

# Imperial College London

MENG INDIVIDUAL PROJECT

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

---

## Optimising Statistical Arbitrage Methods on Cryptocurrencies using Pure Arbitrage as a baseline

---

*Author:*  
Devam Savjani

*Supervisor:*  
Prof. Thomas Lancaster

*Second Marker:*  
Unknown

January 6, 2023

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Objectives . . . . .	3
1.2	Challenges . . . . .	3
1.3	Contributions . . . . .	3
<b>2</b>	<b>Background</b>	<b>4</b>
2.1	Cryptocurrencies . . . . .	4
2.1.1	Blockchain . . . . .	4
2.1.2	Decentralised Finance . . . . .	5
2.2	Arbitrage . . . . .	5
2.3	State of Art . . . . .	5
2.3.1	Pure Arbitrage Techniques . . . . .	6
2.3.2	Statistical Arbitrage Techniques . . . . .	9
<b>3</b>	<b>Project Plan</b>	<b>10</b>
<b>4</b>	<b>Evaluation Plan</b>	<b>11</b>
<b>5</b>	<b>Ethical Issues</b>	<b>12</b>
<b>A</b>	<b>First Appendix</b>	<b>13</b>

# Chapter 1

## Introduction

Hello

**1.1 Objectives**

**1.2 Challenges**

**1.3 Contributions**

## Chapter 2

# Background

### 2.1 Cryptocurrencies

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

#### 2.1.1 Blockchain

The building blocks of cryptocurrencies comes from blockchain. Blockchain is a distributed ledger that stores data, in blocks, in a chain, comprising the data itself as well has a full transaction history [33]. Below shows a diagram of blocks in a blockchain.

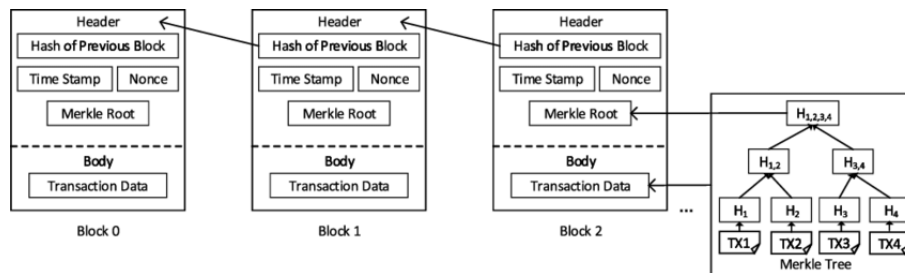


Figure 2.1: Blockchain Diagram [28]

#### Header, Hash of Previous Block and Timestamp

The timestamp and hashes of the block and its' predeccessing block are all used to ensure the ordering of blocks within a chain. By hashing the data to a fixed size, and storing in its succeeding block makes the tampering of chains difficult as it would mean the chain deviates from its old state. In addition to this, by hashing and using Nonce, blockchain employs the Proof-of-Work algorithm to ensure correctness. The Proof-of-Work algorithm is used to confirm and add new transaction to the chain.

#### Nonce

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 which miners use to validate a transaction. A miner first guesses a nonce, 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 was correct, the miner is granted the block [4].

#### Merkle root

A merkle root is also stored in each block to validate transactions in an efficient manner, in terms of storage and searching. A merkle tree is a a tree of hashes where each leaf node is it's data hash

and it's parent node, the hash of their children's hashes. In storing the merkle root, we do not need to directly store each transaction in each block, and also allows a quick search for any malicious alterations in differing blocks [2].

### 2.1.2 Decentralised Finance

One of the first application of blockchain was by Satoshi Nakamoto to create the first 'purely peer-to-peer version of electronic cash' [32]. Nakamoto's solution details the process in which a decentralised, peer to peer approach to verify and track transactions without a centralized institution.

