# POLITECNICO DI TORINO

Corso di Laurea Magistrale

in Engineering and

Management

## Tesi di Laurea Magistrale

Initial Coin Offerings: an Analysis of ICOs Worldwide

Relatore
prof.ssa Elisa Ughetto

Candidato
Enrico Sellitto

Anno Accademico 2019/2020

# Table of Content

# Introduction

Since the birth in 2009 of the Bitcoin, the world started looking to the phenomenon of cryptocurrencies with an increasing interest over years. New cryptocurrencies started to emerge, new dedicated platforms were developed, new technologies were arising, and the cryptocurrencies market gained value and popularity day by day. Together with the Bitcoin, also a new technology was introduced with its birth, i.e. the Blockchain Technology and, more in general, the Distributed Ledger Technologies. With the Distributed Ledger Technologies, the structure of a network with multiple actors involved passed from a centralized system to a decentralized distributed one, where the nodes were able to access and share the information without the need of a central third-party authority.

It is on this background that Initial Coin Offerings started to emerge as a new crowdfunding method based on the Blockchain technology. Initial Coin Offerings (ICOs) are in fact a new alternative method to raise capital, which is revamping the whole concepts of crowdfunding and seed funding. The method relies its functioning on the issuance of a token to investors which are allocating their money on the activity sponsored by the company launching the ICO. When a company is developing a new Blockchain-based project, it offers at a predetermined price and in a specific period (the sale phase of the ICO) a token which can guarantee to the token owner some services on the company's platform. Tokens, which usually do not provide any percentage of company ownership, can be exchanged among investors according to their current market value. The amount raised by the company will be used for the accomplishment of the sponsored project.

ICOs' phenomenon started in 2013 but its popularity increased over years, leading some investors to register consistent gains for their investment, and reaching its peak of popularity in the second half of 2017. On the other side, some ICOs resulted to be scam or frauds because of the weak regulation system governing the whole process. The consequent introduction of a more structured regulation system reduced the diffusion of this crowdfunding method over potential investors. Despite all of this, there are still many companies that are effectively collecting money through token offering, and one of the most challenging theme regarding this topic is related to the key factors that may lead to the success of an ICO: which are the determinants, in terms of development areas, regulating tools (such as protocols that may protect the investors or the company), financial instrument and data, and many others, that would lead to have a greater amount raised by the company?

The aim of this work is to construct a reliable database with the information about already issued ICOs, collecting data from Internet sources, in order to use these data to develop some descriptive statistics and a regression analysis. The focus will be on understanding and emphasizing which may be the most important Coin Offerings' success factors, according to the results obtained by these analyses.

The work will be structured as following: in the first chapter it is faced a discussion about Cryptocurrencies, Blockchain, Distributed Ledger Technologies and ICOs, deepening the birth, the emerge and the develop of these phenomena and the way their mechanisms work. The second chapter of the work focus its attention on the database's construction: this is one of the core sections of the work as here are collected the data that will be used to carry on all the analysis that will be performed during this thesis. In this chapter will be shown the database methodology, listing the different sources used for the database construction, the variables that have been included in the dataset, giving their definition and explaining their purposes, specifying why they have been included in the database.

The third chapter of the work shifts its focus on carrying on some descriptive statistics for some of the variables included in the constructed database. The descriptive statistics will be carried on through the statistical software STATA. All the steps needed for the development of this phase are highlighted and all the outputs obtained from the software will be discussed and analyzed. From the descriptive statistics section is then approached the final chapter of this work, related to the development of a linear regression analysis. Again, the software used for the regression analysis will be STATA. The focus of this section of the thesis is to understand, through the data collected in the previous phases of the work, which are the main factors that may impact, positively or negatively, on the success of a token offering, and which are those factors which do not lead to any variation on the success of an ICO. The main outcomes and results obtained from the analysis will be discussed at the end of the work, specifying which of the variables included in the regression emphasized a positive (or negative) relationship with the ICOs' success and which ones did not. It will be also provided during the development of the project also some qualitative interpretations of the obtained results.

