter	
	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades
0	24.84	83	38.61	250	65.5	276	86.91	310	145.74	378
1	20.7	37	14.38	127	29.54	151	20.37	269	43.94	456
2	22.12	46	30.74	103	49.75	112	92.19	218	164.55	396
3	18.14	31	14.26	86	33.67	98	28.51	231	33.36	425
4	27.65	45	38.44	122	66.23	140	92.71	290	152.26	391
5	23.11	25	25.76	77	45.57	98	31.8	299	57.81	388
6	-2.54	13	-10.28	30	-19.31	49	-10.56	23	-29.71	150
Average	19.14	40	21.70	113.6	38.70	132	48.85	234.3	81.14	369.1

Table 6.8: Returns of each strategy with their optimal parameters

In addition to this, Figure 6.7 shows the account value over time of each of the strategies; it is easy to see that the account values seem to follow a trend; however, the gradients of the return are less drastic in some strategies, i.e. the Kalman Filter. Notably, the number of profitable trades decreased since the beginning of 2023, with the Kalman Filter strategy taking a large loss in January of this year, 2023.

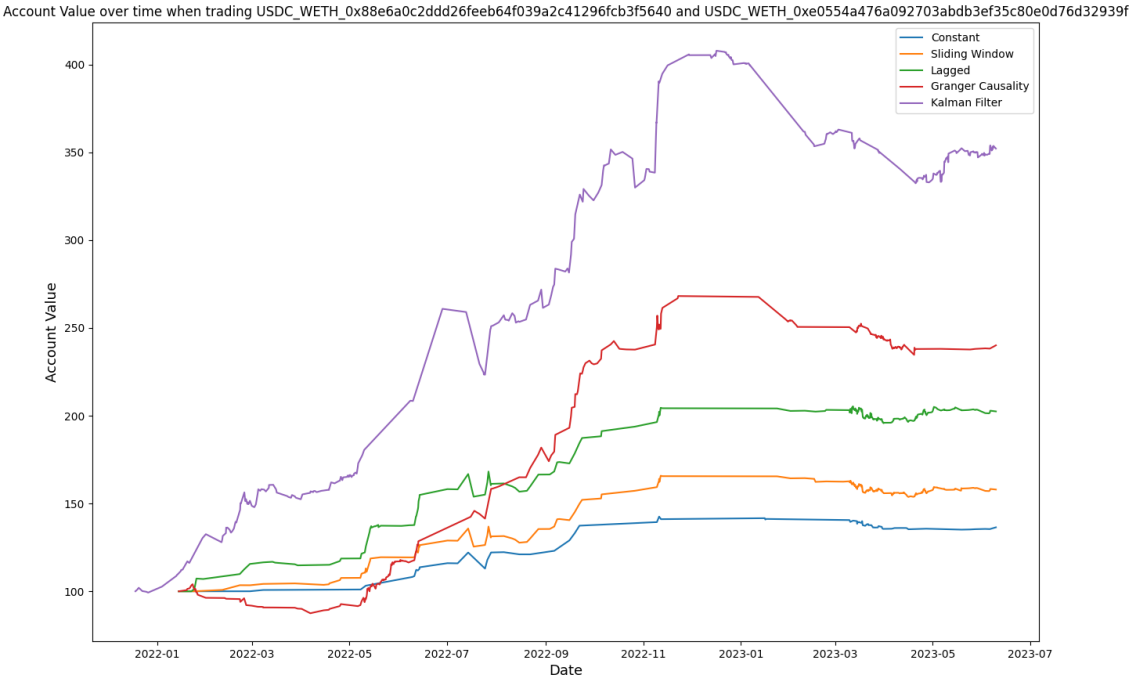


Figure 6.7: Account Value over Time from backtesting Pair 0

Another valuable insight can be seen in Figure 6.8, where the evolution of the hedge ratio is plotted for each strategy. It can be seen that the Kalman Filter, Sliding Window, and Lagged strategies' hedge ratios are reasonably consistent in comparison to the Granger Causality strategy's hedge ratio. However, as seen in Figure 5.2, the Kalman Filter calculates the Hedge Ratio slightly lower and consists of only minor deviations. This indicates a more persistent relationship between the paired assets. The minor deviations observed in the hedge ratio can be attributed to temporary fluctuations in market dynamics or noise in the data. In comparison to the Kalman Filter's hedge ratio estimations, the Granger Causality strategy's estimation shows clearly that the change from Proof-of-Work to Proof-

of-Stake on Ethereum displayed a significant systematic change in the relationship between the two liquidity pools, with the hedge ratio being around 1 before the transition to 0.9 after the transition.

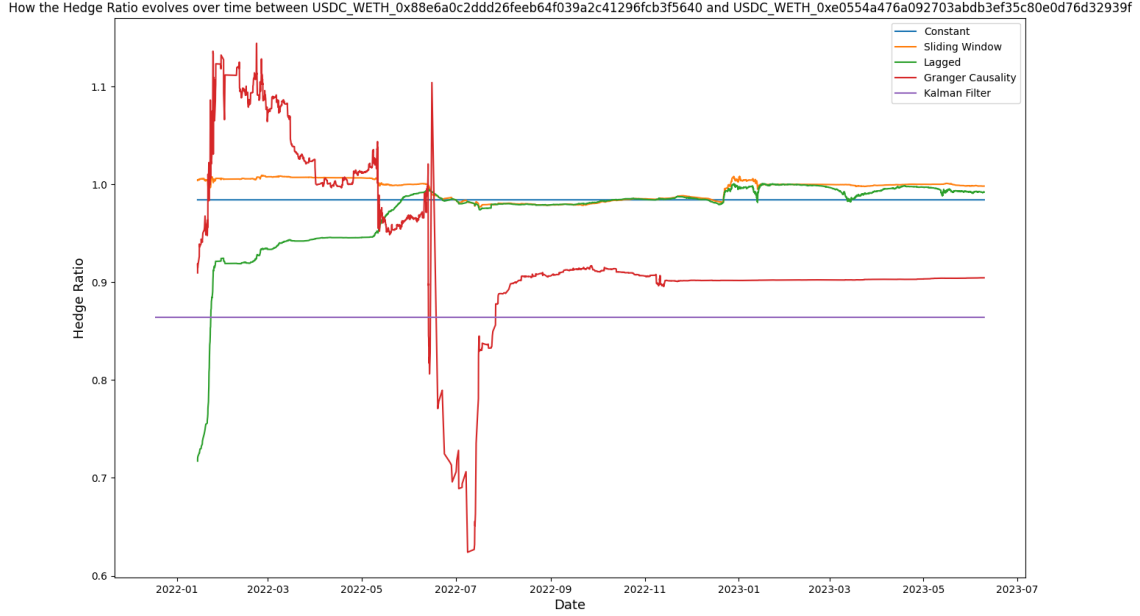


Figure 6.8: The Hedge Ratio over time for Pair 0

6.4.1 Beta

An important metric when evaluating trading strategies or investments used in industry is Beta, β . It measures how correlated the strategy's return is with the market return rate. It is often used as a risk-reward measure, allowing investors to make better investing decisions, balancing the risk and reward. β is defined as $\beta = \frac{Cov(R_M, R_S)}{Var(R_M)}$, where R_M is the market's returns and R_S is the strategy's or stock's returns, the higher the β value, the more correlated the strategy's return is with the market. In contrast, in a lower one, i.e. $\rightarrow -\infty$, the strategy swings the opposing direction to the market. Therefore, a low absolute value is ideal in risk-neutral strategies as it minimises the risk.

Therefore, to analyse the β of each strategy, the conversion from WETH to USD is used as the strategies use WETH as their base currency. As shown in Table 6.9, the β s are very close to 0, indicating that the strategy's dependence on the market returns is very low. The low β values imply that the trading strategy may be designed to generate returns based on factors other than general market movements. It suggests that the strategy relies more on specific signals, indicators, or market inefficiencies rather than broader market conditions. This can be advantageous in certain situations, as it may allow the strategy to perform well even when the overall market is experiencing volatility or downturns.

The comparison of account values and the evolution of ETH price over time are depicted in Figure 6.9. The figure illustrates a limited correlation between the trading strategy's returns and ETH's price. While the returns of the strategy do not closely follow the exact movements in the ETH price, it is evident that significant price fluctuations in ETH impact the strategy's performance. This observation suggests that the trading strategy's returns are influenced, to some extent, by the general trends and movements in the price of ETH. However, it is important to note that the relationship is not entirely linear or directly proportional. Therefore, we can conclude that the impact of ETH price changes on the strategy's performance is limited. It can also be seen that the Constant, Sliding Window and Lagged strategies are less affected by the market compared to the Granger Causality and Kalman Filter strategies.

