Department of Economics and Finance

Master's Degree in Finance

Chair of Econometric Theory

Dynamic Portfolio Allocation:
CRiptocurrency IndeX

Supervisor
Prof. Santucci De Magistris Paolo

Candidate
Francesco Bianco
ID.683181

Co-Supervisor
Prof. Proietti Tommaso

Academic Year 2018-2019

Table of Contents

Abstract

The paper analyse the Blockchain technology and the use of a cryptocurrencies index in the context of asset allocation. The allocation is conducted according to modern portfolio theory, once fixed the target expected returns we compute a time-varying variance-covariance matrix, using a rolling windows, and derived the optimal weights. The simple variance-covariance matrix is then compared to an EWMA variance-covariance matrix. The aim is to provide insights about cryptocurrencies investments. Results show that cryptos have no correlation, or in some cases negative, with the stock market indices. Performances of different portfolios shows that adding crypto-assets in the portfolio allocation strategy can provide boost for returns, improving the Sharpe Ratio of the investment.

1. Introduction

As the popularity for cryptocurrencies (cryptos) grows several studies have been published. Cryptos have caught the attention of many investors (retail but also institutional). Is not yet clear if the popularity of cryptos has arisen in relation to tokens itself (ICO) or for the technology behind them. Undoubtfully, the only possibility to use the blockchain technology is through the usage of tokens, cryptos.

Financial analysts and the public have labelled this phenomenon as a speculative bubble, but what happen if this speculative bubble will turns out to be an innovation in the way we execute transactions each day?

Since its emergence at the start of the decade, blockchain has been heralded as one of the most transformative technologies for financial services. Blockchain hype has led financial institutions to pour money into distributed ledger technology: about $1.7 billion annually as of 2018, per research from Greenwich Associates cited by Bloomberg.

Despite the hype, sentiment around the technology has grown increasingly skeptical as financial institutions struggle to understand the worth of their investments. Incumbents have shuttered some early experiments, and financial institutions executives are beginning to discuss blockchain's prospects in bearish terms. Key difficulties include scaling the technology for commercial application, ongoing regulatory uncertainty, and the difficulty of bringing together competing participants.

Moving forward, it's becoming more clear where exactly blockchain has value, and some players are beginning to make genuine inroads in their adoption and deployment of the technology. Those who are finding success are both pushing back against souring industry sentiment and setting themselves up as industry leaders.

In the banking sector, according to financial institutions, who have explored or are still exploring the blockchain technology, there are several benefits. The banking sector requires high security as it is one of the most attackable fields. In this context, Blockchain can eliminate the threat or the risk of fraud in all areas of banking, and this could equally apply to a trading platform. Furthermore, Blockchain would also address issues such as operational risk and administrative costs as it can be made transparent and immutable. The traceability and the permanent historic record that would exist on Blockchain backing up every asset or item of value that was traded would provide assurance and authenticity all the way through the supply chain. In the end, all the initial skepticisms are slowly disappearing.

The paper is intended to investigate the Blockchain technology and at the same time to analyze an investment strategy dividing the portfolio allocation in real-assets and crypto-assets. The study is conducted with a time-horizon of five years (2015-2020).

The aim of the study is to provide an analysis of cryptos performance and the possible effects of their implementation in portfolio asset allocation. The investment strategy is formulated on the basis of modern portfolio theory then adjustments are made. Implementation of a rolling window and a time-varying VCV matrix are analyzed. Yet, the computation of a VCV matrix following the EWMA method is implemented. To provide more insights about the resulting investment strategy four scenarios are applied, depending on the exposure of the portfolio to the crypto-assets.

The paper is structured in four chapters. Starting from Chapter 2 I am going to introduce the Blockchain's Network and its architecture, what is a Blockchain and how it is structured. We highlight the main features of a Blockchain, advantages and disadvantages, and what makes the Blockchain an invaluable tool for transaction purposes. Starting from the basic definitions this chapter is going to present in a clear and comprehensible way the structure of a Blockchain. We explore different Blockchain's models. We analyze how transactions are made and executed, moreover, all the related characteristics such as encryption and digital signature. At the end, we come to know the main differences between two of the most important protocol of consensus and what is their role in the Blockchain environment. It is important to understand what the technology behind the use of crypto assets is before introducing cryptocurrencies index. In Chapter 3 I am going to present all the inputs that will be used in Chapter 4 for the Portfolio Allocation. Starting from the different asset classes we explore how the Cryptocurrency Index (CriX) is structured and composed. For this purpose, we select asset classes that share some of the main features of the cryptos, keeping in mind that cryptos do not have intrinsic value and do not pay interests/dividends. Our portfolio will be a selection of the main indices in the economy according to the Global Industry Classification Standards (GICS). I have selected one index for each sector of the economy with the aim to reduce the exposure to the unsystematic risk. The idea, as we will see with Markowitz, is to compose a portfolio fully diversified, in such a way we are rewarded only for the systematic risk that we bear. Furthermore, I introduce a Dollar ETF in addition to the actual basket of indices with the aim of catching additional similarities that exist between the cryptos and the currencies. In Chapter 4 I am going to present the methods employed in the portfolio optimization problem. The first paragraph describes the key assumptions of the Markowitz approach and also its drawback in the application. The second paragraph introduce an alternative way in computing the variance-covariance matrix for a list of returns. This method will be used in the portfolio optimization problem to check if there is an improvement in the overall portfolios risk-return trade-off. The third paragraph explain the implementation of a rolling window to compute the realized risk-return of the investment strategy. The last paragraph is used to report all the results obtained in the application of the above-mentioned methods and at the same time explain in more details their application.

2. Blockchain

In this chapter we are going to introduce the Blockchain’s Network and its architecture, what is a Blockchain and how it is structured. We highlight the main features of a Blockchain, advantages and disadvantages, and what makes the Blockchain an invaluable tool for transaction purposes.

Starting from the basic definitions this chapter is going to present in a clear and comprehensible way the structure of a Blockchain. We explore different Blockchain’s models. We analyze how transactions are made and executed, moreover, all the related characteristics such as encryption and digital signature.

At the end, we come to know the main differences between two of the most important protocol of consensus and what is their role in the Blockchain environment.

It is important to understand what the technology behind the use of crypto assets is before introducing cryptocurrencies index.

2.1. Overview and definitions

We can find many definitions of the blockchain. Some focus on its structure, others on the technologies behind it or the implications for business and society. All these aspects are equally important and contribute to give a comprehensive overview of the topic.

The blockchain is a digital ledger, decentralized and distributed over a network, structured as a chain of registers ("blocks") responsible for storing data (from value transactions to entire digital applications).

It is possible to add new blocks of information, but it is not possible to edit or remove blocks previously added to the chain. In this ecosystem, encryption and consent protocols ensure security and immutability.

The result is an open, neutral, reliable and secure system, where our ability to use and trust the system does not depend on the intentions of any individual or institution. The blockchain is more than just a payment infrastructure, a supply chain monitoring system or a digital identity manager.

It is a system with the potential to bring a new level of confidence in applications, introducing a paradigm shift in the way they are implemented and giving us the opportunity to innovate freely.

2.1.1 Ledger, database and blockchain

Starting from the concept of information we can say that one of the main purposes of a blockchain is to save information. The information saved can be of any kind, from a simple transaction of an asset to entire program (smart contract).

At the center of the blockchain there is the concept of recording transactions within the ledger, a traditional ledger with the ability to record transactions of each category of asset, from currencies to real estate properties.

The blockchain is a digital ledger. Ledgers and databases may seem very similar. At the base of both technologies there is in fact the idea of saving data, but while in a database you can enter, delete and edit data, in a ledger you can only add new information. This is made possible by a combination of various factors, including decentralization, cryptography, game theory and other concepts.

Many of the properties that characterize the blockchain make this technology attractive for different scenarios.

For example, a traditional database requires a system of controlled access, the management of which is entrusted directly to known and reliable individuals. A blockchain, on the other hand, can be used by unknown and not "trusted" parties, without the need for any form of access control.

As a result, the blockchain is very useful in scenarios where trust, security and immutability are key requirements.

In a blockchain, the digital ledger is structured as a chain of blocks, each of which is responsible for storing information, such as transaction logs or programs.

2.1.2 The blocks, bricks of the blockchain

Blocks are structures of data added to the blockchain sequentially, one block at a time. Each of them contains a mathematical proof, generated through the use of cryptography, which assumes the sequentiality of the previous block, resulting in a "chain of blocks". The first block of each blockchain is called a "genesis block".

The hash of the block is not usually saved in the block but is calculated whenever it is needed. Depending on how the system is designed, the blocks may have different sizes and store various types of information.

The connection between blocks is generated by means of an encryption function. (cryptographic function of hash), which creates an indissoluble mathematical link between them.

2.1.3 The hash function

This function is used to map data of arbitrary size into data of fixed size. In other words, the input of a hash function can be almost anything (an mp2 file, pdf, spreadsheet, etc...) but the output, called "hash", will always have a finite number of bits.

We can summarize some of the mathematical details of the hash function in the following list:

- The same input always produces the same output (deterministic function), i.e. a hash, which has the shape of a string of letters and numbers.
- Even the slightest change in the input produces a drastic change in the output of the function.
- It is a unidirectional function: it is computationally very easy to generate a hash from any input, but it is very complex to calculate the input from the hash (i.e. calculate the inverse function). There is no way to switch from hash to input except by trying all possible combinations (brute-force method).

SHA-256 is one of the most common hash functions. You can think of a hash as the fingerprint of a digital file.

Since a small change to the input completely alters the hash, once calculated the hash of a file, if the file is modified also the relative hash would undergo some changes.

Blockchain systems make frequent use of the hash function, as it provides a very convenient way to express the entire state of the blockchain in a single string of defined length. For each new block generated, the hash of the previous block is inserted into the input to generate the hash of the new block. In practice, each block contains information, data and hash from the previous block.

Therefore, if someone tries to add, remove or modify some information in any block, they will change the hash of the block and consequently all the subsequent hashes, and this will be immediately evident.
Bitcoin's blockchain currently occupies over 242.4 GB, but the entire blockchain can be represented by a single hash.

To evaluate the current status of a blockchain we do not need to analyze the entire content every time: just look at the hash of the last block. This is extremely useful in evaluating different versions of the same blockchain.

2.2. Blockchain Network

One of the main aims of the blockchain technology is to allow anyone, anywhere in the world, to carry out transactions without the need to rely on a central institution (in the case of monetary transactions, a bank). To do this, the blockchain must be distributed over a network.
We can define a network as a group of interconnected machines that exchange information through communication channels, such as the Internet.
A machine connected to a network is called a node.

The knots in a blockchain

Every machine connected to the blockchain network is a node. It is possible to make a distinction between:

- **Full Node**: downloads and stores locally a complete copy of the blockchain and checks that each transaction and block follow the rules defined by the system. If an anomaly occurs, the block (or transaction) will always be rejected, even if it is considered valid by any other node in the network. A full-node is effectively independent. He doesn't need to trust any other node and follows the rules regardless of everything, propagating valid blocks and transactions, ignoring invalid ones. Using a full-node is the safest way to interact with a blockchain, but it can be quite uncomfortable, as it requires downloading the entire blockchain (in the case of the Bitcoin blockchain we're talking about over 242.4 gigabytes in December 2019).

- **Light-node**: does not store the entire blockchain but only receives the data it needs from a trusted node (a full-node). Consequently, the use of this type of node implies the delegation of trust to a third party (the full-node), in exchange for the simplicity of use. The average user typically uses a light-node and does not have the ability to independently verify the correctness of the data.

A light-node can be, for example, a wallet on a mobile device.

In addition to the decentralization of data, one of the reasons that led to the birth of blockchain technology was the search for a system that was free from the constraints and errors that characterize human beings. A full-node categorically follows the rules imposed by the system, regardless of the decisions of all other nodes.

It follows that it is an intrinsically free system from the problems that have always plagued centralized institutions, such as corruption or a lack of impartiality in the choices made.

2.2.1 Network architecture

The network is a fundamental component in a blockchain system. Based on the network structure and the role of each node, three network models can be identified: centralized, decentralized and distributed.

Centralized and decentralized networks

The degree of centralization is a concept through which it is possible to analyze a system at different levels. We will group the systems according to their centralization from the point of view of architecture, authority and logic.

Architecture

An architecture-level **centralized network** is an infrastructure with a single point of failure that, if compromised, would prevent the entire system from functioning properly.

In a **decentralized network**, resources are distributed and possibly replicated in the network nodes and, consequently, an application is executed by all its participants without generating a single point of possible infrastructural failure. In other words, in order for a decentralized system to stop working, it is necessary to "shut down" all the nodes that compose it.

Authority

A network subject to **centralized authority** is characterized by a central body which controls its data, operations and users. It defines all the rules of the system and has the power to apply these rules to its users, deciding accordingly what is right and what is wrong, requiring the unconditional trust of users.

In a network with **decentralized authority**, there is no central authority and all nodes are considered equal. No one has control of the network and consequently no one can prevent actions or force censorship of content.

A blockchain is characterized by **decentralized authority**. No central authority has control over them.

Logic

A **logically centralized** network must be identified at all times by a single state to function properly. Therefore, it is necessary that all participants agree on the status of the system. There is therefore a single logical state on which all participants agree. The classic example is a global central database in which all data is saved and kept consistent.

In a **logically decentralized** network, there can be several copies of the data and any node can modify its own copy without altering the normal operation of the system. For example, in the case of emails: if I delete an email in my mailbox, I do not delete it in the mailboxes of other people to whom I have sent it.

A blockchain is **logically centralized**. It is always characterized by a single logical state.

Distributed networks

In a distributed network, data and computations are distributed over multiple nodes, but authority can remain centralized. To minimize the risks and complexities of management, distributed networks do not have a single, huge, server or database, but several data centers scattered around the world.

2.2.2 Blockchain Models

The blockchain can be used in any scenario where a logically centralized global state but a distributed and decentralized system structure is required.

A logically centralized state is fundamental for the operation of each blockchain. However, in some cases the authority in a blockchain may not be decentralized but tend to centralize. Specifically, depending on how the authority is managed, there are three models of blockchains: public, authorized and private.

Public blockchain (permissionless)

The public blockchain model is currently the best known and most used. A public blockchain is a system with:

- Decentralized architecture,
- Decentralized authorities,
- Centralized logic.

Decentralization is a key aspect of this model, as any attempt at centralization would introduce a weakness into the system and expose a potential point of failure or control. There is no single authority. Everyone can join the (open) network and there is no possibility of being excluded (resistance to censorship).

An open blockchain does not discriminate on the basis of origin, destination or content (neutral). Each knot has equal rights and responsibilities. Everyone has the possibility to explore and verify each transaction (public and analyzable).

Usually, an open blockchain is also open source, making publicly available and searchable the code that regulates its operation. This allows everyone to check that it is correct or to suggest improvements. When we talk about blockchain, we usually refer to public blockchain.

Private blockchain (permissioned)

Although public blockchains have unique properties, these same properties can make them unsuitable in some contexts, for example in industry.

Private blockchains sacrifice complete decentralization in exchange for control over access permissions and usually better performance.
Private blockchains have a level of access control controlled by one or more authorities.

The level of access verification has the task of deciding who can read/write the data on the blockchain and who can participate in the transaction verification process. The system is only considered reliable if the actors chosen for the verification process are reliable.

Typically, there is a distinction between **completely private blockchain** and **consortium**, where in the former the control and authority is concentrated in a single entity, while in the latter it is distributed among the participants of the network. Since completely private blockchains totally remove decentralization and with it most of the specific advantages of technology, the ones that are most interesting are the blockchain consortiums, since they present themselves as a hybrid solution between public and completely private blockchains.

For governments, institutions or companies, both models may be more convenient, especially when a certain degree of control over the data or participants in the system is required, when a regime of autonomous collaboration between different companies is to be established, or when sensitive data is to be kept confidential.

2.3 Transactions

Encryption is the study of secure communication techniques in a hostile environment (such as the Internet). The blockchain is a system in which cryptography (specifically public key cryptography) occupies a prominent place.

Public key cryptography is a widely used cryptographic system on the Internet and plays a key role in many of the processes involving the blockchain. As we will see, the **addresses** on the blockchain are generated using this cryptographic system and transactions are authenticated using **digital signatures** - i.e. one of the most popular applications of public key cryptography.

The basic idea is to use a pair of keys in mathematical relation between them:

- a randomly generated **private key** that has to be kept secret;
- a **public key** mathematically derived from the private key, which can be shared with anyone.

The keys are nothing more than extremely large numbers, usually represented in hexadecimal (0-9 to represent numbers from zero to nine and 1-f to represent numbers from ten to fifteen).

In Bitcoin, for example, the private key corresponds to a number that occupies 256bit (a sequence of 256 one and zero). The largest number that can be saved in 256 bits is 2^256. To make the idea 2^256 is approximately 10^78, while the number of atoms in the observable universe is approximately 10^80^20. Generating two identical private keys, although mathematically possible, is extremely unlikely.

Generating a public key from a private key is computationally very easy, but reversing this operation is virtually impossible. With the most powerful supercomputers around today, it would take millions of years.

Public key encryption can be used to ensure certain properties such as **encryption**, **authentication**, **integrity**, and **non-repudiation**, in an unsecured environment such as the Internet.

Encryption

Encryption is a process by which a message, or information in general, is encrypted so that only authorized persons can access the original information.

The most illustrative use of encryption is to hide a message so that it cannot be read by unauthorized persons. Once a message is encrypted using a public key algorithm, this message can be transferred through an unsecured channel such as the Internet, but still ensuring the confidentiality of the message.

To do so, the message is encrypted using the **public key** of the person who is to receive this message. In this way, only the owner of the private key connected to that public key will be able to decrypt the message.

However, there is a problem: an encrypted message hides the content from all those who do not have the private key, but it is still possible to **modify** this message without actually understanding what is written there. To solve this problem, other techniques, such as hashing, are used in conjunction with encryption.

Hashing and encryption

Someone might find similarities between hashing and encryption. They actually have very different purposes, although they are often used together.

Encryption encrypts data with the main purpose of ensuring confidentiality. If data is encrypted using public key encryption, it can only be decrypted with the associated private key.

A hash function, on the other hand, is not designed to encrypt a message and cannot be reversed (unidirectional function).

Digital signature

Digital signatures, like traditional signatures, are a way of demonstrating someone's identity without their physical presence, with the difference that mathematics is used instead of a manual signature. Digital signatures are created with a combination of hashing and public key encryption.

With a digital signature you can get:

- **Authentication**: A private key is linked to a specific user. A valid signature unequivocally proves that the message was sent by that user. Authentication does not require to know the true identity of the user, but requires providing information related to his identity (the private key)
- **Integrity**: if a message is digitally signed, any change to the message after the signature invalidates the signature itself (this is a property derived from hashing).
- **Non-Repudiation**: if someone signs a message, they cannot, at a later date, deny that they have signed it.

All these properties are valid as long as the private key remains private.

2.3.1 Execution and Fees

A (valid) transaction is the elementary unit of information that is written on the blockchain.

A valid transaction implies a change of state in the blockchain. The blockchain is logically centralized, i.e. there must be a single state that is considered valid by the network. A transaction generates a new status. Transactions can be monetary, such as sending bitcoins, or involve other digital assets (stocks, property certificates, etc...).

Deterministic transactions

A transaction can be valid and then change the status of the blockchain or be invalid and leave the blockchain in its current state. For this reason, it is said that a transaction is an **atomic operation**, i.e. it cannot generate an intermediate state.

A transaction is immutable. Just as it is not possible for a valid transaction to be rejected, it is not possible to modify a transaction once it has been accepted.

We are normally used to the possibility of cancelling a transaction (for example in banking transactions or with PayPal). It is sometimes a convenient feature, but it presupposes a system that is anything but immutable.

In a blockchain, if you create a valid transaction, it is not possible for anyone to delete, cancel or modify it. The transaction will be executed and will change the status of the blockchain.

However, it is possible to add one or more conditions to a transaction, for example to decide to confirm a payment only after having satisfied specific constraints (as in the case of smart contracts). If the conditions are met, all parties involved in the transaction will know for certain what will happen. It is not possible to change the result.

Create a transaction

The basic requirement for creating a transaction on a blockchain is to have the object of the transaction. In a fully digital system, this is possible thanks to digital signatures.

The digital signature guarantees that:

- The address that created the transaction belongs to the user. The transaction is signed with the user's private key (**authentication**);
- The transaction was not changed after signature (**integrity**);
- The user who owns the private key (used in the transaction) cannot deny having created the transaction (**not repudiation**).

We also remember that in a cryptocurrency transaction there is no physical transfer of money, as these are accounting entries of a digital ledger. A transaction simply records in the ledger the amount transferred from the sender to the recipient.

Once the transaction has been created and signed, it can be propagated to neighboring nodes, which have the task of verifying its validity and deciding whether to propagate it further or not.

The valid transaction is then propagated to the network nodes but is not yet immutably recorded on the distributed ledger (the blockchain).

Confirmations

Before we go into how the nodes decide whether a transaction is valid or not, let's briefly explain what the confirmations are in a transaction.

Transactions are grouped in blocks. Blocks are added to the blockchain sequentially. Each transaction must go through a verification process before it is included in a block. Before being added to a block, a transaction is unconfirmed. Once a transaction is included in a block has 1 confirmation When the next block is created, the same transaction has 2 confirmations, and so on.

The number of transaction confirmations corresponds to the number of blocks subsequent to the one in which the transaction is included.

Once enough confirmations have been obtained, a transaction cannot be cancelled/modified by anyone.

Transaction fees

Usually a transaction includes a commission. Transaction fees correspond to the cost of making a given transaction and are used to reward the miners (we'll see who they are later).

These are therefore commissions that the sender may have to include in his transaction for it to be successful.

Each blockchain has its own system for determining transaction fees. The transaction fee is decided by the sender, may be zero in some cases and is not related to the amount transferred. The commission usually influences the time it takes for a transaction to be confirmed. Especially in times of particular congestion, a zero-commission transaction may require the generation of several blocks before being included and verified accordingly.

2.3.2 Addresses and Wallets

On a blockchain there are no user profiles, but rather addresses. Addresses do not contain encryption but are only identifiers that represent the destination of a transaction.

It is important to remember that a blockchain is just a list of transactions, there is no concept of currency as a physical object that must be kept somewhere. Coins are only accounting items and the final balance of an address is a calculation made by examining all transactions involving that address.

Addresses are identifiers used to transfer digital assets.

The purpose of an address is to enable transactions to (and from) a single entity. It is possible to have numerous addresses that can be shared freely without any security problems, just as it is safe to share a public key.

Generating an address

From a technical point of view, an address is the result of a mathematical operation involving public key cryptography and hashing.

1. First, a private key is generated. It is essential that the private key is generated by a random number, otherwise a critical vulnerability could be created.
2. The private key is derived from the corresponding public key by means of a mathematical process.
3. The public key is passed through a series of cryptographic algorithms (different types of hash functions) to get an address on the blockchain.

Multisignature address (multiple signature)

Multisignature is a technique used to increase transaction security, where more than one signature (i.e. more than one private key) is required to authorize a transaction. A multisignature address is an address associated with more than one private key. Addresses of this type are usually referred to as "m-of-n": at least m out of n total private keys are required to make a transaction. The private keys can all be in the possession of the same person (but obviously kept in different places) or belong to different people.

This creates an address where the property is shared, and it is necessary to have the consent of more people to carry out transactions... This scenario is very common in companies to prevent someone from creating wrong transactions or worse stealing funds.

Wallet

Addresses are generally managed using specific tools called wallets. Unlike traditional portfolios, however, they do not contain money.

A wallet stores the public and private keys of an address and can be seen as "your account". The data corresponding to the addresses are always stored on the blockchain. It is therefore not possible to lose the cryptocurrencies, but only to lose the private keys that give access to those cryptocurrencies.

Usually wallets also provide an interface to track the final balance of all addresses owned by a user and automate certain functions such as signing transactions or suggesting commissions for a transaction. There are three main types of portfolios: **software**, **hardware** or **paper**. In addition, depending on the environment in which these wallets operate, it is possible to make another distinction between **cold storage** and **hot storage**.

Hot storage and cold storage

A hot storage wallet is a wallet that is somehow connected to the Internet, i.e. the private keys have been created or are currently stored on a machine connected to the Internet. On the contrary, a wallet cold storage refers to a wallet whose private keys have never been in contact with the Internet.

A cold storage solution is extremely secure, as it is much more difficult to steal something that is not connected to the Internet.

Paper wallet

A paper wallet is the simplest possible form of cold storage. Basically, it is the private key- address pair printed on a piece of paper. The security of a paper wallet is directly related to the security of the place where the sheet of paper is stored.

Software wallet

A wallet software is an application that can be installed on a computer or smartphone. The private key is encrypted with a password and stored on the machine itself. Wallet software is often chosen for its ease of use. However, if the car on which they are installed were compromised, the private keys could be stolen.

Hardware wallet

A hardware wallet stores private keys in a physical device (hardware). It has great security advantages over wallet software, as private keys are stored in a secure area of the device from which they cannot be extracted. Transactions are signed within the device itself and therefore, even if the wallet were connected to a compromised machine, private keys would remain safe.

Hardware wallets are currently the best compromise in terms of security and ease of use.

Backup and HD wallet

Usually the first time a wallet (hardware or software) is used, a list of words to be saved is communicated. This list, called "passphrase", allows you to restore the wallet and regain access to it. This does not mean that the private keys are stored elsewhere.

In these types of wallets, called **HD** (Hierarchical Deterministic) **wallets**, the passphrase is the point of origin from which private keys are generated. The words that make up the passphrase represent randomness. These wallets implement a system to derive keys from a single starting point known as "seed" (specified in BIP 32 and BIP 39, Bitcoin Improvement Proposal).

The seed allows the user to perform wallet recovery without the need for additional information.

If a computer, hard drive or hardware wallet were destroyed, it would easily be possible to restore private keys on another device by simply re-inserting this seed on another device.

An example of a passphrase could be: "wild never seat speak jazz lumber length oppose ignore house fence invest". It is important to remember that the passphrase is equivalent to private keys and that it must be kept with the same care.

2.4 Consensus and mining

Computers and software are far from being perfect systems: they can crash, be hacked, behave negatively on purpose or even behave in a pseudo-random way. When we connect several computers together in a network, the uncertainty of the final system increases exponentially.

In a blockchain there could be millions of nodes that work independently, and it is not possible to predict how each of these nodes will behave. In a permissionless blockchain you can't trust any entity involved.

Consent

Despite the uncertainty, the knots of a blockchain must come to an agreement on a single state. A blockchain is based on (mathematical) rules but has no rulers. The network has the task of reaching a decision on what happened within the blockchain, through a process called consensus.

Consensus is a general agreement between the members of a given group (the nodes of the blockchain), each of whom has a part of the decision-making power.

In a blockchain the consent is an agreement on what happened and holds the only possible truth about the current state of the blockchain.

Consensus, however, should not be understood as a discrete process where there is no consensus at a moment and the moment after consensus is reached, but rather as a continuous process that involves several participants, each with its own roles and responsibilities. As we will see below, the two main actors in this process are the full-node and the miner. We can say that the consent of a blockchain is the guarantor of the trust we place in this system.

A blockchain uses mathematics, economics and game theory to encourage all actors to reach an agreement on a single state. However, achieving consensus in a distributed and decentralized system remains a very complex issue.

Mining

Mining is a general concept and is not related to any particular blockchain. It can be seen as a process that allows the blockchain network to validate transactions, group them into blocks and add them to the block chain. These operations make it possible to reach distributed consent and make the network secure.

The nodes that take part in the mining process are called miners. More generally, mining can be seen as the decentralized mechanism through which distributed consensus is reached and network security guaranteed.

We talked about consensus as a continuous process in which miner and full node work to add new blocks to the blockchain and verify the validity of these blocks. In detail, a miner is responsible for:

- Together with the nodes, check that the transactions are valid and if so, propagate them to the rest of the network;
- Together with the nodes, check that the new blocks are valid and if so, propagate them to the rest of the network;
- Choose transactions, sort them, and aggregate them into a block.

A full-node is responsible for:

- Check that the transactions are valid and, if so, propagate them to the rest of the network;
- Check that the new blocks are valid and if so, propagate them to the rest of the network.

A full-node therefore contributes to the security of the blockchain by checking the validity of each transaction and each block, so as to ensure that the miners do not "cheat".

A full-node is the safest way to use the blockchain. A full-node will never accept a transaction or block that does not comply with the rules. If a miner creates an invalid block, the other nodes will reject it. When a miner's block is added to the blockchain, it is rewarded for the work done according to the rules defined in the blockchain.

Usually the reward consists of the transaction commissions of the block and eventually, as in the case of Bitcoin, of the cryptocurrencies generated at the addition of a new block.

2.4.1 Proof of Work (PoW)

Proof of Work is a protocol used in the process to reach distributed consensus. In concrete terms, the Proof of Work is based on the search for a number that is computationally difficult to find, but once found it becomes easy for all the other nodes to verify its correctness. In a system that uses PoW, a block is only valid if it contains a valid PoW solution.

Proof of Work is a protocol used to reach distributed consensus in which voting power is based on computational power.

Hash in the PoW

The value to be found in the PoW seems to have all the characteristics of a hash, and in fact the PoW is based on hash algorithms. Let's make a brief recap of the characteristics of a hash function before delving into the mechanism of the PoW.

A hash function takes as input an arbitrary length value and transforms it into an output of defined length. It is also a non-reversible function, i.e. given a hash, the only way to know the input that generated that hash is to try all possible inputs (brute-force).

PoW Mining

In PoW-mining, network nodes compete to solve a complex mathematical problem (an inverse hash with some constraints). Solving this problem is a random process with very low probability and the only way to find a valid PoW is to try all possible combinations until you find the right one.

The first miner to solve the problem has the right to create the next block and earn the reward. Once a new block is created, it is transmitted to the network, waiting for the other nodes to verify its validity. It is very easy for the remaining nodes to check if the solution is correct. If the block is valid, it is forwarded to nearby nodes, otherwise it is ignored.

Mining in a PoW-based system can be summarized as follows:

1. Transactions are created and transmitted to the entire network of nodes.
2. Each miner selects the transactions they want (usually those with the highest commissions) and collects them in a block called candidate block, as it is not yet valid, not having a valid solution to the PoW.
3. Each miner begins to perform calculations to find the solution to the mathematical problem and generate a valid PoW for the block he assembled. For each invalid solution, the miner changes the value of a number, called a nonce, that is added to the PoW input to change the final value of the solution.
4. When a miner generates a valid PoW for the new block, it transmits the block to the network.
5. All nodes in the network check whether the new block is valid or not.
6. If the block is considered valid, the miner wins the block (and the commissions of the transactions contained therein). The new block is forwarded to the network of nodes and added to the blockchain.

Hashrate

In the PoW, when we talk about computing power, we refer to hashrate, because usually the problem to be solved is an inverse hash with some constraints. Hashrate is the number of hashes calculated per second (H/s).

The total hashrate, or network hashrate, is the sum of all the miner hashrates. The probability of a miner finding a valid PoW first is as follows:

$$P = \frac{Hashrate\ of\ the\ miner}{Hashrate\ of\ the\ Network}$$

The hashrate depends on the specific hash algorithm used by the blockchain and the power of the machine used by the miner. Considering the Bitcoin, for example, a person has a hashrate of about 0.00003 H/s, which means that calculating a single hash by hand would take about 9- 10 hours. An ASIC (Application Specific Integrated Circuit) miner can calculate more than 14 TH/s (Tera Hash, one trillion hashes per second). In 2019, the hashrate of the Bitcoin blockchain network amounted to more than 100 million TH/s.

Nonce

Nonce is a value used to vary the input of the hash function used in the calculation of the PoW. This value is changed until the resulting hash meets a specific value called **difficulty**.

Difficulty

A PoW to be considered valid must satisfy a constraint called difficulty. Difficulty is a value that expresses how difficult it is to find a valid PoW.

In practice, you can set the target of the difficulty by requiring that the hash to be found starts (for example) with 5 zeros. That is, it means that the first 5 values of the hash will have to be all 0 to satisfy the target of difficulty. In general, the more we increase the number of zeros, the more difficult it becomes.

One of the checks that nodes make when they receive a new block is to check that the difficulty in the PoW of that block respects the constraints on the difficulty. In the case of Bitcoin, a block generated in October 2018 had 19 zeros.

The difficulty is periodically updated (retargeting) in relation to the hashrate of the network to keep the time necessary for the generation of a block as constant as possible. For example, in Bitcoin the difficulty is adjusted every 2,016 blocks (about 14 days) based on the average time it took to find the previous 2,016 blocks.

Reward

Rewards are provided to encourage the miners to generate new blocks and keep the network secure. Miners who create a new block are rewarded with all the commissions of the transactions included in the block, plus possibly the new coins (crypto currencies) created together with the block (block-reward). Usually the number of new coins created with each block decreases over time, since most crypto currencies have a limit in the maximum number of existing coins (in Bitcoin this limit corresponds to 21 million bitcoins).

The PoW protocol is fair to miners: a miner who owns 5% of the total computing power of the network on average "wins" the PoW and gets the right to create a new block (and earn the reward) 5% of the time.

PoW: pros and cons

The main advantage of the Proof of Work is the strong guarantee of immutability. It is really difficult, if not impossible, to modify a transaction after it has received a sufficient number of confirmations. Remember that the confirmations correspond to the number of blocks added to the blockchain starting from the block in which the transaction is inserted.

Changing a transaction, or the information contained in it, becomes progressively more difficult as new blocks are generated. If a malicious user tries to tamper with a transaction in the X_{t-1} block, the attack may only succeed in one way, i.e. by recalculating the Proof of Work for all of the following blocks ($X_{t-1} - X_t$) before the other miners succeed in undermining the X_t block.

The malicious user to do this must therefore be in possession of an incredible computing power, and all just to tamper with a transaction that occurred 8 blocks earlier (about 1 hour and 20 minutes earlier in the case of Bitcoin).

Thanks to the hash function, editing a block involves recalculating the entire PoW for all the blocks that follow the tampered block.

The further back in time you go, the less likely it is that an attack will be successful. This is why it is advisable to wait for more than one confirmation (6 confirmations in the Bitcoin) to be able to assume with sufficient security the immutability of a transaction.

Part of the community, however, does not think that the PoW is the best method to use in the process to reach consensus and has raised several issues regarding the PoW. The main ones are:

- **Massive energy consumption**. Bitcoin, the largest project using PoW, currently consumes about 0.3% of the world's electricity (over $1 million a day between electricity and mining hardware) and many believe that this situation is not sustainable in the long run. However, the huge energy consumption is the reason why a consensus process based on the Proof of Work is difficult to attack. It is the enormous amount of computing power needed to validate the blockchain that guarantees its immutability. The computing power and the electricity used are the actual proof of the work performed.