# Chapter I

## Scope of the chapter

The main purpose of this chapter is to give to the reader a first general overview regarding Initial Coin Offerings (ICOs) and the technologies behind this innovative Blockchain-based crowdfunding method. After a first quick introduction, this part of the work gives a short presentation about Blockchain, starting from its origin - which are strictly related to the birth of the cryptocurrency Bitcoin – and going through its main features, applications and the way the whole procedure works, highlighting the most innovative implications of the technology. The successive step after this part is a more general and widened overview about Distributed Ledger Technologies (DLTs), which emerged as a new paradigm for the sharing and synchronization of data across a network, and of which Blockchain represents just a subset.

Given a description about the technologies upon where they rely, the focus of this work shifts to Initial Coin Offerings, highlighting their functioning, their most important attributes and their market trends over recent years. Since this method heavily relies on Blockchain technology, the way it works is strictly connected to Blockchain and in general to Distributed Ledger Technologies. After this part, ICOs' regulation systems over the world is emphasized, highlighting the first problems emerged because of a lack of protection for investors, the first rules introduced by governments and the differences in the regulating systems adopted by different public authorities having to face this problem. The last focus of this section regards a short focus towards ICOs' geography, analyzing where and in which external conditions this phenomenon occurs more frequently and which are the main drivers for its diffusion among companies or entrepreneurs and investors.

## 1.1. Initial Coin Offerings and Blockchain: introduction

Initial Coin Offerings (ICOs) have emerged in 2013 as a method in order to carry on entrepreneurial finance, venture capital and other presale crowdfunding activities based on the Blockchain technology structure (Howell et al., 2018). They can be considered also as token offerings, a novel mechanism for entrepreneurial finance. Another definition visualizes ICOs as open calls relying on blockchain technology, through which startup, entrepreneurs or other companies perform their purpose of funding their projects.

The birth of this phenomenon is related to the increasing popularity of cryptocurrencies over years, as it was found out and exploited the opportunity to use cryptos for financing businesses. The project funding occurs by the selling of already existing or new tokens – or more in general through the creation of a legal tender[1] - to a crowd of investors. Tokens are protected digital assets based on Distributed Ledger Technologies with a wide range of purposes, which in general guarantee to the owner the right to use the company's products or services (Li and Mann, 2018; Fisch and Momtaz, 2019, Huang et al., 2019). Despite parallels with Initial Public Offerings (IPOs), where the investments lead to the selling of stock shares and their trading is regulated by superior authorities, tokens generally may not grantees company's ownership rights to its owner, and their trade occurs through unregulated exchange platforms. Risks and rewards offered by tokens may differs a lot from those of equity, because of the high volatility characterizing this type of digital asset (Kaal and Dell'Erba, 2019).

When dealing with ICOs, often these tokens are virtual currencies[2]: in most of the cases they are cryptocurrencies, digital currencies relying on Blockchain and Distributed Ledger Technology (Fisch, 2018).

[1] A legal tender is anything recognized by law to settle a public or private debt or meet a financial obligation, including tax payments, contracts, and legal fines or damages.
[2] A virtual currency is a digital representation of value that can be digitally traded and functions as a medium of exchange, unit of account, or store of value. Virtual tokens or coins may represent other rights as well (FATF Report, 2014).

Before deepening the structure of this innovative crowdfunding method, it may be helpful to have a short overview about what is Blockchain, its origin, how it works and especially its main applications.

## 1.2. Blockchain

### 1.2.1. Origin of the Blockchain

The activities related to Blockchain Technology can be broke down into 3 different, chronologically ordered, categories: Blockchain 1.0, related to virtual currencies' applications, Blockchain 2.0, referring to contracts (i.e. the developing of smart contracts[3]) and Blockchain 3.0, related to other Blockchain applications. This third, last category develops itself in areas such as government, literature, science, arts and others (Swan, 2015). The subdivision just described emphasized that at its earlier stages Blockchain was strictly related to cryptocurrencies. In fact, this technology has its origins in the ones of the Bitcoin, and their whole system was known at the beginning as "Blockchain Bitcoin".