Pool Pair	Strategies' Beta, β - Trading from 14th January 2022 to 9th June 2023				
	Constant	Sliding Window	Lagged	Granger Causality	Kalman Filter
0	-1.57001e-11	-2.37822e-11	-2.34107e-11	-1.17100e-11	-3.36017e-11
1	-1.33728e-11	-2.55276e-11	-2.34361e-11	-1.48089e-11	-2.80161e-11
2	-1.44463e-11	-2.51260e-11	-2.48695e-11	-1.08178e-11	-3.28074e-11
3	-1.30815e-11	-2.34876e-11	-2.48109e-11	-1.12766e-11	-3.32465e-11
4	-1.38961e-11	-2.59124e-11	-2.78297e-11	-1.86356e-11	-3.65904e-11
5	-1.43195e-11	-3.08987e-11	-2.79315e-11	-5.46023e-12	-3.38610e-11
6	3.42127e-12	-4.57712e-11	-3.56508e-11	-9.80665e-11	-1.43624e-10
Average	-1.162784e-11	-2.864367e-11	-2.684845e-11	-2.439650e-11	-4.882103e-11

Table 6.9: β s of each strategy

Trading USDC_WETH_0x88e6a0c2ddd26feeb64f039a2c41296fcb3f5640 and USDC_WETH_0xe0554a476a092703abdb3ef35c80e0d76d32939f

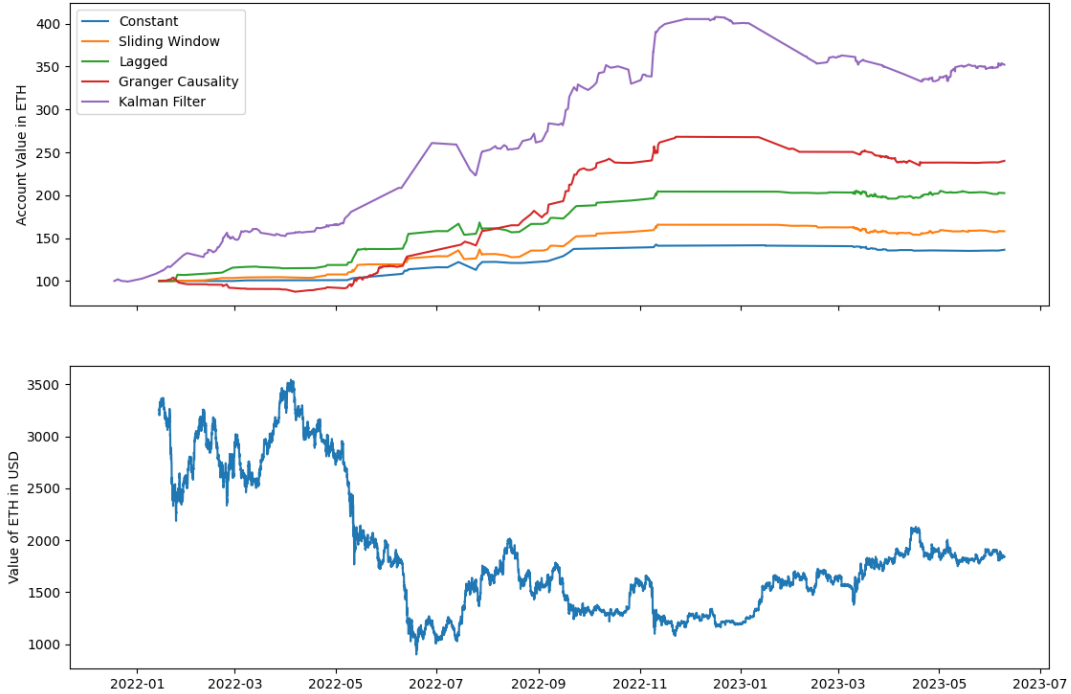


Figure 6.9: Account Value of time with the price of ETH

6.4.2 Volatility & Sharpe Ratio

Another metric commonly used in finance is the Sharpe Ratio; this is because it is important to evaluate how risky the strategies are compared to the risk-free rate. The metric is given by the formula below:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p}$$

Where R_p is the return of a portfolio, R_f is the risk-free rate and σ_p is the standard deviation of the portfolio. A higher Sharpe ratio indicates a better risk-adjusted return, which signifies a higher return than the risk taken. It implies that the investment has achieved a greater excess return compared to the risk-free rate, considering the level of volatility. Conversely, a lower Sharpe ratio suggests a lower risk-adjusted return, indicating a relatively higher level of risk for the given return. The Sharpe ratio also helps investors

compare investments or strategies based on risk-adversity performance. It provides a way to assess whether the return generated by an investment adequately compensates for the level of risk taken.

To calculate each of the strategies' Sharpe Ratios, the Bank of England's interest rate of 4.5% [57]. Table 6.10 displays the Sharpe ratios of each strategy; we see that the Constant Strategy exhibits the highest hedge ratio, followed by the Lagged and Granger Causality strategies, while the Sliding Window and Kalman Filter strategies have the lowest Sharpe ratios. This discrepancy can be attributed to the Kalman Filter strategy having a higher variance, indicating a greater level of risk compared to the Sliding Window, Lagged, and Granger Causality strategies. According to Forbes, a Sharpe ratio > 1 is considered 'good' as it suggests that the investment or strategy has generated a return of more than one unit of risk. Overall, all strategies yield desirable Sharpe ratios, indicating favourable risk-adjusted returns.

Strategy	Volatility	Average Return	Sharpe Ratio
Constant	9.28	19.14	1.46
Sliding Window	16.02	21.70	1.10
Lagged	27.04	38.70	1.29
Granger Causality	38.35	48.85	1.20
Kalman Filter	68.34	81.14	1.17

Table 6.10: Volatility and Sharpe Ratios of each strategy

6.4.3 Comparison with Previous results

Overall, the strategies seem to be a desirable investment for those with enough capital; however, how do these investments stand against other techniques? There has not been any research into statistical arbitrage on decentralised exchanges, with the only research on DEXes evaluating the performance of pure arbitrage, searching for miss-pricing and executing them. Numerous projects found there to be little to no profitable arbitrage opportunities on various decentralised exchanges, including Uniswap, regardless of volume size [58]. This happens as the pair prices are automatically adjusted based on supply and demand, making the opportunities limited, short-lived and costly. In comparison, the results from the strategies implemented in this paper do generate a profit in comparison to the cyclic arbitrage approach. The research into arbitrage opportunities on centralised cryptocurrency exchanges finds there to be more success in the returns, with some yielding a range of expected returns between 10% and 20%, which falls below even the benchmark, Constant, strategy implemented in this project [59, 60].

Other research into statistical approaches to arbitrage has been done on more mainstream assets such as ETFs; however, they have not been experimented on cryptocurrencies. One paper used the Kalman Filter on ETFs and ETNs and found there to be an average of 25.17% returns in 1 year in in-sample backtesting; however, it failed to generate any profit in its out-of-sample results [39]. The implemented Lagged, Kalman Filter, and Granger Causality trading strategies yield a greater APR than this. In addition to this, another piece of research used a combination of machine learning and the Kalman Filter on assets on the BM&FBovespa Exchange, yielding a return of 26.13% in the out-of-sample results [40]. In comparison, the Lagged, Kalman Filter and Granger Causality strategies implemented in this paper have a greater return; however, the strategy implemented by Oliveira and Nóbrega, possesses a greater Sharpe ratio meaning the risk-adjusted returns are more favourable. To summarise, three of the five strategies implemented outperform the current research on both purer forms of arbitrage, i.e. cyclic arbitrage, and the use of Kalman Filter on other types of assets, i.e. ETFs and ETNs.

Std	Pool Pair	Strategy's Annual Percentage Rate (APR) - Trading from 18th December 2021 to 9th June 2023									
		Constant		Sliding Window		Lagged		Granger Causality		Kalman Filter	
		Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades
$\#_{\sigma} = 0.1$	0	3.66	360	19.02	661	41.18	687	17.8	275	50.66	228
	1	-41.23	397	-70.78	660	-63.39	653	-40.2	388	-10.43	307
	2	3.92	454	-1.43	634	27.81	619	24.64	300	54.06	236
	3	-53.35	535	-56.31	572	-59.1	604	-33.24	375	-9.76	301
	4	7.42	375	6.72	759	36.73	669	16.18	296	44.35	227
	5	-40.45	396	-61.12	587	-54.1	593	-37.0	337	-17.72	328
	6	-11.0	72	-68.8	213	-62.28	198	-13.58	68	-4.2	84
$\#_{\sigma} = 0.5$	0	7.84	723	-11.01	1223	62.43	1185	66.93	550	130.2	378
	1	-100.0	832	-100.0	920	-100.0	936	-56.1	650	39.26	456
	2	6.14	688	-0.59	1105	73.22	1025	71.64	630	153.85	396
	3	-84.34	794	-100.0	944	-82.5	1063	-51.1	691	24.18	425
	4	16.97	826	-11.24	1243	66.69	1213	47.5	538	131.36	391
	5	-86.92	837	-100.0	854	-100.0	905	-24.3	503	54.86	388
	6	-27.3	124	-90.13	359	-71.37	301	-3.33	56	-29.04	150
$\#_{\sigma} = 1$	0	-15.08	675	-1.08	864	75.76	865	60.3	586	93.32	442
	1	-100.0	791	-100.0	814	-85.22	1022	-40.6	641	-7.2	505
	2	-14.94	535	4.72	777	86.24	721	75.65	497	103.11	404
	3	-100.0	497	-100.0	657	-64.55	872	-26.11	550	1.42	436
	4	-3.6	689	10.29	799	89.18	822	74.11	558	105.91	460
	5	-100.0	533	-100.0	612	-58.88	785	-25.29	500	26.89	408
	6	-40.18	108	-52.01	169	-34.85	125	-21.08	70	-30.59	158
$\#_{\sigma} = 1.5$	0	3.83	365	30.85	447	83.24	476	79.52	272	100.22	179
	1	-100.0	408	-53.49	509	-23.97	504	7.18	293	48.04	221
	2	-4.17	302	14.67	402	68.45	391	79.98	250	96.97	147
	3	-57.3	389	-44.96	453	-19.47	449	5.83	290	48.86	166
	4	-8.38	324	24.77	428	75.32	426	85.94	273	99.13	201
	5	-100.0	312	-42.04	402	-2.13	379	23.05	259	60.87	175
	6	-31.39	77	-29.45	87	-29.0	89	-13.84	35	-12.98	102
$\#_{\sigma} = 1.75$	0	-7.14	288	24.57	342	81.97	353	61.58	192	98.76	153
	1	-52.98	358	-20.9	293	-5.29	323	12.99	213	45.66	178
	2	-1.71	225	21.43	277	68.39	270	68.62	180	95.11	133
	3	-45.93	309	-14.98	268	-1.01	294	22.19	211	53.56	145
	4	-2.54	241	26.46	294	55.77	291	65.92	201	84.62	152
	5	-41.71	287	-9.73	225	9.79	250	20.48	200	61.97	152
	6	-35.28	76	-19.16	67	-29.96	76	-23.36	37	-16.02	74
$\#_{\sigma} = 2$	0	1.43	226	28.09	271	54.88	268	35.91	141	64.38	116
	1	-30.57	245	5.14	165	19.55	184	4.65	144	35.95	125
	2	-5.94	191	28.81	190	54.98	202	49.3	144	58.95	106
	3	-36.5	228	3.9	160	16.28	178	21.79	154	36.32	121
	4	5.85	206	29.2	190	51.79	201	46.95	149	58.28	125
	5	-19.82	192	7.19	138	21.28	147	12.38	138	42.87	125
	6	-27.18	65	-15.52	49	-17.7	54	-24.53	31	0.67	53
$\#_{\sigma} = 3$	0	1.75	54	18.68	102	17.31	83	27.63	51	17.88	51
	1	-3.91	64	13.87	49	12.0	46	20.85	60	15.28	60
	2	1.32	61	13.41	56	14.79	64	23.26	52	14.02	52
	3	-7.05	73	8.19	46	7.33	48	17.52	55	8.29	58
	4	3.13	67	20.95	57	18.84	64	26.98	60	17.25	57
	5	-4.45	53	13.73	39	12.17	39	23.27	51	17.39	57
	6	-1.17	13	-3.24	20	-2.18	25	3.19	5	8.01	16