## 2.2 Arbitrage

Arbitrage is the process in which a trader simultaneously buys and sells an asset in order to take advantage of a market inefficiency [5]. Arbitrage is also possible in other types of securities by finding price inefficiencies in the prices of options, forward contracts and other exotics.

Sources have shown that the word "*Arbitrage*" has been used as early as the Renaissance era where surviving documents showed a large amount of bills being exchanged [37]. There has also been some evidence to suggest that arbitrage was used as early as the Greek and Roman eras. Objects such as Sumerian cuneiform tablets show trade of ancient bills however we cannot come to strong conclusions of this. Early forms of arbitrage would likely to have been purchasing a commodity then transporting them to a foreign land and selling them at a higher price. This type of arbitrage is called commodity arbitrage and is still applicable today. With the example above, transporting the goods takes a significant amount of time to the merchant, trader, which could cause variations in the price, however in the modern day this has been reduced and with electronic exchanges this time to buy and sell is very small. This means inefficiencies in the market, where a trader can profit purely by buying and selling, should not exist. This is called 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 [1]. 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 [23, 39].

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, mean reversion, which uses historical means to analyse if assets have deviated from the mean to then hope that the prices will converge to its mean, 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 analyze data and trades based on statistical analysis. This project shows how these opportunities can be exploited both in a pure manner as well as using statistical methods.

## 2.3 State of Art

To better understand the project and to be able to research into something new and novel it is important to understand the current state of art, i.e. previous research on the topic. 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 [15, 9, 22, 8, 14, 19, 35]. Furthermore, there has been very limited research in comparing statistical strategies and pure methods of arbitrage of cryptocurrencies. Despite this, there has been plentiful research on arbitrage as a whole as it is immensely profitable, as a result of this

people/institutions tend to keep their newly found research secret. Of the published research, I have looked into the arbitrage techniques that are used. As arbitrage can be highly profitable, it can be found in countless types of assets, such as options, stocks, bonds and many other types of products. Research into all types of products exist going into the theory and practical aspects of each [30, 40]. The most similar type of asset class to cryptocurrencies is fiat currencies, such as the US Dollar and the Great British Pound Sterling. The research in arbitrage in foreign exchanges show that using a triangular/cyclic arbitrage is highly profitable and effective [7, 6, 24].

### 2.3.1 Pure Arbitrage Techniques

As I am using pure arbitrage as a baseline, thus choosing the most optimal strategy is most ideal to better understand the impacts of the optimizations and the statistical strategies themselves. There has been some research on pure arbitrage strategies by finding cyclic opportunities on both centralized and decentralised exchanges, which I go further into in this section.

As previously mentioned, research into this topic is still in its infancy thus which means a very thin slice of exploration on the subject matter. Majority of the research has been into the arbitrage on centralized exchanges, [29, 12, 36]. These all find massive inefficiencies within these exchanges by finding arbitrage opportunities. One of the more in depth pieces of research, Igor Makarov's and Antoinette Schoar's Trading and arbitrage in cryptocurrency markets [29] finds a large violation in the Law of One Price by finding price discrepancies between the same cryptocurrencies depending on different geolocations. The paper uses 34 exchanges in 19 differing countries, each exchange is grouped accordingly to its location and base currencies, leaving China, Japan, Korea, US, Europe and another group for that uses the Tether, USDT. Within each group the arbitrage index is calculated to compare the maximum difference in prices between exchanges within the same exchange group. This is done by calculating the volume-weighted average price at each exchange, then dividing the maximum price by the minimum price thus if the arbitrage index is 1, then there does not exist an arbitrage opportunity. It is shown that the arbitrage index is over 1 most of the time in all regions thus show a large amount of arbitrage opportunities across different exchanges, with opportunities lasting for as long as several weeks. It is also shown that the arbitrage spreads are consistent and correlated between regions and countries. Although the paper goes into some detail about how one can go about implementing such strategies and it's complications, it didn't implement them thus provides a simply theoretical hypothesis that may or may not work in practice.