- **Hard to climb**. The PoW is one of the bottlenecks in the ability to scale the system. Many argue that slow transactions and high commissions are blocking the large-scale adoption of the blockchain. However, it is possible to make this type of blockchain scalable without modifying the consent algorithm, adopting off-chain solutions (in the case of Bitcoin or similar, we talk about Lightning Network) or changing the size of the block (for example Bitcoin Cash).

- **He's vulnerable to a 51% attack**. If a miner reached 51% of the total computing power of the network, it would (theoretically) be able to create blocks faster than all the remaining miners together. It may therefore happen that the miner in question is able to reverse or modify some of its transactions (double spending) or to block the confirmation of new transactions (censorship of transactions). However, if a miner could successfully execute a 51% attack, he would still not be able to modify the old transactions, since he would have to recalculate the PoW of all subsequent blocks while the other honest miners continue to undermine on the correct blockchain. Such an attack would require the use of an incredible amount of resources for the attacker. If someone actually managed to put together more than 51% of the computing power, it would be much more profitable for him to follow the rules of the blockchain.

- **Geographical discrimination, economies of scale and centralization**. At the moment, most of the miners are concentrated in places where the cost of electricity and temperatures are low (to save on electricity and cooling systems). In addition, economies of scale are used to negotiate cheaper prices for both electricity and mining equipment. This often results in a centralization of the mining process, leading to the concentration of mines in a few geographical areas or to their joining together to share computing power.

2.4.2 Proof of Stake (PoS)

The Proof of Stake is another protocol used in the process to reach distributed consensus. The purpose of the Proof of Stake is the same as that of the PoW, but the process of reaching the final goal is different.

Unlike the Proof of Work, in which miner that solve mathematical problems are rewarded, in the Proof of Stake validators are alternated (validators, can be considered the equivalent of the miners in the PoW) chosen in advance based on the amount of crypto currencies in their possession for the relevant blockchain, also known as stakes.

Proof of Stake is a protocol used to reach distributed consensus in which each token has one vote.

PoS Mining (staking)

In the PoS-mining, instead of the computing power possessed, the tokens possessed are used. Users with tokens can "point" (Staking) their own tokens (technically, pointing means temporarily blocking the tokens until the staking process ends) to have in return the right to confirm the transactions of a block (become a validator) and receive a reward.

The creator of a new block is then chosen in advance using a combination of different parameters, depending on the type of algorithm used. Some parameters may be the number of tokens (stakes), or the time the validator was in possession of those tokens. Like the PoW, the PoS protocol is also fair for validators: a validator who owns 5% of the total amount of tokens, on average gets the right to create a new block (and earn the reward) 5% of the time.

It can therefore be said that:

$$Voting\ power\ = \frac{Validator's\ Stake}{Total\ Network\ Stake}$$

Compared to PoW, the Proof of Stake is more efficient, as there is no need to perform complex calculations for each new block. PoS supporters say that compared to PoW, PoS has the following advantages:

- **Attacks are more expensive**. The PoS is also theoretically vulnerable to a 51% attack. An attacker, in this case, will not need 51% of the total hashrate but 51% of the total tokens. However, if an attacker tried to buy 51% of the tokens, the market would react with a rare increase in the price of the token. Moreover, people with many tokens have less incentive to attack the blockchain, since an attack would have the counterproductive consequence of destroying the trust in that blockchain, and consequently the value of that token.

- **Cheaper**. As there are no electricity and hardware costs for mining, all people can afford to participate in the network, reducing the current centralization of PoW-based systems.

- **Punishment**. It is possible to create economic disincentives for malevolent actors, for example by destroying their stakes.

- **Loyalty**. Miners are encouraged to stay on the same blockchain. If they wanted to participate in the PoS on another blockchain, they would have to change the tokens in their possession. In the PoW, however, if the currency you are undermining is no longer profitable, you can simply change blockchain.

2.4.3 Fork

Often in the context of blockchain we hear about fork, but it is not always clear what is a fork and what it actually involves. Although the term is often used to indicate the division of a blockchain (chain-split), in reality it contains a set of different possible scenarios.

A fork is a situation in which one of the following things happens:

- Different nodes have **temporarily** different opinions on the transaction history, but the rules of the blockchain (**regular forks**) remain unchanged. In this case there is a temporary division of the blockchain (the consent on the transaction history is temporarily lost).

- The rules of the blockchain have changed **backwards** and all nodes share the same transaction history (**soft fork**). There is no division of the blockchain.
- The rules of the blockchain have changed in a **non-retrocompatible** manner but all nodes are updated to the new rules and share the same transaction history (**hard fork**). There is no division of the blockchain.
- The rules of the blockchain have changed in a **non-retrocompatible** way and different nodes have different opinions on the rules of the blockchain, not sharing the same transaction history (**hard fork with chain split**). There is a division of the blockchain (the consent on the transaction history is permanently lost).

A regular fork does not change the rules of consent. Soft forks and hard forks, on the other hand, imply a modification of these rules. If the new rules are less rigid (not backwards compatible) you get a hard fork. If the rules become stricter (compatible with previous versions) you get a soft fork.

3. Dataset

In this chapter we are going to present all the inputs that will be used in the Chapter 4 for the Portfolio Allocation. Starting from the different asset classes we explore how the Cryptocurrency Index (CriX) is structured and composed. For this purpose, we select asset classes that share some of the main features of the cryptos, keeping in mind that cryptos do not have intrinsic value and do not pay interests/dividends.

Our portfolio will be a selection of the main indices in the economy according to the Global Industry Classification Standards (GICS). I have selected one index for each sector of the economy with the aim to reduce the exposure to the unsystematic risk. The idea, as we will see with Markowitz, is to compose a portfolio fully diversified, in such a way we are rewarded only for the systematic risk that we bear.

Furthermore, we introduce a Dollar ETF in addition to the actual basket of indices with the aim of catching additional similarities that exist between the cryptos and the currencies.

3.1 Asset Classes

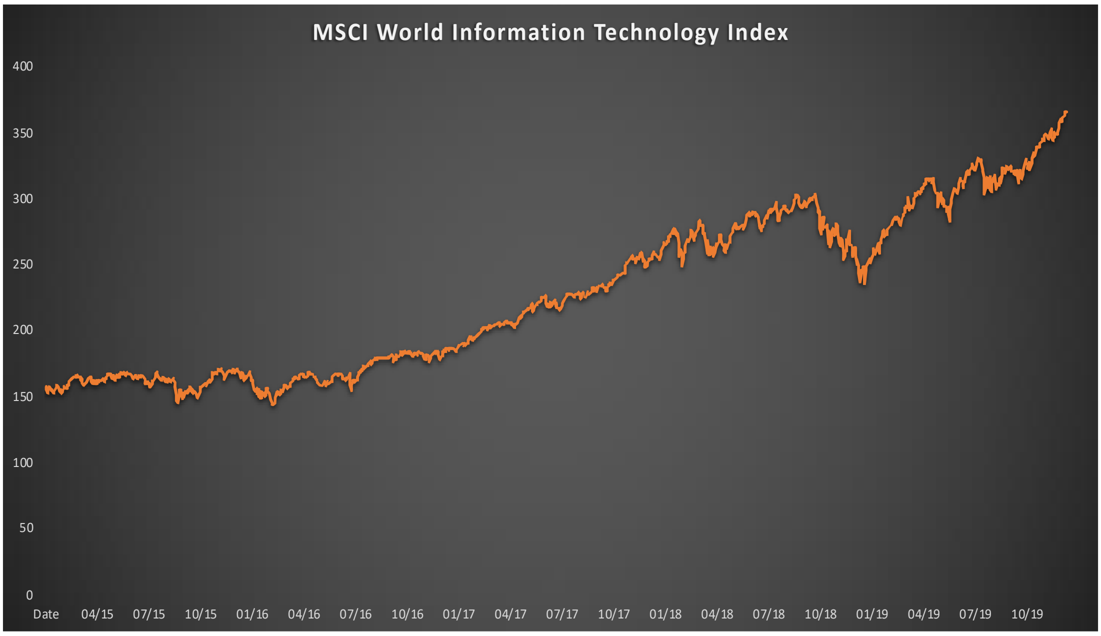
Plenty of papers and journals has described the returns from the cryptos as being the highest possible if compared to real world assets. To be fair, we have to say that also the global economies have done a great job in the last five years.

Below I describe, according to the Global Industry Classification Standards (GICS), each different index I am going to use as input to the portfolios optimization problem in chapter 4.

The MSCI World Indices described below are designed to capture the large and mid-cap segments across 23 Developed Markets countries.

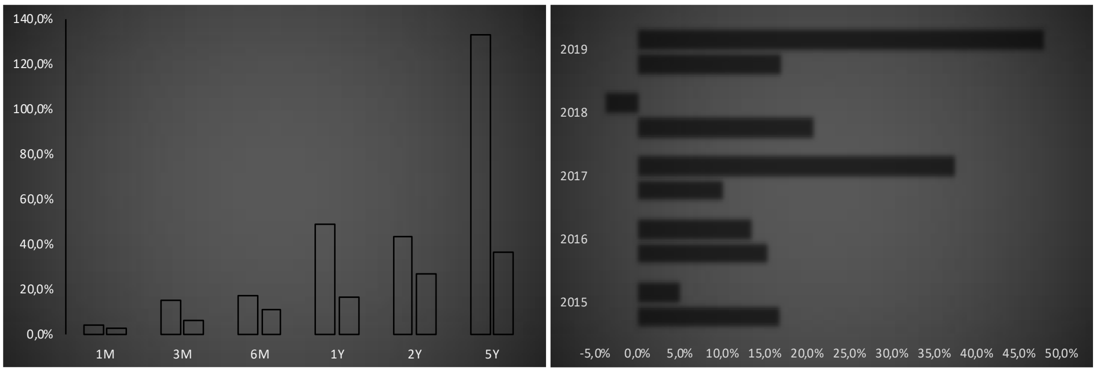
MSCI World Information Technology Index

The technology sector is the category of stocks relating to the research, development and/or distribution of technologically based goods and services.



Source: MSCI

Index Performance and Composition



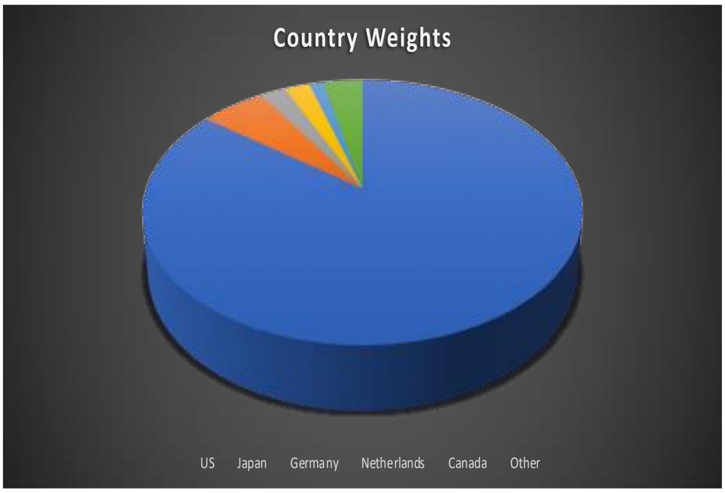
Source: MSCI

Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	4.01%	15.22%	17.08%	49.24%	43.38%	132.86%
Std Dev	3.08%	6.40%	11.28%	16.82%	26.72%	36.43%
SR	1.25	2.30	1.43	2.81	1.47	3.37

Annualized	2015	2016	2017	2018	2019
Return	4.94%	13.34%	37.39%	-3.81%	47.88%
Std Dev	16.69%	15.28%	10.04%	20.69%	16.91%
SR	0.18	0.74	3.52	-0.28	2.71

Index Characteristics **Top Constituents**

	MSCI World I.T.		Wts. (%)
Constituents	169	Apple	17.11
	Mkt Cap (USD Millions)	Microsoft Corp	14.75
Index	7,756,298.35	Visa	4.18
Largest	1,327,057.21	Mastercard	3.48
Smallest	2,352.26	Intel Corp	3.42
Average	45,895.26	Cisco Systems	2.65
Median	14,118.49	Adobe	2.06

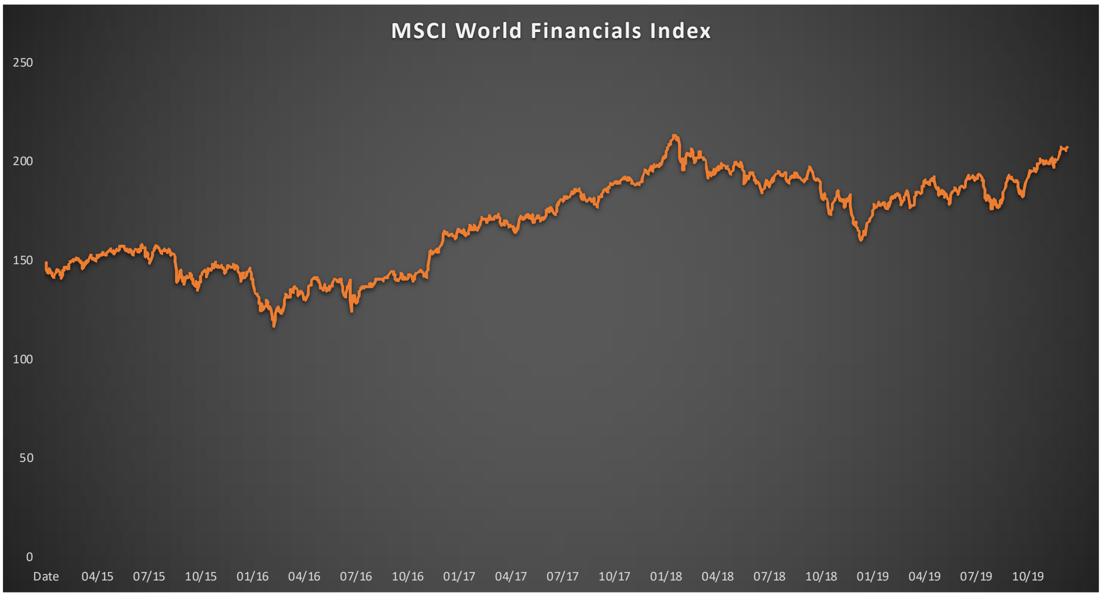


Country	Weights
United States	85,24%
Japan	5,63%
Germany	2,26%
Netherlands	2,19%
Canada	1,22%
Other	3,46%

Source: MSCI

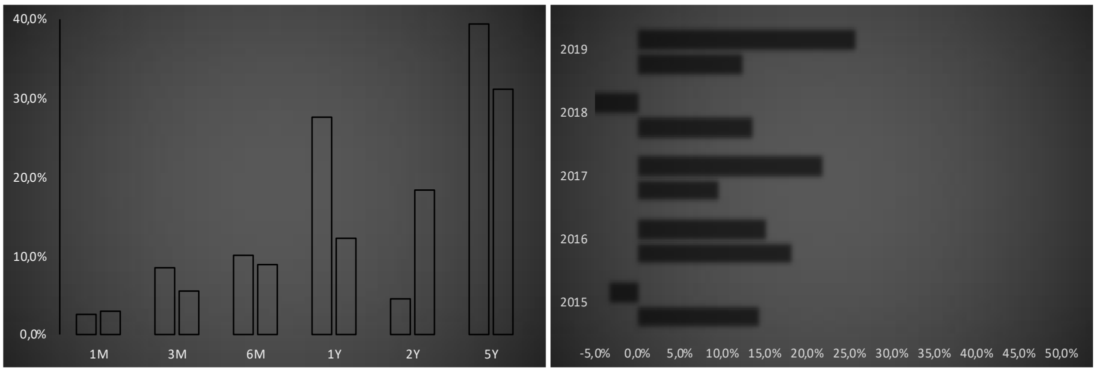
MSCI World Financial Index

The financial sector is a section of the economy made up of firms and institutions that provide financial services to commercial and retail customers. This sector comprises a broad range of industries including banks, investment companies, insurance companies, and real estate firms.



Source: MSCI

Index Performance and Composition

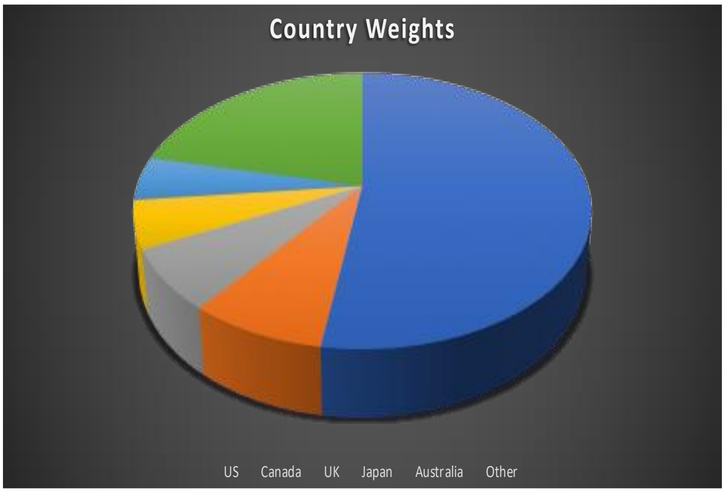


Source: MSCI

Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	2.60%	8.57%	10.16%	27.57%	4.49%	39.24%
Std Dev	2.94%	5.64%	8.80%	12.29%	18.30%	31.00%
SR	0.83	1.43	1.04	2.08	0.03	0.94

Annualized	2015	2016	2017	2018	2019
Return	-3.23%	15.14%	21.74%	-17.07%	25.61%
Std Dev	14.25%	18.14%	9.58%	13.50%	12.32%
SR	-0.37	0.72	2.06	-1.41	1.92

Index Characteristics		**Top Constituents**	
	MSCI World Financials		Wts. (%)
Constituents	249	JPMorgan Chase & Co	6.36
	Mkt Cap (USD Millions)	Bank of America	4.44
Index	7,009,648.88	Berkshire Hathaway	4.26
Largest	445,729.41	Wells Fargo & Co	3.21
Smallest	1,791.84	Citigroup	2.57
Average	28,151.20	HSBC Holdings	2.26
Median	13,401.86	AIA Group	1.81

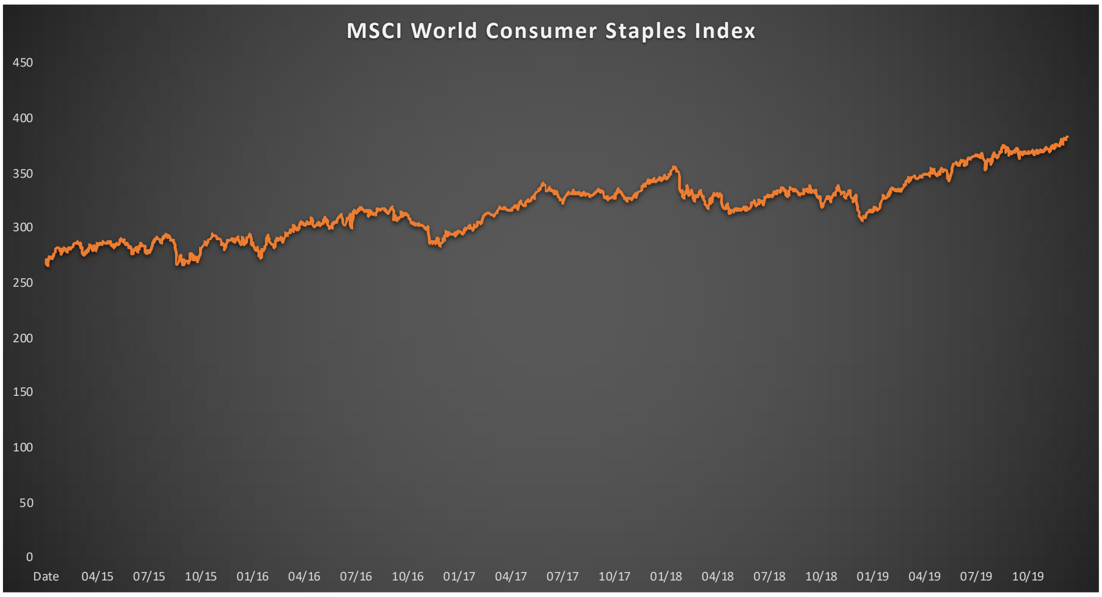


Country	Weights
United States	52.43%
Canada	8.21%
U.K.	7.12%
Japan	5.6%
Australia	5.11%
Other	21.53%

Source: MSCI

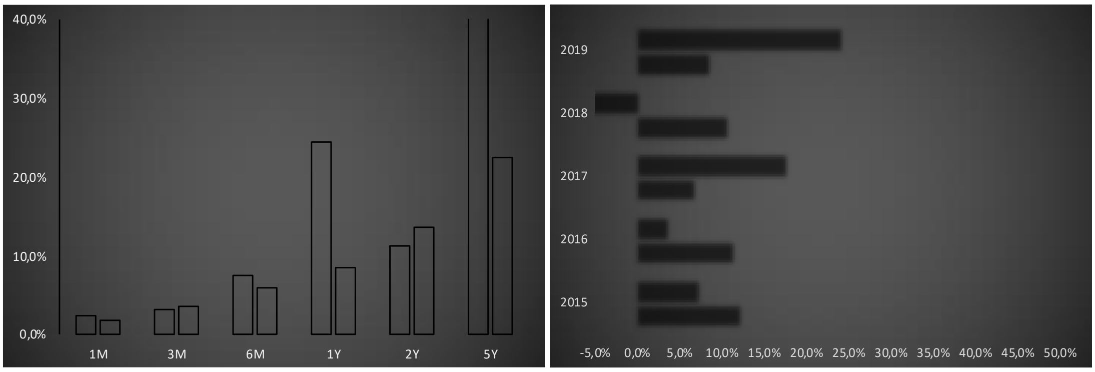
MSCI World Consumer Staples Index

Consumer staples are essential products that people are unable—or unwilling—to cut out of their budgets regardless of their financial situation. Consumer staples are considered to be non-cyclical, meaning that they are always in demand no matter how well the economy is - or is not - performing.



Source: MSCI

Index Performance and Composition



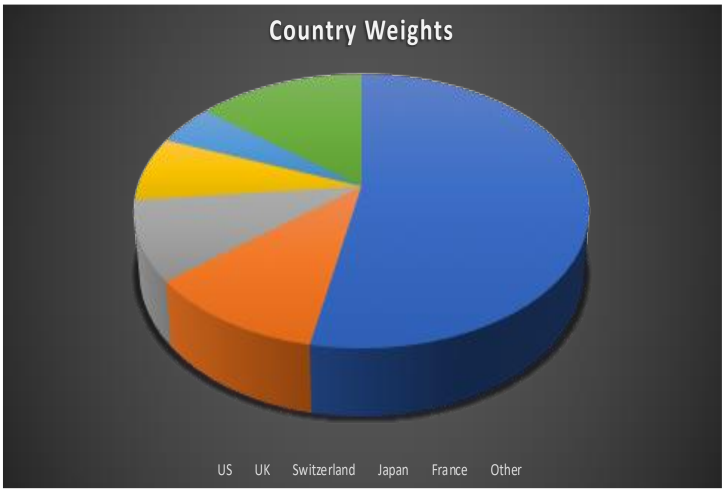
Source: MSCI

Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	2.46%	3.24%	7.51%	24.48%	11.20%	41.43%
Std Dev	1.88%	3.62%	6.03%	8.43%	13.58%	22.51%
SR	1.22	0.76	1.08	2.67	0.53	1.40

Annualized	2015	2016	2017	2018	2019
Return	7.21%	3.52%	17.55%	-9.71%	24.08%
Std Dev	12.23%	11.28%	6.69%	10.58%	8.44%
SR	0.43	0.13	2.32	-1.11	2.62

Index Characteristics **Top Constituents**

	MSCI World C. Staples		Wts. (%)
Constituents	121	Nestle	8.70
	Mkt Cap (USD Millions)	Procter & Gamble Co	8.44
Index	3,703,442.54	Coca Cola	6.07
Largest	322,017.12	PepsiCo	5.16
Smallest	1,930.98	Walmart	4.58
Average	30,606.96	Philip Morris	3.57
Median	11,827.97	CostCo Wholesale	3.49

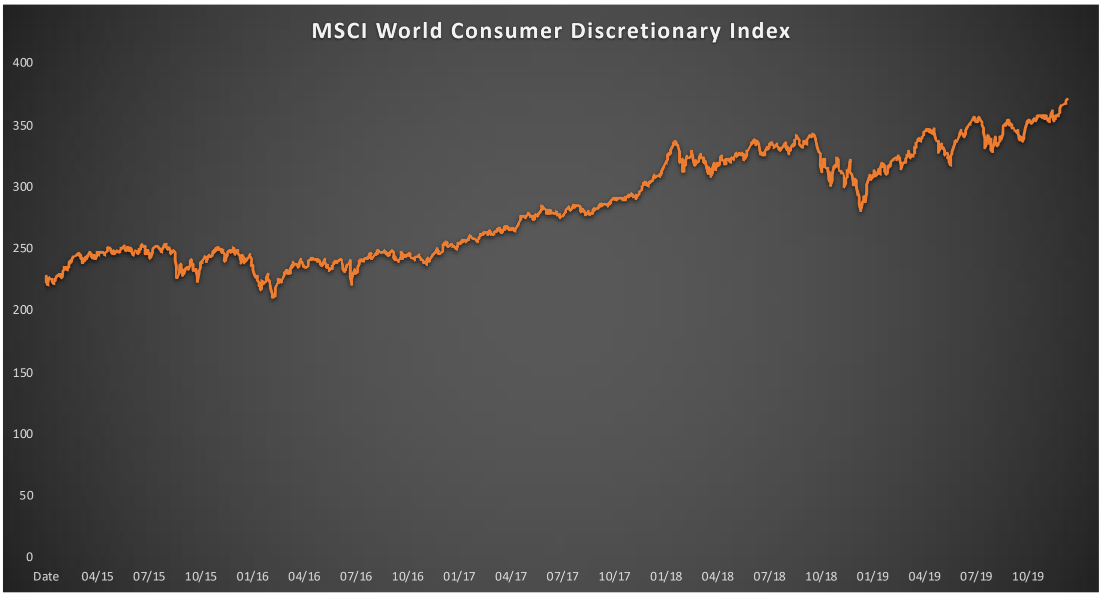


Country	Weights
United States	53.04%
U.K.	11%
Switzerland	9.34%
Japan	7.64%
France	4.85%
Other	14.13%

Source: MSCI

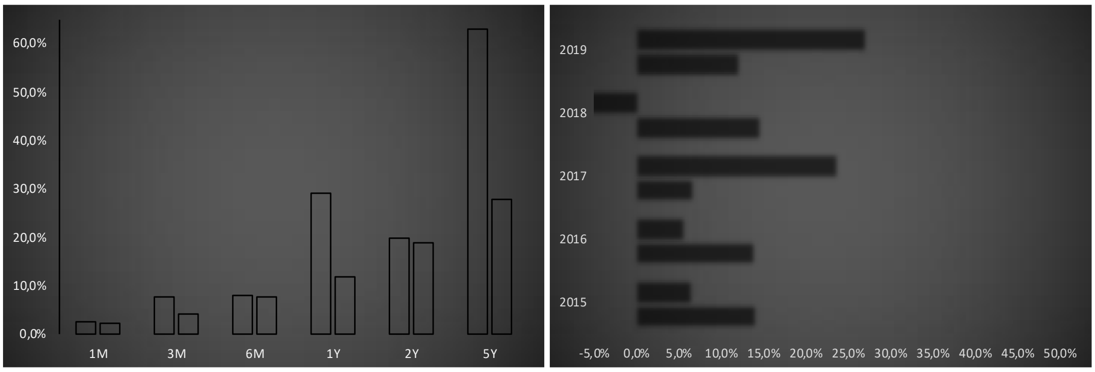
MSCI World Consumer Discretionary Index

Consumer discretionary is the term given to goods and services that are considered non-essential by consumers, but desirable if their available income is sufficient to purchase them. The purchase of consumer discretionary goods is also influenced by the state of the economy, which can affect consumer confidence.



Source: MSCI

Index Performance and Composition



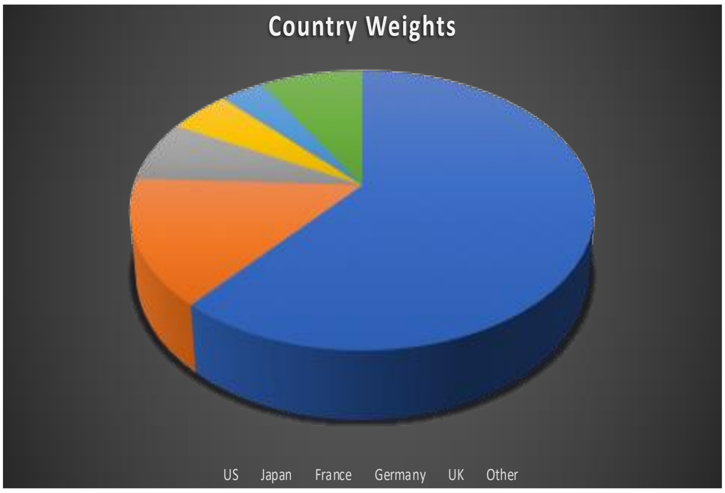
Source: MSCI

Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	2.70%	7.72%	8.04%	29.12%	19.93%	62.98%
Std Dev	2.44%	4.38%	7.91%	11.87%	18.75%	27.89%
SR	1.04	1.65	0.89	2.28	0.85	1.90

Annualized	2015	2016	2017	2018	2019
Return	6.33%	5.52%	23.59%	-6.41%	26.92%
Std Dev	13.93%	13.73%	6.57%	14.48%	11.91%
SR	0.31	0.26	3.29	-0.58	2.09

Index Characteristics **Top Constituents**

	MSCI World C. Discr.		Wts. (%)
Constituents	200	Amazon	16.93
	Mkt Cap (USD Millions)	Home Depot	5.21
Index	4,590,427.53	Toyota Motor C.	3.53
Largest	776,938.40	McDonald’s Corp	3.27
Smallest	1,494.32	LVMH Moet Hennessy	2.82
Average	22,952.14	Nike	2.76
Median	9,577.04	Starbucks Corp	2.29

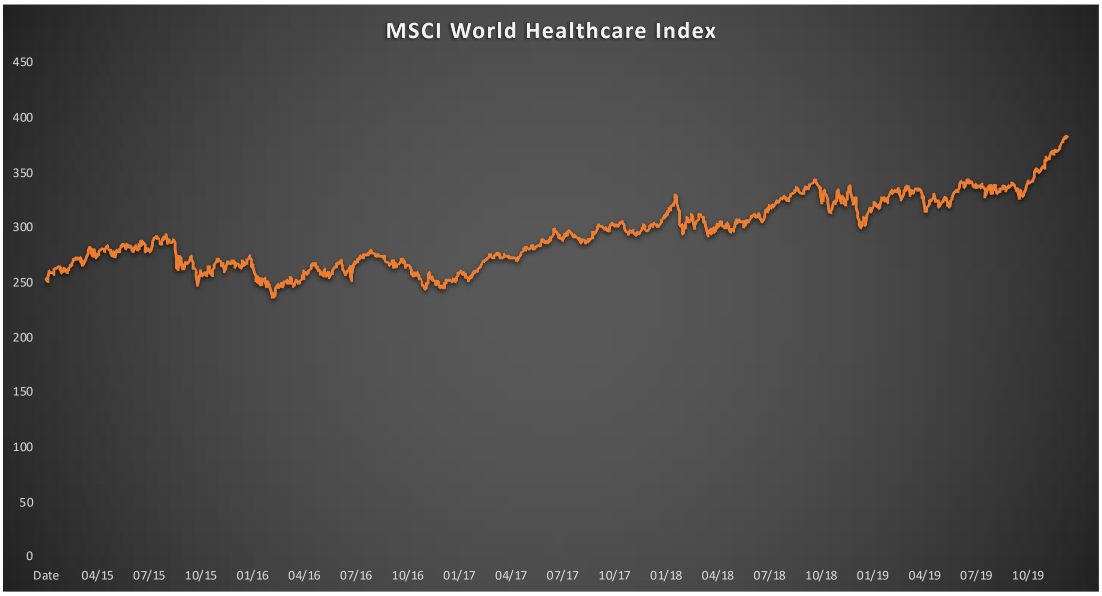


Country	Weights
United States	61.18%
Japan	14.6%
France	7.07%
Germany	5.03%
U.K.	3.63%
Other	8.49%

Source: MSCI

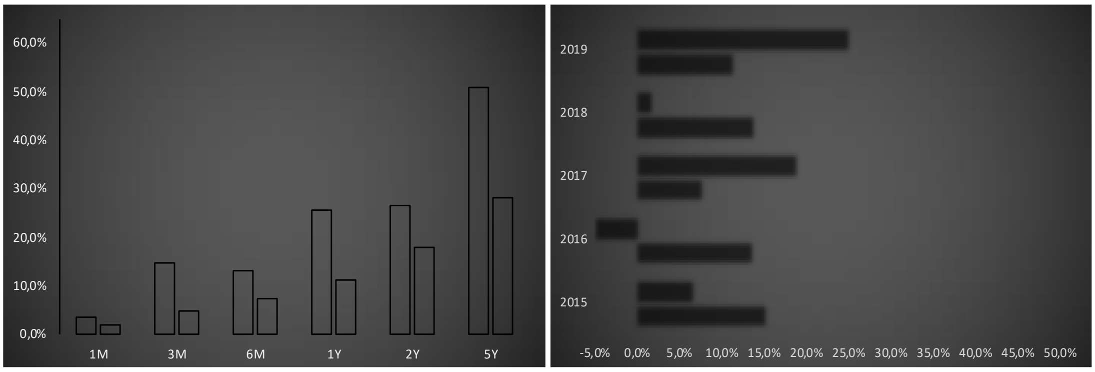
MSCI World Health Care Index

The healthcare sector consists of businesses that provide medical services, manufacture medical equipment or drugs, provide medical insurance, or otherwise facilitate the provision of healthcare to patients.