Table 6.2: Returns of various standard deviation thresholds of each strategy

Gas Price Threshold	Pool Pair	Strategy's Annual Percentage Rate (APR) - Trading from 18th December 2021 to 9th June 2023									
		Constant		Sliding Window		Lagged		Granger Causality		Kalman Filter	
		Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades
60 th percentile = 4.7e-08	0	7.97	193	26.18	221	30.36	242	30.83	124	51.35	105
	1	-23.52	253	-0.69	178	-9.49	205	7.4	139	26.12	124
	2	8.71	134	30.05	171	26.33	169	30.82	103	51.46	77
	3	-22.44	214	0.58	165	-8.34	188	7.65	134	31.41	90
	4	14.69	150	34.45	178	24.69	184	31.4	123	47.09	90
	5	-12.54	191	6.98	128	3.85	150	15.33	124	38.32	96
	6	-27.3	53	-4.83	30	-15.56	46	-25.36	26	-6.35	50
70 th percentile = 6.36e-08	0	12.85	216	33.49	255	33.74	277	33.2	146	58.28	120
	1	-24.53	283	-2.14	209	-17.44	243	4.14	161	29.83	140
	2	12.4	156	33.28	199	25.97	200	36.35	123	61.2	94
	3	-23.07	240	0.35	193	-14.67	219	4.35	162	35.85	107
	4	14.6	175	33.64	215	23.48	219	38.4	152	59.88	111
	5	-13.43	220	6.63	152	-3.77	179	13.97	148	42.8	116
	6	-26.12	59	-6.02	37	-17.86	56	-17.84	30	-8.94	58
80 th percentile = 8.83e-08	0	12.79	239	37.89	282	42.17	298	55.47	159	48.36	127
	1	-27.28	310	-3.55	235	-17.09	267	17.17	178	19.48	148
	2	12.72	182	37.34	224	34.2	222	63.04	141	50.22	102
	3	-27.36	266	-1.32	220	-14.55	247	14.51	177	24.88	116
	4	14.38	196	37.34	241	30.37	240	63.65	164	50.73	122
	5	-14.74	239	8.61	177	-3.93	201	25.26	162	32.34	123
	6	-25.48	60	-12.8	46	-16.09	59	-16.77	32	-8.39	59
90 th percentile = 1.29e-07	0	7.19	261	35.32	307	89.98	323	52.91	176	82.2	146
	1	-40.49	336	-9.77	264	5.2	294	10.93	197	40.46	167
	2	5.12	201	31.92	249	73.17	244	63.31	161	80.48	122
	3	-34.87	287	-6.13	243	7.72	272	15.23	195	46.48	133
	4	8.38	217	32.28	267	66.24	264	58.31	187	82.52	139
	5	-29.62	265	-0.58	201	22.09	226	20.87	183	55.01	144
	6	-30.43	69	-15.39	56	-18.0	70	-18.27	34	-9.18	67
100 th percentile = 2.13e-05	0	-2.2	288	30.8	342	87.79	353	66.76	192	100.13	153
	1	-47.23	358	-16.28	293	0.91	323	16.58	213	46.78	178
	2	2.63	225	28.29	277	74.07	270	75.52	180	97.43	133
	3	-42.06	309	-9.95	268	4.71	294	25.33	211	52.93	145
	4	1.72	241	29.97	294	62.84	291	69.21	201	86.18	152
	5	-37.04	287	-4.83	225	14.37	250	23.51	200	62.31	152
	6	-33.66	76	-17.93	67	-27.78	76	-21.93	37	-14.77	74

Table 6.4: Returns of each strategy when using different gas price thresholds

Initial Investment	Pool Pair	Strategy's Annual Percentage Rate (APR) - Trading from 18th December 2021 to 9th June 2023									
		Constant		Sliding Window		Lagged		Granger Causality		Kalman Filter	
		Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades
1ETH = \$3,946.26	0	-100.0	3	-100.0	4	-100.0	2	-100.0	2	-100.0	3
	1	-100.0	3	-100.0	4	-100.0	2	-100.0	2	-100.0	3
	2	-100.0	3	-100.0	4	-100.0	2	-100.0	3	-100.0	3
	3	-100.0	3	-100.0	4	-100.0	2	-100.0	3	-100.0	3
	4	-100.0	3	-100.0	4	-100.0	2	-100.0	4	-100.0	3
	5	-100.0	3	-100.0	5	-100.0	2	-100.0	2	-100.0	3
	6	-100.0	4	-100.0	3	-100.0	2	-100.0	2	-100.0	2
5ETH = \$19,731.30	0	-100.0	72	-100.0	133	-100.0	226	-100.0	144	-92.99	146
	1	-100.0	67	-100.0	104	-100.0	162	-100.0	30	-100.0	160
	2	-100.0	79	-100.0	138	-100.0	199	-100.0	132	-87.15	122
	3	-100.0	71	-100.0	135	-100.0	165	-100.0	112	-100.0	121
	4	-100.0	63	-100.0	135	-100.0	174	-100.0	129	-88.55	139
	5	-100.0	67	-100.0	114	-100.0	147	-100.0	100	-97.5	144
	6	-100.0	37	-96.31	56	-100.0	70	-100.0	8	-100.0	35
10ETH = \$39,462.60	0	-100.0	241	-73.29	307	-5.49	323	-9.44	176	28.26	146
	1	-100.0	216	-100.0	227	-100.0	291	-56.99	197	-7.98	167
	2	-72.85	201	-55.37	249	-3.76	244	-1.41	161	28.45	122
	3	-100.0	231	-100.0	240	-73.99	272	-50.8	195	1.99	133
	4	-80.04	217	-64.22	267	-14.27	264	-11.29	187	30.22	139
	5	-100.0	221	-75.2	201	-48.32	226	-42.44	183	6.61	144
	6	-52.11	69	-37.11	56	-42.18	70	-35.21	34	-30.05	67
25ETH = \$98,656.50	0	-18.93	261	7.87	307	62.82	323	33.34	176	67.25	146
	1	-71.42	336	-33.85	264	-17.26	294	-5.83	197	28.16	167
	2	-15.96	201	7.51	249	49.6	244	45.58	161	66.77	122
	3	-58.44	287	-26.86	243	-12.37	272	-3.11	195	32.29	133
	4	-15.09	217	6.05	267	42.03	264	37.49	187	65.93	139
	5	-53.22	265	-19.49	201	2.9	226	4.43	183	40.5	144
	6	-38.19	69	-22.01	56	-26.43	70	-24.02	34	-15.81	67
50ETH = \$197,313.00	0	-0.76	261	27.61	307	82.45	323	48.52	176	78.51	146
	1	-49.33	336	-16.59	264	-1.54	294	6.43	197	37.68	167
	2	-1.27	201	25.19	249	67.31	244	59.25	161	77.32	122
	3	-41.96	287	-11.83	243	3.3	272	10.75	195	43.43	133
	4	0.85	217	25.11	267	59.95	264	53.16	187	78.63	139
	5	-37.12	265	-6.58	201	17.6	226	17.37	183	51.84	144
	6	-33.3	69	-17.81	56	-20.8	70	-20.18	34	-11.58	67
75ETH = \$295,969.50	0	4.54	261	32.75	307	87.19	323	51.42	176	80.97	146
	1	-43.77	336	-12.39	264	2.51	294	9.31	197	39.03	167
	2	2.97	201	29.4	249	70.56	244	61.79	161	79.32	122
	3	-37.01	287	-8.16	243	6.7	272	13.5	195	45.19	133
	4	6.18	217	29.66	267	64.16	264	56.71	187	81.31	139
	5	-32.1	265	-2.36	201	20.4	226	19.21	183	53.72	144
	6	-31.37	69	-16.19	56	-18.93	70	-18.9	34	-9.98	67
100ETH = \$394,626.00	0	7.19	261	35.32	307	89.98	323	52.91	176	82.2	146
	1	-40.49	336	-9.77	264	5.2	294	10.93	197	40.46	167
	2	5.12	201	31.92	249	73.17	244	63.31	161	80.48	122
	3	-34.87	287	-6.13	243	7.72	272	15.23	195	46.48	133
	4	8.38	217	32.28	267	66.24	264	58.31	187	82.52	139
	5	-29.62	265	-0.58	201	22.09	226	20.87	183	55.01	144
	6	-30.43	69	-15.39	56	-18.0	70	-18.27	34	-9.18	67

55
Table 6.5: Returns of each strategy when using different initial investment volumes (Note: Ethereum to USD conversion rate is from 18th December 2021)