Cristian Pauna implements an arbitrage strategy in [36]. 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, stores them in a relational database which then triggers a server that is used to generate trading signals.

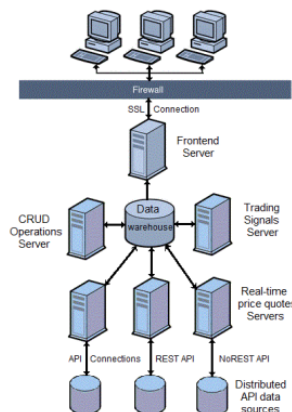


Figure 2.2: Arbitrage system architecture [36]

Now looking at DEXs, research done by Stephen Bryne in his paper, [11] introduces other methods that are commonly used however, not been academically been explored. Methods using properties of ethereum such as using smart contracts for flash loans and buying and selling the same asset in large volumes within the same block in the chain, leading to large profits. The paper details into the technical implementations and issues such as security and reliability, however it fails to provide analysis and quantitative results on the different methods used to exploit arbitrage opportunities and their differences.

As previously mentioned triangular and cyclic arbitrage is one of the most used and purest forms of arbitrage to implement and analyse, [10] 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.

---

**Algorithm 1** Maximum Profit Route Searching (R)

---

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

---

```

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

```

---

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 infact the larger portion size the revenue decreases as the token pair prices are adjusted based on supply/demand.
2. Transaction fees - Each exchange has their own transaction fee.
3. Other considerations such as price slippage - Exchanges have different liquidity levels which depends on the usage and liquidity providers that the exchange employs.

It is found that using this strategy out of the exchanges; Uniswap, 1inch, Kyberswap and Bancor. 1inch was the only exchange that generated a profit whereas the others lose money. The results are shown below on the revenues recieved on each platform experimented on.

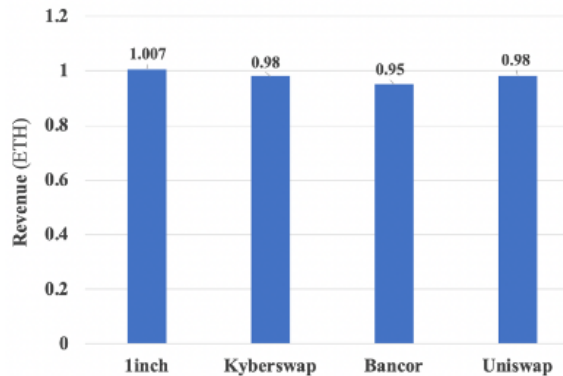


Figure 2.3: Trading revenues same token routes within different exchange [10]

Another paper that implemented and evaluated a cyclic arbitrage opportunity is [41]. The research consists of proposing a theoretical arbitrage model and further evaluation on real transac-

tional 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  $\delta_1$  of  $A_1$  to  $\delta_2$  of  $A_2$

*Trade 2:* Exchange  $\delta_2$  of  $A_2$  to  $\delta_3$  of  $A_3$

...

*Trade  $n$ :* Exchange  $\delta_n$  of  $A_n$  to  $\delta'_1$  of  $A_1$

It is important to note that  $\delta_i = \delta_{i+1}$  where  $i \in [1..n)$  and when  $i = n$ ,  $\delta_n = \delta'_1$ , i.e. the output of a 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 obviously depend on how liquid the exchange is, thus the liquidity pools of each possible trading pair is 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  $A_j$ . [41]*

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 in order 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 analysis, it is found that between 4th May 2020 to 15th April 2021, there were countless exploitable arbitrage opportunities and in fact grew to 1,750 in the 11 months that it was tested on. 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 [20]. Furthermore, it is found that 287,241 of the 292,606 arbitrages executed started with ETH, 85% of the arbitrages used a cycle of length 3. The total revenue of the cyclic arbitrage was 34,429 ETH. However, gas fees accounts 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, the paper explored 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 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 do not effect the market price that may effect the outcome of the arbitrage.