Source: MSCI

Index Performance and Composition



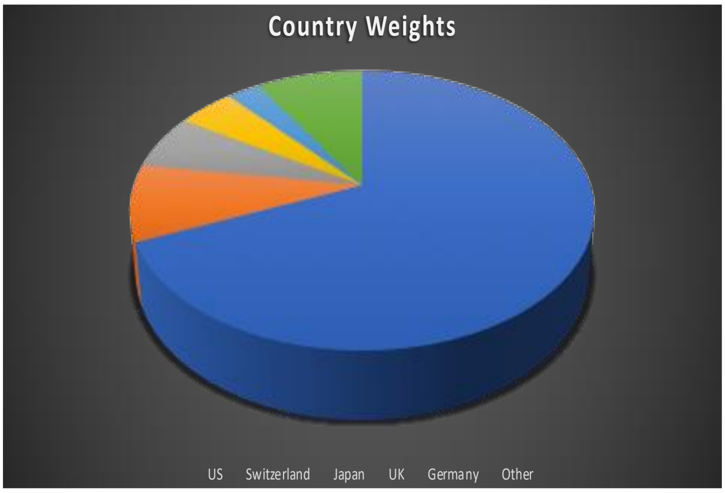
Source: MSCI

Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	3.47%	14.62%	13.20%	25.59%	26.57%	50.75%
Std Dev	2.01%	4.93%	7.41%	11.32%	17.81%	28.09%
SR	1.64	2.86	1.65	2.08	1.27	1.45

Annualized	2015	2016	2017	2018	2019
Return	6.51%	-4.87%	18.77%	1.63%	24.91%
Std Dev	15.10%	13.64%	7.56%	13.80%	11.29%
SR	0.30	-0.50	2.22	-0.03	2.03

Index Characteristics **Top Constituents**

	MSCI World Healthcare		Wts. (%)
Constituents	147	Johnson & Johnson	6.63
	Mkt Cap (USD Millions)	United Health Group	4.80
Index	5,808,057.73	Merck & Co	4.01
Largest	384,975.08	Roche Holding Genuss	3.92
Smallest	1,421.47	Pfizer	3.73
Average	39,510.60	Novartis	3.51
Median	15,283.59	Abbott Lab.	2.64

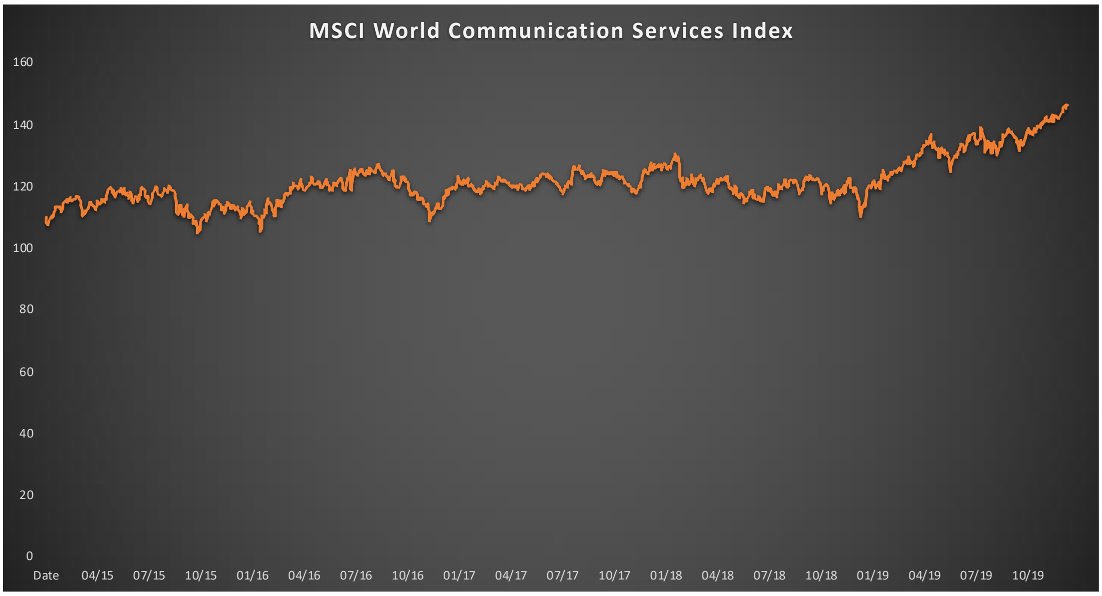


Country	Weights
United States	68.58%
Switzerland	8.84%
Japan	6.3%
U.K.	4.7%
Germany	2.73%
Other	8.84%

Source: MSCI

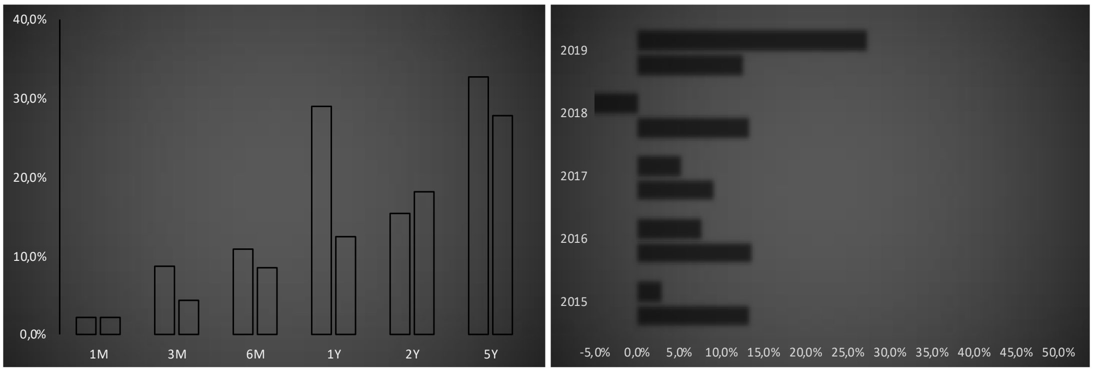
MSCI World Communication Services Index

The telecommunication sector is made up of companies that make communication possible on a global scale, whether it is through the phone or the Internet, through airwaves or cables, through wires or wirelessly.



Source: MSCI

Index Performance and Composition

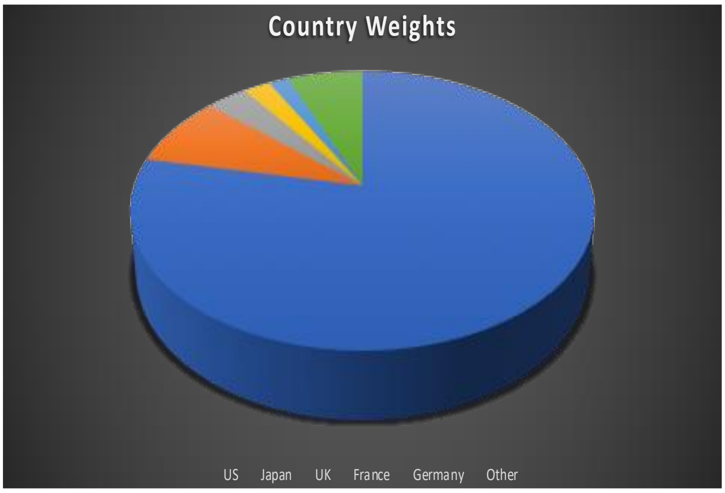


Source: MSCI

Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	2.31%	8.75%	10.82%	28.94%	15.40%	32.72%
Std Dev	2.25%	4.37%	8.49%	12.38%	18.18%	27.73%
SR	0.96	1.89	1.16	2.18	0.63	0.82

Annualized	2015	2016	2017	2018	2019
Return	2.87%	7.54%	5.15%	-10.37%	27.16%
Std Dev	13.20%	13.49%	9.05%	13.26%	12.43%
SR	0.07	0.41	0.35	-0.93	2.02

Index Characteristics		**Top Constituents**	
	MSCI World Comm.Ser.		Wts. (%)
Constituents	104	Facebook A	13.16
	Mkt Cap (USD Millions)	Alphabet C	11.14
Index	3,751,189.09	Alphabet A	10.70
Largest	493,774.67	AT&T	7.61
Smallest	1,750.52	Disney (Walt)	6.95
Average	36,069.13	Verizon Comm.	6.77
Median	9,041.32	Comcast Corp	5.44

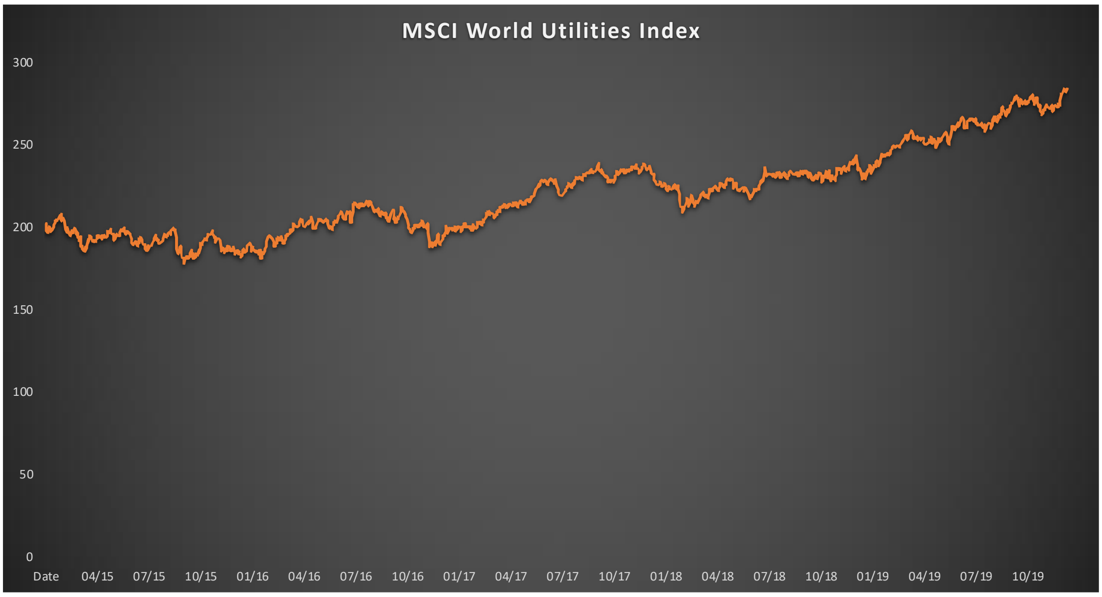


Country	Weights
United States	78.21$
Japan	8.31%
U.K.	3.38%
France	2.23%
Germany	1.63%
Other	6.24%

Source: MSCI

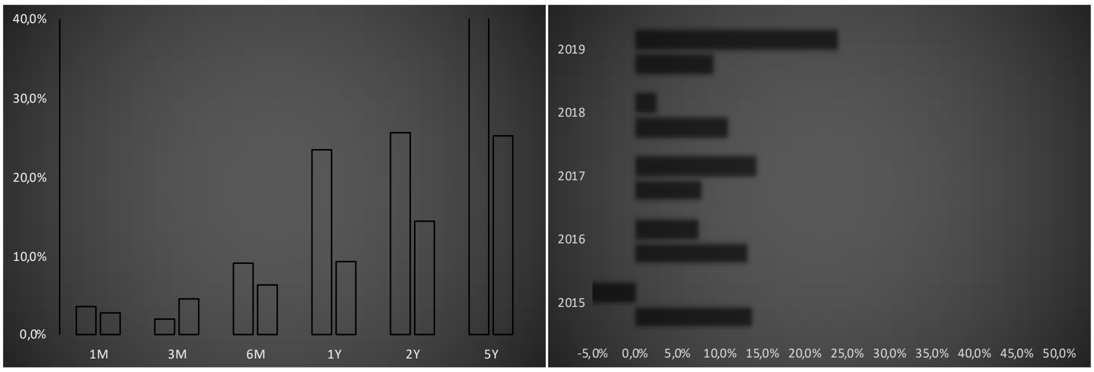
MSCI World Utilities Index

The utilities sector refers to a category of companies that provide basic amenities, such as water, electricity, and natural gas. They are part of the public service landscape and therefore heavily regulated. Investors typically treat utilities as long-term holdings and use them to inject steady income in their portfolios.



Source: MSCI

Index Performance and Composition



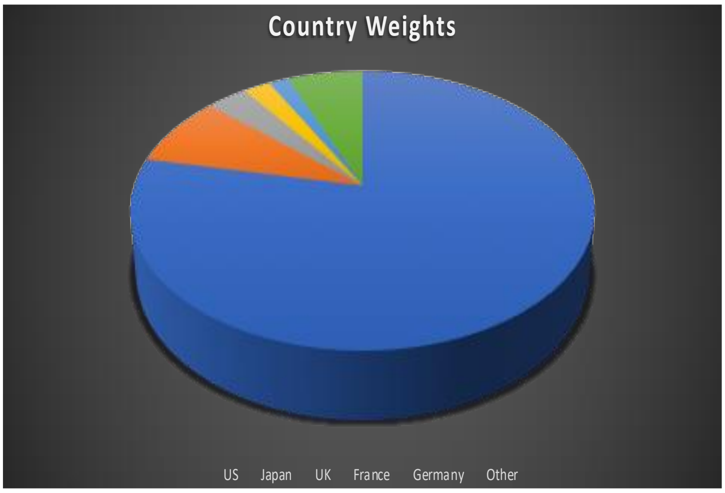
Source: MSCI

Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	3.67%	1.99%	9.05%	23.50%	25.48%	40.56%
Std Dev	2.79%	4.60%	6.41%	9.29%	14.33%	25.14%
SR	1.25	0.32	1.26	2.31	1.50	1.22

Annualized	2015	2016	2017	2018	2019
Return	-6.51%	7.38%	14.31%	2.47%	23.87%
Std Dev	13.70%	13.30%	7.78%	10.96%	9.25%
SR	-0.62	0.40	1.58	0.04	2.37

Index Characteristics **Top Constituents**

	MSCI World Utilities		Wts. (%)
Constituents	83	Nextera Energy	7.59
	Mkt Cap (USD Millions)	Dominion Energy	4.46
Index	1,527,664.33	Southern Company	4.36
Largest	116,019.47	Duke Energy Corp	4.35
Smallest	2,204.69	Enel	4.23
Average	18,405.59	Iberdrola	4.08
Median	10,269.58	American Electric Power	3.05

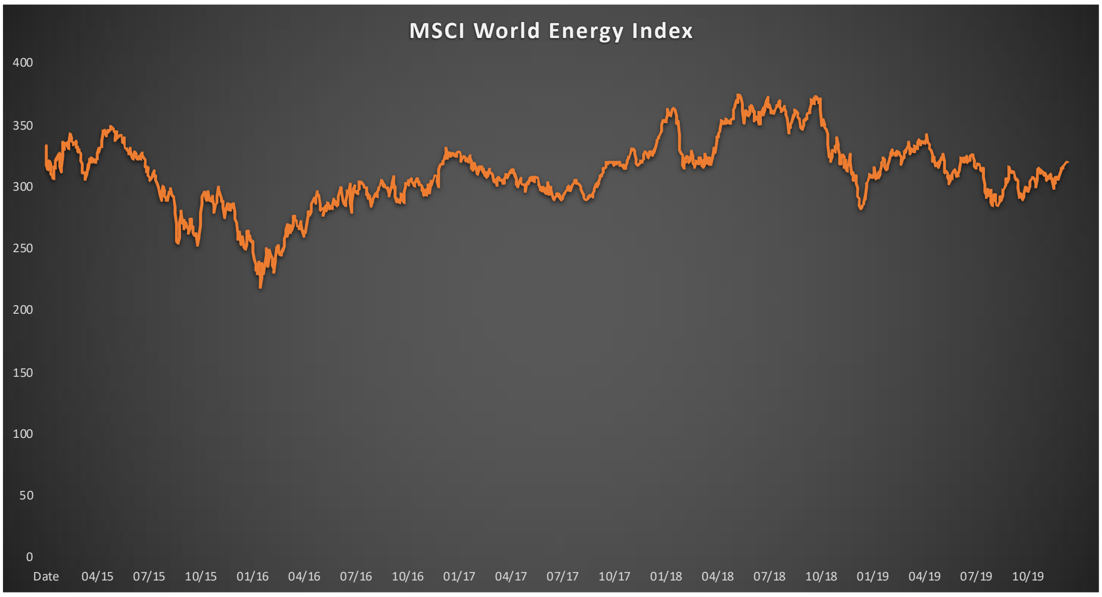


Country	Weights
United States	60.26%
Spain	6.1%
U.K.	5.65%
Italy	5.54%
Japan	3.8%
Other	18.64%

Source: MSCI

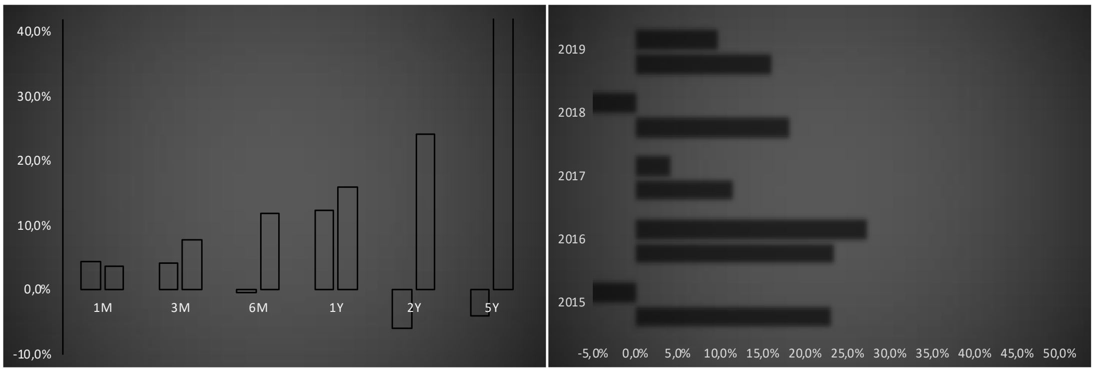
MSCI World Energy Index

The energy sector is a category of stocks that relate to producing or supplying energy. The energy sector or industry includes companies involved in the exploration and development of oil or gas reserves, oil and gas drilling, and refining. The energy industry also includes integrated power utility companies such as renewable energy.



Source: MSCI

Index Performance and Composition



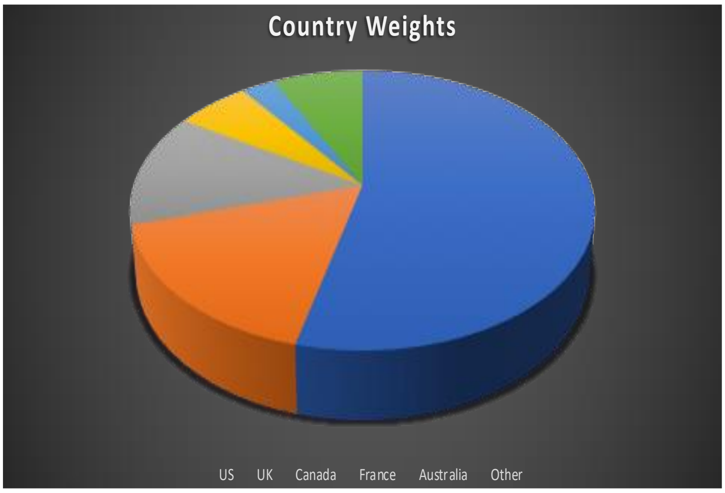
Source: MSCI

Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	4.24%	4.14%	-0.41%	12.26%	-6.05%	-4.00%
Std Dev	3.63%	7.62%	11.81%	15.95%	24.13%	42.33%
SR	1.12	0.48	-0.12	0.64	-0.42	-0.33

Annualized	2015	2016	2017	2018	2019
Return	-22.87%	27.28%	4.17%	-16.94%	9.75%
Std Dev	23.03%	23.31%	11.42%	18.11%	15.97%
SR	-1.08	1.08	0.19	-1.05	0.49

Index Characteristics **Top Constituents**

	MSCI World Energy		Wts. (%)
Constituents	67	Exxon Mobil Corp	13.42
	Mkt Cap (USD Millions)	Chevron Corp	10.40
Index	2,199,416.60	Total	6.03
Largest	295,246.60	BP	5.79
Smallest	1,809.04	Royal Dutch Shell A	5.76
Average	32,827.11	Royal Duch Shell B	5.04
Median	14,113.40	Enbridge	3.66

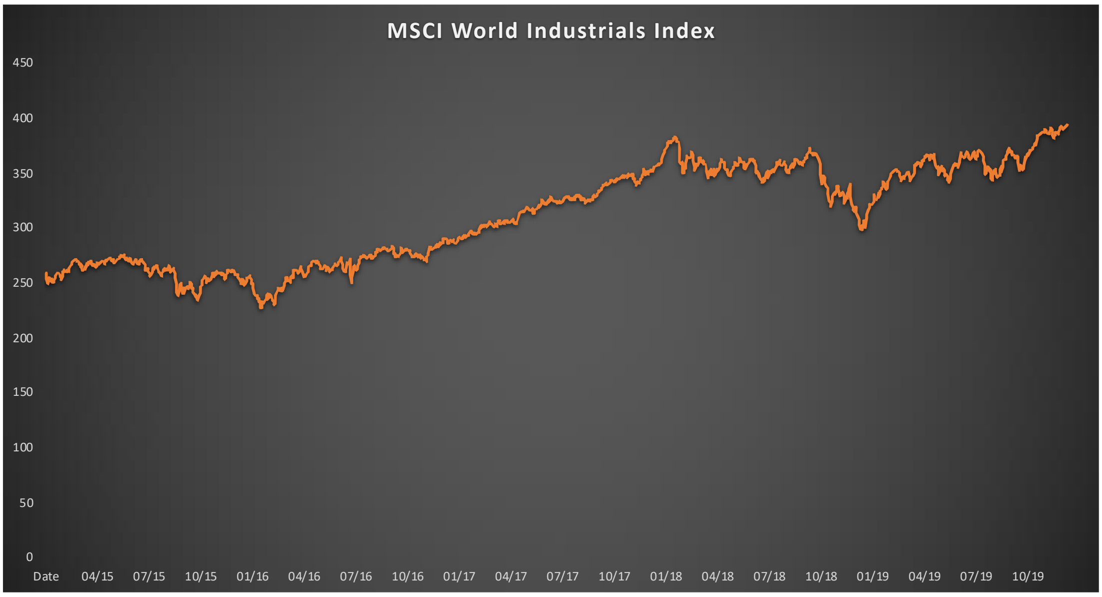


Country	Weights
United States	53.91%
U.K.	16.59%
Canada	13.22%
France	6.03%
Australia	2.8%
Other	7.45%

Source: MSCI

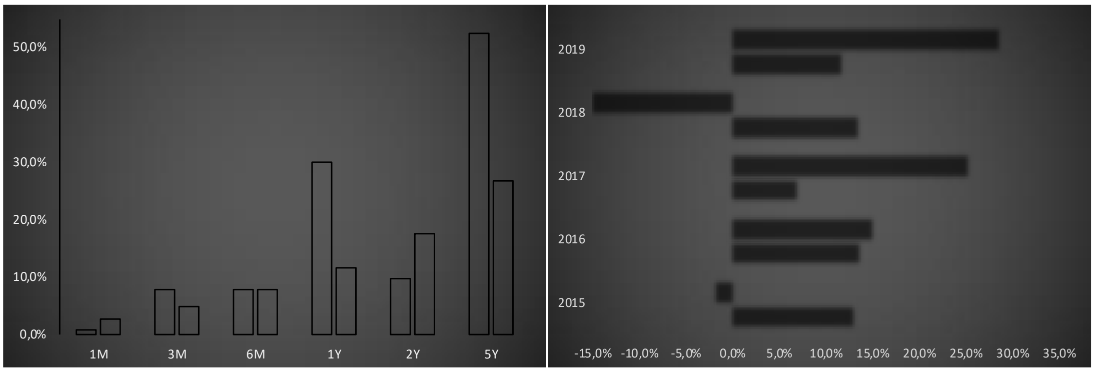
MSCI World Industrials Index

The industrial goods sector is a category of stocks of companies who produce capital goods used in construction and manufacturing. Businesses in the industrial goods sector make and sell machinery, equipment, and supplies that are used to produce other goods rather than sold directly to consumers.



Source: MSCI

Index Performance and Composition



Source: MSCI

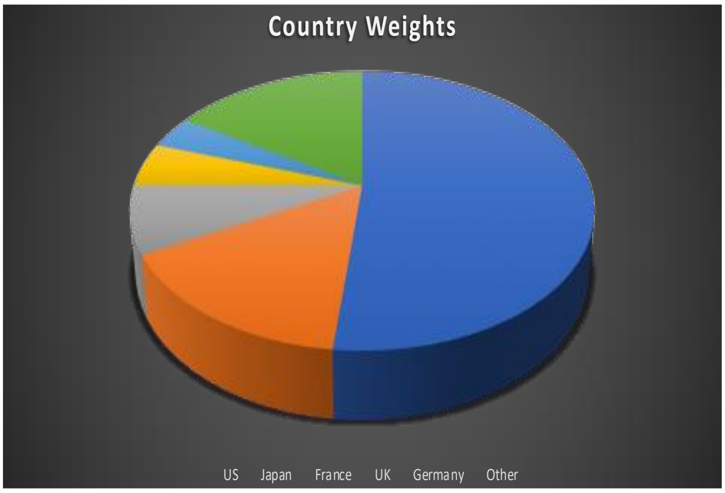
Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	0.89%	8.02%	7.99%	30.06%	9.81%	52.49%
Std Dev	2.66%	4.98%	7.98%	11.64%	17.75%	26.69%
SR	0.27	1.51	0.88	2.41	0.33	1.59

Annualized	2015	2016	2017	2018	2019
Return	-1.74%	15.00%	25.09%	-14.94%	28.49%
Std Dev	12.93%	13.48%	6.83%	13.31%	11.68%
SR	-0.29	0.96	3.38	-1.27	2.27

Index Characteristics

Top Constituents

	MSCI World Industrials		Wts. (%)
Constituents	272	Boeing Co	3.55
	Mkt Cap (USD Millions)	Union Pacific Corp	2.60
Index	4,904,525.79	Honeywell International	2.60
Largest	174,142.99	United Technologies Corp	2.50
Smallest	1,674.35	3M Co	2.07
Average	18,031.34	Siemens	2.04
Median	9,176.00	Lockheed Martin Corp	2.02

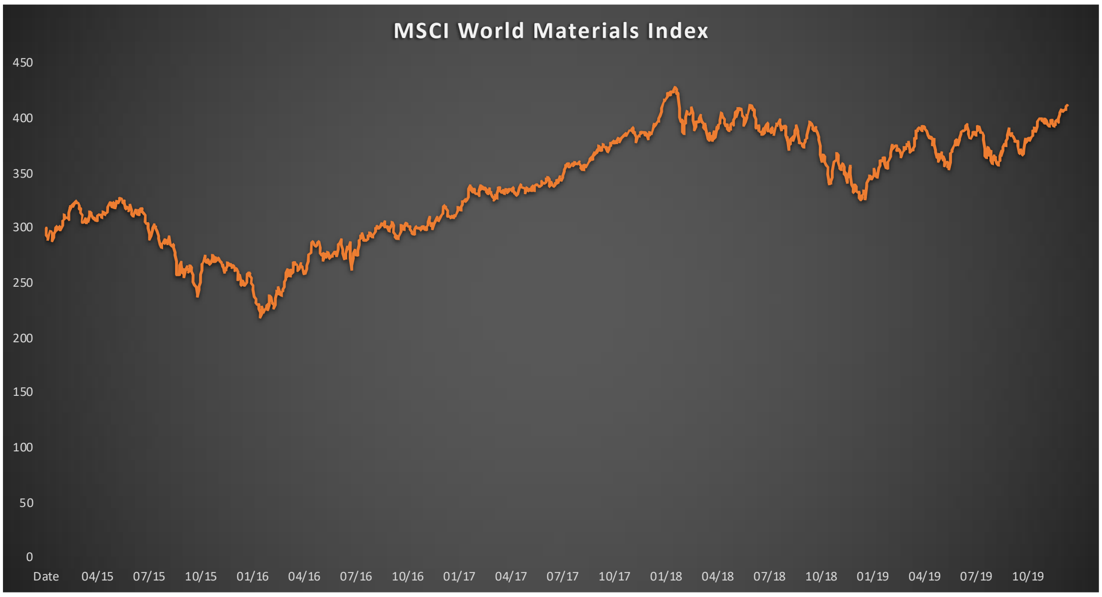


Country	Weights
United States	51.70%
Japan	15.47%
France	7.86%
U.K.	5.07%
Germany	3.81%
Other	16.10%

Source: MSCI

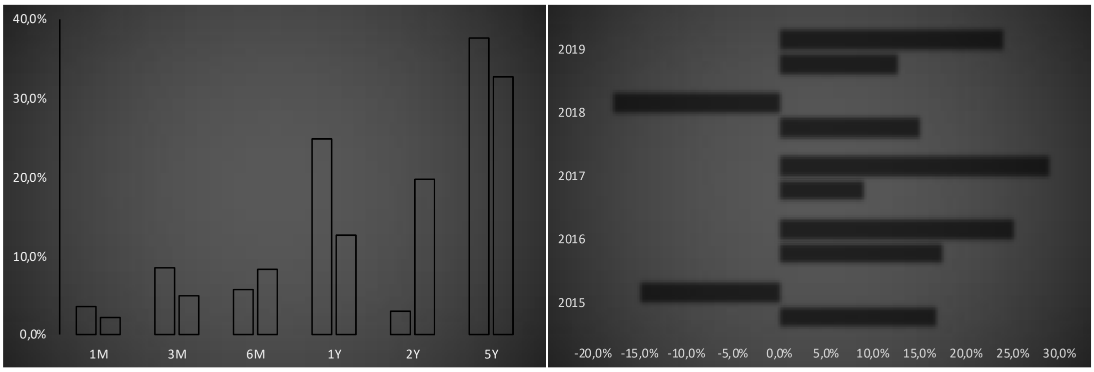
MSCI World Materials Index

The basic materials sector is a category of stocks for companies involved in the discovery, development, and processing of raw materials. The sector includes companies engaged in mining and metal refining, chemical products, and forestry products. They are sensitive to changes in the business cycle.



Source: MSCI

Index Performance and Composition



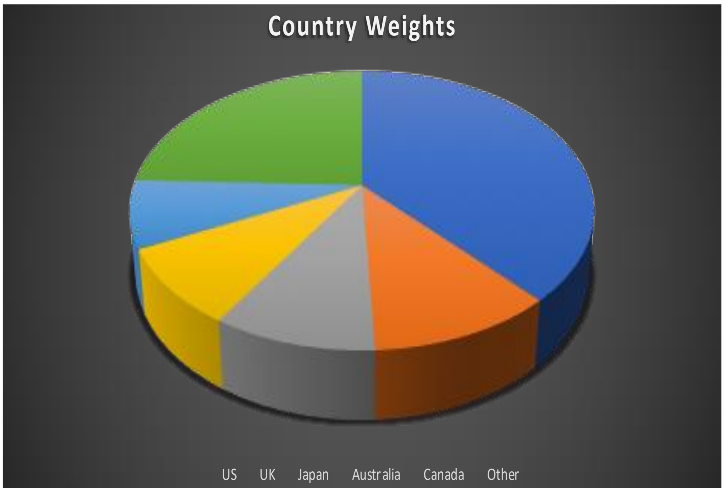
Source: MSCI

Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	3.52%	8.53%	5.70%	24.71%	3.02%	37.47%
Std Dev	2.12%	5.05%	8.33%	12.67%	19.71%	32.62%
SR	1.58	1.59	0.56	1.79	-0.05	0.84

Annualized	2015	2016	2017	2018	2019
Return	-14.96%	25.13%	28.85%	-17.81%	24.04%
Std Dev	16.81%	17.51%	9.02%	14.99%	12.71%
SR	-1.01	1.32	2.98	-1.32	1.73

Index Characteristics **Top Constituents**

	MSCI World Materials		Wts. (%)
Constituents	127	Linde	5.87
	Mkt Cap (USD Millions)	BHP Group	4.11
Index	1,961,449.85	BASF	3.54
Largest	115,091.93	RIO Tinto Plc	3.44
Smallest	1,598.45	Air Liquide	3.41
Average	15,444.49	Air Products & Chemicals	2.64
Median	8,716.38	Ecolab	2.55

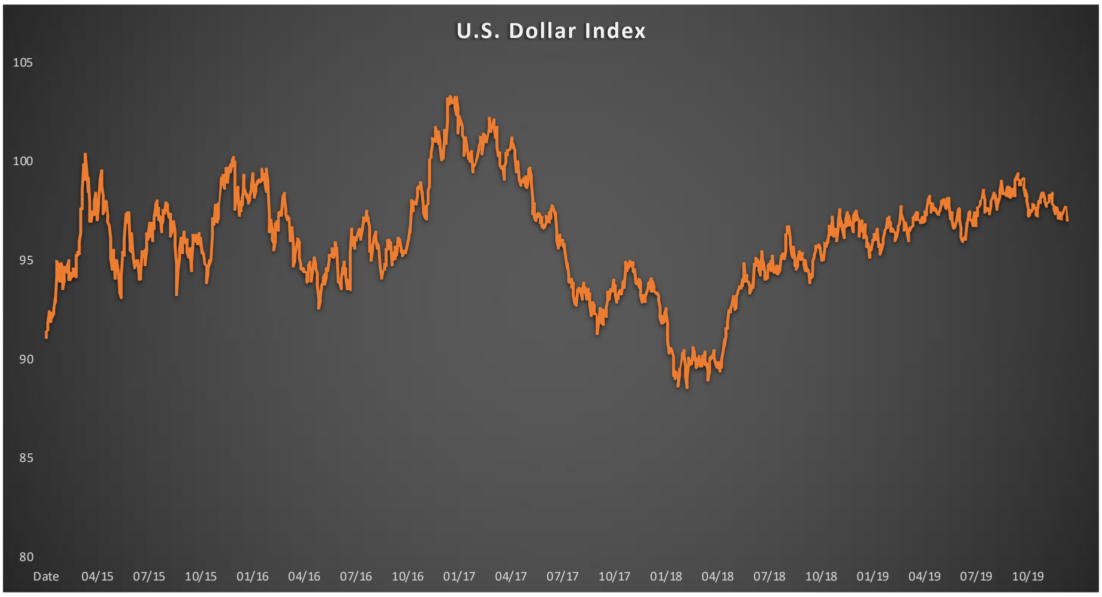


Country	Weights
United States	38.22%
U.K.	11.02%
Japan	9.66%
Australia	8.88%
Canada	7.82%
Other	24.39%

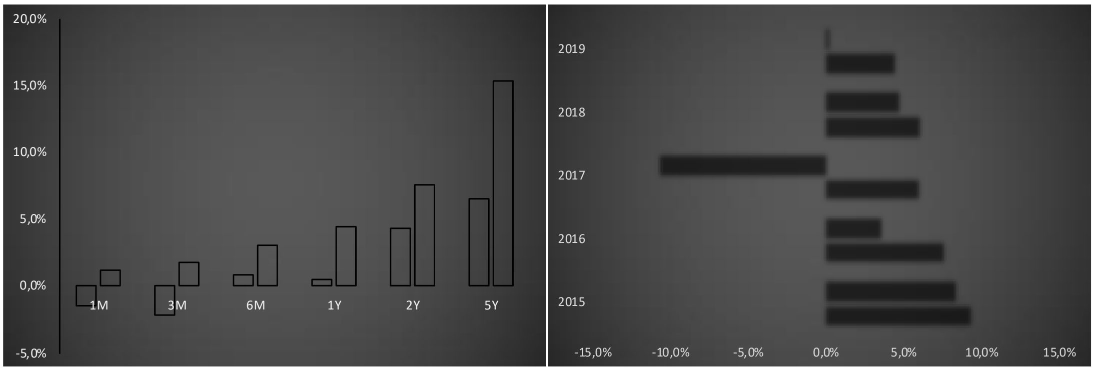
Source: MSCI

U.S. Dollar Index

The U.S. Dollar Index (USDX, DXY) is an index of the value of the United States dollar relative to a basket of foreign currencies, generally this basket is made up with U.S. trade partners' currencies. When the dollar is "stronger" (in value) when compared to other currencies the index increases in value.