Window Size	Pool Pair	Strategy's Annual Percentage Rate (APR) - Trading from 9th June 2022 to the 9th June 2023									
		Constant		Sliding Window		Lagged		Granger Causality		Kalman Filter	
		Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades	Return %	# of Trades
7 days	0	22.15	404	51.04	329	35.35	357	76.7	238	24.5	358
	1	-36.93	328	-11.1	273	-28.35	304	-1.24	316	-26.28	293
	2	21.97	365	46.2	336	36.53	332	81.15	234	15.32	362
	3	-37.98	323	-18.41	278	-27.55	293	-8.08	315	-35.33	306
	4	26.56	377	44.91	351	40.59	344	83.15	255	23.02	334
	5	-26.07	272	-6.36	231	-21.73	249	15.63	268	-21.63	239
	6	-28.77	63	-21.75	67	-23.42	65	-15.56	31	-27.23	67
14 days	0	20.66	383	11.37	293	15.97	352	40.32	182	10.84	353
	1	-31.35	318	-24.58	249	-38.7	287	-15.16	246	-34.38	277
	2	17.73	327	10.61	283	5.99	311	38.21	186	6.87	311
	3	-29.1	299	-35.08	268	-40.28	297	-20.64	256	-43.54	312
	4	20.21	346	11.44	278	5.44	308	47.6	212	12.56	305
	5	-18.27	251	-13.4	192	-29.71	242	-8.71	200	-28.11	231
	6	-24.32	42	-17.15	45	-5.46	38	-38.92	32	-19.27	45
30 days	0	21.69	289	39.05	260	75.61	270	55.53	104	10.56	323
	1	-32.04	318	-9.96	219	-4.39	238	21.46	133	-42.09	287
	2	19.44	208	39.02	203	66.67	193	57.53	85	3.43	232
	3	-26.53	274	-7.43	200	0.83	219	24.28	117	-39.38	241
	4	15.11	245	36.06	217	52.06	206	60.94	104	8.39	257
	5	-12.89	214	2.11	153	14.63	168	35.09	105	-30.14	218
	6	-9.55	24	0.45	26	11.62	16	-23.8	21	-5.29	24
60 days	0	15.48	314	48.26	275	68.23	277	81.41	108	21.07	351
	1	-28.11	303	-0.49	180	20.55	172	42.31	135	-34.98	259
	2	22.46	100	40.3	92	67.59	83	77.2	50	13.71	147
	3	-6.78	179	13.53	112	39.52	96	56.05	68	-22.18	154
	4	25.22	170	47.69	139	67.06	120	83.31	106	13.68	209
	5	7.16	141	25.9	80	48.05	76	48.54	135	-0.08	97
	6	-7.56	23	2.66	25	14.11	15	-20.72	17	-3.22	23
90 days	0	16.47	238	47.24	289	64.28	295	87.26	71	30.47	287
	1	-17.53	208	7.99	149	26.87	144	63.9	75	-15.38	176
	2	28.07	56	42.6	42	71.06	33	92.17	36	14.67	46
	3	10.76	78	30.95	45	57.68	35	67.29	53	0.61	59
	4	32.23	137	55.23	106	76.85	103	93.66	73	30.24	126
	5	12.98	91	33.68	54	55.49	56	76.41	36	7.44	61
	6	-7.56	23	2.66	25	14.11	15	-20.72	17	-3.22	23

Table 6.6: Returns of various Window Sizes

Chapter 7

Conclusion

In conclusion, the project's objective was to analyse the profitability and the investment opportunities that arise using a statistical arbitrage method, mean reversion, on decentralised exchanges. Five strategies were implemented to best estimate the hedge ratio; constant hedge ratio strategy using ordinary least squares (OLS), sliding window using OLS, lagged using OLS, Granger Causality test-based OLS model, and Kalman Filter-based hedge ratio estimation, where the constant strategy acted as a benchmark. For analysis of the performance of each strategy, a backtesting system was implemented to accurately simulate the market environments through the use of data directly obtained from the Ethereum blockchain. Furthermore, the infrastructure for a live trading system was also implemented in order to be able to run the strategy effectively on the Ethereum network.

Upon analysis using the optimal parameters for each of the strategies, the Kalman Filter results in the best annual percentage return of 81.14%, Granger Causality with a return of 48.85%, Lagged has a return of 38.70% and the Sliding Window and Constant (benchmark) having returns of 21.70% and 19.14% respectively. It can be seen that the former strategies perform well above the benchmark, making them appealing to investors. Furthermore, the returns are also not correlated to the market rate furthering its appeal to investors due to its low $|\beta|$ values.

Overall, the strategies exhibit strong performance compared to the current state of the art. However, it is essential to note that a significant initial investment, approximately 50 ETH equivalent to 86,491.50 USD (based on the conversion rate of 17th June 2023), is required to generate a substantial profit. This is primarily due to the transaction costs of executing trades on the Ethereum network. As a result, with higher initial investment, the impact of gas fees becomes relatively insignificant, leading to more profitable outcomes rather than losses. Investors with enough capital would be able to employ these strategies, making the strategies unattractive to the average investor; however, larger institutions may find these strategies appealing with the high return and good Sharpe ratio.

7.1 Future Work

7.1.1 Running the Strategies Live

The next step would be to execute and assess the real-time performance of the trading strategies. By running the strategies live, their effectiveness and profitability can be thoroughly examined, providing valuable insights into their practical application and potential for generating returns.

7.1.2 Different Decentralised Exchanges

To further the research, an intriguing avenue for further exploration would be to assess statistical arbitrage opportunities in other decentralized exchanges (DEXes). By examining the performance of mean reversion strategies across various DEX platforms, valuable insights can be gained regarding the decentralized trading ecosystem and the potential profitability of such trading approaches.

7.1.3 Different Tokens

Furthermore, it is important to acknowledge the limitations of the liquidity pool selection process, specifically regarding the availability of easily borrowable tokens. The current approach relied on tokens readily accessible through Aave, thus imposing a constraint on the pool of eligible tokens. However, as the decentralized ecosystem grows, including additional tokens and lending protocols would facilitate borrowing a wider range of altcoins, thereby unlocking a multitude of untapped arbitrage possibilities.

7.1.4 Different Blockchains

The Ethereum blockchain was chosen as the primary focus of this project due to its extensive ecosystem and diverse functionalities. However, with the rise in popularity of various other blockchains for different reasons, it becomes intriguing to explore the performance of the same trading strategies when applied to different blockchain networks. Evaluating the returns from running the strategies on alternative blockchains can provide valuable insights into the comparative profitability and potential opportunities across different blockchain ecosystems.

Appendix A

Additional Background

A.1 Pure Arbitrage

As mentioned in Chapter 2.3, research into this topic is still in its infancy thus which means a very thin slice of exploration on the subject matter. The majority of the research has been into the arbitrage on centralised exchanges [61, 59, 60]. Cristian Pauna investigates and implements an arbitrage strategy in [60]. The paper details the technical details of arbitrage trading from the data and the system architecture used. Pauna finds complications such as requesting data from multiple exchanges, converting the data such that it is homogeneous and also managing server load. Pauna presents the architecture such that the servers request data from the necessary exchanges, aggregating prices in a relational database which then triggers a server that is used to generate trading signals.

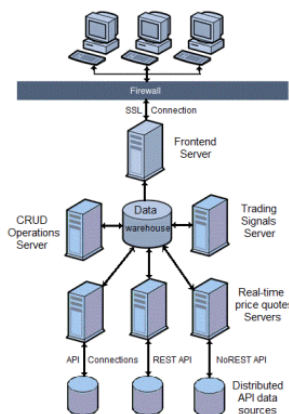


Figure A.1: Arbitrage system architecture [60]

Triangular and cyclic arbitrage is one of the most used and purest forms of arbitrage to implement and analyse, [58] explores triangular arbitrage on decentralised exchanges. Algorithm 1 is the algorithm used to find the most profitable arbitrage route on a particular platform, once this is calculated, it is compared with other routes on other platforms. Initially, the system converts the base token into another token and converts it back into the base token, using only one token is used as a middle route, then using the algorithm below, increases the number of middle tokens.

On evaluating the performance of the strategy on differing platforms depended on three main features of each exchange:

1. Portion size - Depending on how much the “trader” invested revenues differed and

Algorithm 1 Maximum Profit Route Searching (R)

Input: T (token list), P (price graph), n (current route)

```
for  $i = 1, \dots, T$  do
   $r = \text{get\_profit}(n + i)$ 
  for  $j = 1, \dots, P[i]$  do
     $p = \max(r, R(T, P, n_j))$ 
  end for
end for
return  $p$ 
```

with the larger portion size, the revenue decreases as the token pair prices are adjusted based on supply/demand.

2. Transaction fees - Each exchange has its own transaction fee.
3. Other considerations such as price slippage - Exchanges have different liquidity levels which depend on the usage and liquidity providers that the exchange employs.

Figure A.2 displays the revenues obtained by same trading token route, $\text{ETH} \rightarrow \text{MKR} \rightarrow \text{OMG} \rightarrow \text{USDT} \rightarrow \text{ETH}$. As we can see upon applying the strategy on multiple exchanges; Uniswap, 1inch, Kyberswap and Bancor, 1 inch was the only exchange that generated a profit whereas the others lose money [58].

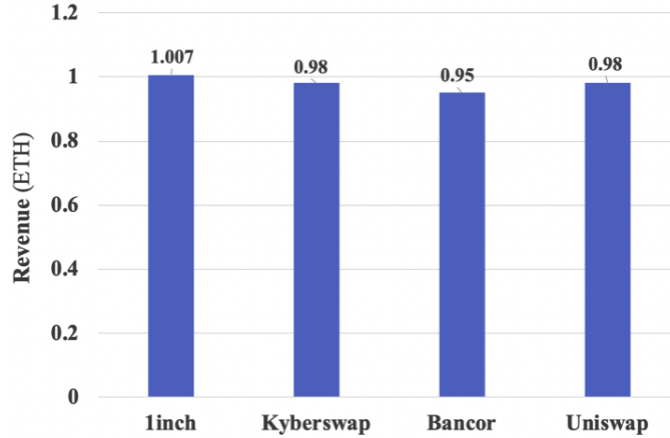


Figure A.2: Trading profits same token routes within different exchanges [58]

Another paper that implemented and evaluated a cyclic arbitrage opportunity is [54]. The research consists of proposing a theoretical arbitrage model and further evaluation of real transactional data. The arbitrage model used is simple to understand, as it searches for a cyclic transaction between n tokens, A_1, A_2, \dots, A_n is a sequence of n trades:

Trade 1: Exchange δ_1 of A_1 to δ_2 of A_2

Trade 2: Exchange δ_2 of A_2 to δ_3 of A_3

...