In 2008 an unidentified person – or a group of people - who used the fictitious name of Satoshi Nakamoto, published the whitepaper "Bitcoin: A Peer-to-Peer[4] Electronic Cash System" where he - or they - highlighted the idea of a "purely peer-to-peer version of electronic cash [that] would allow online payments to be sent directly from one party to another one without going through a financial institution" (Nakamoto, 2008). Some months later this publication, in the beginning of 2009, the first units of Bitcoin were

---

[3] "Smart contracts, in the context of DLT, are programs that are written on the underlying distributed ledger and are executed automatically by nodes on the network. Any instruction that could be executed by a computer could theoretically be run by a smart contract. Transactions or data recorded on the distributed ledger trigger the smart contract and the actions taken are in turn recorded in the ledger" (Natarajan et al., 2017). Smart contracts are discussed in 1.3.1.

[4] "A peer-to-peer system is a computer network which enables peers to share the network resources, computational power and data storage, without relying on a central authority. Most commonly, peer-to-peer systems form overlay networks deployed in the Internet and are used for file sharing, real-time data streaming and computationally intensive tasks." (Galuba & Girdzijauskas, 2009)

released[5] and so born the first cryptocurrency. Nowadays, the Bitcoin is the world's most famous cryptocurrency, reaching its peak of market value in the end of 2017 (when its evaluation achieved a value of almost USD 20.000)[6].

The technology upon where the crypto relied was named Blockchain because of the structure[7] utilized by Nakamoto in order to organize and store information and transactions[8]. After the develop of the Bitcoin, many other ways to manage and store information through a P2P structure widened, leading to the term "Distributed Ledger Technology" (DLT) to refer to this category of technologies, of which Blockchain represents just a subset (Natarajan et al., 2017).

The problem Satoshi Nakamoto was facing in 2008 mattered with the need of financial institutions as trusted centralized third party during an online transaction, which was leading to avoid the possibility to have non-reversible transactions, to limit the applications of online payments and to increase transaction costs for the parties involved. The necessity of a third party in an online trading was necessary in order to avoid double-spending[9], which would lead to a consistent devaluation of the asset exchanged for the trade. Nakamoto's innovative solution relied over the utilization of cryptography instead of the trusted third party (which usually was a bank), and on a structure whose irreversibility of transactions would easily lead to a better protection from frauds: Nakamoto proposed "a solution to the double-spending problem using a peer-to-peer distributed timestamp[10] server to generate computational proof of the chronological order of transactions" (Nakamoto, 2008).

---

[5] The first Bitcoin transaction occurred on January 12, 2009, when Hal Finney (a computer programmer and developer) downloaded the Bitcoin software the same day it was released and received from Nakamoto 10 Bitcoins (Peterson, 2014). It has been discussed that Finney may have been the person behind Satoshi Nakamoto, but there are no proofs supporting this hypothesis.

[6] Data from https://www.coindesk.com/price/bitcoin

[7] Blockchain is structured as a chain of chronologically ordered blocks, where each block contains information about transactions.

[8] Most commonly a transaction is considered as a data structure that represents the transfer of value between two users in the blockchain network (Bashir, 2018)

[9] The double-spending problem relies on the fact that a node inside a network may spend twice the same currency because of the absence of a digital unique asset representing the currency itself. This is of course inconsistent with development of a digital asset (Chohan, 2017).

[10] A timestamp is a sequence of characters representing an encoded information which identifies when a transaction (or more in general an event) occurred.

The Bitcoin relies so on a decentralized distributed structure where all the transactions are recorded in a public ledger which is available for all the nodes[11] in the network. The Blockchain is Bitcoin's public ledger, where all Bitcoin's transactions are continuously recorded, in a linear and chronological order. Every node owns and can access in any moment a copy of the ledger containing all the transactions' history (Swan, 2015). In such a structure, no nodes can exercise authority over the others and there is no more a single central authority that is enrolled to manage all the information flow of the network.