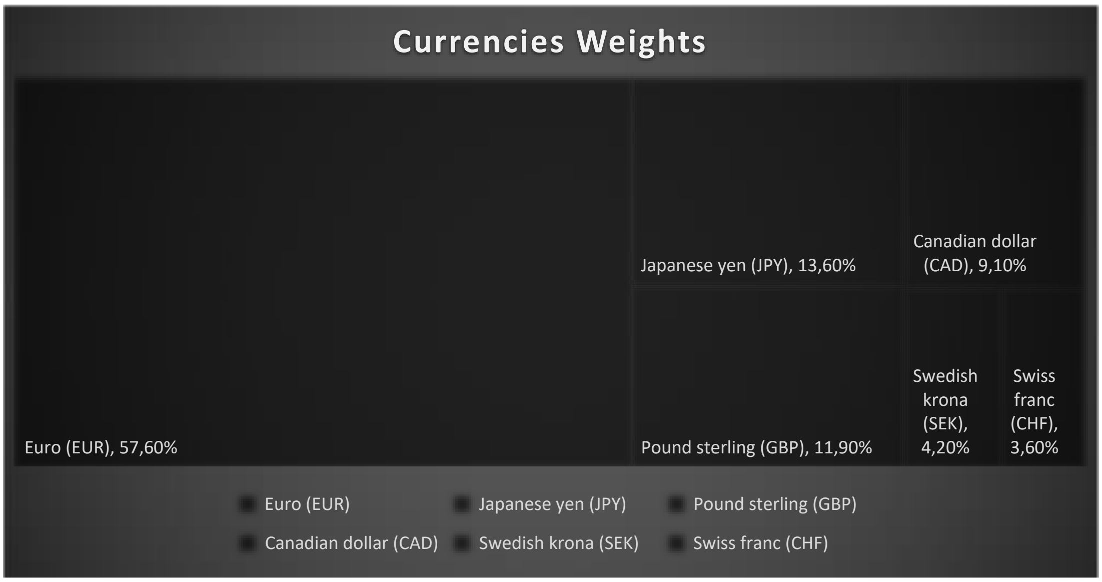
Index Performance and Composition



Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	-1.42%	-2.11%	0.83%	0.55%	4.28%	6.52%
Std Dev	1.18%	1.81%	3.00%	4.48%	7.51%	15.35%
SR	-1.35	-1.45	-0.06	-0.32	0.04	-0.23

Annualized	2015	2016	2017	2018	2019
Return	8.29%	3.56%	-10.57%	4.70%	0.20%
Std Dev	9.29%	7.58%	5.90%	6.02%	4.45%
SR	0.68	0.21	-2.16	0.45	-0.40

The Index is a weighted geometric mean of the dollar's value relative to the following select currencies:



3.2 Cryptocurrency Index (CriX)

In the financial industry already exist important benchmarks like the S&P500 and DAX. These indices describe the composition and trend of certain segments of the financial markets.

Index providers decide on a fixed number of index constituents which will represent the market segment. It is a huge challenge to set this fixed number and develop the rules to find the constituents, especially since markets change and this has to be taken into account. A method relying on the AIC is proposed to quickly react to market changes giving the possibility to create an index, referred to as CRIX, for the cryptocurrency market.

Nowadays more and more companies have started offering digital payment systems. Smartphones have evolved into a digital wallet. Own currencies for the digital market were therefore just a matter of time. The idea of letting companies offer concurrent currencies seemed for a long time scarcely probable, but the invention of the Blockchain has made it possible.

Cryptocurrencies (abbr. cryptos) have surfaced and opened up an angle towards this new level of economic interaction. Since the appearance of bitcoins, several new cryptos have spread through the Web and offered new ways of proliferation.

Obviously, the crypto market is fanning out and shows clear signs of acceptance and deepening liquidity, so that a closer look at its general moves and dynamics is called for.

Elendner et al. (2016) studied the top 10 cryptocurrencies by market capitalization and found that their returns are weakly correlated which each other. This brings to the conclusion that Bitcoin, even though it dominates the market in terms of its market capitalization, does not lead the market. The movements of other cryptocurrencies are important too, when one analyzes the market of cryptocurrencies.

By designing CRIX, a market index (benchmark), will enable each interested party to study the performance of the crypto market as a whole or single cryptos. Studying the stochastic dynamics of CRIX will allow to create ETFs or contingent claims.

Before introducing Index Construction some definitions need to be defined. First, the term benchmark.

Definition 1. *A benchmark is a measure which consists of a selection of cryptos that are representing the market.*

Index providers construct their indices by following this definition with a fixed number of constituents (FTSE, S&P). But markets change which should cause the chosen number of index constituents to be altered too. While trying to mimic the movements of an innovative market like the crypto market, one is confronted with a frequently changing market structure. Therefore, a different approach is necessary that enables to react to changes in the market structure. A dynamic methodology guaranteeing the diversity of an index at any time is to be constructed. Furthermore, the benchmark is meant to be investable. Regarding the portfolio choice, we define the following selection definition.

Definition 2. *Between investment portfolios with equal performance, the one with the least assets is preferable.*

Following both definitions, for the crypto market CRIX, a CRyptocurrency IndeX, has been established. The CRIX is computed by evaluating the differences in the log returns of the market against a selection of possible benchmarks. Studies figure out that the AIC works well to evaluate the differences. It penalizes the index for the number of constituents, so definitions 1 and 2 are met. For the calculation of the respective likelihoods, a non-parametric approach using the Epanechnikov (1969) kernel is applied. We proof the impact of the value of an asset in the market on the AIC method, thus we are applying a top-down approach to select the assets for the benchmarks to choose from.

3.2.1 Index Construction

The basic idea of any price index is to weight the prices of its constituent goods by the quantities of the goods purchased or consumed. The Laspeyres index takes the value of a basket of *k* assets and compares it against a base period:

$$P_{0t}^{L}(k) = \frac{\sum_{i=1}^{k} P_{it} Q_{i0}}{\sum_{i=1}^{k} P_{i0} Q_{i0}} \tag{1}$$

with P_{it} the price of asset *i* at time *t* and Q_{i0} the quantity of asset *i* at time 0 (the base period).

For market indices (such as S&P500 or DAX) the quantity Q_{i0} is the number of shares of the asset *i* in the base period. Multiplied with its corresponding price, the market capitalization results, hence the constituents of the index are weighted by their market capitalizations. But markets change. A company which was representative for market developments yesterday might no longer be important today. On top of that, companies can go bankrupt, a corporation can raise the number of its outstanding shares, or trading in it can become infrequent. All these situations must produce a change in the index structure, so that the market is still adequately represented. Hence companies have to drop out of the index and have to be replaced by others.

The Index rules determine in which cases such an event happens. The formula of Laspeyres (1) cannot handle such events entirely because a change of constituents will result in a change in the index value that is not due to price changes. Therefore, established price indices like DAX or S&P500, and the newly founded index $CRIX(k)$, a CRyptocurrency IndeX, use the adjusted formula of Laspeyres,

$$CRIX_t(k,\beta) = \frac{\sum_{i=1}^{k} \beta_{i,t_l^-} P_{it} Q_{it_l^-}}{Divisor(k)_{t_l^-}} \tag{2}$$

with P , Q and *i* defined as before, β_{i,t_l^-} the adjustment factor of asset *i* found at time point, t_l^-, *l* indicates that this is the *l*-th adjustment factor, and , t_l^- the last time point when Q_{i,t_l} , $Divisor(k)_{t_l^-}$ and β_{i,t_l^-} were updated.

In the classical setting, β_{i,t_l^-} is defined to be β_{i,t_l^-} = 1 for all *i* and *l*. Anyhow, some indices use β_{i,t_l^-} to achieve *maximal* weighting rules. The *Divisor* ensures that the index value of CRIX has a predefined value on the starting date. It is defined as

$$Divisor(k,\beta)_0 = \frac{\sum_{i=1}^{k} \beta_{i0} P_{i0} Q_{i0}}{starting\ value} \tag{3}$$

The starting value could be any possible number, commonly 100, 1000 or 10000. It ensures that a positive or negative development from the base period will be revealed. Whenever changes to the structure of CRIX occur, the *Divisor* is adjusted in such a way that only price changes are reflected by the index. Defining k_1 and k_2 as number of constituents, it results

$$\frac{\sum_{i=1}^{k_1} \beta_{i,t_{l-1}^-} P_{i,t-1} Q_{i,t_{l-1}^-}}{Divisor(k_1,\beta)_{t_{l-1}^-}} = CRIX_{t-1}(k_1,\beta) = CRIX_t(k_2,\beta) = \frac{\sum_{j=1}^{k_2} \beta_{i,t_l^-} P_{j,t} Q_{j,t_l^-}}{Divisor(k_2,\beta)_{t_l^-}} \tag{4}$$

In indices like FTSE, S&P500 or DAX the number of index members is fixed, k_1 = k_2,). As long as the goal behind these indices is the reflection of the price development of the selected assets, this is a straightforward approach. These indices are also indicators for the development of the market as a whole. The question is whether the included assets are representing the market.

Since the constituents are chosen using a top-down approach, meaning that the biggest companies by market capitalization are included, the intuitive answer is yes. But it leaves a sour taste that additional assets may describe the market more appropriately.

One may object by referring to total market indices that are providing a full description. But financial praxis has shown that smaller indices like DAX30 and S&P500 receive more attention in evaluating the movements of their corresponding markets. It is therefore appealing to know which are the representative assets in a market and which smaller number of index constituents eases the handling of a tracking portfolio. Additionally, one may be concerned that an index would include illiquid and non-investable assets which makes the management of a tracking portfolio even more difficult. This is indeed a problem in the crypto-currencies market. Some cryptos have a fairly high market capitalization while their respective trading volume is very low. An asset which is not frequently traded cannot add enough information to a market index to display market changes and is difficult to trade for an investor.

These thoughts raise the question which value of k is "optimal" for building an investable benchmark for the market. Additionally, especially young and innovative markets may change their structure over time. Therefore, a quantification of an accurate crypto benchmark with sparse number of constituents is asked for. Since the crypto market shows a frequently changing market structure with a huge number of illiquid cryptos, we apply a time varying index selection structure.

3.2.2 Dynamic Index Construction

This section is dedicated to describing the composition rule which is used to find the number of index members—the spine of CRIX. Since CRIX will be a benchmark for the crypto market, the dimension and evaluation of the market has to be defined:

Definition 3. *The total market (TM) consists of all cryptos in the crypto universe. Its value is the combined market value of the cryptos.*

To compare the TM with a benchmark candidate, it will be normalized by a Divisor,

$$TM(K)_t = \frac{\sum_{i=1}^{K} P_{it} Q_{it_l^-}}{Divisor(K)_{t_l^-}} \tag{5}$$

with *K* the number of all cryptos in the crypto universe. Note that no adjustment factor is used for TM(*K*)*t*. Further define the log returns:

$$\varepsilon(K)_t^{TM} = \log\{TM(K)_t\} - \log\{TM(K)_{t-1}\} \tag{6}$$

$$\varepsilon(k,\beta)_t^{CRIX} = \log\{CRIX(k,\beta)_t\} - \log\{CRIX(k,\beta)_{t-1}\} \tag{7}$$

Where $CRIX(k,\beta)_t$ is the CRIX with *k* constituents at time point *t*.
The goal is to optimize *k* and β so that a sparse but accurate approximation in terms of

$$\min_{k,\beta} \left|\left|\varepsilon(k,\beta)\right|\right|^2 = \left|\left|\varepsilon(K)^{TM} - \varepsilon(k,\beta)^{CRIX}\right|\right|^2, \tag{8}$$

is achieved. We chose a squared loss function in (8), since it heavily penalizes deviations. The expected squared loss is defined as

$$\boldsymbol{E}\left(\left|\left|\varepsilon(k,\beta)\right|\right|^2\right) = \int_{-\infty}^{\infty} \left|\left|\varepsilon(k,\beta)\right|\right|_2^2 f\{\varepsilon(k,\beta)\}\, d\varepsilon(k,\beta) \tag{9}$$

The density, *f*, is estimated nonparametrically with an Epanechnikov kernel, since according to Härdle et al. (2004) the Epanechnikov (1969) kernel shows a good balance between variance optimization and numerical performance. In nonparametric estimation with an Epanechnikov kernel, Epa, the estimator of *f* is derived by

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} Epa\left(\frac{x - x_i}{h}\right), \quad Epa(u) = \frac{3}{4}(1 - u^2)\mathbf{I}(|u| \leq 1)$$

where *h* is the bandwidth.

Since this is not in our interest, the choice of the density smoothing parameter, *h*, is performed under Mean Integrated Squared Error, MISE. As already mentioned, the AIC will be used later to choose the index.

Since the value of $TM(K)_t$ is unknown and not measurable due to a lack of information, the total market index will be defined and used as a proxy for the $TM(K)$. The definition is inspired by total market indices like S&P (2015) and Wilshire Associates (2015). They use all stocks for which prices are available.

Definition 4. *The total market index (TMI) contains all cryptos in the crypto universe for which prices are available. The cryptos are weighted by their market capitalization.*

This change (5) to

$$TMI_t(k_{max}) = \frac{\sum_{i=1}^{k_{max}} P_{it} Q_{it_l^-}}{Divisor(k_{max})_{t_l^-}}$$

with $k_{\max}$ the maximum number of cryptos with available prices and (8) to

$$\min_{k,\beta} \left|\left|\hat{\varepsilon}(k,\beta)\right|\right|^2 = \left|\left|\varepsilon(k_{max})^{TM} - \varepsilon(k,\beta)^{CRIX}\right|\right|^2 \tag{10}$$
$$s.t.: 1 \leq k \leq k^u$$
$$k^u \in [1, k_{max}]$$
$$s \in [1, k_{max} - k]$$
$$\beta^{1*(k+s)} = (1, \dots, 1, \beta_{k+1}, \dots, \beta_{k+s})^T$$
$$\beta_{k+1}, \dots, \beta_{k+s} \in (-\infty, \infty).$$

We introduced several constraints with (10). We will search for an index under the classical approach of Laspeyres, where $\beta = 1$. We include $\beta_{k+1}, \dots, \beta_{k+s}$, to evaluate if adding s more assets to the index explains the difference between $\varepsilon(k_{max})^{TM}$ and $\varepsilon(k,\beta)^{CRIX}$ better. The first k assets won't be adjusted by a parameter, so no parameter estimation is necessary. This makes the first term a constant. The parameters of the next s assets have to be estimated.

A number of criteria are applicable.

We'll evaluate which criteria to use for our purpose. Since CRIX is to be a benchmark model, all possible models under certain restrictions for the number of parameters are included in the test set, $\Theta_{AIC} = \{CRIX(k_1, \beta), CRIX(k_2, \beta), \ldots\}$, where $k_1,\ k_2, \ldots$ are predefined values. Recall that the intention behind CRIX is to discover the best model to describe the data (benchmark) under a squared loss function.

Define the loss function in (9) for $\hat{\varepsilon}(k, \beta)$,

$$R_T\{\hat{\varepsilon}(k, \beta)\} = \boldsymbol{E}\left(\left|\left|\hat{\varepsilon}(k, \beta)\right|\right|^2\right) \tag{11}$$

and define the number of constituents which minimize the risk in $R_T\ (k, \beta)$ as k^* and s^* for the model set Θ, Shibata (1983).

For this paragraph, consider $\hat{\varepsilon}(k, \beta) \sim\ N(0, \hat{\sigma}(k, \beta)^2)$. k^* and s^* will be interpreted as the number of constituents which balance the bias and variance, define

$$H_T\{\hat{\varepsilon}(k, \beta), s\} = \left|\left|\hat{\varepsilon}(k, \beta)\right|\right|^2 + s\hat{\sigma}(k, \beta)^2 \tag{12}$$

Mean efficiency shall be defined as

$$eff(\Theta) = \frac{H_T(\hat{\varepsilon}(k^*, \beta), s^*)}{R_T(\Theta)} \tag{13}$$

A criterion is defined to be asymptotic mean efficient if

$$a.eff(\Theta) = \lim_{T \to \infty} inf \frac{H_T(\hat{\varepsilon}(k^*, \beta), s^*)}{R_T(\Theta)} = 1 \tag{14}$$

This result holds if the number of constituents, k^* and s^*, increases with T, Shibata (1983). Of course, this result was derived under the assumption of normally distributed errors. Since we are estimating the distribution non parametrically, this result might not hold. For example, Boisbunon et al. (2013) investigate that the result for gaussian distributed errors should hold for spherically symmetric and elliptically contoured distributions too. This leads us to the conclusion that asymptotic optimality might be still given in this case. We oracle so, because for infinitely many observations, the nonparametric estimator tends to the true distribution, Härdle et al. (2004).

The assumption of k^* and s^* increasing with *T* is plausible in this case since longer time horizons *T* would include cryptos which aren't part of shorter ones due to bankruptcy or since they haven't been found yet. Both lead to more complexity. It follows that all of the asymptotically optimal criteria would lead to a mean efficient model choice in terms of squared risk for a given selection of models which fits the intention to discover a best model. It remains to find the suitable one.

Define the characteristic function as

$$\varphi(t) = \int_{-\infty}^{\infty} \exp(\mathbf{i}tx)\, f(x)dx \qquad (15)$$

with **i** ∈ C and *t* ∈ R. The Fourier inversion theorem states (Shephard (1991)):

Theorem 1. *Suppose g and φ are integrable in the Lebesgue sense and*

$$\varphi(t) = \int_{-\infty}^{\infty} \exp(\mathbf{i}tx)\, g(x)dx \qquad (16)$$

then

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\mathbf{i}tx)\, \varphi(t)dt \qquad (17)$$

holds everywhere.

The moment generating function is defined as

$$M(t) = \int_{-\infty}^{\infty} \exp(tx)\, f(x)dx \qquad (18)$$

If the moments generating function exists, it holds

$$\varphi(t) = M(\mathbf{i}t). \qquad (19)$$

We see that the characteristic function depends on the moment generating function of $\hat{\varepsilon}$. Most of the asymptotically optimal criteria depend on the empirical versions of the first two moments of $\hat{\varepsilon}$.

Just the AIC uses the full distribution via the likelihood and therefore all the moments. This makes its information basis richer. For the derivation of the number of index members of CRIX, we will use the AIC, because it uses the most information compared to the other asymptotically optimal criteria: it is the only one which depends on the likelihood.

The maximum likelihood, derived by

$$L\{\hat{\varepsilon}(k,\beta)\} = \max_{\beta} \prod_{t} f\{\hat{\varepsilon}(k,\beta)_t\} \tag{20}$$

where *f*, in (9), represents the density of the $\varepsilon(k,\beta)_t$ over all *t*. The AIC is defined to be

$$AIC\{\hat{\varepsilon}(k,\beta), s\} = -2logL\{\hat{\varepsilon}(k,\beta)\} + s * 2 \tag{21}$$

Akaike (1998).

To decide with AIC which number *k* should be used, a procedure was created which compares the squared difference between log returns of the TMI, see Definition 4, and several candidate indices,

$$\left|\left|\hat{\varepsilon}(k_j,\beta)\right|\right|^2 = \left|\left|\varepsilon(k_{max})^{TM} - \varepsilon(k_j,\beta)^{CRIX}\right|\right|^2 \tag{22}$$

where $\varepsilon(k_j,\beta)^{CRIX}$ is the log return of CRIX version with k_j constituents and $\hat{\varepsilon}(k_j,\beta)$ is the respective difference. The candidate indices, $CRIX(k_j,\beta)$, have different numbers of constituents which fulfill $k_1 < k_2 < k_3 < \cdots$, where $k_j = k_1 + s(j-1)$. Therefore, the number of constituents between the indices are equally spaced. By definition both information criteria evaluate the differences, $\hat{\varepsilon}(k_j,\beta)$, between the candidates and the TMI with the respective likelihood $L\{\hat{\varepsilon}(k_j, \beta), s\}$, see Equation (21). This procedure implies that the IC method evaluates if s more assets add information to CRIX. If so, these assets are added to the intercept and the next s assets are tested for. We expect assets with a higher market capitalization to have a higher influence on the AIC, so we formulated the following theorem:

Theorem 2. *The rate of improvement of the AIC depends on the relative value of an asset in the market.*

Therefore, we will follow the common practice to include the assets with the highest market capitalization in the index,

$$\operatorname*{argmax}_i \sum_{j=1}^{k} P_{j,i,t_l^-} Q_{j,i,t_l^-}, \qquad i \in \{1, \dots, K\}.$$

Thus, we apply a top-down approach to decide about the number of index constituents. For the sorting of the index constituents by highest market capitalization, we just rely on the closing data of the last day of a month. We chose to do so, since the next periods CRIX will just depend on Q_{i,t^-}, (2), and not on data which lie further in the past.

Since the differences between the $TMI(k_{max})$ and $CRIX(k_j, \beta)$ are caused over time by the missing time series in $CRIX(k_j, \beta)$, the independence assumption of the $\hat{\varepsilon}(k_j, \beta)$ for all j cannot be fulfilled by construction. But Györfi et al. (1989) give arguments that under certain conditions, the rate of convergence is essentially the same as for an independent sample. Since the same data are used to estimate f^j and the information criterion, a “leave-one-out” cross-validation procedure is performed in order to have in-sample data for the calculation of the density and pseudo-out-of-sample data for the information criterion, hence new observations; see Boisbunon et al. (2013).

3.2.3 Crix Rules

The constituents of the indices are regularly checked so that the corresponding index always represents its asset universe well. It is common to do this on a quarterly basis. In case of CRIX this reallocation is much faster. In the past, coins have shown a very volatile behavior, not just in the manner of price volatility. In some weeks, many occur out of nothing in the market and many others vanish from the market even when they were before very important, e.g., Auroracoin. This calls for a faster reallocation of the market benchmark than on a quarterly basis. The monthly reallocation makes sure that CRIX catches the momentum of the cryptocurrency market well. Therefore, on the last day of every month, the cryptos which had the highest market capitalization on the last day in the last month will be checked and the first k will be included in CRIX for the coming month.

Since a review of an index is commonly performed on a quarterly basis the number of index members of CRIX will be checked on a quarterly basis too. The described procedure (paragraph 3.2.2) will be applied to the observations from the last three months on the last day of the third month after the markets closed. The number of index constituents, *k*, will be used for the next three months. Thus, CRIX corresponds to a monthly rebalanced portfolio which number of constituents is reviewed quarterly.

The number of constituents is recalculated quarterly to ensure an up-to-date fit to the current market situation. The reallocation period of the CriX is 1 month, at this time point the liquidity will be checked again. It may happen that a crypto has a high market capitalization but is not traded frequently. Two measures are applied which are modified versions of the liquidity rules from the STOXX Japan 600 and the AEX Family. The applied rules are the following:

1. *0.25* percentile of *ADTV*

$$ADTV_i \geq ADTV_{0.25}$$

where $ADTV_{0.25}$ is the 0.25 percentile of the *ADTV* distribution of all cryptos in the last period and $ADTV_i$ is the *ADTV* of a single crypto.

2. *0.25* percentile of Average Daily Traded Coins (*ADTC*)

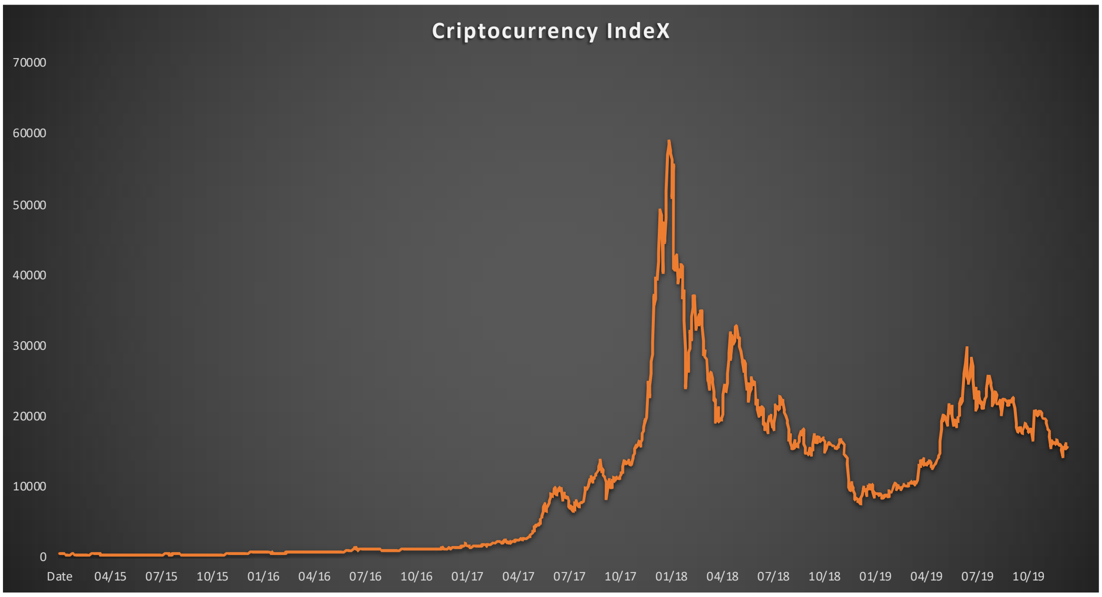
$$ADTC_i \geq ADTC_{0.25}$$

where $ADTC_{0.25}$ is the *0.25* percentile of the *ADTC*s of all cryptos in the last period and $ADTC_i$ is the *ADTC* of a single crypto.
If a crypto fulfills at least one of the two rules, it is eligible for the CRIX set of constituents.

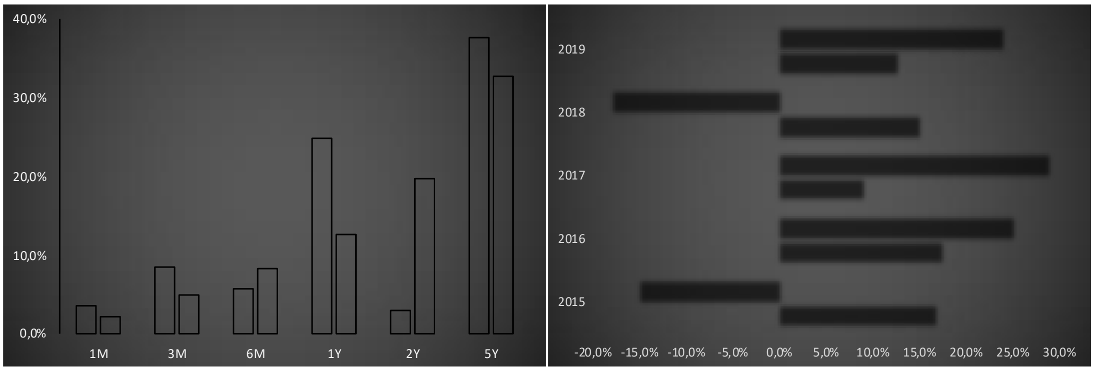
It may happen that some data are missing for some of the analyzed time series. If an isolated missing value occurs alone in the dataset, meaning that the values before and after it are not missing, then Missing At Random (MAR) is assumed. This assumption means that just observed information cause the missingness, Horton and Kleinman (2007). The Last-Observation-Carried-Forward (LOCF) method is then applied to fill the gap for the application of the AIC. We did not choose a different approach since a regression or imputation may alter the data in the wrong direction. By LOCF, we imply no change and just do not exclude the crypto. If two or more data are missing in a row, then the MAR assumption may be violated, therefore no method is applied. The corresponding time series is then excluded from the computation in the derivation period. If data are missing during the computation of the index values, the LOCF method is applied too.

This is done to make the index insensitive to this crypto at this time point. CRIX should mimic market changes, therefore an imputation or regression method for the missing data would distort the view of the market.

3.2.4 CriX Composition and Performance

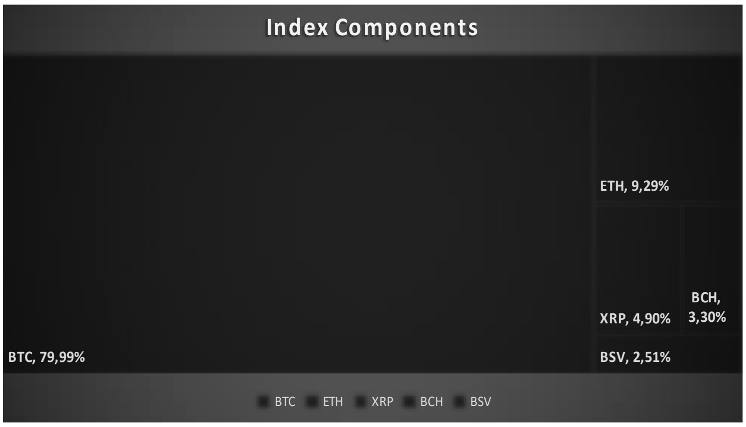


Index Performance and Composition



Buy-Hold	1M	3M	6M	1Y	2Y	5Y
Return	-2.65%	-12.10%	-47.66%	62.35%%	-66.97%	2422.5%
Std Dev	14.72%	33.85%	49.15%	68.50%	110.15%	164.37%
SR	-0.19	-0.37	-0.99	0.88	-0.64	14.68

Annualized	2015	2016	2017	2018	2019
Return	16.92%	131.72%	2546.1%	-81.38%	60.31%
Std Dev	64.19%	50.18%	89.21%	85.77%	68.02%
SR	0.23	2.59	28.52	-0.97	0.86



Crypto	Weights
BTC	79.99%
ETH	9.29%
XRP	4.90%
BCH	3.30%
BSV	2.51%

4. Dynamic Portfolio Allocation: Markowitz Approach

In this Chapter I am going to present the methods employed in the portfolio optimization problem. The first paragraph describes the key assumptions of the Markowitz approach and also its drawback in the application. The second paragraph introduce an alternative way in computing the variance-covariance matrix for a list of returns. This method is used in the portfolio optimization problem to check if there is an improvement in the overall portfolios risk-return. The last paragraph is used to report all the results obtained in the application of the above-mentioned methods and at the same time explain in more details their application.

4.1 Introduction

In the early 1960s, the investment community talked about risk, but there was no specific measure for the term. To build a portfolio model, however, investors had to quantify their risk variable. The basic portfolio model was developed by Harry Markowitz (1952, 1959), who derived the expected rate of return for a portfolio of assets and an expected risk measure. Markowitz showed that the variance of the rate of return was a meaningful measure of portfolio risk under a reasonable set of assumptions. More important, he derived the formula for computing the variance of a portfolio. This portfolio variance formula not only indicated the importance of diversifying investments to reduce the total risk of a portfolio but also showed how to effectively diversify. The Markowitz model is based on several assumptions regarding investor behavior:

1. Investors consider each investment alternative as being represented by a probability distribution of expected returns over some holding period.
2. Investors maximize one-period expected utility, and their utility curves demonstrate diminishing marginal utility of wealth.
3. Investors estimate the risk of the portfolio on the basis of the variability of expected returns.
4. Investors base decisions solely on expected return and risk, so their utility curves are a function of expected return and the expected variance (or standard deviation) of returns only.

5. For a given risk level, investors prefer higher returns to lower returns. Similarly, for a given level of expected return, investors prefer less risk to more risk.

Under these assumptions, a single asset or portfolio of assets is considered to be efficient if no other asset or portfolio of assets offers higher expected return with the same (or lower) risk or lower risk with the same (or higher) expected return.

The model is an ex-ante model of portfolio analysis. This means that to implement the Markowitz model the expected return, variance and covariance must be estimated. Typically, the procedure for obtaining these inputs is to calculate the historical values ex-post. Using historical returns, you can easily calculate the other two parameters by assigning an equal weight to each period observed in the market. However, using ex post data to estimate ex-ante parameters of the portfolio can lead to disappointing results.

One of the first problems that can lead to failure of this method is due to the estimate of the risk measure. The Markowitz's portfolio theory uses data with equal weights. Doing this does not take into account the dynamics of the market structure. One of the ways to reduce the estimate of these errors is to use exponential weighted return and variances. Exponentially weighted data assign greater weight to more recent observations, taking into account the dynamic structure of the market. This is also one of the reasons why I decided to implement an EWMA model to calculate the variance-covariance matrix of returns. If we look at the graph of cryptos it is easy to notice a peak due to a potential bubble.

It’s clear from recent studies that most of the financial academic literature has focused on modeling the covariance of the securities returns. Beyond the academic literature, there are also many industrial contributions as RiskMetrics (1996). JP Morgan and Reuters introduced a methodology for the determination and diversification of the market risk of portfolios, method that immediately became popular.

The aim is to use the matrix of the covariance calculated with the EWMA method. The covariance matrix thus obtained is used as input for the Markowitz theory.

Subsequently, for illustrative purposes, we are going to see which the real differences are produced by the two methods (unweighted and weighted covariance matrix) on the efficient frontier. In addition, we will comment if the EWMA method can be better in determining a portfolio composed of both real assets that Cryptos.

In the next section we introduce the portfolio selection process Markowitz (1952).

4.2 Portfolio Optimization

The modern portfolio theory is based on the idea that investors seek high returns from investments trying to minimize their risk. Although, expecting higher returns with lower risk level may seem contradictory. That is why building a portfolio requires a trade-off between risk and return.

Investors choose how much of their wealth to distribute in each financial instrument, in this way, diversifying their financial exposure. Mean-variance optimization developed by Markowitz (1952) can be used to determine how an investor distributes his wealth between the various financial instruments.