Trade n: Exchange δ_n of A_n to δ'_1 of A_1

It is important to note that $\delta_i = \delta_{i+1}$, i.e. the output of trade is equivalent to the input of the next. The revenues within a cycle are defined as $\delta_{i+1} - \delta_i$, and the overall profit is $\delta'_1 - \delta_1$. This is not as simple as the revenues depend on how liquid the exchange is,

thus the liquidity pools of each possible trading pair are hugely important. Therefore, the paper proposes a theorem, below:

Theorem 1 *For a given cycle $A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n \rightarrow A_1$ with n tokens, there exists an arbitrage opportunity for the cyclic transaction if the product of exchange rates $\frac{a_{2,1}a_{3,2}\dots a_{1,n}}{a_{1,2}a_{2,3}\dots a_{n,1}} > \frac{1}{r_1^n r_2^n}$ where $a_{i,j}$ denotes the liquidity of token A_i in the liquidity pool with token A_j . [54]*

In addition to the theorem, to obtain an optimal strategy we need to compute the optimal trading volume of a cycle, $A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n \rightarrow A_1$. The paper proposes the optimal trading volume to be $\delta_a^{op} = \frac{\sqrt{r_1 r_2 a' a - a}}{r_1}$ where $a = \frac{a'_{1,n} a_{n,1}}{a_{n,1} + r_1 r_2 a'_{n,1}}$ and $a' = \frac{r_1 r_2 a'_{1,n} a_{n,1}}{a_{n,1} + r_1 r_2 a'_{n,1}}$. Thus to calculate such arbitrage opportunities knowing the liquidity of tokens in other tokens' liquidity pools, algorithm 2 infers the direction and volumes to trade to get the optimal revenue.

Algorithm 2 Computing the equivalent liquidity of the cycle

```

 $a'_{1,n} \leftarrow a_{1,2}$ 
 $a'_{n,1} \leftarrow a_{2,1}$ 
for  $i$  from 2 to  $n - 1$  do
   $a'_{1,n} \leftarrow \frac{a'_{1,n} a_{i,i+1}}{a_{i,i+1} + r_1 r_2 a'_{n,1}}$ 
   $a'_{n,1} \leftarrow \frac{r_1 r_2 a'_{1,n} a_{i+1,i}}{a_{i,i+1} + r_1 r_2 a'_{n,1}}$ 
end for

```

After analyzing Ethereum block data and applying this strategy to identify the number of arbitrage opportunities, it was found that between May 4, 2020, and April 15, 2021, there were numerous exploitable and profitable arbitrage opportunities. These opportunities grew consistently to reach 1,750 in 11 months, as depicted in Figure A.3. Only cycles with length 3 were experimented with and only cycles including ETH as 80% of the liquidity pools on Uniswap include ETH and another cryptocurrency [62]. Furthermore, it is found that 287,241 of the 292,606 arbitrages executed started with ETH, and 85% of the arbitrages used a cycle of length 3. The total revenue of the cyclic arbitrage was 34,429 ETH. However, gas fees account for 24.6% of the total revenue leaving an approximate 25,971 ETH profit.

The paper then delves into the implementation of the smart contract, and explores how both *sequential* and *atomic* implementations would affect the revenue and execution of the contracts. It was found that 52.3% of the arbitrages that were executed sequentially generated a loss, likely due to the fact that, when one submits n orders, the n blockchain transactions are executed sequentially, meaning some external transactions can be inserted between these transactions. Thus using atomic transactions avoids this issue of external transactions does not affect the market price that may affect the outcome of the arbitrage.

Furthermore, the authors of the paper also investigated the performance differences between using private smart contracts and public contracts. Deploying a smart contract that calls Uniswap functions, i.e. a private smart contract, is intuitively better and achieves a higher success rate of a lower bound of 52% and a higher bound of 90% in comparison to calling a public Uniswap smart contract which has a success rate of 27.3%. Overall the paper provides an insightful look into cyclic arbitrage in DEXes and highlights important decisions made such as liquidity calculations and smart contracts while comparing the performance of different options available.



Figure A.3: Number of exploitable opportunities in Uniswap V2 over time. The purple line represents the number of cycles that provide revenue higher than 0.0001 ETH. The green represents the number of cycles whose revenue is under 0.001 ETH. The blue line represents the number of cycles whose revenue is under 0.01 ETH. [54]

A.2 Optimal Portfolio Design for Mean Reversion

There has been further research into optimizing mean reversion, one of which was to use the successive convex approximation method on the mean reverting portfolio design [63]. The paper initially proposes the mean reversion portfolio:

- For each asset, the price at time t is denoted as p_t and its corresponding log-price $y_t \triangleq \log(p_t)$, its vector form of M assets $\mathbf{y}_t \triangleq [y_{1,t}, \dots, y_{M,t}]^T$.
- The log-price spread is given by $y_t \triangleq \beta^T \mathbf{y}_t$, where $\beta \triangleq [\beta_1, \dots, \beta_M]^T$ denotes the hedge ratios.
- The cointegration space with N relations is defined by $\mathbf{B} \triangleq [\beta_1, \dots, \beta_N]$, thus the N spreads are $s_t \triangleq \mathbf{B}^T \mathbf{y}_t$.
- For these N spreads, the portfolio weight matrix is denoted as $\mathbf{w} \triangleq [w_1, \dots, w_N]^T$.
- The auto-covariance matrix for the spreads s_t is defined as $M_i \triangleq \text{Cov}(s_t, s_{t+i}) = \mathbb{E}[(s_t - \mathbb{E}[s_t])(s_{t+i} - \mathbb{E}[s_{t+i}])^T]$

Now that we have defined everything required, we can now formalize the problem. The general problem of mean reversion portfolio design problem is formalized by:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && F(\mathbf{w}) \triangleq U(\mathbf{w}) + \mu V(\mathbf{w}) + \gamma S(\mathbf{w}) \\ & \text{subject to} && \mathbf{w} \in \left\{ \mathbf{w} \mid \|\mathbf{B}\mathbf{w}\|_0 \leq L \right\}, \quad \text{where } L \text{ is the total leveraged investment} \end{aligned}$$

- μ defines the trade-off between the mean reversion measure and the variance preference.
- γ defines the regularization parameter of how sparse we would like the cointegration space to be.

Where the Mean Reversion term:

$$U(\mathbf{w}) \triangleq \xi \frac{\mathbf{w}^T \mathbf{H} \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} + \zeta \left(\frac{\mathbf{w}^T \mathbf{M}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} \right)^2 + \eta \sum_{i=2}^p \left(\frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} \right)^2$$

And the variance term:

$$V(\mathbf{w}) \triangleq \begin{cases} 1/\mathbf{w}^T \mathbf{M}_0 \mathbf{w} & \text{VarInv}(\mathbf{w}) \\ 1/\sqrt{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} & \text{StdInv}(\mathbf{w}) \\ -\mathbf{w}^T \mathbf{M}_0 \mathbf{w} & \text{VarNeg}(\mathbf{w}) \\ -\sqrt{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} & \text{StdNeg}(\mathbf{w}) \end{cases}$$

The variance term can be represented in any of the four forms.

And the asset selection term:

$$S(\mathbf{w}) \triangleq \|\mathbf{B}\mathbf{w}\|_0 = \sum_{m=1}^M \text{sgn}(|[\mathbf{B}\mathbf{w}]_m|)$$

This asset selection criterion is not necessary however as trading incurs a cost, selecting all of the assets is costly, thus selecting a subset of assets to trade is more profitable. To formalize this goal, we would like to minimize the cointegration space thus we use the ℓ_0 norm.

The paper then goes on to solve the optimization problem using the successive convex approximation (SCA) method [64]. The SCA method takes an optimization problem in the form of:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{x} \in \mathcal{X} \end{aligned}$$

Where $\mathcal{X} \subseteq \mathbb{R}^N$ is convex and $f(\mathbf{x})$ is non-convex. The SCA method involves starting at an initial point $\mathbf{x}^{(0)}$ and solving a series of subproblems of surrogate functions $\tilde{f}(\mathbf{x}; \mathbf{x}^{(k)})$ over the set \mathcal{X} . The sequence $\{\mathbf{x}^{(k)}\}$ is generated by:

$$\begin{cases} \hat{\mathbf{x}}^{(k+1)} = \underset{\mathbf{x} \in \mathcal{X}}{\text{argmin}} \tilde{f}(\mathbf{x}; \mathbf{x}^{(k)}) \\ \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \gamma^{(k)}(\hat{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}) \end{cases}$$

The first step is to generate a descent direction and then update the variable with a step size of $\gamma^{(k)}$. After applying this method to the MRP problem and further analysis of the paper, the following algorithm is proposed and used to solve the MRP design problem:

Algorithm 3 SCA-Based Algorithm for The Optimal MRP Design Problem

Require: $\mathbf{H}, \mathbf{M}_i, \mu, \gamma, \mathbf{B}, L$ and τ

- 1: Set $k = 0, \gamma^{(0)}$ and $\mathbf{w}^{(0)}$
 - 2: **repeat**
 - 3: Compute $\mathbf{A}^{(k)}$ and $\mathbf{b}^{(k)}$
 - 4: $\hat{\mathbf{w}}^{(k+1)} = \underset{\mathbf{w} \in \mathcal{W}}{\text{argmin}} \mathbf{w}^T \mathbf{A}^{(k)} \mathbf{w} + \mathbf{b}^{(k)T} \mathbf{w}$
 - 5: $\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \gamma^{(k)}(\hat{\mathbf{w}}^{(k+1)} - \mathbf{w}^{(k)})$
 - 6: $k \leftarrow k + 1$
 - 7: **until** convergence
-

However, 4 line is a convex problem and has no closed-form solution thus to solve this subproblem using the ADMM method, this is done by introducing an auxiliary variable $\mathbf{z} = \mathbf{B}\mathbf{w}$.