### 1.2.2. How does Blockchain work

Blockchain is a revolutionary and disruptive technological innovation with potentials in providing fresh capital to innovative ventures. As it was earlier anticipated, it is a decentralized distribution ledger which heavily relies on cryptography.

Nodes inside a Blockchain structure can be miners - that create new blocks - or block signers - which validate transactions – according to the consensus mechanism inside the network. When a new transaction occurs, a node creates it and signs it with its private key[12]. The transaction is then propagated, through a flooding[13] protocol, to nodes that, according to the predetermined mechanism, should validate it. After its validation is completed, the transaction is aggregated with the others in a block of information and so added to the ledger of each of the participants. This block is then added to an already existing and chronologically ordered chain of previously validated blocks. The new block is timestamped through hashes[14] and contains pieces of information of the previously added blocks, so that an attempt of alteration of the existing and newly-created block

[11] A node represents an actor inside a network. In distributed networks, a node may be a client, a server or a peer (a peer may act sometimes as a client and sometimes as a server, according to the situation).
[12] To a user are usually given a public address, where the funds are deposited, and a unique private key, in order to access his address. The private key signature confirms to the other actors in the network that the transaction has been done from a certain user.
[13] Flooding is a computer networking technique in which every incoming packet of information is sent out on every outgoing link except the one from where arrived (Tanenbaum & Wetherall, 2011).
[14] Hashes are the result of a cryptographic algorithm and will be later described.

would require the alteration also of all the previous blocks, which would be something nearly impossible in a decentralized environment as it is Blockchain (Bashir, 2018; Ducas & Wilner, 2017). The information contained in the data block is so shared across the whole network, and the data are structured in an encrypted way so that their details are not publicly available. The ledger is replicated for each of the member of the network and so each participant has a copy of it always accessible and available. (Natarajan et al., 2018).

The technology utilizes the hashing system, which is a cryptographic algorithm that transforms any arbitrary-length string (composed by any character) into a hash, a string composed by numbers and letters and of a prespecified length. The hashing process is irreversible, i.e. it does not allow to understand from the obtained hash the original string that generated it: two very similar strings, which have just few little differences each other, may generate completely different hashes. The hash contributes to provide to the blockchain system security of data and immutability of the shared register. (Adhami et al.2018)

To the hashed value added to the block it might be also added a nonce ('number used once'), which is a pseudo-random number[15] that, in communication protocols, does not allow old communication protocols to be used another time. Nonce are used in order to protect from replay attack, which can be defined as "an attack on a security protocol using replay of messages from a different context into the intended […] [one], thereby fooling the honest participant(s) into thinking they have successfully completed the protocol run" (Malladi et al, 2002.; Bertani, 2014).

The consensus mechanism in Blockchain structure is known as Proof-of-Work (PoW) algorithm. It has been adopted by Bitcoin and Ethereum: this algorithm relies on the fact that each node must find a hash value inferior than a certain number (which represents the difficulty level set by the network). This process is called mining. The first node

[15] A pseudo-random number is a sequence of numbers obtained through an algorithm that can generate it with properties that nearly approximate the ones of sequences of random numbers. (www.khanacademy.org/computing/computer-science/cryptography/crypt/v/random-vs-pseudorandom-number-generators)

finding a winning hash – a hash that comply with the requirements set by the network - can add the proposed block and claim the mining reward[16] (Baliga, 2017). In case of two nodes finding a winning hash at the same time, the one with the higher difficulty score (i.e. the one with the longest chain) will be validated (Natarajan et al., 2017).

As it was earlier anticipated, from Blockchain's birth widened a group of technologies known as Distributed Ledger Technologies (DLTs) of which Blockchain represents just a subset. The focus of this work will now shift from a Blockchain-perspective to a more general one, regarding DLTs.

## 1.3. Distributed Ledger Technology

"Distributed Ledger Technology refers to a novel and fast-evolving approach to recording and sharing data across multiple data stores (or ledgers). This technology allows for transactions and data to be recorded, shared, and synchronized across a distributed network of different network participants." (Natarajan et al., 2018).