The proportion of financial instruments in a portfolio depends not only from the mean or the variance of their returns but also from their relationship, covariance. For this purpose, the mean and variance of returns together with their covariance are calculated as portfolio optimization input. The Markowitz's portfolio theory uses a scheme with equal weights to calculate the parameters listed above. Once the input parameters are obtained, risk and return for each portfolio of financial instruments are calculated.

4.2.1 Matrix Algebra

When the research is done selecting a big quantity of data it can be difficult to compute portfolio expected returns, variances and covariances using algebra. Matrix algebra is a great simplification of calculations.

The portfolio return:

$$r_p = w'E[r] = w'\mu = w_1\mu_1 + w_2\mu_2 + \cdots + w_n\mu_n,$$

where:

n = number of assets,

w = vector of weights (n x 1),

μ = vector of mean returns (n x 1).

The variance of the portfolio:

$$\sigma_p^2 = var(w'R) = w'\Sigma w = (w_1, \dots, w_n) \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \cdots & \sigma_n^2 \end{bmatrix} \begin{pmatrix} w_1 \\ \cdots \\ w_n \end{pmatrix}$$

The goal of portfolio optimization is to find a combination of assets (weights) that minimize the standard deviation of portfolio returns for each level of expected return. In other words, a combination of assets that maximize the expected return of the portfolio for each level of risk.

The optimization problem faces certain constraints, a budget constraint and a short-selling constraint. However, we can summarize the portfolio choice problem:

$$\min_{w} \sigma_p^2 = w'\Sigma w \quad s.t.$$

$$\mu_p = w'\mu = \mu_0$$

$$w'1 = 1$$

$$w_i \geq 0; \qquad for\ any\ i = 1,2, \dots, n$$

By changing the level of expected return and solving iteratively the model we can draw the efficient frontier. The efficient frontier is composed of portfolios, combinations of financial instruments, which have the higher expected return given the same level of risk, or a lower risk given the same level of expected return. No portfolio below the efficient frontier can be considered optimal. No portfolio on the efficient frontier dominates another portfolio on the efficient frontier, the choice about the level of risk-return remains to the investor.

4.3 Exponential Weighted Scheme (EWMA)

We can measure variance historically or implicitly (implied volatility). When measuring historically, the easiest method is a simple variance.

Exponentially weighted moving average is one of the extensions how to measure historical volatility. This method put more weights on recent observations, and it let these current observations to make a bigger influence on the forecasted volatility comparing it with older observations. In EWMA model the latest data has the highest weights and weights for previous data decline exponentially over time.

There are two advantages of EWMA model when compared to simple historical models and simple moving average (MA) model which puts the same weights to the all data points.

The first advantage is that in the real-world volatility is affected more by recent events comparing it with some event in the past and EWMA at the same time gives more attention to those recent events. At the same time simply moving average model weights recent event as same as event in the past and this can lead to misleading too low volatility forecast results if, for example, specific shock suddenly drops out of the sample or vice versa if specific shock is in the sample for a long period of time.

The second advantage is that "the effect on volatility of a single given observation declines at an exponential rate as weights attached to recent events fall".

Exponentially weighted moving average model can be computed in several ways. One of them is the following:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)u_{t-1}^2$$

with $0 < \lambda < 1$, where σ_t^2 is the variance at time *t* and u_t^2 are the square root of returns at time *t*. For the exercise λ is set equals to 0.86, as suggested from theory.

Additionally, we can use the matrix notation

$$\boldsymbol{D}_t = \lambda \boldsymbol{D}_{t-1} + (1-\lambda)(\boldsymbol{u}_{t-1}\boldsymbol{u}'_{t-1}),$$

where $\boldsymbol{D}_t$ is the variance covariance matrix at time *t*.
We will use the EWMA method also in comparison to the "un-weighted" covariance matrix under Markowitz. The aim is to check if weighting more recent observations this will lead to a better optimal portfolio in the mean-variance optimization.

4.4 Preliminaries and Process

In this paragraph I am going to list the results from the methods employed in the portfolio optimization and at the same time providing additional information about the process. Both methods follow the minimization problem as explained in (4.2). The basic difference between the Markowitz Approach and the EWMA Approach is the way in which we compute the variance-covariance matrix of the assets return.

We have seen that, under Markowitz, the variance-covariance matrix is "un-weighted", it takes into consideration all the past and actual observations assigning the same weight to each one of them. In this way, past and present shocks are weighted the same. As opposed to this approach, the EWMA method allow us to put a different weight on past and actual observations. Present observations are weighted more than past observation, depending on the parameter λ, also called the "smoothing" parameter.

For the purpose of this exercise, all the portfolios have a target return of 0.08 (8%) annually. The choice of the target return is made under the assumption that the majority of the indices employed have securities listed in United States. Thus, I decided to follow the average return of the S&P500, which in the last century delivered approximately between 8-9% return annually.

To implement a **dynamic approach** I have divided between the expected returns and the variance-covariance matrices estimated and realized. The difference among them lies in the way they have been calculated. The expected returns and variance **estimated** of the portfolios are computed as

$$\mu_{p,t} = \boldsymbol{w}_t \boldsymbol{\mu}_{t,i}$$

$$\sigma^2_{p,t} = \boldsymbol{w}_t \boldsymbol{\Sigma}_{\mathrm{t}} \boldsymbol{w}'_t$$

or, under the EWMA approach

$$\sigma^2_{p,t} = \boldsymbol{w}_t \mathbf{D}_{\mathrm{t}} \boldsymbol{w}'_t$$

The expected returns and variance **realized** of the portfolios are computed as

$$\hat{\mu}_{p,t} = \widehat{\boldsymbol{w}}_{t-1} \boldsymbol{\mu}_{t,i}$$

$$\hat{\sigma}^2_{p,t} = \widehat{\boldsymbol{w}}_{t-1} \boldsymbol{\Sigma}_{\mathrm{t}} \widehat{\boldsymbol{w}}'_{t-1}$$

or under the EWMA approach

$$\hat{\sigma}^2_{p,t} = \widehat{\boldsymbol{w}}_{t-1} \mathbf{D}_{\mathrm{t}} \widehat{\boldsymbol{w}}'_{t-1}$$

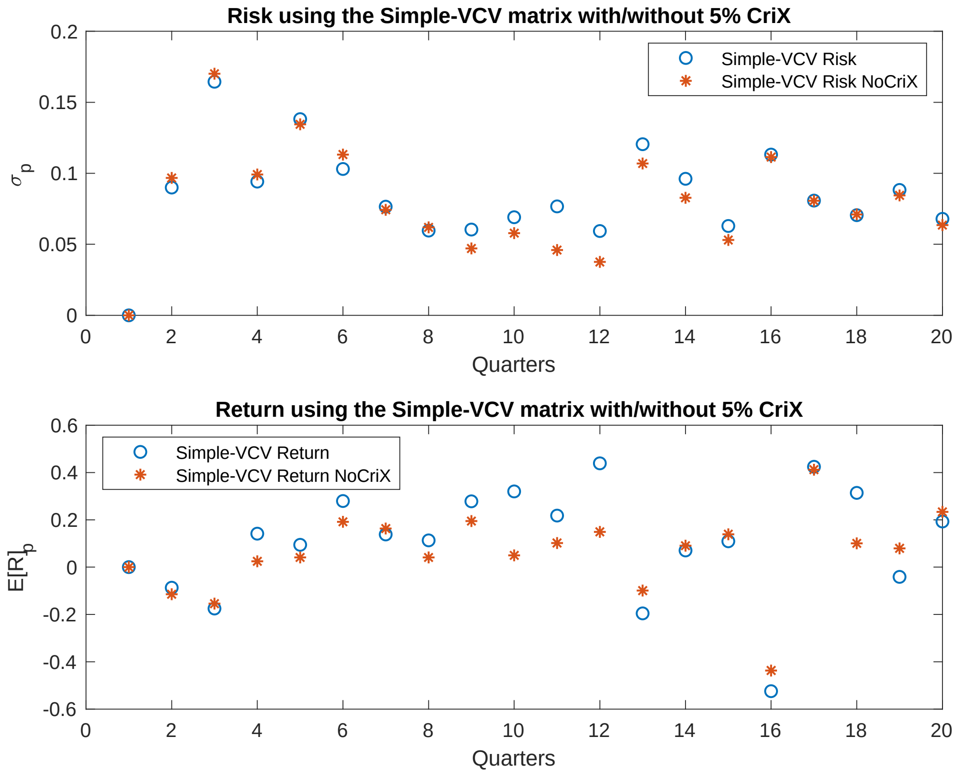
It means that the realized returns and variance-covariance matrices are obtained multiplying the expected returns and the variance-covariance matrices (at time *t*) by weights (of time $t-1$) that are the results of the portfolio optimization problem. In other words, I stored the portfolio weights derived from the portfolio optimization problem (*t*) and applied to the following expected return and variance-covariance matrix (*t+1*). In order to implement this procedure I have used a rolling window of 3 months (Quarterly results). In this way, we do not rely on historical ex-post data to estimate ex-ante parameters, rather with the use of a rolling window we are considering a time-varying VCV matrix.

4.5 Results

In this section I classified the results with the aim to provide more insights on the VCV matrix involved in the portfolio minimization problem. I divided between the Simple-VCV Matrix computed as seen in section 4.2 and the EWMA-VCV Matrix described in section 4.3. Furthermore, the section is structured on the basis of the exposure we have towards the cryptos. The analysis has been conducted on 4 different types of allocation. The starting point is at 5% (at the lower bound) of exposure and rise to 10% and 15% which is the upper bound of the minimization problem. The last type of allocation is an extreme case where the total real-assets weight is fixed at 55% keeping the exposure to cryptos at 45%. All the analysis are conducted from 01/2015 to 01/2020, a 5 years' time-horizon. The standard deviations and returns are expressed in annual terms.

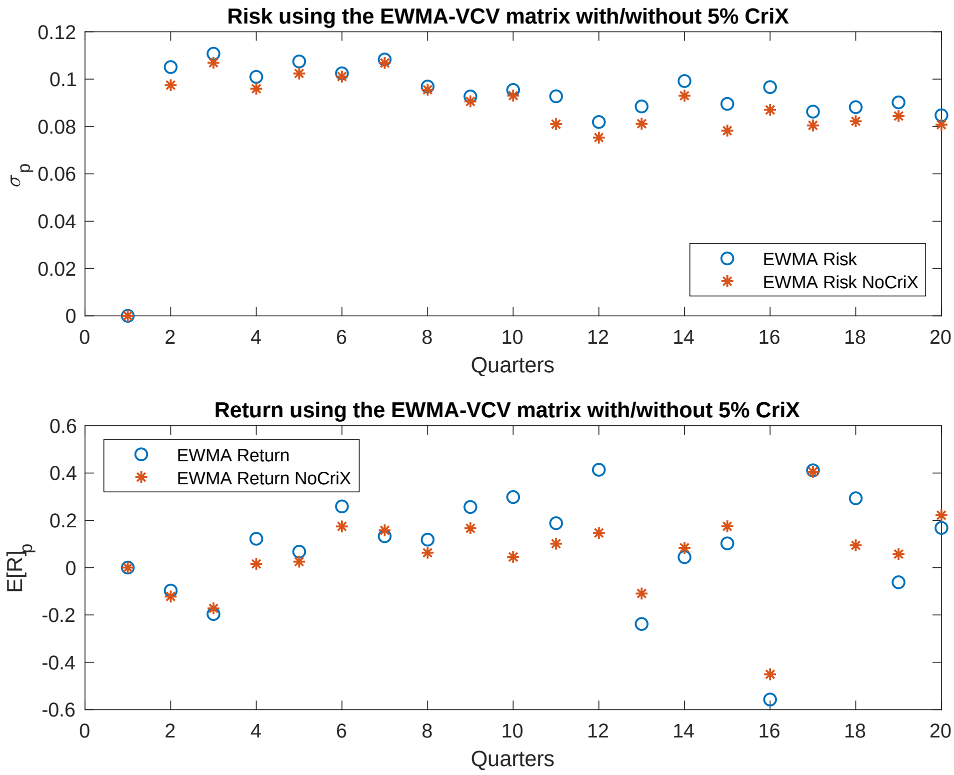
Portfolio Allocation – 5% Exposure

Figure 1 – Simple-VCV Matrix with and without CriX



		Simple-VCV Matrix			
		r_{crix}	σ_{crix}	r_{nocrix}	σ_{nocrix}
2015	Q1	0	0	0	0
	Q2	-0.0870	0.0899	-0.1146	0.0967
	Q3	-0.1745	0.1644	-0.1542	0.1701
	Q4	0.1413	0.0941	0.0247	0.0991
2016	Q1	0.0943	0.1381	0.0409	0.1344
	Q2	0.2795	0.1030	0.1914	0.1132
	Q3	0.1380	0.0764	0.1631	0.0743
	Q4	0.1131	0.0596	0.0410	0.0619
2017	Q1	0.2778	0.0604	0.1945	0.0470
	Q2	0.3198	0.0690	0.0498	0.0578
	Q3	0.2180	0.0766	0.1020	0.0459
	Q4	0.4389	0.0593	0.1491	0.0375
2018	Q1	-0.1952	0.1205	-0.0991	0.1069
	Q2	0.0706	0.0961	0.0907	0.0827
	Q3	0.1093	0.0628	0.1391	0.0530
	Q4	-0.5238	0.1130	-0.4366	0.1114
2019	Q1	0.4241	0.0807	0.4118	0.0806
	Q2	0.3138	0.0705	0.1007	0.0710
	Q3	-0.0412	0.0882	0.0795	0.0844
	Q4	0.1926	0.0678	0.2336	0.0635

Figure 2 – EWMA-VCV Matrix with and without CriX

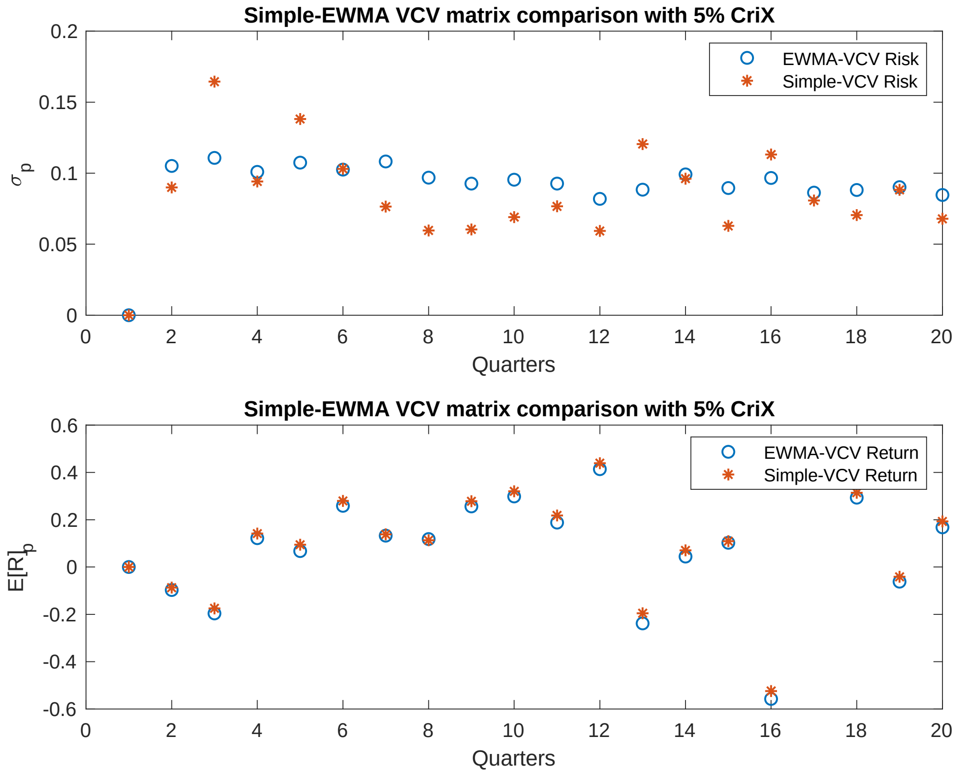


EWMA-VCV Matrix

		r_{crix}	σ_{crix}	r_{nocrix}	σ_{nocrix}
2015	Q1	0	0	0	0
	Q2	-0.0972	0.1051	-0.1217	0.0974
	Q3	-0.1961	0.1107	-0.1722	0.1069
	Q4	0.1219	0.1009	0.0158	0.0959
2016	Q1	0.0676	0.1074	0.0254	0.1024
	Q2	0.2583	0.1024	0.1741	0.1011
	Q3	0.1327	0.1083	0.1576	0.1068
	Q4	0.1182	0.0968	0.0628	0.0955
2017	Q1	0.2564	0.0926	0.1662	0.0905
	Q2	0.2982	0.0954	0.0452	0.0930
	Q3	0.1877	0.0927	0.1011	0.0809
	Q4	0.4137	0.0818	0.1466	0.0753

2018	Q1	-0.2380	0.0884	-0.1092	0.0811
	Q2	0.0444	0.0992	0.0838	0.0929
	Q3	0.1027	0.0895	0.1745	0.0782
	Q4	-0.5575	0.0966	-0.4507	0.0870
2019	Q1	0.4110	0.0863	0.4050	0.0804
	Q2	0.2933	0.0881	0.0944	0.0822
	Q3	-0.0618	0.0901	0.0574	0.0844
	Q4	0.1675	0.0846	0.2214	0.0807

Figure 3 – Comparison between Simple and EWMA-VCV Matrix



Simple-EWMA VCV matrix comparison with 5% CriX

Simple-EWMA VCV matrix comparison with 5% CriX

Figure 4 – Comparison between Simple and EWMA-VCV Matrix

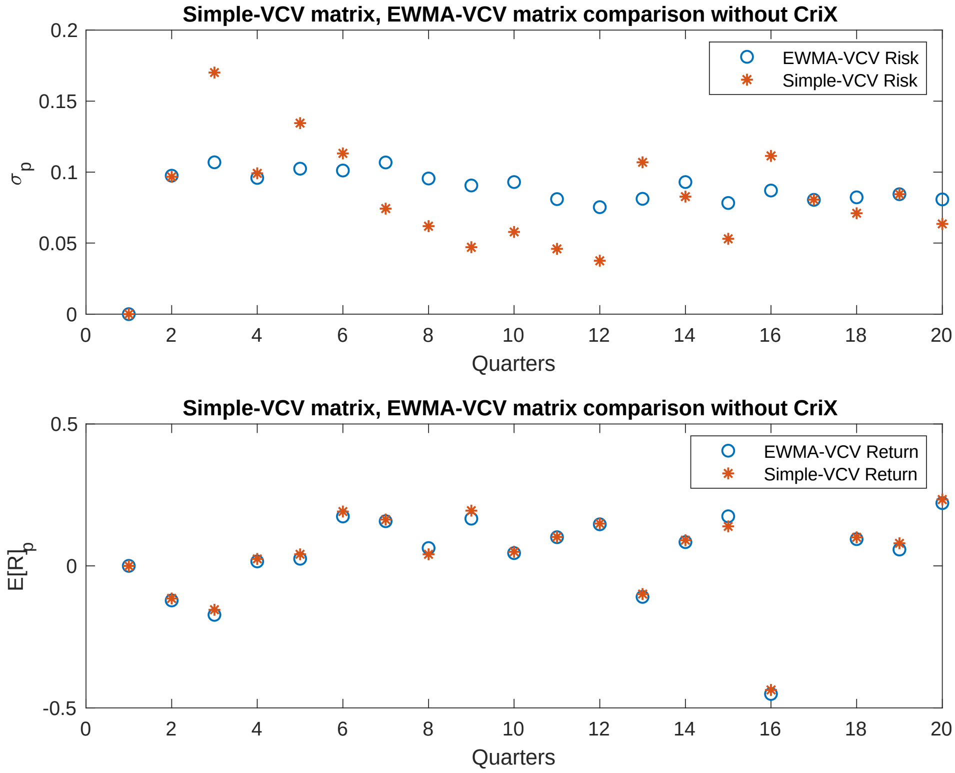


Figure 5 – SR obtained under the Simple and EWMA-VCV Matrix

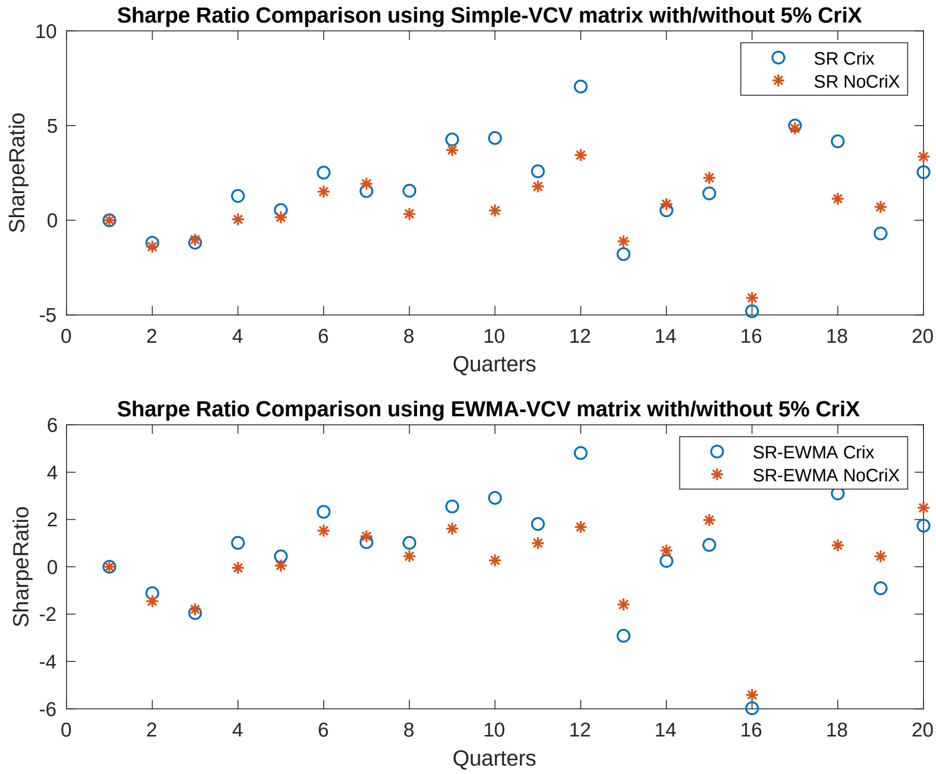
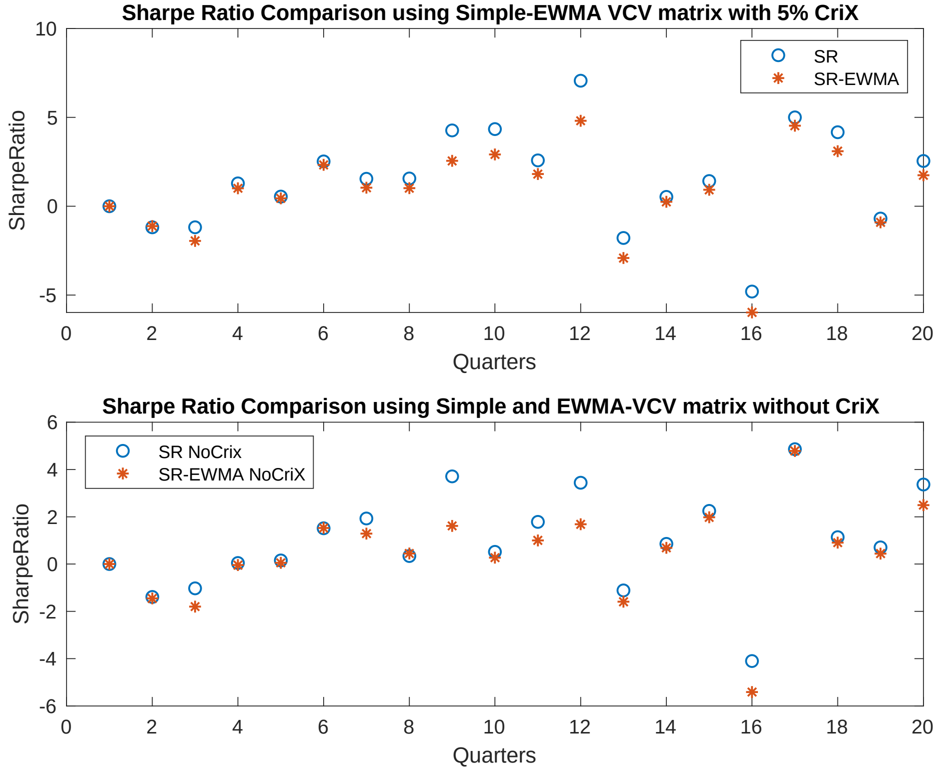


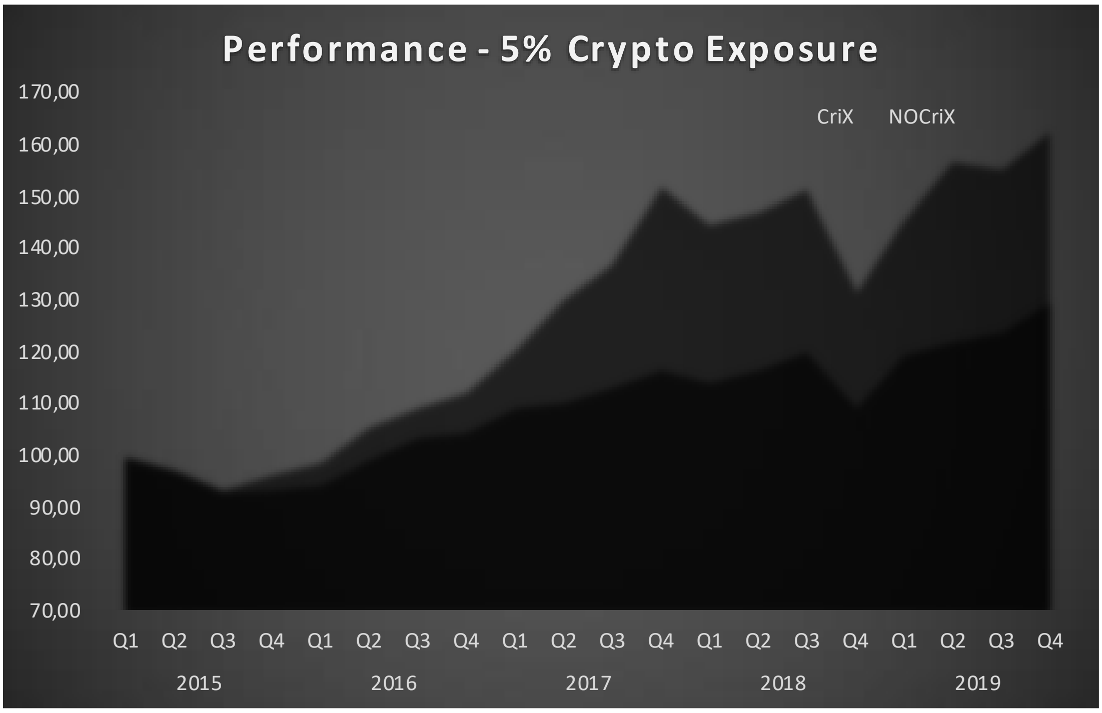
Figure 6 – SR comparison between Simple and EWMA-VCV Matrix



		Simple-VCV Matrix		EWMA-VCV Matrix	
		SR_{crix}	SR_{nocrix}	SR_{crix}	SR_{nocrix}
2015	Q1	0	0	0	0
	Q2	-1.19	-1.39	-1.12	-1.45
	Q3	-1.18	-1.02	-1.95	-1.80
	Q4	1.29	0.05	1.01	-0.04
2016	Q1	0.54	0.16	0.44	0.05
	Q2	2.52	1.51	2.33	1.52
	Q3	1.54	1.92	1.04	1.29
	Q4	1.56	0.34	1.01	0.45
2017	Q1	4.27	3.71	2.55	1.61
	Q2	4.34	0.51	2.91	0.27
	Q3	2.58	1.78	1.81	1.00
	Q4	7.06	3.44	4.81	1.68

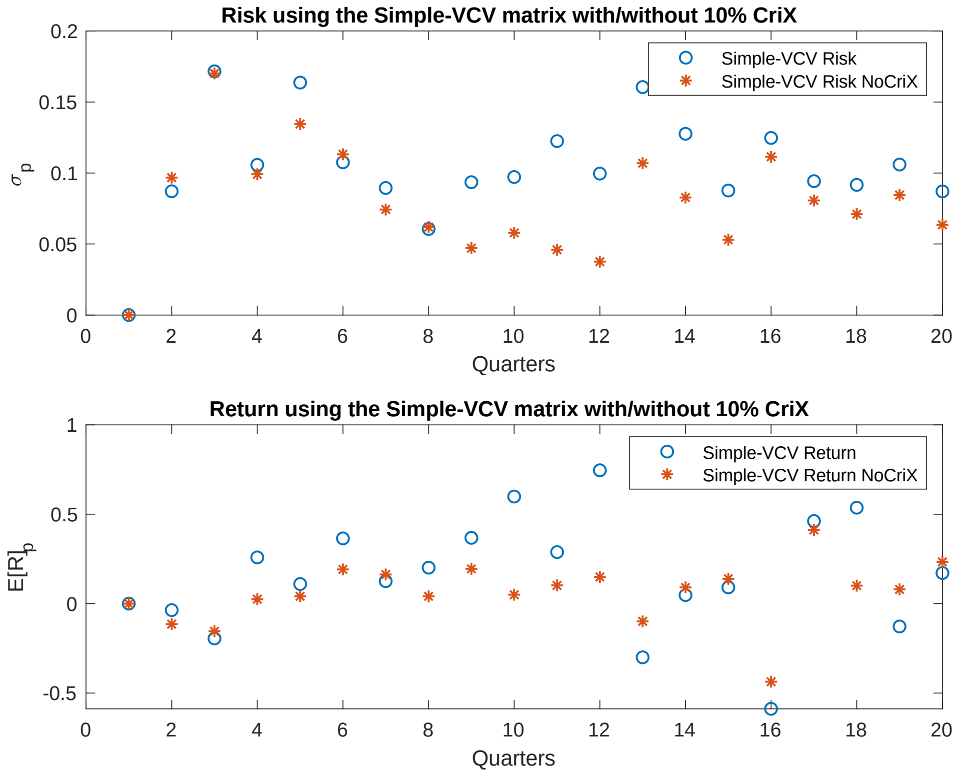
2018	Q1	-1.78	-1.11	-2.92	-1.59
	Q2	0.53	0.85	0.25	0.68
	Q3	1.42	2.24	0.92	1.97
	Q4	-4.80	-4.10	-5.98	-5.41
2019	Q1	5.00	4.86	4.53	4.78
	Q2	4.17	1.13	3.10	0.90
	Q3	-0.69	0.70	-0.91	0.44
	Q4	2.54	3.36	1.74	2.49

Figure 7 – Performance of 100€ invested in portfolio



Portfolio Allocation – 10% Exposure

Figure 8 – Simple-VCV Matrix with and without CriX

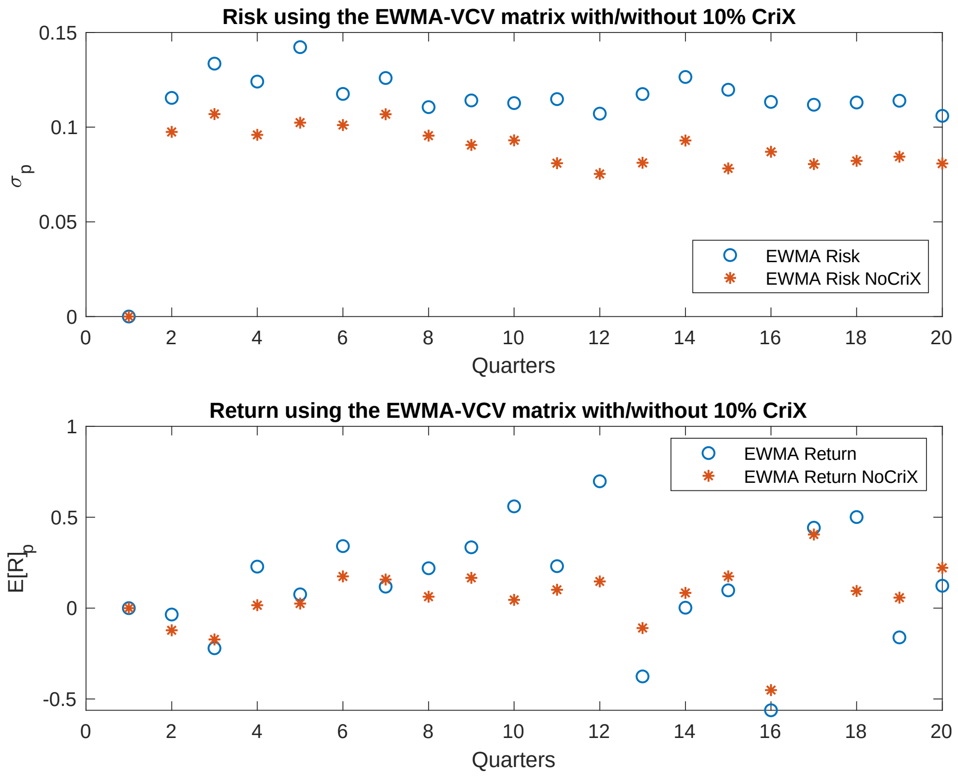


Simple-VCV Matrix

		r_{crix}	σ_{crix}	r_{nocrix}	σ_{nocrix}
2015	Q1	0	0	0	0
	Q2	-0.0361	0.0871	-0.1146	0.0967
	Q3	-0.1944	0.1715	-0.1542	0.1701
	Q4	0.2582	0.1057	0.0247	0.0991
2016	Q1	0.1096	0.1636	0.0409	0.1344
	Q2	0.3646	0.1075	0.1914	0.1132
	Q3	0.1255	0.0894	0.1631	0.0743
	Q4	0.2015	0.0606	0.0410	0.0619
2017	Q1	0.3679	0.0935	0.1945	0.0470
	Q2	0.5990	0.0971	0.0498	0.0578
	Q3	0.2886	0.1224	0.1020	0.0459
	Q4	0.7453	0.0995	0.1491	0.0375

2018	Q1	-0.3001	0.1604	-0.0991	0.1069
	Q2	0.0474	0.1275	0.0907	0.0827
	Q3	0.0907	0.0876	0.1391	0.0530
	Q4	-0.5883	0.1247	-0.4366	0.1114
2019	Q1	0.4620	0.0942	0.4118	0.0806
	Q2	0.5366	0.0916	0.1007	0.0710
	Q3	-0.1273	0.1059	0.0795	0.0844
	Q4	0.1711	0.0870	0.2336	0.0635