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{z}}{\text{minimize}} && \mathbf{w}^T \mathbf{A} \mathbf{w} + \mathbf{b}^T \mathbf{w} \\ & \text{subject to} && \|\mathbf{z}\|_1 \leq B, \mathbf{B}\mathbf{w} - \mathbf{z} = \mathbf{0} \end{aligned}$$

This is then summarized into Algorithm 4:

Algorithm 4 An ADMM-Based Algorithm for Problem on line 4 in Algorithm 3

Require: $\mathbf{A}, \mathbf{b}, \mathbf{B}, B, \rho$

- 1: Set $\mathbf{w}^{(0)}, \mathbf{z}^{(0)}, \mathbf{u}^{(0)}$ and $k = 0$
 - 2: **repeat**
 - 3: $\mathbf{w}^{(k+1)} = -(2\mathbf{A} + \rho\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{b} + \rho\mathbf{B}^T(\mathbf{u}^{(k)} - \mathbf{z}^{(k)}))$
 - 4: $\mathbf{h}^{(k)} = \mathbf{B}\mathbf{w}^{(k+1)} + \mathbf{u}^{(k)}$
 - 5: $\mathbf{z}^{(k+1)} = \Pi_{\mathcal{C}}(\mathbf{h}^{(k)})$
 - 6: $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \mathbf{B}\mathbf{w}^{(k+1)} - \mathbf{z}^{(k+1)}$
 - 7: $k \leftarrow k + 1$
 - 8: **until** convergence
-

After all of this analysis, the authors of the paper, [65, 63], ran simulations on real data comparing underlying spread. It found that it resulted in consistent profits as shown in Figure A.4.

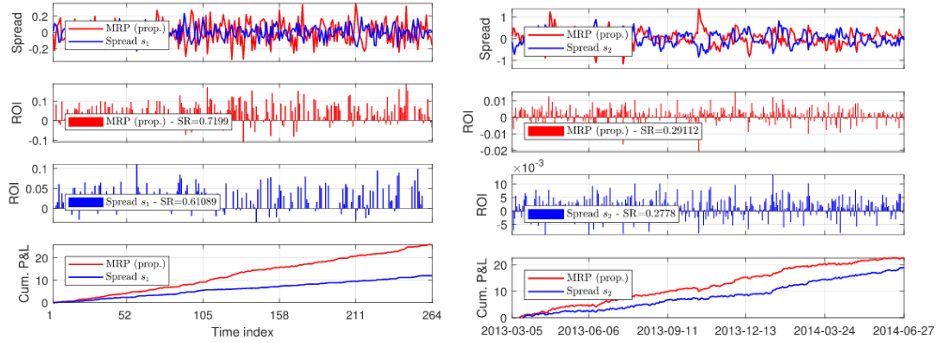


Figure A.4: A mean-reversion trading based on real data [63]

Overall, this research successfully formulizes, solves the optimization problem mathematically, and goes further to implement the algorithms to solve the problem programmatically. In addition, the author compares the implementation with other benchmark algorithms, showing that it results in a greater P&L and Sharpe ratio.

Appendix B

Supporting Tables

B.1 Exerpt of the liquidity_pools Table

pool address	token0	token1	volume in USD	created at timestamp	feetier
0x88e6a0c2d...	USDC	WETH	375230561243.465	1620250931	500
0x8ad599c3...	USDC	WETH	70454095868.0967	1620169800	3000
0x11b815efb...	WETH	USDT	62385006691.8387	1620251172	500
0x3416cf6c7...	USDC	USDT	57192593471.8346	1636825557	100
0x4585fe772...	WBTC	WETH	49170385539.9928	1620246230	500
0x4e68ccd3e...	WETH	USDT	30135014933.0963	1620232628	3000
0x60594a40...	DAI	WETH	26075053939.434	1620237823	500
0xcbcdf9626...	WBTC	WETH	21870989841.1326	1620158974	3000
0x5777d92f2...	DAI	USDC	16143305036.8948	1636771503	100
0x7858e59e0...	USDC	USDT	15473402409.0591	1620159478	500
0x99ac8ca70...	WBTC	USDC	12568187132.1649	1620241995	3000
0xc2e9f25be...	DAI	WETH	12519316091.9979	1620159368	3000
0xe0554a476...	USDC	WETH	9381529300.20357	1636926269	100
0x6c6bc977e...	DAI	USDC	7219493916.70291	1620158293	500
0xac4b3dac...	APE	WETH	6621032721.87519	1647516735	3000
0x8c54aa2a3...	FEI	USDC	6206853090.73714	1621839430	500
0x53dd58b3...	sOHM	gOHM	0	1652914688	10000
0xbc90c4de...	SHIB	NSTIC	0	1652910391	100
0xaa1297b0...	BUSD	DPC	0	1674655715	100
0x75087e533...	DAI	ICAP	0	1652899566	10000
0xc65a68019...	stkAAVE	FRAX	0	1652817927	10000
0x94589b18...	VVV	SOL	0	1674701603	10000
0xf42f0def92...	APEFI	ApeUSD	0	1652808878	3000
0xb9ba65f15...	FRAX	ApeUSD	0	1652808680	500
0x1dfb167f1...	GHD	WETH	0	1652784327	3000

Table B.1: A selection of liquidity pools on Uniswap

Appendix C

Supporting Algorithms

C.1 Gas Price Collection

Algorithm 5 Retrieval of hourly gas prices where min_time & max_time are arguments

```
rows_set  $\leftarrow$  {}
for timestamp from  $min\_time$  to  $max\_time + (60 \times 60)$ ,  $step\_size = (60 \times 60)$  do
    found_result  $\leftarrow$  False
    window_size_in_minutes  $\leftarrow$  5
    while not found_result do
        transaction_data  $\leftarrow$  GraphQL Query with arguments: {“min_time” :
timestamp - (60 * window_size_in_minutes), “max_time” : timestamp + (60 *
window_size_in_minutes)}
        if transaction_data! = 0 then
            transaction_data_sorted = sorted(transaction_data, key = lambda x :
abs(timestamp - int(x[‘timestamp’])))
            rows_set.update(timestamp : (timestamp, transaction_data_sorted[0][‘gasPrice’]))
            found_result  $\leftarrow$  True
        else
            window_size_in_minutes  $\leftarrow$  window_size_in_minutes + 5
        end if
    end while
    insert rows_set[timestamp] into gas_prices table
end for
```

Appendix D

Supporting Code Snippets

D.1 Data Collection GraphQL Queries

D.1.1 Uniswap Pool Historical Price Data Collection

```
1 query ($id: ID!, $prev_max_time: Int!) {  
2   pool(id: $id) {  
3     poolHourData(where: {periodStartUnix_gt: $prev_max_time}, orderBy:  
4       periodStartUnix, first: 1000) {  
5       id  
6       token0Price  
7       token1Price  
8       periodStartUnix  
9       liquidity  
10      feesUSD  
11    }  
12  }  
}
```

Listing D.1: GraphQL query to collect for pool pricing data

D.1.2 Uniswap Pool Live Price Data Collection

```
1 query ($id: ID!) {  
2   pool(id: $id) {  
3     token0 {  
4       symbol  
5     }  
6     token1 {  
7       symbol  
8     }  
9     token0Price  
10    token1Price  
11    liquidity  
12  }  
13 }
```

Listing D.2: GraphQL query to collect for live pricing data

D.1.3 Uniswap Historical Gas Price Data Collection

```
1 query ($min_time: Int!, $max_time: Int!) {  
2   transactions(where: {timestamp_gt: $min_time, timestamp_lt: $max_time},  
3     first:1000, orderBy: timestamp, orderDirection: asc) {  
4     id  
5     timestamp  
6     gasPrice  
7   }  
}
```

```
7 }
```

Listing D.3: GraphQL query to collect for gas pricing data

D.1.4 Uniswap Live Gas Price Data Collection

```
1 query {  
2   transactions(first:1, orderBy: timestamp, orderDirection:desc) {  
3     gasPrice  
4     timestamp  
5   }  
6 }
```

Listing D.4: GraphQL query to collect for live gas price data

D.1.5 Aave Historical Lending Pool Data Collection

```
1 query ($symbol: String!, $prev_max_time: Int!) {  
2   reserves(where: {symbol: $symbol}) {  
3     id  
4     symbol  
5     lifetimeBorrows  
6     baseLTVasCollateral  
7     reserveLiquidationThreshold  
8     borrowHistory(  
9       where: {timestamp_gt: $prev_max_time}, first: 1000, orderBy:  
10      timestamp, orderDirection: asc) {  
11        id  
12        timestamp  
13        borrowRate  
14      }  
15    }  
16 }
```

Listing D.5: GraphQL query to retrieve for lending pricing data

D.2 Backtesting Code Snippets

D.2.1 Order Execution Code

Buying ETH

```
1 amount_to_swap = self.account['WETH'] * order[1]  
2 self.account['WETH'] = self.account['WETH'] - amount_to_swap  
3 self.account['ETH'] = self.account['ETH'] + amount_to_swap
```

Listing D.6: Execution code for Buying ETH

Swapping Tokens

```
1 if is_for_token0:  
2   for swap_for_token0 in swaps:  
3     swap_token, swap_volume = swap_for_token0  
4     self.account['WETH'] = self.account['WETH'] - (swap_volume * prices  
5     [f'P{swap_token[1]}'])  
6     self.account[swap_token] = self.account[swap_token] + (swap_volume  
7     * (1 - swap_fees[swap_token]))  
8 else:  
9   for swap_for_token1 in swaps:  
10    swap_token, swap_volume = swap_for_token1
```

```

9         self.account[swap_token] = self.account[swap_token] - (swap_volume
/ prices[f'P{swap_token[1]}'])
10        self.account['WETH'] = self.account['WETH'] + (swap_volume * (1 -
swap_fees[swap_token]))

```

Listing D.7: Execution code for Swapping Tokens on a specified liquidity pool

Opening a Buy Position

```

1 self.account['WETH'] = self.account['WETH'] - (volume * buy_price)
2 self.account[token] = self.account[token] + (volume * (1 - swap_fees[token
]))