Furthermore, the authors of the paper also investigated the performance differences of 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 success rate of 27.3%. Overall the paper provides an insightful look into cyclic arbitrage in DEXs and highlights important decisions made such as liquidity calculations and smart contracts, while comparing the performance of differing options available.



### 2.3.2 Statistical Arbitrage Techniques

1. [21] -
2. [3] -
3. [34] -
4. [25] -
5. [26] -
6. [17] -
7. [13] -
8. [43] -
9. [] -
10. [38] -
11. [] -
12. [] -
13. [31] -
14. [18] -
15. [42] -
16. [16] -
17. [27] -

## Chapter 3

# Project Plan

## Chapter 4

# Evaluation Plan

## Chapter 5

# Ethical Issues

## Appendix A

### First Appendix

# Bibliography

- [1] Law of One Price: Definition, Example, Assumptions.
- [2] Merkle Tree in Blockchain: What is it and How does it work | Simplilearn.
- [3] *Advanced Studies of Financial Technologies and Cryptocurrency Markets*. Springer Singapore, Singapore, 1st ed. 2020. edition, 2020.
- [4] Components of Blockchain Network, April 2021.
- [5] What is arbitrage?, Jul 2021.
- [6] Yukihiro Aiba, Naomichi Hatano, Hideki Takayasu, Kouhei Marumo, and Tokiko Shimizu. Triangular arbitrage as an interaction among foreign exchange rates. *Physica A: Statistical Mechanics and its Applications*, 310(3-4):467–479, 2002.
- [7] Q Farooq Akram, Dagfinn Rime, and Lucio Sarno. Arbitrage in the foreign exchange market: Turning on the microscope. *Journal of International Economics*, 76(2):237–253, 2008.
- [8] Susan Athey, Ivo Parashkevov, Vishnu Sarukkai, and Jing Xia. Bitcoin pricing, adoption, and usage: Theory and evidence. 2016.
- [9] Zeta Avarikioti, Lioba Heimbach, Yuyi Wang, and Roger Wattenhofer. Ride the lightning: The game theory of payment channels. In *International Conference on Financial Cryptography and Data Security*, pages 264–283. Springer, 2020.
- [10] Naratorn Boonpeam, Warodom Werapun, and Tanakorn Karode. The arbitrage system on decentralized exchanges. In *2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 768–771. IEEE, 2021.
- [11] Stephen Byrne. An exploration of novel trading and arbitrage methods within decentralised finance.
- [12] Tommy Crépelrière, Matthias Pelster, and Stefan Zeisberger. Arbitrage in the Market for Cryptocurrencies, December 2022.
- [13] Heather E. Dempsey. Market Inefficiency: Pairs Trading with the Kalman Filter. December 2017.
- [14] David Easley, Maureen O’Hara, and Soumya Basu. From mining to markets: The evolution of bitcoin transaction fees. *Journal of Financial Economics*, 134(1):91–109, 2019.
- [15] Ittay Eyal. The miner’s dilemma. In *2015 IEEE Symposium on Security and Privacy*, pages 89–103. IEEE, 2015.
- [16] Gianna Figá-Talamanca, Sergio Focardi, and Marco Patacca. Common dynamic factors for cryptocurrencies and multiple pair-trading statistical arbitrages. *Decisions in economics and finance*, 44(2):863–882, 2021.
- [17] Thomas Fischer, Christopher Krauss, and Alexander Deinert. Statistical arbitrage in cryptocurrency markets. *Journal of Risk and Financial Management*, 12(1):31, Feb 2019.
- [18] Ahmet Goncu and Erdiñç Akyildirim. Statistical arbitrage with pairs trading. *International Review of Finance*, 16(2):307–319, 2016.