Figure 9 – EWMA-VCV Matrix with and without CriX



		EWMA-VCV Matrix			
		r_{crix}	σ_{crix}	r_{nocrix}	σ_{nocrix}
2015	Q1	0	0	0	0
	Q2	-0,0356	0,1154	-0,1217	0,0974
	Q3	-0,2207	0,1335	-0,1722	0,1069
	Q4	0,2282	0,1241	0,0158	0,0959
2016	Q1	0,0754	0,1422	0,0254	0,1024
	Q2	0,3410	0,1175	0,1741	0,1011
	Q3	0,1183	0,1259	0,1576	0,1068
	Q4	0,2193	0,1106	0,0628	0,0955
2017	Q1	0,3345	0,1141	0,1662	0,0905
	Q2	0,5593	0,1126	0,0452	0,0930
	Q3	0,2308	0,1148	0,1011	0,0809
	Q4	0,6976	0,1071	0,1466	0,0753
2018	Q1	-0,3761	0,1174	-0,1092	0,0811
	Q2	0,0019	0,1265	0,0838	0,0929
	Q3	0,0973	0,1197	0,1745	0,0782
	Q4	-0,5619	0,1133	-0,4507	0,0870
2019	Q1	0,4423	0,1118	0,4050	0,0804
	Q2	0,5012	0,1130	0,0944	0,0822
	Q3	-0,1614	0,1139	0,0574	0,0844
	Q4	0,1228	0,1060	0,2214	0,0807

Figure 10 – Comparison between Simple and EWMA-VCV Matrix

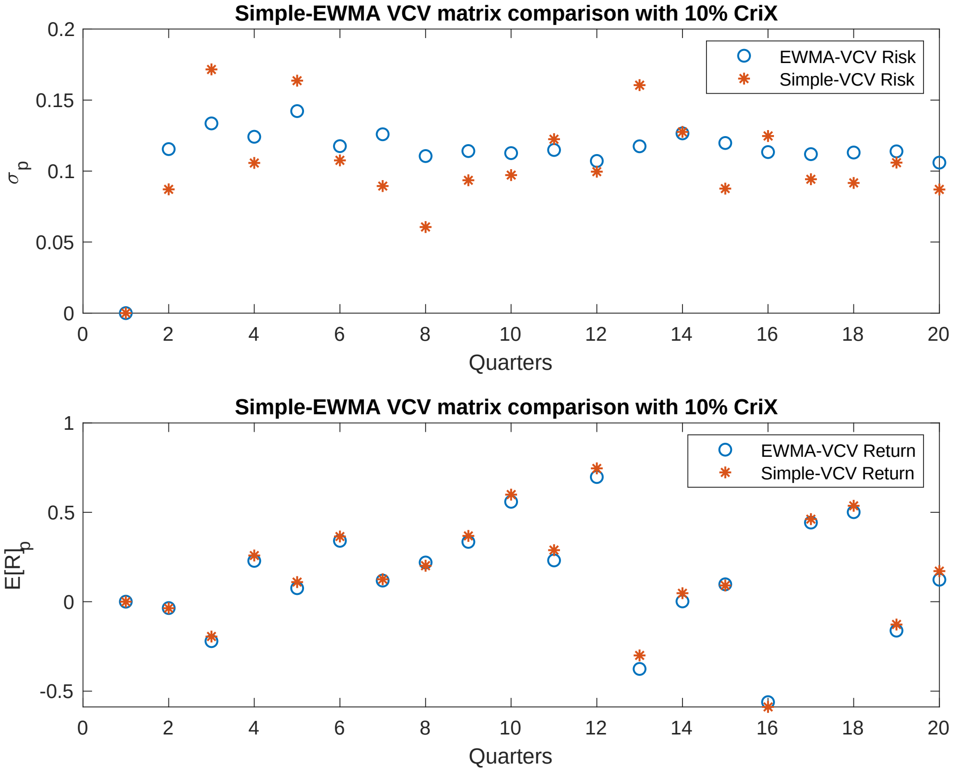


Figure 11 – Comparison between Simple and EWMA-VCV Matrix

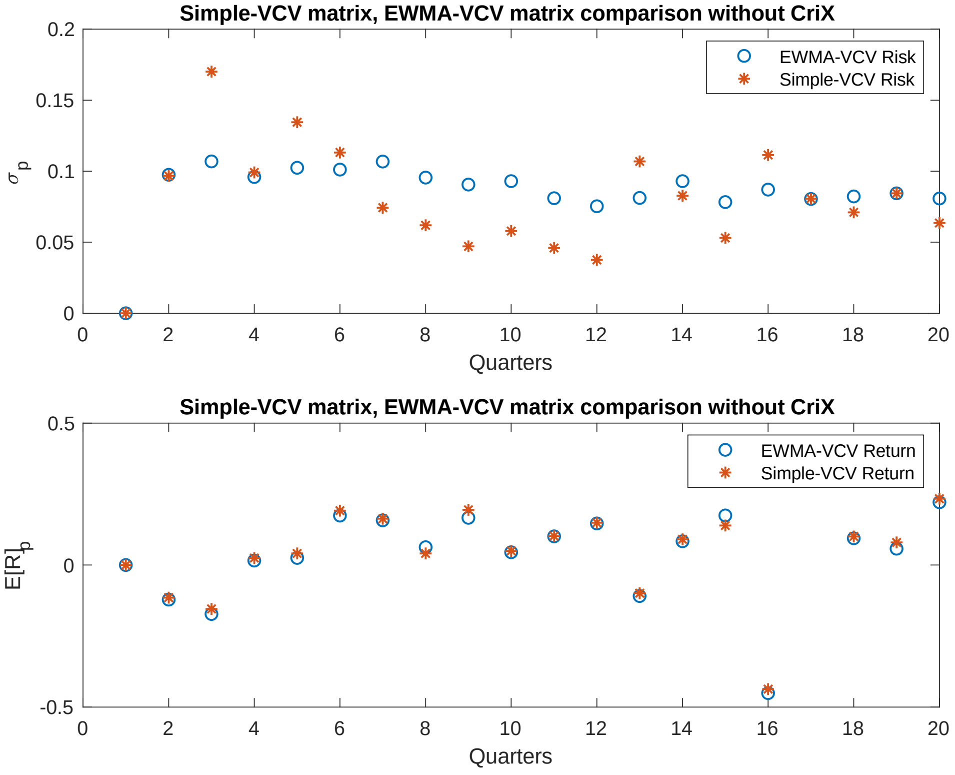


Figure 12 – SR obtained under the Simple and EWMA-VCV Matrix

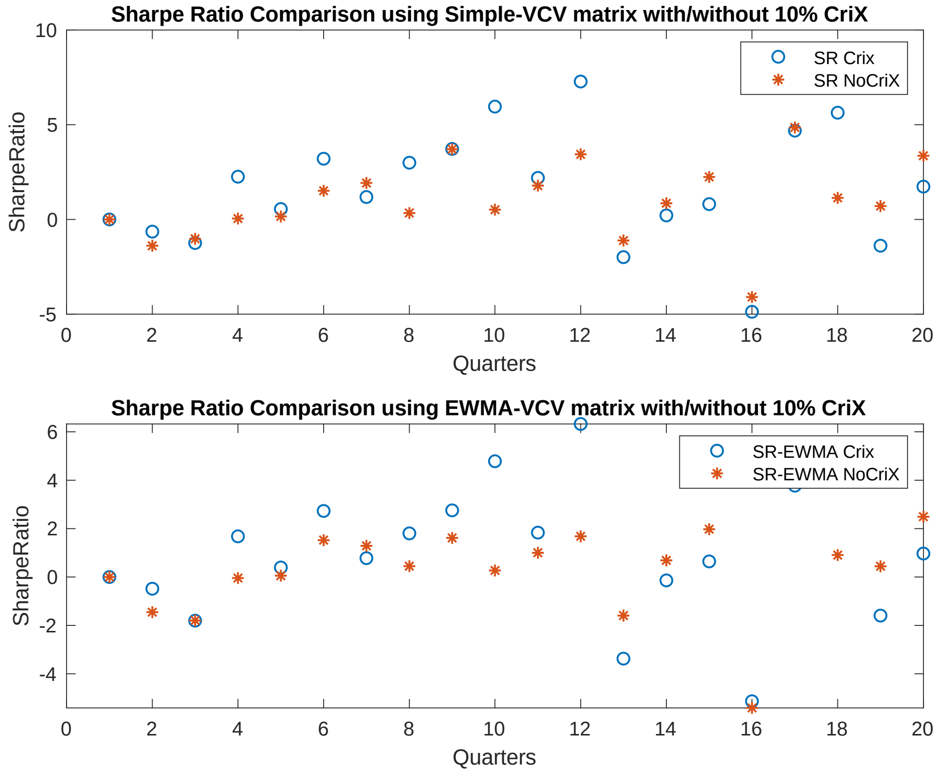
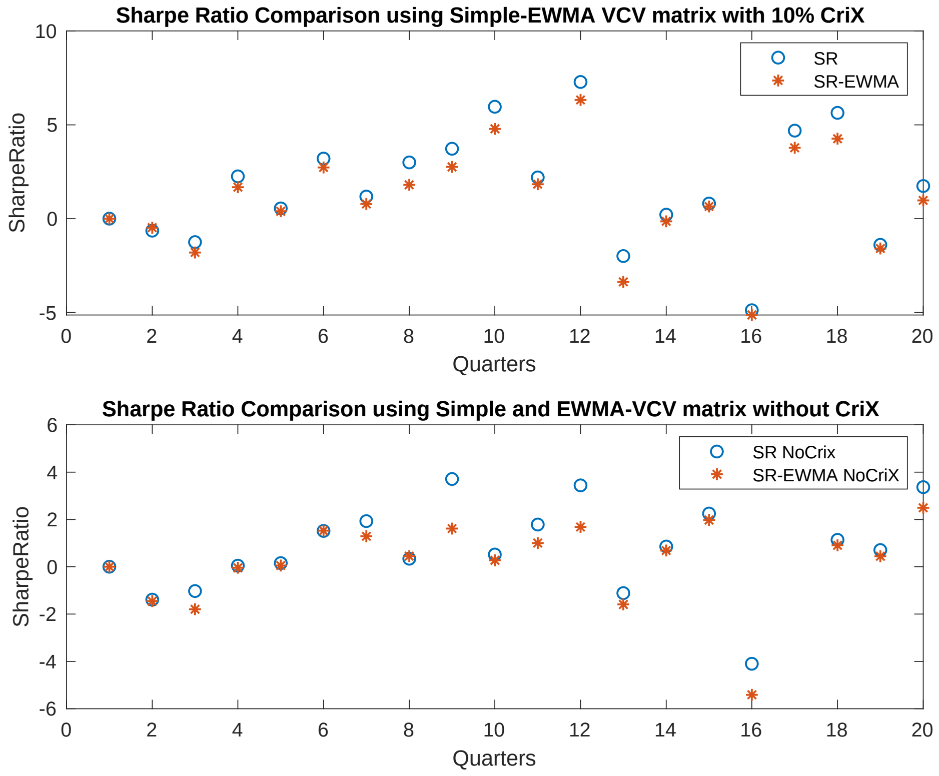
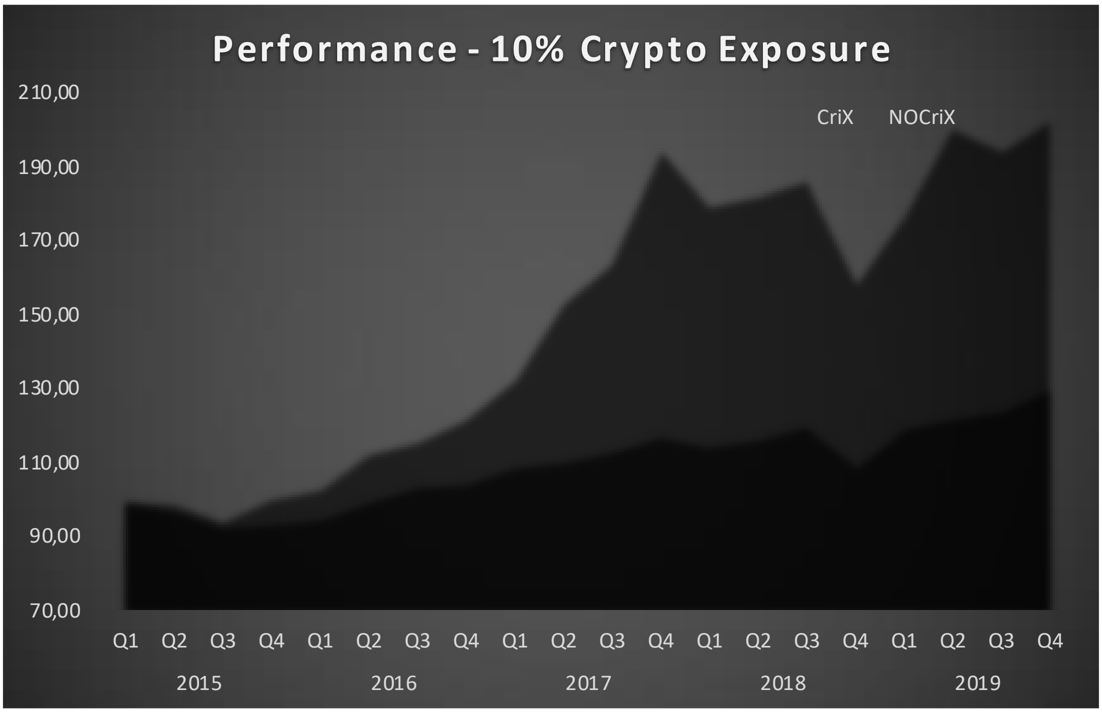


Figure 13 – SR comparison between Simple and EWMA-VCV Matrix



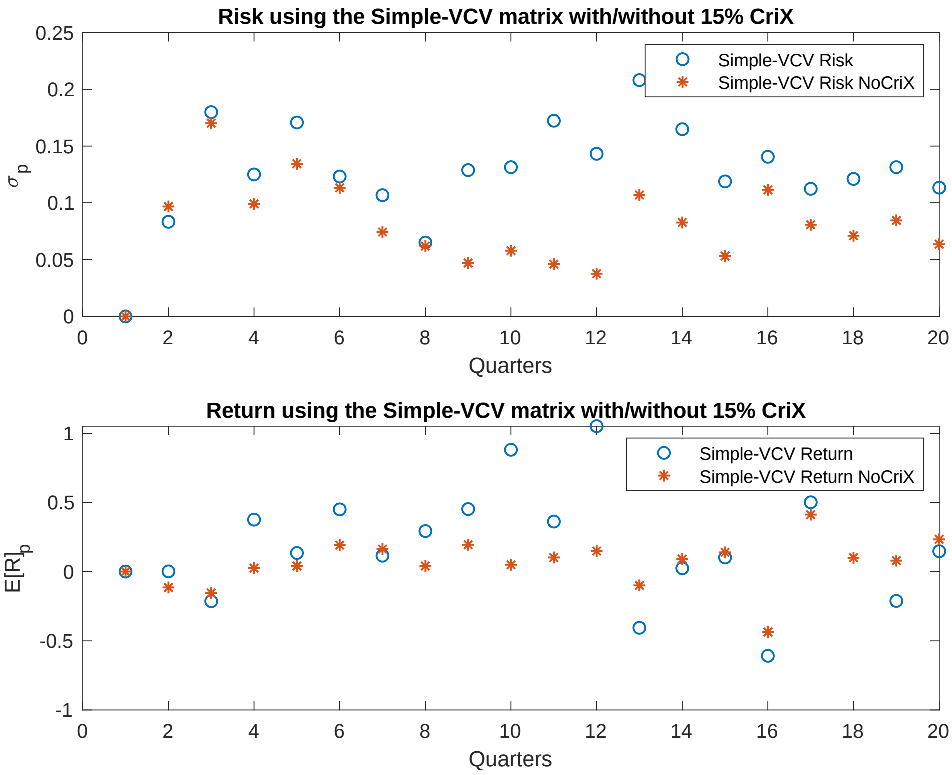
		Simple-VCV Matrix		EWMA-VCV Matrix	
		SR_{crix}	SR_{nocrix}	SR_{crix}	SR_{nocrix}
2015	Q1	0	0	0	0
	Q2	-0,64	-1,39	-0,48	-1,45
	Q3	-1,25	-1,02	-1,80	-1,80
	Q4	2,25	0,05	1,68	-0,04
2016	Q1	0,55	0,15	0,39	0,05
	Q2	3,20	1,51	2,73	1,52
	Q3	1,18	1,92	0,78	1,29
	Q4	2,99	0,34	1,80	0,45
2017	Q1	3,72	3,71	2,75	1,61
	Q2	5,96	0,51	4,79	0,27
	Q3	2,19	1,78	1,83	1,00
	Q4	7,28	3,44	6,32	1,68
2018	Q1	-1,99	-1,11	-3,37	-1,59
	Q2	0,21	0,85	-0,14	0,68
	Q3	0,81	2,24	0,65	1,97
	Q4	-4,88	-4,09	-5,13	-5,41
2019	Q1	4,69	4,85	3,77	4,78
	Q2	5,63	1,13	4,26	0,90
	Q3	-1,39	0,70	-1,59	0,44
	Q4	1,73	3,36	0,97	2,49

Figure 14 – Performance of 100€ invested in portfolio



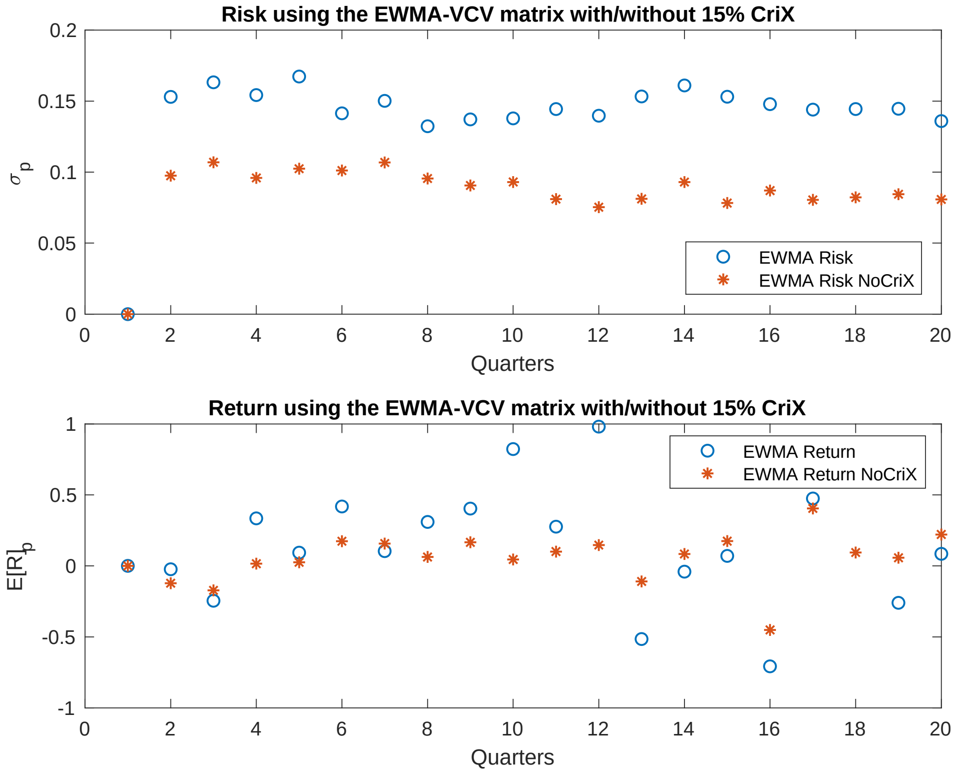
Portfolio Allocation – 15% Exposure

Figure 15 – Simple-VCV Matrix with and without CriX



		Simple-VCV Matrix			
		r_{crix}	σ_{crix}	r_{nocrix}	σ_{nocrix}
2015	Q1	0	0	0	0
	Q2	0.0012	0.0832	-0.1146	0.0967
	Q3	-0.2139	0.1799	-0.1542	0.1701
	Q4	0.3755	0.1250	0.0247	0.0991
2016	Q1	0.1341	0.1707	0.0409	0.1344
	Q2	0.4503	0.1231	0.1914	0.1132
	Q3	0.1149	0.1067	0.1631	0.0743
	Q4	0.2933	0.0650	0.0410	0.0619
2017	Q1	0.4524	0.1287	0.1945	0.0470
	Q2	0.8803	0.1313	0.0498	0.0578
	Q3	0.3614	0.1722	0.1020	0.0459
	Q4	1.0516	0.1431	0.1491	0.0375
2018	Q1	-0.4060	0.2080	-0.0991	0.1069
	Q2	0.0244	0.1648	0.0907	0.0827
	Q3	0.1023	0.1189	0.1391	0.0530
	Q4	-0.6088	0.1405	-0.4366	0.1114
2019	Q1	0.5010	0.1123	0.4118	0.0806
	Q2	0.7594	0.1210	0.1007	0.0710
	Q3	-0.2120	0.1313	0.0795	0.0844
	Q4	0.1479	0.1133	0.2336	0.0635

Figure 16 – EWMA-VCV Matrix with and without CriX



EWMA-VCV Matrix

		r_{crix}	σ_{crix}	r_{nocrix}	σ_{nocrix}
2015	Q1	0	0	0	0
	Q2	-0.0236	0.1530	-0.1217	0.0974
	Q3	-0.2450	0.1632	-0.1722	0.1069
	Q4	0.3348	0.1542	0.0158	0.0959
2016	Q1	0.0933	0.1673	0.0254	0.1024
	Q2	0.4189	0.1414	0.1741	0.1011
	Q3	0.1038	0.1501	0.1576	0.1068
	Q4	0.3102	0.1322	0.0628	0.0955
2017	Q1	0.4040	0.1371	0.1662	0.0905
	Q2	0.8226	0.1378	0.0452	0.0930
	Q3	0.2762	0.1443	0.1011	0.0809
	Q4	0.9812	0.1397	0.1466	0.0753

2018	Q1	-0.5152	0.1533	-0.1092	0.0811
	Q2	-0.0402	0.1610	0.0838	0.0929
	Q3	0.0705	0.1531	0.1745	0.0782
	Q4	-0.7070	0.1479	-0.4507	0.0870
2019	Q1	0.4749	0.1440	0.4050	0.0804
	Q2	0.7093	0.1444	0.0944	0.0822
	Q3	-0.2595	0.1446	0.0574	0.0844
	Q4	0.0853	0.1359	0.2214	0.0807

Figure 17 – Comparison between Simple and EWMA-VCV Matrix

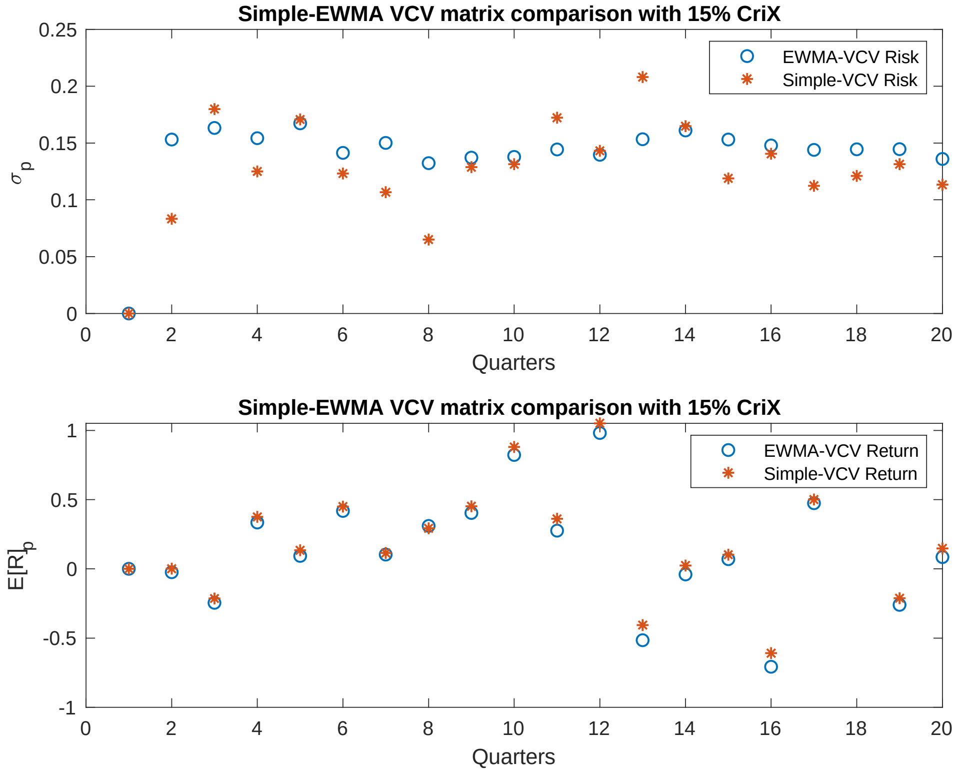


Figure 18 – Comparison between Simple and EWMA-VCV Matrix

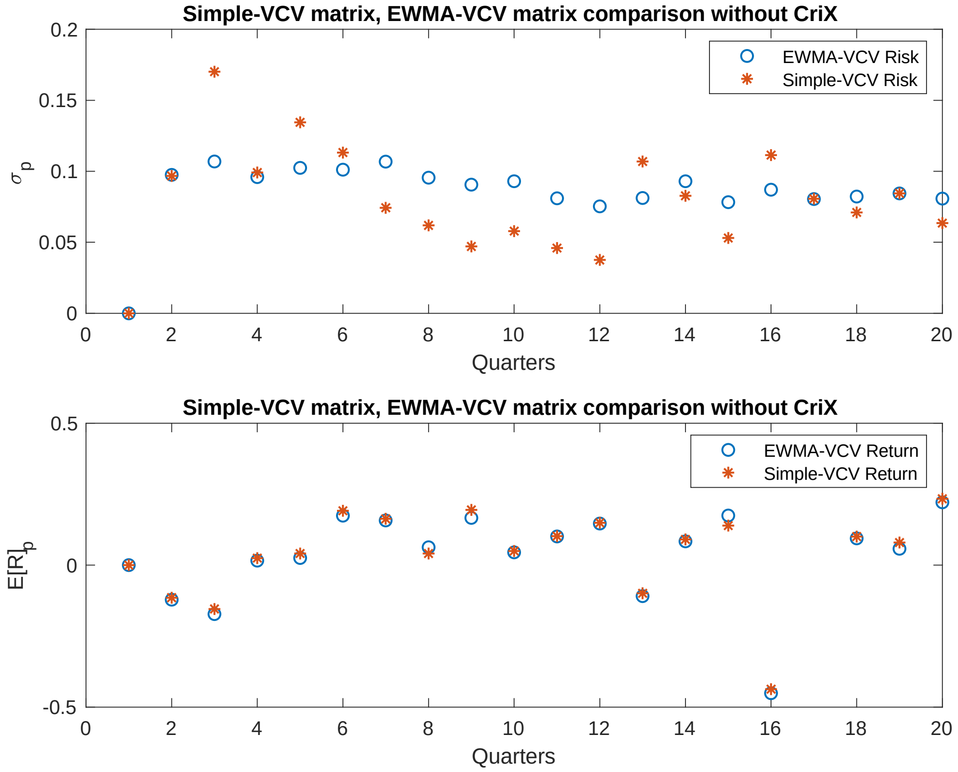


Figure 19 – SR obtained under the Simple and EWMA-VCV Matrix

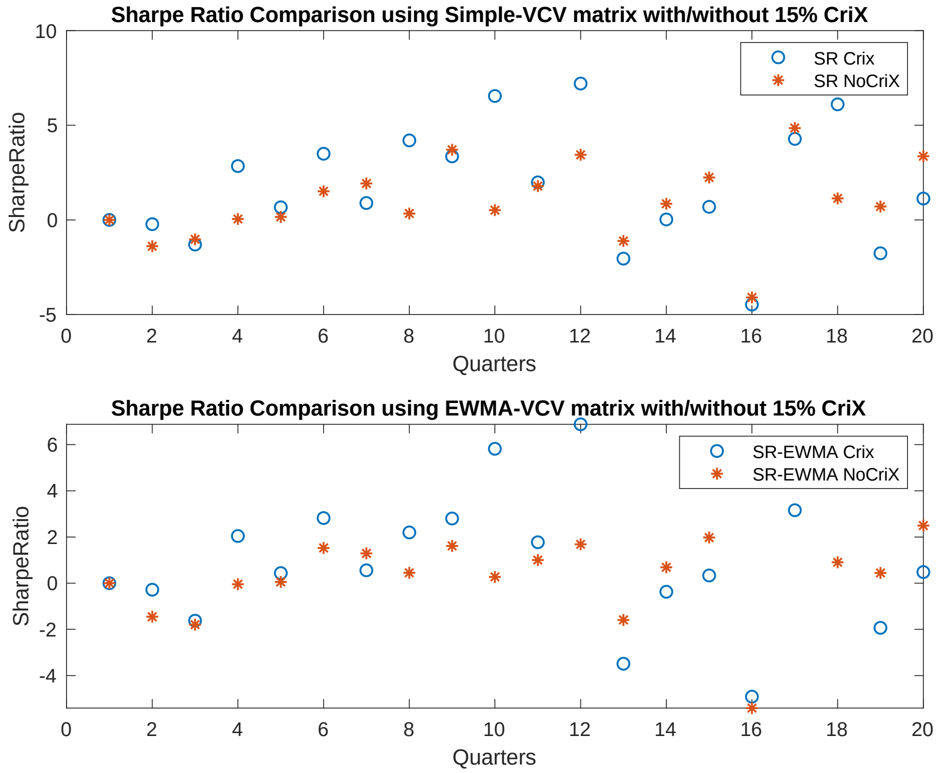
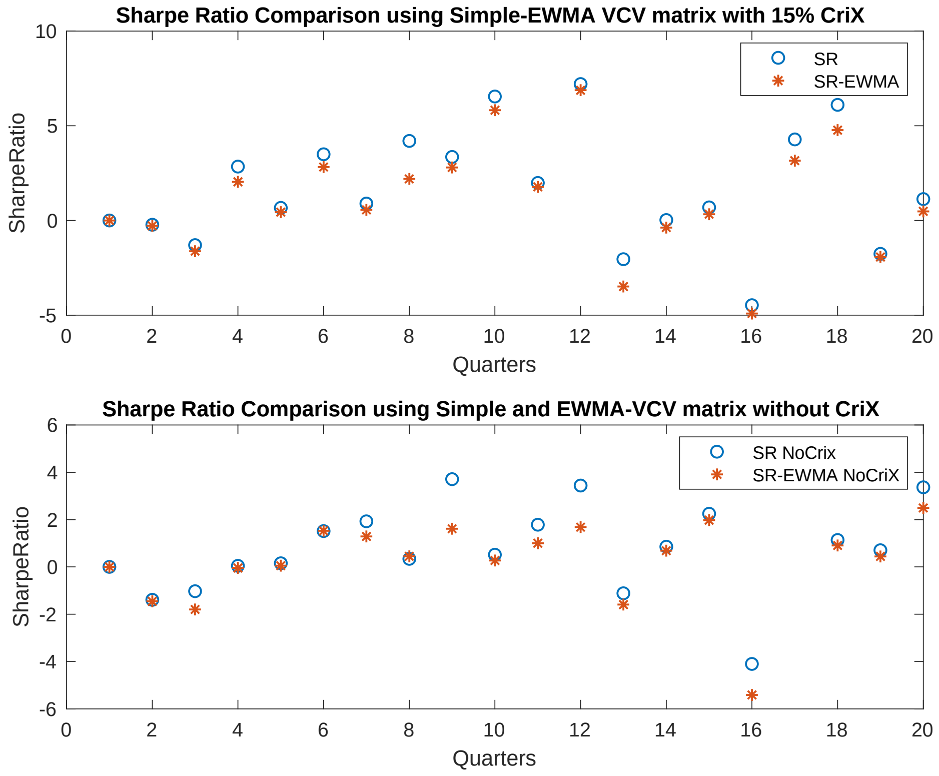


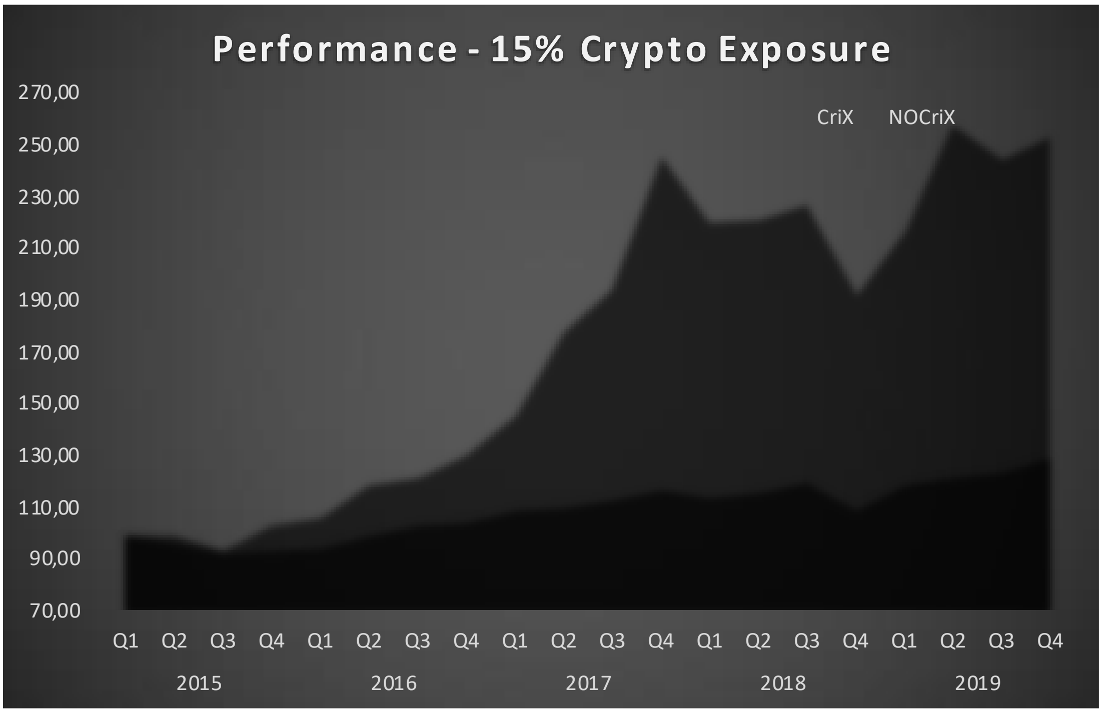
Figure 20 – SR comparison between Simple and EWMA-VCV Matrix