```

Listing D.8: Execution code for opening a buy positon

Closing a Buy Position

```

1 buy_token, bought_price, buy_volume, buy_timestamp = self.open_positions['
BUY'][buy_id]
2 volume_bought = buy_volume * (1 - swap_fees[buy_token])
3 self.account[buy_token] = self.account[buy_token] - volume_bought
4 self.account['WETH'] = self.account['WETH'] + (prices[f'P{buy_token[1]}'] *
(volume_bought * (1 - swap_fees[buy_token])))

```

Listing D.9: Execution code for closing a buy positon

Opening a Sell Position

```

1 # Borrow token and Deposit collateral
2 amount_to_move_to_collateral_WETH = ((volume * sell_price) / ltv_eth)
3 self.account[token] = self.account[token] + volume
4 self.account['WETH'] = self.account['WETH'] -
    amount_to_move_to_collateral_WETH
5 self.account['collateral_WETH'] = self.account['collateral_WETH'] +
    amount_to_move_to_collateral_WETH
6
7 # Swap borrowed tokens to WETH
8 self.account[token] = self.account[token] - volume
9 self.account['WETH'] = self.account['WETH'] + (volume * (1 - swap_fees[
token]) * sell_price)

```

Listing D.10: Execution code for opening a sell positon

Closing a Sell Position

```

1 sell_token, sold_price, sell_volume, sell_timestamp = self.open_positions['
SELL'][sell_id]
2
3 # Swap WETH back to Token
4 volume_required_to_return = sell_volume
5 previous_timestamp = sell_timestamp
6
7 for apy_idx in apy[sell_token].index:
8     local_apy = apy[sell_token].loc[apy_idx]['borrow_rate']
9     number_of_seconds = apy[sell_token].loc[apy_idx]['timestamp'] -
previous_timestamp
10    secondly_yield = (1 + local_apy)**(1 / (365*24*60*60))
11    volume_required_to_return *= secondly_yield ** number_of_seconds
12    previous_timestamp = apy[sell_token].loc[apy_idx]['timestamp']
13
14 self.account['WETH'] = self.account['WETH'] - (sold_price *
volume_required_to_return / (1 - swap_fees[sell_token]))

```



```

15 self.account[sell_token] = self.account[sell_token] +
    volume_required_to_return
16
17 # Return Borrowed Tokens and Collateral
18 self.account[sell_token] = self.account[sell_token] -
    volume_required_to_return
19 self.account['WETH'] = self.account['WETH'] + self.account['collateral_WETH']
20 self.account['collateral_WETH'] = 0

```

Listing D.11: Execution code for closing a sell position

Withdrawing Collateral

```

1 self.account['WETH'] = self.account['WETH'] + withdraw_amount
2 self.account['collateral_WETH'] = self.account['collateral_WETH'] -
    withdraw_amount

```

Listing D.12: Execution code for withdrawing a selected amount of collateral

Depositing Additional Collateral

```

1 self.account['WETH'] = self.account['WETH'] - deposit_amount
2 self.account['collateral_WETH'] = self.account['collateral_WETH'] +
    deposit_amount

```

Listing D.13: Execution code for depositing a selected amount of collateral

Deducting Gas Fees on OPEN and CLOSE Orders

```

1 if len([order[0] for order in signal if order[0] == 'OPEN']) == 2:
2     if strategy.should_batch_trade:
3         self.account['ETH'] = self.account['ETH'] - (
            GAS_USED_BY_OPEN_BUY_AND_SELL_POSITION * gas_price_in_eth)
4     else:
5         self.account['ETH'] = self.account['ETH'] - ((GAS_USED_BY_SWAP +
            GAS_USED_BY_SWAP + GAS_USED_BY_BORROW) * gas_price_in_eth)
6
7 if len([order[0] for order in signal if order[0] == 'CLOSE']) == 2:
8     if strategy.should_batch_trade:
9         self.account['ETH'] = self.account['ETH'] - (
            GAS_USED_BY_CLOSE_BUY_AND_SELL_POSITION * gas_price_in_eth)
10    else:
11        self.account['ETH'] = self.account['ETH'] - ((GAS_USED_BY_SWAP +
            GAS_USED_BY_SWAP + GAS_USED_BY_REPAY) * gas_price_in_eth)

```

Listing D.14: Execution code for deducting gas fees on OPEN and CLOSE orders

D.2.2 Strategy Hedge Ratio Calculation Code

Constant Strategy

```

1 model = sm.OLS(history_p1, sm.add_constant(history_p2))
2 results = model.fit()
3 # Gradient of the OLS i.e.  $X = results.params[0] + results.params[1] * p2\_token1\_price$ 
4 hedge_ratio = results.params[1]

```

Listing D.15: Calculation of the hedge ratio of the Constant Strategy

Sliding Window Strategy

```
1 def update_hedge_ratio(self):
2     p1, p2 = self.history_p1[-window_size:], self.history_p2[-window_size:]
3     model = sm.OLS(p2, sm.add_constant(p1))
4     results = model.fit()
5     self.hedge_ratio = results.params[1]
```

Listing D.16: Calculation of the hedge ratio in `update_hedge_ratio` of the Sliding Window Strategy

Lagged Strategy

```
1 def update_hedge_ratio(self):
2     maxlag = min(self.lag, min(self.window_size_in_hours, len(self.
3     history_p1))-1)
4     p1, p2 = self.history_p1[-(self.window_size_in_hours+maxlag):], self.
5     history_p2[-self.window_size_in_hours:]
6     lagged_p1 = sm.tsa.lagmat(p1, maxlag=maxlag)[-self.window_size_in_hours
7     ::]
8     model = sm.OLS(p2, sm.add_constant(lagged_p1))
9     results = model.fit()
10    self.hedge_ratio = results.params[1]
```

Listing D.17: Calculation of the hedge ratio in `update_hedge_ratio` of the Lagged Strategy

Granger Causality Test Strategy

```
1 def update_hedge_ratio(self):
2     p1, p2 = self.history_p1, self.history_p2
3     data = pd.DataFrame({'Asset1': p1, 'Asset2': p2})
4     granger_results = statsmodels.tsa.stattools.grangercausalitytests(data,
5     maxlag=[1], verbose=False)
6     self.hedge_ratio = granger_results[1][1][1].params[0]
```

Listing D.18: Calculation of the hedge ratio in `update_hedge_ratio` of the Granger Causality Test Strategy

Kalman Filter Strategy

```
1 def update_hedge_ratio(self):
2     x, y = self.history_p2[-1], self.history_p1[-1]
3     state_means_stepwise, state_covs_stepwise = self.kf.filter_update(
4     filtered_state_mean=self.means_trace[-1],
5     filtered_state_covariance=self.covs_trace[-1],
6     observation=x,
7     observation_matrix=np.array([[y, 1]]))
8
9     self.means_trace.append(state_means_stepwise)
10    self.covs_trace.append(state_covs_stepwise)
```

Listing D.19: Calculation of the hedge ratio in `update_hedge_ratio` of the Kalman Filter Strategy

D.2.3 Live Trading Code Snippets

Solidity Code for Opening and Closing Positions

```

1 function openBuySellPositions(address buyPoolAddress, bool buyZeroForOne,
    int256 buyAmount, address sellTokenAddress, uint256 sellAmount, uint256
    collateralAmount, address sellPoolAddress, bool sellZeroForOne)
    external {
2     swapExactUsingPool(buyPoolAddress, buyZeroForOne, buyAmount);
3     borrowToken(sellTokenAddress, sellAmount, collateralAmount);
4     swapExactUsingPool(sellPoolAddress, sellZeroForOne, int256(
        sellAmount));
5 }
6
7 function closeBuySellPositions(address buyPoolAddress, bool buyZeroForOne,
    int256 buyAmount, address sellTokenAddress, uint256 sellAmount, uint256
    collateralAmount, address sellPoolAddress, bool sellZeroForOne
8 ) external {
9     swapExactUsingPool(buyPoolAddress, buyZeroForOne, buyAmount);
10    swapExactUsingPool(sellPoolAddress, sellZeroForOne, int256(
        sellAmount));
11    repayBorrowedToken(sellTokenAddress, sellAmount, collateralAmount);
12 }

```

Bibliography

- [1] “I lost millions through cryptocurrency trading addiction’,” *BBC News*, May 2021. [Online]. Available: <https://www.bbc.com/news/uk-scotland-57268024>
- [2] C. Gondek, “What Big Companies Are Investing In Cryptocurrency?” [Online]. Available: <https://originstamp.com/blog/what-big-companies-are-investing-in-cryptocurrency/>
- [3] “History of bitcoin exchanges and trading - Bit2Me Academy.” [Online]. Available: <https://academy.bit2me.com/en/history-exchanges-bitcoin-trading/>
- [4] “DEX to CEX Spot Trade Volume.” [Online]. Available: <https://www.theblock.co/data/decentralized-finance/dex-non-custodial/dex-to-cex-spot-trade-volume>
- [5] M. A. Egiyi and G. N. Ofoegbu, “Cryptocurrency and climate change: An overview,” *International Journal of Mechanical Engineering and Technology (IJMET)*, vol. 11, no. 3, pp. 15–22, 2020.
- [6] M. Nofer, P. Gomber, O. Hinz, and D. Schiereck, “Blockchain,” *Business & Information Systems Engineering*, vol. 59, no. 3, pp. 183–187, 2017.
- [7] “Ethereum Whitepaper.” [Online]. Available: <https://ethereum.org>
- [8] Y.-C. Liang, *Blockchain for Dynamic Spectrum Management*, 01 2020, pp. 121–146.
- [9] “Merkle Tree in Blockchain: What is it and How does it work | Simplilearn.” [Online]. Available: <https://www.simplilearn.com/tutorials/blockchain-tutorial/merkle-tree-in-blockchain>
- [10] “Consensus mechanisms.” [Online]. Available: <https://ethereum.org>
- [11] “Proof-of-stake (PoS).” [Online]. Available: <https://ethereum.org>
- [12] V. Buterin *et al.*, “A next-generation smart contract and decentralized application platform,” *white paper*, vol. 3, no. 37, pp. 2–1, 2014.
- [13] “What are smart contracts on blockchain? | IBM.” [Online]. Available: <https://www.ibm.com/topics/smart-contracts>
- [14] “Introduction to smart contracts.” [Online]. Available: <https://ethereum.org/en/developers/docs/smart-contracts/>
- [15] “Smart Contracts - Industrial IoT Use Case Profile | IoT ONE Digital Transformation Advisors.” [Online]. Available: <https://www.iotone.com/usecase/smart-contracts/u88>
- [16] “Ethereum Virtual Machine (EVM).” [Online]. Available: <https://ethereum.org/en/developers/docs/evm/>
- [17] G. Wood, “Ethereum: A secure decentralised generalised transaction ledger.” [Online]. Available: <https://ethereum.github.io/yellowpaper/paper.pdf>