- [19] Campbell R Harvey. Cryptofinance. *Available at SSRN 2438299*, 2016.
- [20] Lioba Heimbach, Ye Wang, and Roger Wattenhofer. Behavior of liquidity providers in decentralized exchanges. *arXiv preprint arXiv:2105.13822*, 2021.
- [21] Jianfeng Huang. Triangular arbitrage across forex and cryptocurrency markets during the covid-19 crisis: a mrs-ar approach. *Applied economics letters*, 29(15):1352–1357, 2022.
- [22] Gur Huberman, Jacob D Leshno, and Ciamac Moallemi. Monopoly without a monopolist: An economic analysis of the bitcoin payment system. *The Review of Economic Studies*, 88(6):3011–3040, 2021.
- [23] Peter Isard. How far can we push the "law of one price"? *The American Economic Review*, 67(5):942–948, 1977.
- [24] Takatoshi Ito, Kenta Yamada, Misako Takayasu, and Hideki Takayasu. Free lunch! arbitrage opportunities in the foreign exchange markets. Technical report, National Bureau of Economic Research, 2012.
- [25] Christopher Krauss. Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys*, 31(2):513–545, 2017.
- [26] Christopher Krauss, Xuan Anh Do, and Nicolas Huck. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500. *European Journal of Operational Research*, 259(2):689–702, 2017.
- [27] Ladislav Kristoufek and Elie Bouri. Exploring sources of statistical arbitrage opportunities among bitcoin exchanges. *Finance Research Letters*, 51:103332, 2023.
- [28] Ying-Chang Liang. *Blockchain for Dynamic Spectrum Management*, pages 121–146. 01 2020.
- [29] Igor Makarov and Antoinette Schoar. Trading and arbitrage in cryptocurrency markets. *Journal of financial economics*, 135(2):293–319, 2020.
- [30] Jingyua Mo. A Theoretical Model of Cross-market Arbitrage.
- [31] Simão. Moraes Sarmiento. *A Machine Learning based Pairs Trading Investment Strategy*. SpringerBriefs in Computational Intelligence. Springer International Publishing, Cham, 1st ed. 2021. edition, 2021.
- [32] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. May 2009.
- [33] Michael Nofer, Peter Gomber, Oliver Hinz, and Dirk Schiereck. Blockchain. *Business & Information Systems Engineering*, 59(3):183–187, 2017.
- [34] Jarley P. Nóbrega and Adriano 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)*, pages 1294–1299, 2014.
- [35] Emiliano Pagnotta and Andrea Buraschi. An equilibrium valuation of bitcoin and decentralized network assets. *Available at SSRN 3142022*, 2018.
- [36] Cristian PAUNA. Arbitrage trading systems for cryptocurrencies. design principles and server architecture. *Informatica economica*, 22(2/2018):35–42, 2018.
- [37] Geoffrey Poitras. Origins of arbitrage. *Financial History Review*, 28(1):96–123, 2021.
- [38] Andrew Pole. *Statistical arbitrage : algorithmic trading insights and techniques*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2015 - 2007.
- [39] J.David Richardson. Some empirical evidence on commodity arbitrage and the law of one price. *Journal of International Economics*, 8(2):341–351, 1978.
- [40] Nataliya Spiricheva, Vladislav Baklykov, Mihail Tuhbadshin, and Oleg Mukhnovskiy. Implementation of stock market arbitrage. In *2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, pages 0055–0058, 2019.

- [41] Ye Wang, Yan Chen, Haotian Wu, Liyi Zhou, Shuiguang Deng, and Roger Wattenhofer. Cyclic Arbitrage in Decentralized Exchanges, January 2022. arXiv:2105.02784 [cs, q-fin].
- [42] Ziping Zhao, Rui Zhou, and Daniel P Palomar. Optimal mean-reverting portfolio with leverage constraint for statistical arbitrage in finance. *IEEE transactions on signal processing*, 67(7):1681–1695, 2019.
- [43] Ziping Zhao, Rui Zhou, Zhongju Wang, and Daniel P. Palomar. Optimal portfolio design for statistical arbitrage in finance. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, pages 801–805, 2018.