		Simple-VCV Matrix		EWMA-VCV Matrix	
		SR_{crix}	SR_{nocrix}	SR_{crix}	SR_{nocrix}
2015	Q1	0	0	0	0
	Q2	-0.22	-1.39	-0.28	-1.45
	Q3	-1.30	-1.02	-1.62	-1.79
	Q4	2.84	0.05	2.04	-0.04
2016	Q1	0.67	0.15	0.44	0.05
	Q2	3.49	1.51	2.82	1.52
	Q3	0.89	1.92	0.56	1.28
	Q4	4.20	0.33	2.19	0.44
2017	Q1	3.36	3.71	2.80	1.61
	Q2	6.55	0.51	5.82	0.27
	Q3	1.98	1.78	1.77	1.00
	Q4	7.20	3.44	6.88	1.68

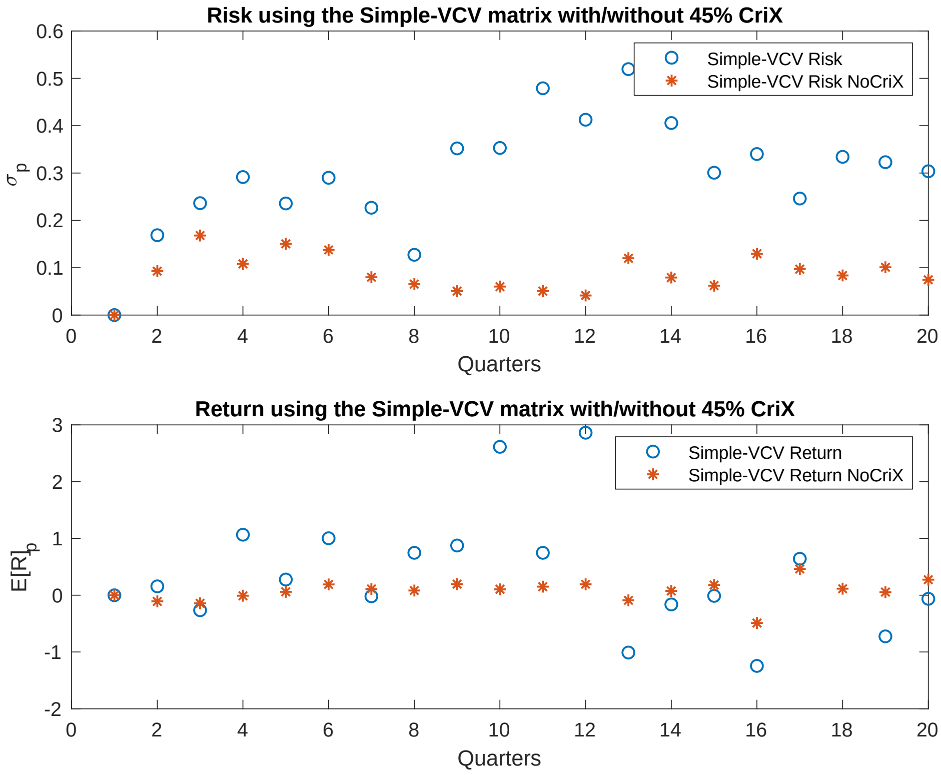
2018	Q1	-2.04	-1.11	-3.49	-1.59
	Q2	0.02	0.85	-0.37	0.68
	Q3	0.69	2.24	0.33	1.97
	Q4	-4.47	-4.10	-4.91	-5.41
2019	Q1	4.28	4.85	3.15	4.78
	Q2	6.11	1.13	4.77	0.90
	Q3	-1.76	0.70	-1.93	0.44
	Q4	1.12	3.36	0.48	2.49

Figure 21 – Performance of 100€ invested in portfolio



Portfolio Allocation – 45% Exposure

Figure 22 – Simple-VCV Matrix with and without CriX

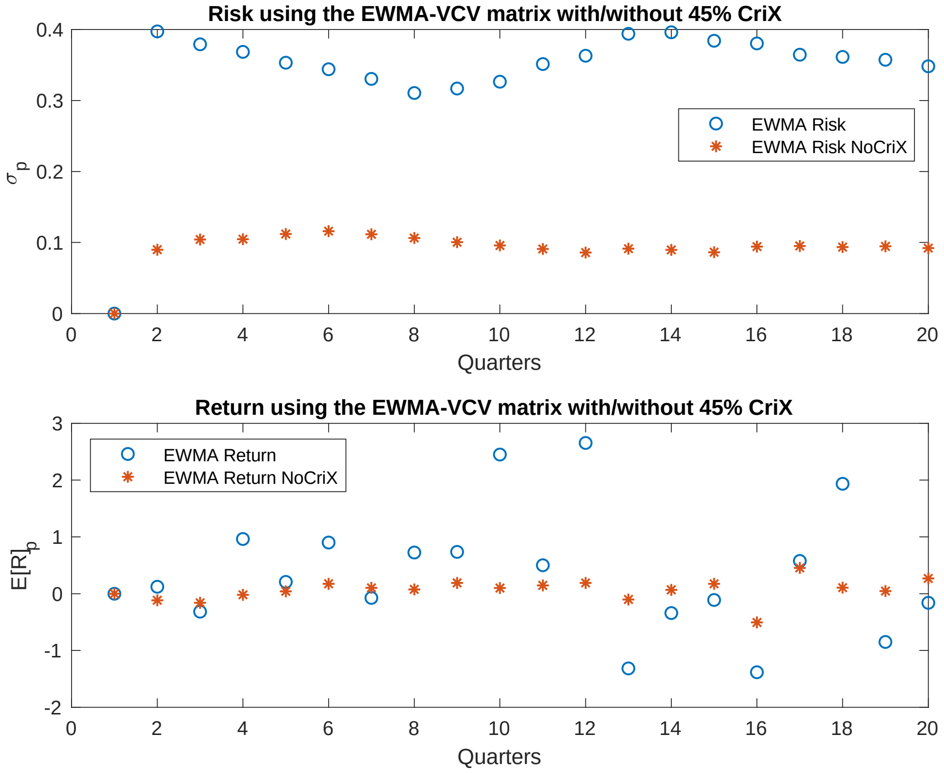


Simple-VCV Matrix

		r_{crix}	σ_{crix}	r_{nocrix}	σ_{nocrix}
2015	Q1	0	0	0	0
	Q2	0.1566	0.1685	-0.1085	0.0926
	Q3	-0.2642	0.2364	-0.1417	0.1679
	Q4	1.0662	0.2915	-0.0090	0.1080
2016	Q1	0.2752	0.2357	0.0601	0.1504
	Q2	1.0031	0.2899	0.1903	0.1376
	Q3	-0.0176	0.2265	0.1098	0.0798
	Q4	0.7474	0.1272	0.0830	0.0653
2017	Q1	0.8746	0.3519	0.1952	0.0504
	Q2	2.6116	0.3531	0.1045	0.0599
	Q3	0.7480	0.4790	0.1501	0.0505
	Q4	2.8608	0.4126	0.1929	0.0413

2018	Q1	-1.0081	0.5196	-0.0898	0.1199
	Q2	-0.1615	0.4055	0.0752	0.0791
	Q3	-0.0111	0.3008	0.1793	0.0621
	Q4	-1.2441	0.3402	-0.4897	0.1295
2019	Q1	0.6413	0.2460	0.4623	0.0970
	Q2	2.0738	0.3341	0.1138	0.0835
	Q3	-0.7246	0.3228	0.0553	0.1009
	Q4	-0.0640	0.3037	0.2739	0.0745

Figure 23 – EWMA-VCV Matrix with and without CriX



		EWMA-VCV Matrix			
		r_{crix}	σ_{crix}	r_{nocrix}	σ_{nocrix}
2015	Q1	0	0	0	0
	Q2	0.1214	0.3973	-0.1152	0.0899
	Q3	-0.3177	0.3792	-0.1604	0.1043
	Q4	0.9628	0.3684	-0.0192	0.1047
2016	Q1	0.2083	0.3531	0.0419	0.1119
	Q2	0.9007	0.3440	0.1755	0.1160
	Q3	-0.0733	0.3305	0.1036	0.1116
	Q4	0.7247	0.3106	0.0763	0.1063
2017	Q1	0.7371	0.3168	0.1914	0.1004
	Q2	2.4500	0.3264	0.1002	0.0958
	Q3	0.5004	0.3514	0.1469	0.0908
	Q4	2.6549	0.3630	0.1900	0.0857
2018	Q1	-1.3146	0.3940	-0.1013	0.0913
	Q2	-0.3398	0.3962	0.0681	0.0896
	Q3	-0.1093	0.3842	0.1741	0.0863
	Q4	-1.3829	0.3803	-0.5059	0.0942
2019	Q1	0.5765	0.3644	0.4539	0.0951
	Q2	1.9365	0.3614	0.1067	0.0935
	Q3	-0.8497	0.3574	0.0468	0.0945
	Q4	-0.1609	0.3482	0.2686	0.0921

Figure 24 – Comparison between Simple and EWMA-VCV Matrix

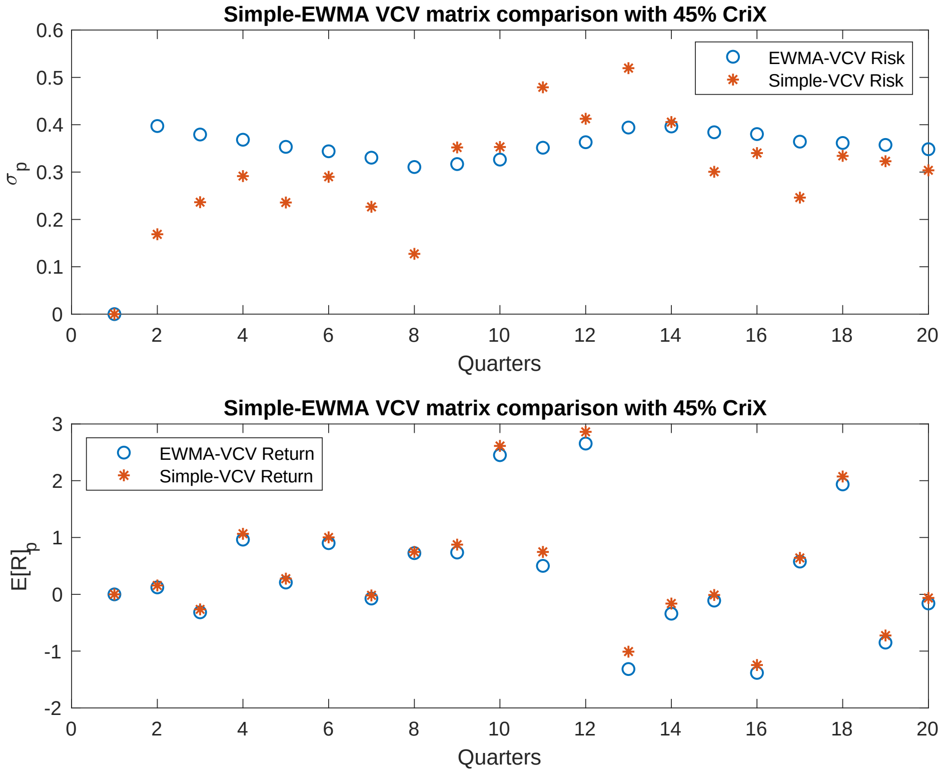


Figure 25 – Comparison between Simple and EWMA-VCV Matrix

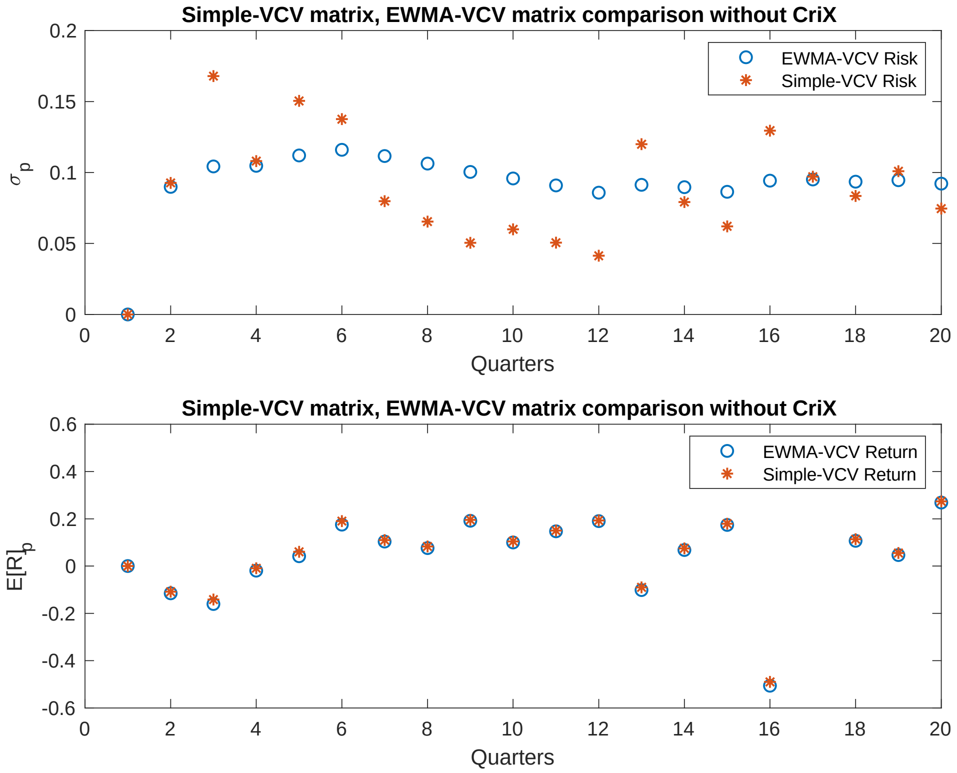


Figure 26 – SR obtained under the Simple and EWMA-VCV Matrix

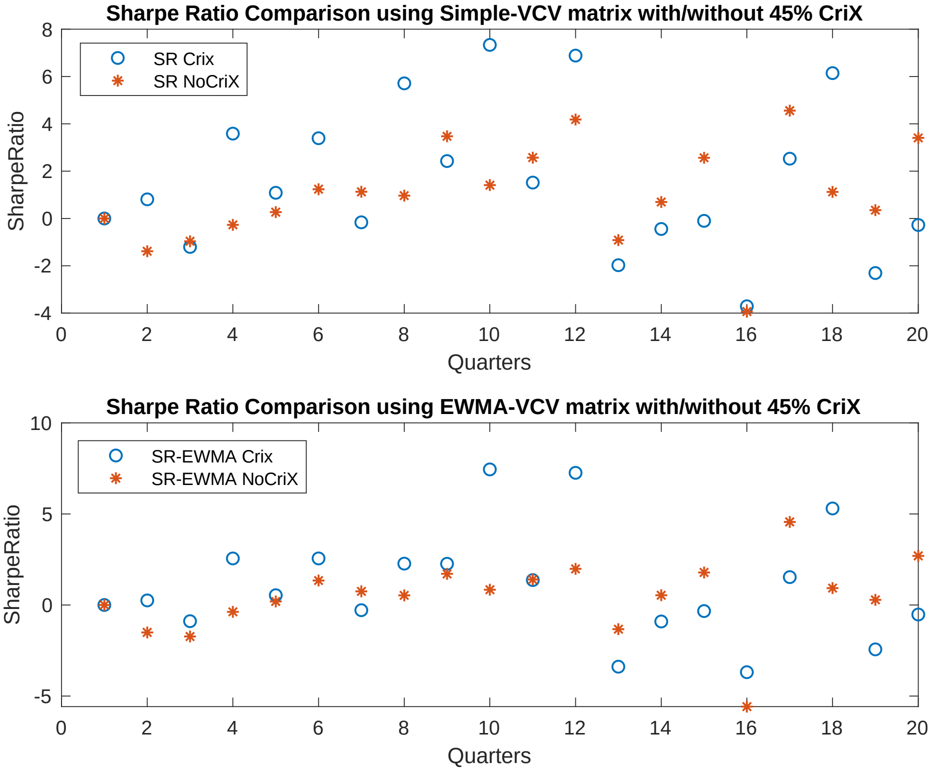
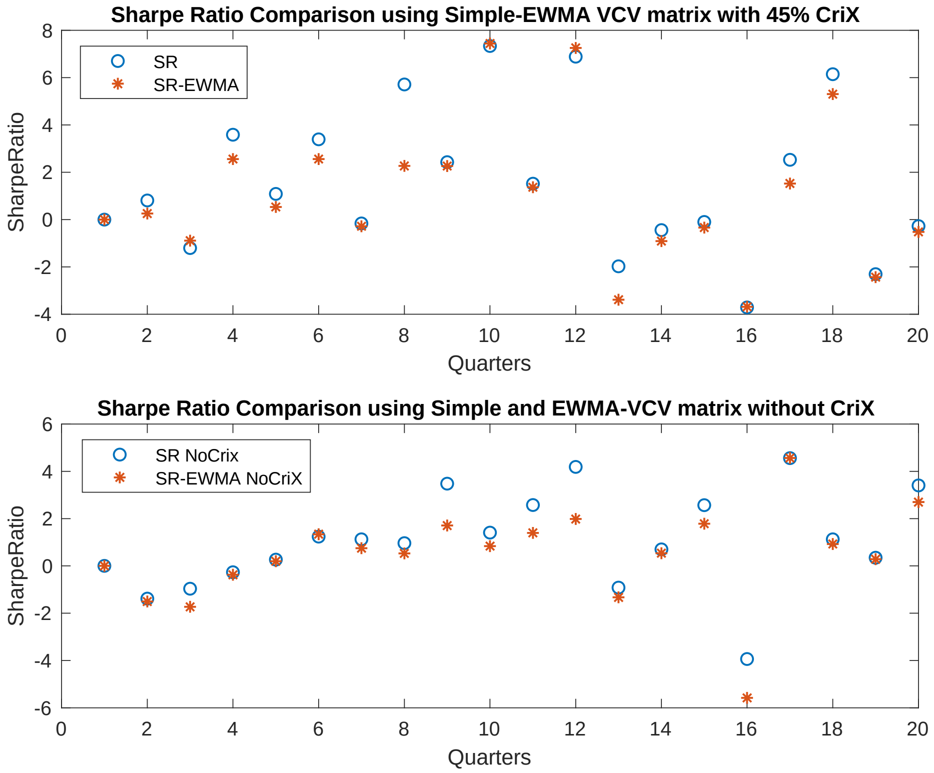
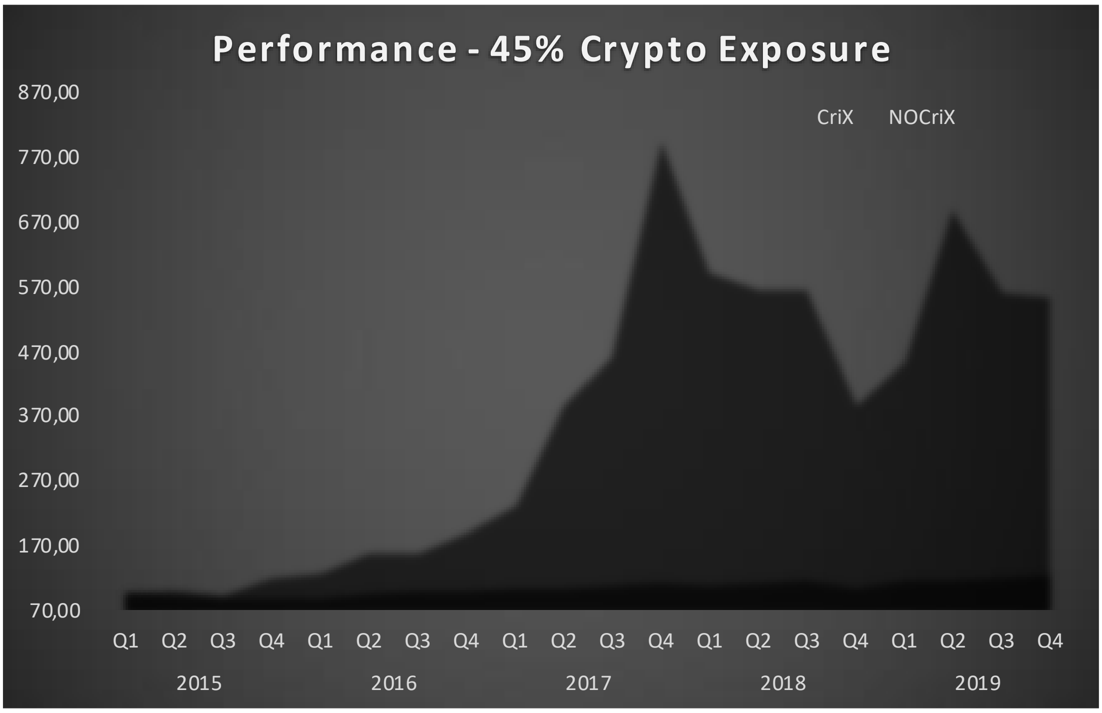


Figure 27 – SR comparison between Simple and EWMA-VCV Matrix



		Simple-VCV Matrix		EWMA-VCV Matrix	
		SR_{crix}	SR_{nocrix}	SR_{crix}	SR_{nocrix}
2015	Q1	0	0	0	0
	Q2	0.81	-1.38	0.25	-1.50
	Q3	-1.20	-0.96	-0.89	-1.73
	Q4	3.59	-0.27	2.56	-0.37
2016	Q1	1.08	0.27	0.53	0.19
	Q2	3.39	1.24	2.56	1.34
	Q3	-0.16	1.12	-0.28	0.75
	Q4	5.71	0.96	2.27	0.53
2017	Q1	2.43	3.47	2.26	1.70
	Q2	7.34	1.41	7.44	0.84
	Q3	1.52	2.57	1.36	1.39
	Q4	6.88	4.18	7.25	1.98
2018	Q1	-1.98	-0.91	-3.38	-1.32
	Q2	-0.45	0.70	-0.91	0.53
	Q3	-0.10	2.56	-0.33	1.78
	Q4	-3.71	-3.93	-3.69	-5.58
2019	Q1	2.52	4.56	1.52	4.56
	Q2	6.14	1.12	5.30	0.93
	Q3	-2.31	0.35	-2.43	0.28
	Q4	-0.27	3.40	-0.52	2.70

Figure 28 – Performance of 100€ invested in portfolio



5. Conclusions

The dissertation highlights the main features of the Blockchain. This study tried to answer the questions: “Is still growing the Blockchain technology?” ,“What happens when implementing cryptos in portfolio asset allocation?”, “Are there benefits from diversification?”, “Can cryptos boost the returns of a portfolio providing upside potential?”.

Even if it is not yet clear how and when this technology will be implemented there are some indicators that, in my opinion, should be analyzed to understand if the Blockchain technology is growing. From the study, we have seen that the Cryptocurrency Index (CriX) have experienced the highest peak around December 2017/January 2018, the market has labelled this event as a “speculative bubble”.

Even if we share the idea of a market bubble, meaning that the cryptos market were “overvalued” at that time, the cryptos still delivered positive returns from the crash onwards. Nevertheless, as I said, some indicators may help us understand what’s going on in the entire Blockchain in the last five years.

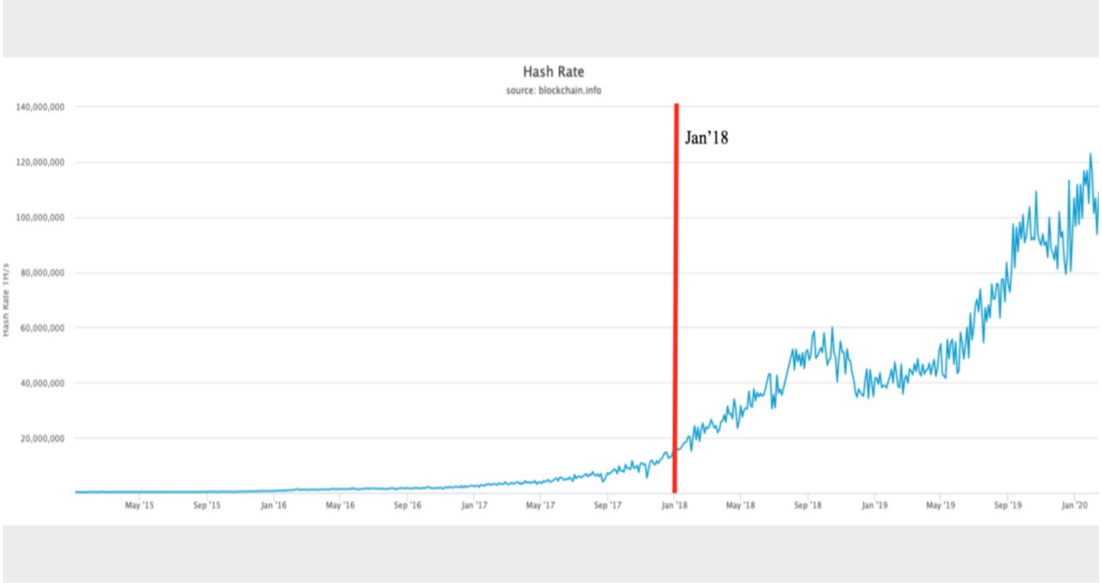


Figure 29 – Hash Rate

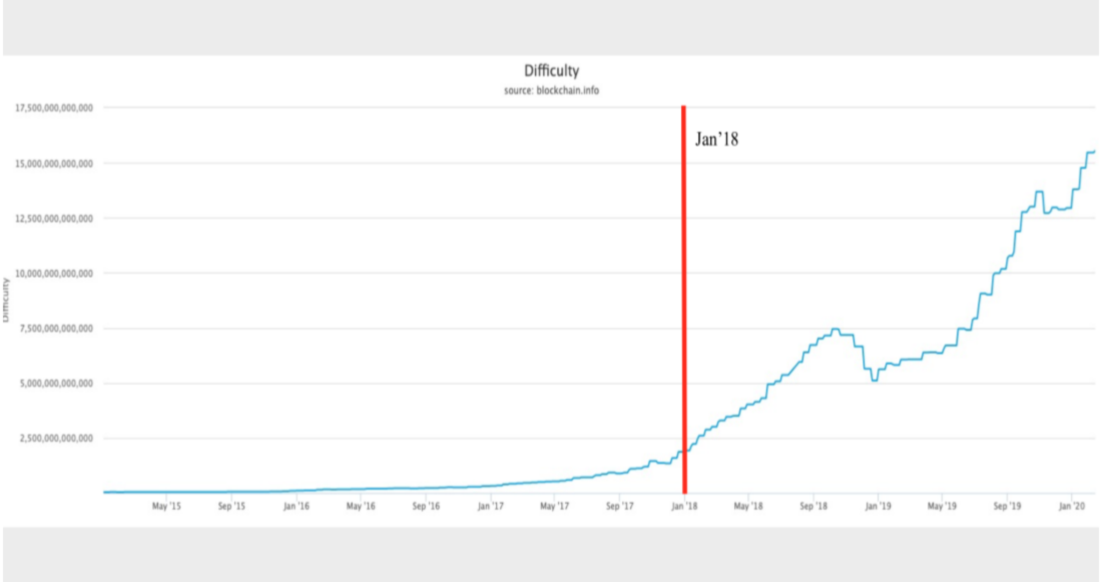


Figure 30 – Difficulty

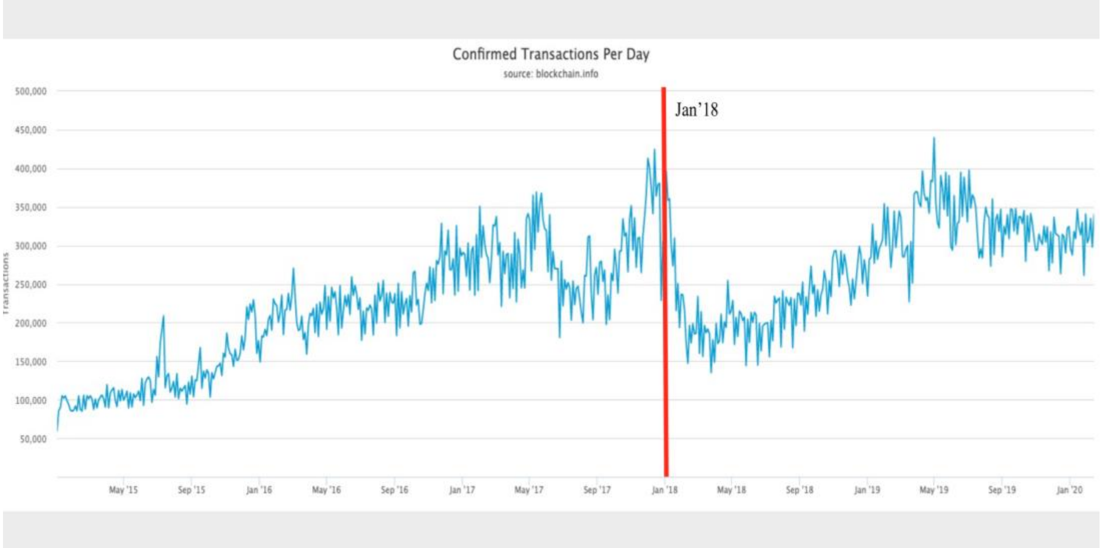


Figure 31 – Transactions Per Day

The mining information can be summarized by Figure 29-30. Even if the CriX drop of more than 50/60% the hash rate and the difficulty still increased exponentially. The network activity can be summarized by Figure 31. After the crash, the Blockchain lost trust from the market, with a decrease in the daily transactions, but recovering from the market crash there are more and more clients that execute transactions daily on the Blockchain. Qualitative speaking, in my opinion, the Blockchain technology is spreading, as also banks are testing its application for future development, has the potential to improve the way we process and execute transactions and may provide solutions to the actual banking system problems.

The actual study is intended to integrate the research conducted by SFOX, a cryptocurrency trading platform, on the correlation between the cryptos and the stock market. The research highlights that cryptos (BTC, ETH, BCH, LTC, BSV) exhibit negative correlation with the S&P500 (ranging from -0.4 to 0), while gold correlation was around -0.18. A case could be made that traders may increasingly view cryptocurrencies, especially bitcoin, as a hedge against global markets, perhaps even more so than gold.

In the light of this, interesting is the way cryptos perform in a portfolio allocation context. In first instance, the analysis show a correlation between cryptos and market indices close to 0, in some cases negative. Clearly, adding CriX produce some benefits. Assets that are not perfectly correlated, in our case uncorrelated or even negative, provide a greater benefit of diversification reducing the overall variance of the portfolio.

An optimal measure used to analyze and compare the investment strategies in CriX is the Sharpe Ratio. We can conclude that allocating 5,10 or 15 percent of our portfolio to cryptos produce a SR which is always greater than the SR of the same investment performed without the CriX. Results are confirmed also under the EWMA method.

6. References

- Gianluca Chiap, Jacopo Ranalli, Raffaele Bianchi. Blockchain, tecnologia e applicazioni per il business. Milano (Italia). Editore Ulrico Hoepli Milano. 2019.

- Roberto Garavaglia. Tutto su blockchain. Capire la tecnologia e le nuove opportunità. Milano (Italia). Editore Ulrico Hoepli Milano. 2018.

- Trimborn, Simon and H$\ddot{a}$rdle, Wolfgang Karl. "CRIX an Index for blockchain based currencies". November 3,2016.

- Horasanh, Mehmet and Fidan, Neslihan. "Portfolio Selection by using time varying covariance matrices". Journal of Economic and Social Research 9(2), 1-22.

Dynamic Portfolio Allocation: CRiptocurrency IndeX
Summary

Chapter 1 - Introduction

As the popularity for cryptocurrencies (cryptos) grows several studies have been published. Cryptos have caught the attention of many investors (retail but also institutional). Is not yet clear if the popularity of cryptos has arisen in relation to tokens itself (ICO) or for the technology behind them. Undoubtfully, the only possibility to use the blockchain technology is through the usage of tokens, cryptos.

Financial analysts and the public have labelled this phenomenon as a speculative bubble, but what will happen if this speculative bubble will turns out to be an innovation in the way we execute transactions each day?

Since its emergence at the start of the decade, blockchain has been heralded as one of the most transformative technologies for financial services. Blockchain hype has led financial institutions to pour money into distributed ledger technology: about $1.7 billion annually as of 2018, per research from Greenwich Associates cited by Bloomberg.

Despite the hype, sentiment around the technology has grown increasingly skeptical as financial institutions struggle to understand the worth of their investments. Incumbents have shuttered some early experiments, and financial institutions executives are beginning to discuss blockchain's prospects in bearish terms. Key difficulties include scaling the technology for commercial application, ongoing regulatory uncertainty, and the difficulty of bringing together competing participants.

Moving forward, it's becoming more clear where exactly blockchain has value, and some players are beginning to make genuine inroads in their adoption and deployment of the technology. Those who are finding success are both pushing back against souring industry sentiment and setting themselves up as industry leaders.

In the banking sector, according to financial institutions, who have explored or are still exploring the blockchain technology, there are several benefits. The banking sector requires high security as it is one of the most attackable fields. In this context, Blockchain can eliminate the threat or the risk of fraud in all areas of banking, and this could equally apply to a trading platform. Furthermore, Blockchain would also address issues such as operational risk and administrative costs as it can be made transparent and immutable. The traceability and the permanent historic record that would exist on Blockchain backing up every asset or item of value that was traded would provide assurance and authenticity all the way through the supply chain. In the end, all the initial skepticisms are slowly disappearing.

The paper is intended to investigate the Blockchain technology and at the same time to analyze an investment strategy dividing the portfolio allocation in real-assets and crypto-assets. The study is conducted with a time-horizon of five years (2015-2020).

The aim of the study is to provide an analysis of cryptos performance and the possible effects of their implementation in portfolio asset allocation. The investment strategy is formulated on the basis of modern portfolio theory then adjustments are made. Implementation of a rolling window and a time-varying VCV matrix are analyzed. Yet, the computation of a VCV matrix following the EWMA method is implemented. To provide more insights about the resulting investment strategy four scenarios are applied, depending on the exposure of the portfolio to the crypto-assets.

Chapter 2 - Blockchain

The blockchain is a digital ledger, decentralized and distributed over a network, structured as a chain of registers ("blocks") responsible for storing data (from value transactions to entire digital applications or smart contracts). It is possible to add new blocks of information, but it is not possible to edit or remove blocks previously added to the chain. In this ecosystem, encryption and consent protocols ensure security and immutability.