- [18] “Components of Blockchain Network,” Apr. 2021. [Online]. Available: <https://www.geeksforgeeks.org/components-of-blockchain-network/>
- [19] “Gas and fees.” [Online]. Available: <https://ethereum.org/en/developers/docs/gas/>
- [20] “What Is a DEX (Decentralized Exchange)? | Chainlink.” [Online]. Available: <https://chain.link/education-hub/what-is-decentralized-exchange-dex>
- [21] “Decentralized Exchange Platforms in Crypto Trading.” [Online]. Available: <https://www.gemini.com/cryptopedia/decentralized-exchange-crypto-dex>
- [22] “What is arbitrage?” Jul 2021. [Online]. Available: <https://online.hbs.edu/blog/post/what-is-arbitrage>
- [23] “Law of One Price: Definition, Example, Assumptions.” [Online]. Available: <https://www.investopedia.com/terms/l/law-one-price.asp>
- [24] P. Isard, “How far can we push the “law of one price?”” *The American Economic Review*, vol. 67, no. 5, pp. 942–948, 1977. [Online]. Available: <http://www.jstor.org/stable/1828075>
- [25] J. Richardson, “Some empirical evidence on commodity arbitrage and the law of one price,” *Journal of International Economics*, vol. 8, no. 2, pp. 341–351, 1978. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0022199678900272>
- [26] I. Eyal, “The miner’s dilemma,” in *2015 IEEE Symposium on Security and Privacy*. IEEE, 2015, pp. 89–103.
- [27] Z. Avarikioti, L. Heimbach, Y. Wang, and R. Wattenhofer, “Ride the lightning: The game theory of payment channels,” in *International Conference on Financial Cryptography and Data Security*. Springer, 2020, pp. 264–283.
- [28] G. Huberman, J. D. Leshno, and C. Moallemi, “Monopoly without a monopolist: An economic analysis of the bitcoin payment system,” *The Review of Economic Studies*, vol. 88, no. 6, pp. 3011–3040, 2021.
- [29] S. Athey, I. Parashkevov, V. Sarukkai, and J. Xia, “Bitcoin pricing, adoption, and usage: Theory and evidence,” 2016.
- [30] D. Easley, M. O’Hara, and S. Basu, “From mining to markets: The evolution of bitcoin transaction fees,” *Journal of Financial Economics*, vol. 134, no. 1, pp. 91–109, 2019.
- [31] C. R. Harvey, “Cryptofinance,” *Available at SSRN 2438299*, 2016.
- [32] E. Pagnotta and A. Buraschi, “An equilibrium valuation of bitcoin and decentralized network assets,” *Available at SSRN 3142022*, 2018.
- [33] C. Krauss, “Statistical arbitrage pairs trading strategies: Review and outlook,” *Journal of Economic Surveys*, vol. 31, no. 2, pp. 513–545, 2017. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/joes.12153>
- [34] H. Rad, R. K. Y. Low, and R. Faff, “The profitability of pairs trading strategies: distance, cointegration and copula methods,” *Quantitative Finance*, vol. 16, no. 10, pp. 1541–1558, 2016. [Online]. Available: <https://doi.org/10.1080/14697688.2016.1164337>
- [35] Y.-x. Lin, M. Michael, and G. Chandra, “Loss protection in pairs trading through minimum profit bounds: A cointegration approach,” *Journal of Applied Mathematics and Decision Sciences*, vol. 2006, 08 2006.

- [36] B. Alsadik, “Chapter 10 - kalman filter,” in *Adjustment Models in 3D Geomatics and Computational Geophysics*, ser. Computational Geophysics, B. Alsadik, Ed. Elsevier, 2019, vol. 4, pp. 299–326. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128175880000106>
- [37] G. Vidyamurthy, *Pairs Trading: quantitative methods and analysis*. John Wiley & Sons, 2004, vol. 217.
- [38] “How a Kalman filter works, in pictures.” [Online]. Available: <https://www.bzarg.com/p/how-a-kalman-filter-works-in-pictures/>
- [39] H. E. Dempsey, “Market Inefficiency: Pairs Trading with the Kalman Filter,” Dec. 2017. [Online]. Available: <http://www.hedempsey.com/papers/Pairs%20Trading%20with%20a%20Kalman%20Filter.pdf>
- [40] J. P. Nóbrega and A. L. I. Oliveira, “A combination forecasting model using machine learning and kalman filter for statistical arbitrage,” in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2014, pp. 1294–1299.
- [41] C. Krauss, X. A. Do, and N. Huck, “Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500,” *European Journal of Operational Research*, vol. 259, no. 2, pp. 689–702, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0377221716308657>
- [42] S. Moraes Sarmiento, *A Machine Learning based Pairs Trading Investment Strategy*, 1st ed., ser. SpringerBriefs in Computational Intelligence. Cham: Springer International Publishing, 2021.
- [43] T. G. Fischer, C. Krauss, and A. Deinert, “Statistical arbitrage in cryptocurrency markets,” *Journal of Risk and Financial Management*, vol. 12, no. 1, 2019. [Online]. Available: <https://www.mdpi.com/1911-8074/12/1/31>
- [44] D. Ju-Long, “Control problems of grey systems,” *Systems & Control Letters*, vol. 1, no. 5, pp. 288–294, 1982. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S016769118280025X>
- [45] R. Engle, “Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models,” *Journal of Business & Economic Statistics*, vol. 20, no. 3, pp. 339–350, 2002.
- [46] G. Figá-Talamanca, S. Focardi, and M. Patacca, “Common dynamic factors for cryptocurrencies and multiple pair-trading statistical arbitrages,” *Decisions in economics and finance*, vol. 44, no. 2, pp. 863–882, 2021.
- [47] B. M. Blau, T. Griffith, and R. J. Whitby, “Comovement in the cryptocurrency market,” *Economics Bulletin*, vol. 40, no. 1, p. 1, 2020.
- [48] “DeFi Exchange Data and Charts for DEXs, AMMs and Swaps.” [Online]. Available: <https://www.theblock.co/data/decentralized-finance/dex-non-custodial>
- [49] “Top 6 DeFi Lending Platforms,” Apr. 2023. [Online]. Available: <https://beincrypto.com/learn/defi-lending-platforms/>
- [50] F. Schär, “Decentralized finance: On blockchain-and smart contract-based financial markets,” *FRB of St. Louis Review*, 2021.
- [51] “How Uniswap works | Uniswap.” [Online]. Available: <https://docs.uniswap.org/contracts/v2/concepts/protocol-overview/how-uniswap-works>

- [52] “Risk Parameters.” [Online]. Available: <https://docs.aave.com/risk/asset-risk/risk-parameters>
- [53] “The Graph.” [Online]. Available: <https://thegraph.com>
- [54] Y. Wang, Y. Chen, H. Wu, L. Zhou, S. Deng, and R. Wattenhofer, “Cyclic Arbitrage in Decentralized Exchanges,” Jan. 2022, arXiv:2105.02784 [cs, q-fin]. [Online]. Available: <http://arxiv.org/abs/2105.02784>
- [55] “Granger Causality Test - an overview | ScienceDirect Topics.” [Online]. Available: <https://www.sciencedirect.com/topics/social-sciences/granger-causality-test>
- [56] “Uniswap v3 pool to exchange between usdc and weth,” 2023. [Online]. Available: <https://etherscan.io/address/0xe0554a476a092703abdb3ef35c80e0d76d32939f#analytics>
- [57] “Interest rates and Bank Rate,” May 2023. [Online]. Available: <https://www.bankofengland.co.uk/monetary-policy/the-interest-rate-bank-rate>
- [58] N. Boonpeam, W. Werapun, and T. Karode, “The arbitrage system on decentralized exchanges,” in *2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. IEEE, 2021, pp. 768–771.
- [59] T. Crépellière, M. Pelster, and S. Zeisberger, “Arbitrage in the Market for Cryptocurrencies,” Rochester, NY, Dec. 2022. [Online]. Available: <https://papers.ssrn.com/abstract=3606053>
- [60] C. PAUNA, “Arbitrage trading systems for cryptocurrencies. design principles and server architecture,” *Informatica economica*, vol. 22, no. 2/2018, pp. 35–42, 2018.
- [61] I. Makarov and A. Schoar, “Trading and arbitrage in cryptocurrency markets,” *Journal of financial economics*, vol. 135, no. 2, pp. 293–319, 2020.
- [62] L. Heimbach, Y. Wang, and R. Wattenhofer, “Behavior of liquidity providers in decentralized exchanges,” *arXiv preprint arXiv:2105.13822*, 2021.
- [63] Z. Zhao, R. Zhou, and D. P. Palomar, “Optimal mean-reverting portfolio with leverage constraint for statistical arbitrage in finance,” *IEEE transactions on signal processing*, vol. 67, no. 7, pp. 1681–1695, 2019.
- [64] Q. Tran-Dinh, W. Michiels, and M. Diehl, “An inner convex approximation algorithm for bmi optimization and applications in control,” *Proceedings of the IEEE Conference on Decision and Control*, 02 2012.
- [65] Z. Zhao, R. Zhou, Z. Wang, and D. P. Palomar, “Optimal portfolio design for statistical arbitrage in finance,” in *2018 IEEE Statistical Signal Processing Workshop (SSP)*, 2018, pp. 801–805.