Blocks are structures of data added to the blockchain sequentially, one block at a time. Each of them contains a mathematical proof, generated through the use of cryptography, which assumes the sequentiality of the previous block, resulting in a "chain of blocks". The first block is called a "genesis block". The hash of the block is not usually saved in the block but is calculated whenever it is needed. The connection between blocks is generated by means of an encryption function (cryptographic function of hash), which creates an indissoluble mathematical link between them.

The hash function is used to map data of arbitrary size (the input can be almost anything) into data of fixed size (the output, called "hash", is a finite number). The main features of an hash function are:

- The same input always produce the same output (deterministic function),
- Even the slightest change in the input produces a drastic change in the output of the function,
- It is a unidirectional function: is easy to generate an hash from any input but it is very complex to calculate the input from the hash.

The hash is like the fingerprint of a digital file. Blockchain systems use the hash function as it provides a very convenient way to express the entire state of the blockchain in a single string of defined length. For each new block generated, the hash of the previous block is inserted into the input to generate the hash of the new block. In practice, each block contains information, data and hash from the previous block.

One of the aims of the blockchain technology is to allow anyone, anywhere in the world, to carry out transactions without the need to rely on a central institution. To do this the blockchain must be distributed over a network, which is a group of interconnected machines that exchange information. A machine connected to a network is called a node (Full Node or Light Node).

The network is a fundamental component in a blockchain system. Based on the network structure and the role of each node, three network models can be identified: centralized, decentralized (from a point of view of the architecture, authority and logic) and distributed.

The public blockchain model is currently the best known and most used. A public blockchain is a system with: decentralized architecture, decentralized authorities, centralized logic. Decentralization is a key aspect of this model, as any attempt at centralization would introduce a weakness into the system and expose a potential point of failure or control. There is no single authority. Everyone can join the (open) network and there is no possibility of being excluded (resistance to censorship). Each knot has equal rights and responsibilities. Everyone has the possibility to explore and verify each transaction (public and analyzable).

In this hostile environment is important the study of a secure communication technique like encryption. The blockchain is a system in which cryptography (specific public key cryptography) occupies a prominent place. The addresses on the blockchain are generated using this cryptographic system and transactions are authenticated using digital signatures. The basic idea is to use a pair of keys in mathematical relation between them:

- a randomly generated private key that has to be kept secret;
- a public key mathematically derived from the private key, which can be shared with anyone.

Public key encryption can be used to ensure certain properties such as encryption, authentication, integrity and non-repudiation in an unsecured environment such as the Internet. Encryption encrypts data with the main purpose of ensuring confidentiality. If data is encrypted using public key encryption, it can only be decrypted with the associated private key.

Digital signatures are a way of demonstrating someone’s identity without their physical presence, with the difference that mathematics is used instead of a manual signature. Digital signatures are created with a combination of hashing and public key encryption. With a digital signature you can get:

- Authentication: a private key is linked to a specific user. A valid signature unequivocally proves that the message was sent by that user.
- Integrity: if a message is digitally signed, any change to the message after the signature invalidates the signature itself.
- Non-Repudiation: if someone signs a message, they cannot, at a later date, deny that they have signed it.

All this property are valid as long as the private key remains private.

A valid transaction is the elementary unit of information written on the blockchain and implies a change of state in the blockchain. Transactions can be monetary (sending bitcoins) or involve other digital assets (stocks, property certificates, etc.).

A transaction can change the status of the blockchain or be invalid and leave the blockchain in its current state. A transaction is immutable. Just as it is not possible for a valid transaction to be rejected, it is not possible to modify a transaction once it has been accepted. However, it is possible to add one or more conditions to a transaction.

The basic requirement for creating a transaction on a blockchain is to have the object of the transaction. In a fully digital system, this is possible thanks to digital signatures. Moreover, we also remember that in a cryptocurrency transaction there is no physical transfer of money, as these are accounting entries of a digital ledger. A transaction simply records in the ledger the amount transferred from the sender to the recipient.

Once the transaction has been created and signed, is propagated to neighboring nodes, which have the task of verifying its validity and deciding whether to propagate it further or not. The valid transaction is then propagated to the network nodes but is not yet immutably recorded on the distributed ledger (the blockchain).

Before nodes decide whether a transaction is valid or not transactions are grouped in blocks. Blocks are added to the blockchain sequentially. Before being added to a block, a transaction is unconfirmed. Once a transaction is included in a block has 1 confirmation. When the next block is created the same transaction has 2 confirmations, and so on. The number of transaction confirmations corresponds to a number of blocks subsequent to the one in which the transaction is included. Once enough confirmations have been obtained, a transaction cannot be cancelled/modified by anyone.

Usually a transaction includes a commission. Transaction fees correspond to the cost of making a given transaction and are used to reward the miners. The fee is not related to the amount transferred. The commission usually influences the time it takes for a transaction to be confirmed.

On a blockchain there are no user profiles, but rather addresses. Addresses do not contain encryption but are only identifiers that represent the destination of a transaction. It is important to remember that a blockchain is just a list of transactions, there is no concept of currency as a physical object that must be kept somewhere. Coins are only accounting items and the final balance of an address is a calculation made by examining all transactions involving that address.

The purpose of an address is to enable transactions to (and from) a single entity. From a technical point of view, an address is the result of a mathematical operation involving public key cryptography and hashing.

1. First, a private key is generated. It is essential that the private key is generated by a random number, otherwise a critical vulnerability could be created.
2. The private key is derived from the corresponding public key by means of a mathematical process.
3. The public key is passed through a series of cryptographic algorithms (different types of hash functions) to get an address on the blockchain.

Addresses are generally managed using specific tools called wallets. A wallet stores the public and private keys of an address and can be seen as “your account”. The data corresponding to the addresses are always stored on the blockchain. It is therefore not possible to lose the cryptocurrencies, but only to lose the private keys that give access to those cryptocurrencies. There are three main types of portfolios: software, hardware or paper. In addition, depending on the environment in which these wallets operate is possible to make another distinction between cold and hot storage.

Computers and software are far from being perfect systems: they can crash, be hacked, behave negatively on purpose or even behave in a pseudo-random way. When we connect several computers together in a network, the uncertainty of the final system increases exponentially.

In a blockchain there could be millions of nodes that work independently, and it is not possible to predict how each of these nodes will behave. In a permissionless blockchain you can't trust any entity involved. Despite the uncertainty, the knots of a blockchain must come to an agreement on a single state. A blockchain is based on (mathematical) rules but has no rulers. The network has the task of reaching a decision on what happened within the blockchain, through a process called consensus.

Consensus is a general agreement between the members of a given group (the nodes of the blockchain), each of whom has a part of the decision-making power. In a blockchain the consent is an agreement on what happened and holds the only possible truth about the current state of the blockchain.

Consensus, however, should not be understood as a discrete process where there is no consensus at a moment and the moment after consensus is reached, but rather as a continuous process that involves several participants, each with its own roles and responsibilities.

A blockchain uses mathematics, economics and game theory to encourage all actors to reach an agreement on a single state. We can say that the consent of a blockchain is the guarantor of the trust we place in this system.

Mining is a general concept and is not related to any particular blockchain.
It can be seen as a process that allows the blockchain network to validate transactions, group them into blocks and add them to the block chain. These operations make it possible to reach distributed consent and make the network secure.

The nodes that take part in the mining process are called miners. More generally, mining can be seen as the decentralized mechanism through which distributed consensus is reached and network security guaranteed.

We talked about consensus as a continuous process in which miner and full node work to add new blocks to the blockchain and verify the validity of these blocks.

In detail, full nodes and miners are responsible to check that the transactions and new blocks are valid and if so, propagate them to the rest of the network; and to choose transactions, sort them, and aggregate them into a block.

A full-node therefore contributes to the security of the blockchain by checking the validity of each transaction and each block, so as to ensure that the miners do not "cheat".

A full-node is the safest way to use the blockchain. A full-node will never accept a transaction or block that does not comply with the rules. If a miner creates an invalid block, the other nodes will reject it. When a miner's block is added to the blockchain, it is rewarded for the work done according to the rules defined in the blockchain.
Usually the reward consists of the transaction commissions of the block and eventually, as in the case of Bitcoin, of the cryptocurrencies generated at the addition of a new block.

Proof of Work (PoW)

Proof of Work is a protocol used in the process to reach distributed consensus. In concrete terms, the Proof of Work is based on the search for a number that is computationally difficult to find, but once found it becomes easy for all the other nodes to verify its correctness. In a system that uses PoW, a block is only valid if it contains a valid PoW solution.

Proof of Work is a protocol used to reach distributed consensus in which voting power is based on computational power.

PoW Mining

In PoW-mining, network nodes compete to solve a complex mathematical problem (an inverse hash with some constraints). Solving this problem is a random process with very low probability and the only way to find a valid PoW is to try all possible combinations until you find the right one.
The first miner to solve the problem has the right to create the next block and earn the reward. Once a new block is created, it is transmitted to the network, waiting for the other nodes to verify its validity. It is very easy for the remaining nodes to check if the solution is correct. If the block is valid, it is forwarded to nearby nodes, otherwise it is When a miner generates a valid PoW for the new block, it transmits the block to the network.

In the PoW, when we talk about computing power, we refer to hashrate, because usually the problem to be solved is an inverse hash with some constraints. Hashrate is the number of hashes calculated per second (H/s).

The total hashrate, or network hashrate, is the sum of all the miner hashrates. The probability of a miner finding a valid PoW first is as follows:

$$P = \frac{Hashrate\ of\ the\ miner}{Hashrate\ of\ the\ Network}$$

The hashrate depends on the specific hash algorithm used by the blockchain and the power of the machine used by the miner. Considering the Bitcoin, for example, a person has a hashrate of about 0.00003 H/s, which means that calculating a single hash by hand would take about 9- 10 hours. An ASIC (Application Specific Integrated Circuit) miner can calculate more than 14 TH/s (Tera Hash, one trillion hashes per second). In 2019, the hashrate of the Bitcoin blockchain network amounted to more than 100 million TH/s.

A PoW to be considered valid must satisfy a constraint called difficulty. Difficulty is a value that expresses how difficult it is to find a valid PoW. In practice, you can set the target of the difficulty by requiring that the hash to be found starts (for example) with 5 zeros. That is, it means that the first 5 values of the hash will have to be all 0 to satisfy the target of difficulty. In general, the more we increase the number of zeros, the more difficult it becomes.

One of the checks that nodes make when they receive a new block is to check that the difficulty in the PoW of that block respects the constraints on the difficulty. In the case of Bitcoin, a block generated in October 2018 had 19 zeros.

The difficulty is periodically updated (retargeting) in relation to the hashrate of the network to keep the time necessary for the generation of a block as constant as possible. For example, in Bitcoin the difficulty is adjusted every 2,016 blocks (about 14 days) based on the average time it took to find the previous 2,016 blocks.

Rewards are provided to encourage the miners to generate new blocks and keep the network secure. Miners who create a new block are rewarded with all the commissions of the transactions included in the block, plus possibly the new coins (crypto currencies) created together with the block (block-reward). Usually the number of new coins created with each block decreases over time, since most crypto currencies have a limit in the maximum number of existing coins (in Bitcoin this limit corresponds to 21 million bitcoins).

The PoW protocol is fair to miners: a miner who owns 5% of the total computing power of the network on average "wins" the PoW and gets the right to create a new block (and earn the reward) 5% of the time.

The main advantage of the Proof of Work is the strong guarantee of immutability. It is really difficult, if not impossible, to modify a transaction after it has received a sufficient number of confirmations. Remember that the confirmations correspond to the number of blocks added to the blockchain starting from the block in which the transaction is inserted.

Changing a transaction, or the information contained in it, becomes progressively more difficult as new blocks are generated. If a malicious user tries to tamper with a transaction in the X_{t-1} block, the attack may only succeed in one way, i.e. by recalculating the Proof of Work for all of the following blocks ($X_{t-1} - X_t$) before the other miners succeed in undermining the X_t block. The malicious user to do this must therefore be in possession of an incredible computing power, and all just to tamper with a transaction that occurred 8 blocks earlier (about 1 hour and 20 minutes earlier in the case of Bitcoin). Thanks to the hash function, editing a block involves recalculating the entire PoW for all the blocks that follow the tampered block. The further back in time you go, the less likely it is that an attack will be successful.

Part of the community, however, does not think that the PoW is the best method to use in the process to reach consensus and has raised several issues regarding the PoW. The main ones are:

- **Massive energy consumption**. Bitcoin, the largest project using PoW, currently consumes about 0.3% of the world's electricity (over $1 million a day between electricity and mining hardware) and many believe that this situation is not sustainable in the long run. However, the huge energy consumption is the reason why a consensus process based on the Proof of Work is difficult to attack. It is the enormous amount of computing power needed to validate the blockchain that guarantees its immutability. The computing power and the electricity used are the actual proof of the work performed.

- **Hard to climb**. The PoW is one of the bottlenecks in the ability to scale the system. Many argue that slow transactions and high commissions are blocking the large-scale adoption of the blockchain. However, it is possible to make this type of blockchain scalable without modifying the consent algorithm, adopting off-chain solutions (in the case of Bitcoin or similar, we talk about Lightning Network) or changing the size of the block (for example Bitcoin Cash).

- **He's vulnerable to a 51% attack**. If a miner reached 51% of the total computing power of the network, it would (theoretically) be able to create blocks faster than all the remaining miners together. It may therefore happen that the miner in question is able to reverse or modify some of its transactions (double spending) or to block the confirmation of new transactions (censorship of transactions). However, if a miner could successfully execute a 51% attack, he would still not be able to modify the old transactions, since he would have to recalculate the PoW of all subsequent blocks while the other honest miners continue to undermine on the correct blockchain. Such an attack would require the use of an incredible amount of resources for the attacker. If someone actually managed to put together more than 51% of the computing power, it would be much more profitable for him to follow the rules of the blockchain.

- **Geographical discrimination, economies of scale and centralization**. At the moment, most of the miners are concentrated in places where the cost of electricity and temperatures are low (to save on electricity and cooling systems).

Proof of Stake (PoS)
The Proof of Stake is another protocol used in the process to reach distributed consensus. The purpose of the Proof of Stake is the same as that of the PoW, but the process of reaching the final goal is different.
Unlike the Proof of Work, in which miner that solve mathematical problems are rewarded, in the Proof of Stake validators are alternated (validators, can be considered the equivalent of the miners in the PoW) chosen in advance based on the amount of crypto currencies in their possession for the relevant blockchain, also known as stakes.

Proof of Stake is a protocol used to reach distributed consensus in which each token has one vote.

PoS Mining (staking)

In the PoS-mining, instead of the computing power possessed, the tokens possessed are used. Users with tokens can "point" (Staking) their own tokens (technically, pointing means temporarily blocking the tokens until the staking process ends) to have in return the right to confirm the transactions of a block (become a validator) and receive a reward.

The creator of a new block is then chosen in advance using a combination of different parameters, depending on the type of algorithm used. Some parameters may be the number of tokens (stakes), or the time the validator was in possession of those tokens.

Like the PoW, the PoS protocol is also fair for validators: a validator who owns 5% of the total amount of tokens, on average gets the right to create a new block (and earn the reward) 5% of the time.

It can therefore be said that:

$$Voting\ power\ = \frac{Validator's\ Stake}{Total\ Network\ Stake}$$

Compared to PoW, the Proof of Stake is more efficient, as there is no need to perform complex calculations for each new block. PoS supporters say that compared to PoW, PoS has the following advantages:

- **Attacks are more expensive**. The PoS is also theoretically vulnerable to a 51% attack. An attacker, in this case, will not need 51% of the total hashrate but 51% of the total tokens. However, if an attacker tried to buy 51% of the tokens, the market would react with a rare increase in the price of the token. Moreover, people with many tokens have less incentive to attack the blockchain, since an attack would have the counterproductive consequence of destroying the trust in that blockchain, and consequently the value of that token.

- **Cheaper**. As there are no electricity and hardware costs for mining, all people can afford to participate in the network, reducing the current centralization of PoW-based systems.

- **Punishment**. It is possible to create economic disincentives for malevolent actors, for example by destroying their stakes.

- **Loyalty**. Miners are encouraged to stay on the same blockchain. If they wanted to participate in the PoS on another blockchain, they would have to change the tokens in their possession. In the PoW, however, if the currency you are undermining is no longer profitable, you can simply change blockchain.

Chapter 3 - Dataset

Our portfolio will be a selection of the main indices in the economy according to the Global Industry Classification Standards (GICS). I have selected one index for each sector of the economy with the aim to reduce the exposure to the unsystematic risk. The idea, as we will see with Markowitz, is to compose a portfolio fully diversified, in such a way we are rewarded only for the systematic risk that we bear. Furthermore, we introduce a Dollar ETF in addition to the actual basket of indices with the aim of catching additional similarities that exist between the cryptos and the currencies.

Plenty of papers and journals has described the returns from the cryptos as being the highest possible if compared to real world assets. To be fair, we have to say that also the global economies have done a great job in the last five to ten years.

CriX

In the financial industry already exist important benchmarks like the S&P500 and DAX. These indices describe the composition and trend of certain segments of the financial markets. Index providers decide on a fixed number of index constituents which will represent the market segment. It is a huge challenge to set this fixed number and develop the rules to find the constituents, especially since markets change and this has to be taken into account. A method relying on the AIC is proposed to quickly react to market changes giving the possibility to create an index, referred to as CRIX, for the cryptocurrency market.

Nowadays more and more companies have started offering digital payment systems. Smartphones have evolved into a digital wallet. Own currencies for the digital market were therefore just a matter of time. The idea of letting companies offer concurrent currencies seemed for a long time scarcely probable, but the invention of the Blockchain has made it possible.

Cryptocurrencies (abbr. cryptos) have surfaced and opened up an angle towards this new level of economic interaction. Since the appearance of bitcoins, several new cryptos have spread through the Web and offered new ways of proliferation.

Obviously, the crypto market is fanning out and shows clear signs of acceptance and deepening liquidity, so that a closer look at its general moves and dynamics is called for.

By designing CRIX, a market index (benchmark), will enable each interested party to study the performance of the crypto market as a whole or single cryptos. Studying the stochastic dynamics of CRIX will allow to create ETFs or contingent claims.

Chapter 4 – Dynamic Portfolio Allocation: Markowitz Approach

In the early 1960s, the investment community talked about risk, but there was no specific measure for the term. To build a portfolio model, however, investors had to quantify their risk variable. The basic portfolio model was developed by Harry Markowitz (1952, 1959), who derived the expected rate of return for a portfolio of assets and an expected risk measure. Markowitz showed that the variance of the rate of return was a meaningful measure of portfolio risk under a reasonable set of assumptions:

1. Investors consider each investment alternative as being represented by a probability distribution of expected returns over some holding period.
2. Investors maximize one-period expected utility, and their utility curves demonstrate diminishing marginal utility of wealth.
3. Investors estimate the risk of the portfolio on the basis of the variability of expected returns.
4. Investors base decisions solely on expected return and risk, so their utility curves are a function of expected return and the expected variance (or standard deviation) of returns only.
5. For a given risk level, investors prefer higher returns to lower returns. Similarly, for a given level of expected return, investors prefer less risk to more risk.

More important, he derived the formula for computing the variance of a portfolio. This portfolio variance formula not only indicated the importance of diversifying investments to reduce the total risk of a portfolio but also showed how to effectively diversify.

The model is an ex-ante model of portfolio analysis. This means that to implement the Markowitz model the expected return, variance and covariance must be estimated. Typically, the procedure for obtaining these inputs is to calculate the historical values ex-post. Using historical returns, you can easily calculate the other two parameters by assigning an equal weight to each period observed in the market. However, using ex post data to estimate ex-ante parameters of the portfolio can lead to disappointing results.

One of the first problems that can lead to failure of this method is due to the estimate of the risk measure. The Markowitz's portfolio theory uses data with equal weights. Doing this does not take into account the dynamics of the market structure. One of the ways to reduce the estimate of these errors is to use exponential weighted return and variances. Exponentially weighted data assign greater weight to more recent observations, taking into account the dynamic structure of the market. This is also one of the reasons why I decided to implement an EWMA model to calculate the variance-covariance matrix of returns.

The aim is to use the VCV matrix calculated with the EWMA method as input for the Markowitz theory. Subsequently, for illustrative purposes, we are going to see which are the real differences produced by the two methods (unweighted and weighted covariance matrix).

The modern portfolio theory is based on the idea that investors seek high returns from investments trying to minimize their risk. Investors choose how much of their wealth to distribute in each financial instrument, in this way, diversifying their financial exposure. Mean-variance optimization developed by Markowitz (1952) can be used to determine how an investor distributes his wealth between the various financial instruments.
The Markowitz's portfolio theory uses a scheme with equal weights to calculate the parameters listed above. Once the input parameters are obtained, risk and return for each portfolio of financial instruments are calculated.

When the research is done selecting a big quantity of data it can be difficult to compute portfolio expected returns, variances and covariances using algebra. Matrix algebra is a great simplification of calculations.
The portfolio return:

$$r_p = w'E[r] = w'\mu = w_1\mu_1 + w_2\mu_2 + \cdots + w_n\mu_n,$$

where:
n = number of assets,
w = vector of weights (n x 1),
μ = vector of mean returns (n x 1).

The variance of the portfolio:

$$\sigma_p^2 = var(w'R) = w'\Sigma w$$

The goal of portfolio optimization is to find a combination of assets (weights) that minimize the standard deviation of portfolio returns for each level of expected return. In other words, a combination of assets that maximize the expected return of the portfolio for each level of risk.

The optimization problem faces certain constraints, a budget constraint and a short-selling constraint. However, we can summarize the portfolio choice problem:

$$\min_{w} \sigma_p^2 = w'\Sigma w \quad s.t.$$
$$\mu_p = w'\mu = \mu_0$$
$$w'1 = 1$$

$$w_i \geq 0; \quad for\ any\ i = 1,2,\dots,n$$

We can measure variance historically or implicitly (implied volatility). When measuring historically, the easiest method is a simple variance.

Exponentially weighted moving average is one of the extensions how to measure historical volatility. This method put more weights on recent observations, and it let these current observations to make a bigger influence on the forecasted volatility comparing it with older observations. In EWMA model the latest data has the highest weights and weights for previous data decline exponentially over time.

There are two advantages of EWMA model when compared to simple historical models and simple moving average (MA) model which puts the same weights to the all data points.
The first advantage is that in the real-world volatility is affected more by recent events comparing it with some event in the past and EWMA at the same time gives more attention to those recent events. At the same time simply moving average model weights recent event as same as event in the past and this can lead to misleading too low volatility forecast results if, for example, specific shock suddenly drops out of the sample or vice versa if specific shock is in the sample for a long period of time.
The second advantage is that "the effect on volatility of a single given observation declines at an exponential rate as weights attached to recent events fall".

Exponentially weighted moving average model can be computed in several ways. One of them is the following:

$$\boldsymbol{D}_t = \lambda \boldsymbol{D}_{t-1} + (1-\lambda)(\boldsymbol{u}_{t-1}\boldsymbol{u}'_{t-1}),$$

with $0 < \lambda < 1$, where $\boldsymbol{D}_t$ is the variance covariance matrix at time *t* and u_t are the square root of returns at time *t*. For the exercise λ is set equals to 0.86, as suggested from theory.

We will use the EWMA method also in comparison to the "un-weighted" covariance matrix under Markowitz. The aim is to check if weighting more recent observations this will lead to a better optimal portfolio in the mean-variance optimization.

Process
For the purpose of this exercise, all the portfolios have a target return of 0.08 (8%) annually. The choice of the target return is made under the assumption that the majority of the indices employed have securities listed in United States. Thus, I decided to follow the average return of the S&P500, which in the last century delivered approximately between 8-9% return.

To give the exercise a **dynamic approach** I have divided between the expected returns and the variance-covariance matrices estimated and realized. The difference among them lies in the way they have been calculated. The expected returns and variance **estimated** of the portfolios are computed as

$$\mu_{p,t} = \boldsymbol{w}_t \boldsymbol{\mu}_{t,i}$$
$$\sigma_{p,t}^2 = \boldsymbol{w}_t \boldsymbol{\Sigma}_{\text{t}} \boldsymbol{w}'_t$$

or, under the EWMA approach

$$\sigma_{p,t}^2 = \boldsymbol{w}_t \mathbf{D}_{\mathrm{t}} \boldsymbol{w}_t'$$

The expected returns and variance **realized** of the portfolios are computed as

$$\hat{\mu}_{p,t} = \widehat{\boldsymbol{w}}_{t-1} \boldsymbol{\mu}_{t,i}$$
$$\hat{\sigma}_{p,t}^2 = \widehat{\boldsymbol{w}}_{t-1} \boldsymbol{\Sigma}_{\mathrm{t}} \widehat{\boldsymbol{w}}_{t-1}'$$

or under the EWMA approach

$$\hat{\sigma}_{p,t}^2 = \widehat{\boldsymbol{w}}_{t-1} \mathbf{D}_{\mathrm{t}} \widehat{\boldsymbol{w}}_{t-1}'$$

It means that the realized returns and variance-covariance matrices are obtained multiplying the expected returns and the variance-covariance matrices (at time *t*) by weights (of time $t-1$) that are the results of the portfolio optimization problem. In other words, I stored the portfolio weights derived from the portfolio optimization problem (*t*) and applied to the following expected return and variance-covariance matrix (*t+1*). In order to implement this procedure I have used a rolling window of 3 months (Quarterly results). In this way, we do not rely on historical ex-post data to estimate ex-ante parameters, rather with the use of a rolling window we are considering a time-varying VCV matrix.

Results

In this section I classified the results with the aim to provide more insights on the VCV matrix involved in the portfolio minimization problem. I divided between the Simple-VCV Matrix and the EWMA-VCV Matrix. Furthermore, the results are listed on the basis of their exposure towards the cryptos. The analysis has been conducted on 4 different types of allocation. The starting point is at 5% (at the lower bound) of exposure and rise to 10% and 15% which is the upper bound of the minimization problem. The last type of allocation is an extreme case where the total real-assets weight is fixed at 55% keeping the exposure to cryptos at 45%. All the analysis are conducted from 01/2015 to 01/2020, a 5 years' time-horizon.

Chapter 5 – Conclusions

The dissertation highlights the main features of the Blockchain. This study tried to answer the questions: "Is still growing the Blockchain technology?" ,"What happens when implementing cryptos in portfolio asset allocation?", "Are there benefits from diversification?", "Can cryptos boost the returns of a portfolio providing upside potential?".

Even if it is not yet clear how and when this technology will be implemented there are some indicators that, in my opinion, should be analyzed to understand if the Blockchain technology is growing. From the study, we have seen that the Cryptocurrency Index (CriX) have experienced the highest peak around December 2017/January 2018, the market has labelled this event as a "speculative bubble".

Even if we share the idea of a market bubble, meaning that the cryptos market were "overvalued" at that time, the cryptos still delivered positive returns from the crash onwards. Nevertheless, as I said, some indicators may help us understand what's going on in the entire Blockchain in the last five years.

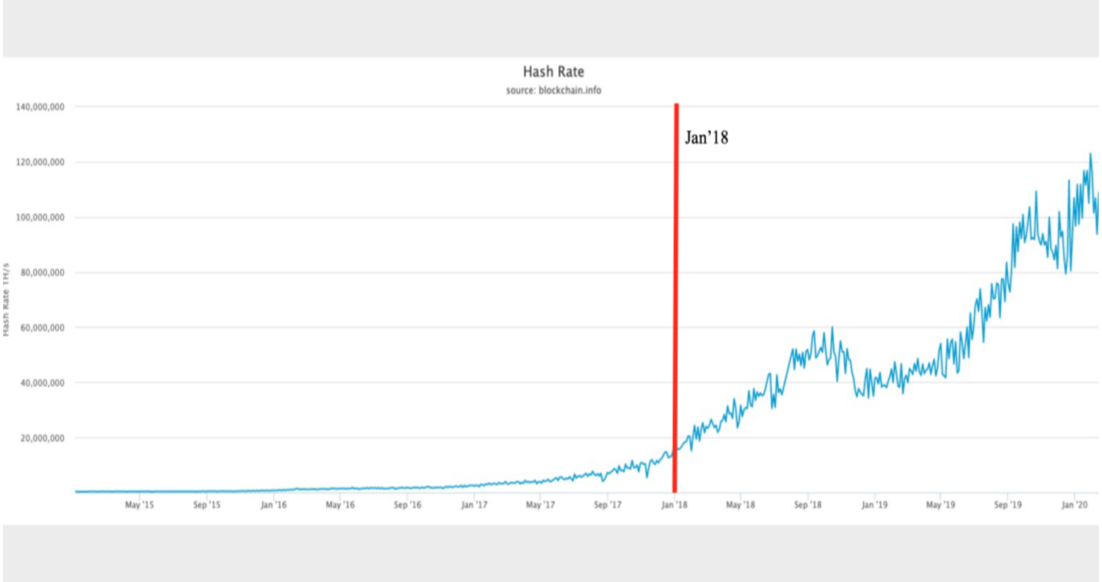


Figure 29 – Hash Rate

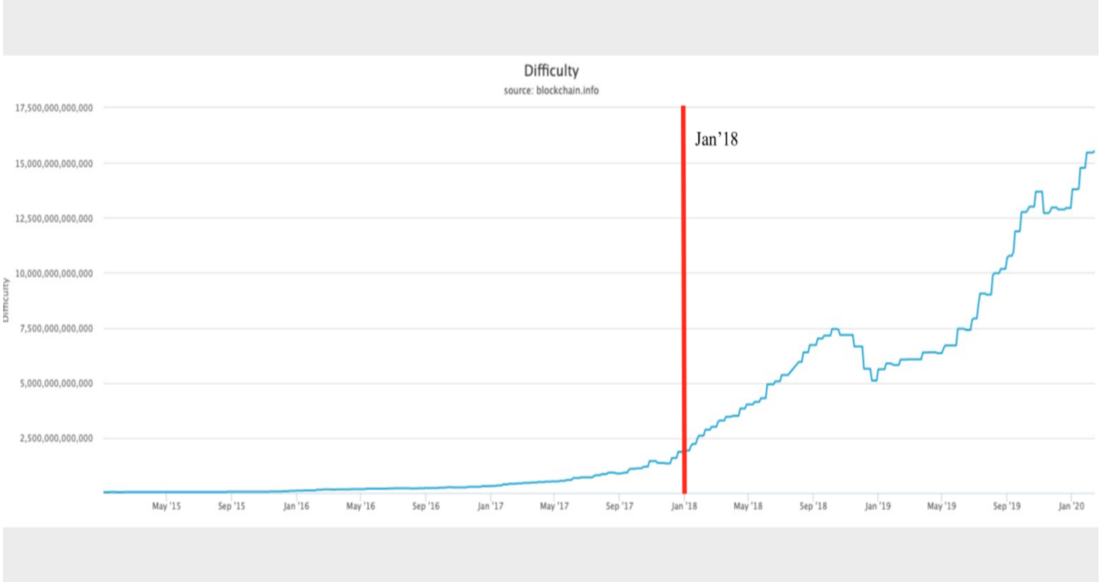


Figure 30 – Difficulty

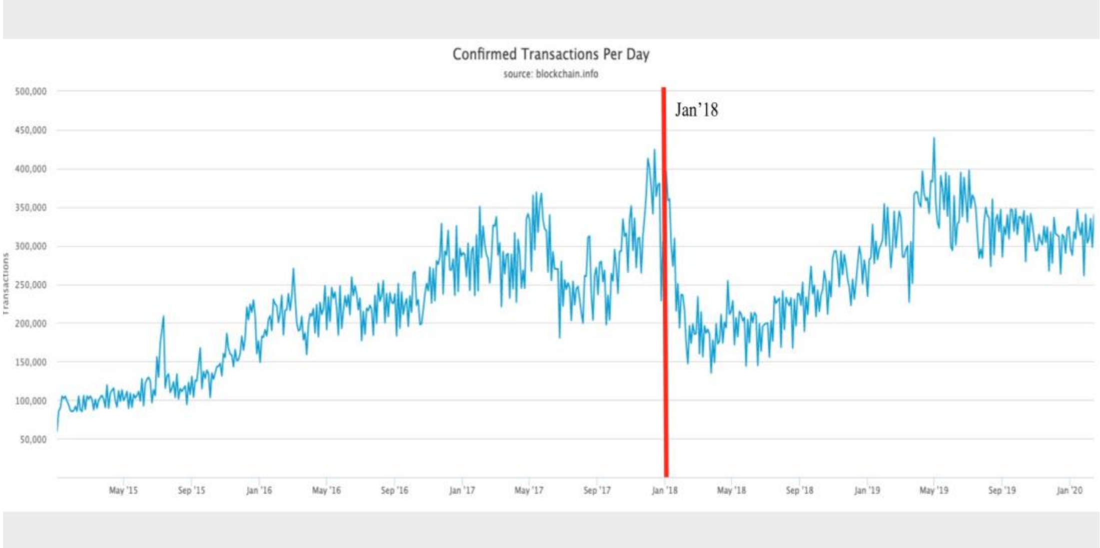


Figure 31 – Transactions Per Day

The mining information can be summarized by Figure 29-30. Even if the CriX drop of more than 50/60% the hash rate and the difficulty still increased exponentially. The network activity can be summarized by Figure 31. After the crash, the Blockchain lost trust from the market, with a decrease in the daily transactions, but recovering from the market crash there are more and more clients that execute transactions daily on the Blockchain. Qualitative speaking, in my opinion, the Blockchain technology is spreading, as also banks are testing its application for future development, has the potential to improve the way we process and execute transactions and may provide solutions to the actual banking system problems.

The actual study is intended to integrate the research conducted by SFOX, a cryptocurrency trading platform, on the correlation between the cryptos and the stock market. The research highlights that cryptos (BTC, ETH, BCH, LTC, BSV) exhibit negative correlation with the S&P500 (ranging from -0.4 to 0), while gold correlation was around -0.18. A case could be made that traders may increasingly view cryptocurrencies, especially bitcoin, as a hedge against global markets, perhaps even more so than gold.

In the light of this, interesting is the way cryptos perform in a portfolio allocation context. In first instance, the analysis show a correlation between cryptos and market indices close to 0, in some cases negative. Clearly, adding CriX produce some benefits. Assets that are not perfectly correlated, in our case uncorrelated or even negative, provide a greater benefit of diversification reducing the overall variance of the portfolio.

An optimal measure used to analyze and compare the investment strategy in CriX is the Sharpe Ratio. We can conclude that allocating 5,10 or 15 percent of our portfolio to cryptos produce a SR which is always greater than the SR of the same investment performed without the CriX. Results are confirmed also under the EWMA method.