Before deepening the structure of a DLT, it is helpful to have an overview about centralized and decentralized systems.

In a centralized system there is a central node, which represents a central authority, and a one-to-many relationship between the central node and the others included in the network. The central single entity manages and controls all the information flow, from validating transactions to allowing other actors to access data. In such a system, if the controlling entity shuts down, all the transactions would be terminated and so they could not be processed (Greeshma & Shoney, 2017).

In a decentralized system there is no dependency from a single superior node, but the control can be shared among many nodes which can exercise authority over the others of the network. This situation is comparable to have many centralized systems inside a

[16] "Each 'miner' that produces a valid proof-of-work in the Bitcoin network receives Bitcoins as a reward […], which serves as an economic incentive to maintain system integrity." (Natarajan et al, 2018)

bigger one, where there is not a master node superior to all the other participants of the network, but many of them with less power. (Bashir, 2017)

The core concept of a Distributed Ledger Technology is that all the participants of the network has access to the same shared ledger at any time. The idea born with the creation of the Bitcoin after Nakamoto's paper. The concept relied on the fact that each of the participants owns a copy of the ledger. The Blockchain is a type of DLT, so not all DLTs are Blockchain. DLT can be thought both in a centralized or decentralized way, even if it was born with an orientation towards a decentralized system. It is possible in fact to make distinction between public or private ledger (in terms of access for the actors) and permissioned or permissionless ledger (in terms of roles of the actors).

It is possible to talk about private ledger when there are several copies of the database among the nodes of the network, but these copies are not available to all of them. Even if the database is distributed among several nodes, there might be a centralization of the authority, with a superior entity enrolled to allow or not the database access to a part of the actors involved. In a private ledger so just a trusted group of nodes has access to the ledger. A private ledger is always permissioned, while a permissioned ledger may be public or private.

If the ledger is accessible to every nodes of the network, it is a public ledger. However, some users may just be able to create data, while others would maintain the database adding blocks or validating transactions: in this case the ledger is public and permissioned. In this type of distributed ledger, the actors allowed to maintain the database are charged by the network's owner.

The last case is the one where every node of the network owns a copy of the ledger (public) and, according to the consensus mechanism, every node can contribute to the maintenance of the data (Natarajan et al., 2018; Ølnes et al., 2017). In this last scenario the ledger is public and permissioned.

### 1.3.1. A DLT Application: Smart Contracts

DLTs have seen, since their origin and over the years, a growing popularity because of their applications in cryptocurrencies, but their utilization's fields are much more widened. The possibility to have smart contracts through DLTs allow these technologies to have applications over several business areas.

Even if smart contracts gained popularity after the development of Blockchain and DLTs, the term was at first used for the first time by Nick Szabo[17] in 1997, when in a paper he highlighted the idea of a smart contract through a vending machine: the vending machine controls the ownership of an asset, and after the input of inserting a coin into the machine, the ownership of the asset is transferred to the buyer (Szabo, 1997; Natarajan et al., 2017).

Smart contracts are systems that automatically moves digital assets basing on pre-specified rules. DLT systems provide a platform that, through code instructions, allow smart contracts to manage real assets and their ownership without the need of a third party to be involved. During a transaction, there is no more need of trust towards the counterparty or towards a third party involved, because of this innovative application of DLT. This is possible for the ability of the nodes to execute the code instructions. The instructions given to the software must be precise and the effectiveness of the instrument strongly relies on the quality of the code: the software should be able to analyze every situation and to act according the input instruction given by the developers (Buterin, 2013; Natarajan et al., 2017; Swan, 2015).

Ethereum represented one of the first and, nowadays, one of the most advanced organizations regarding this type of DLT application.

[17] Nick Szabo is a computer scientist and cryptographer known for researches in digital contracts and currencies. In 1998 he was the first to think about a digital currency, which he named "bit gold" and was considered a precursor of Bitcoin (Peck, 2012). Some authors, such as Dominic Frisby, thinks that Nick Szabo is the person behind Satoshi Nakamoto, even if Frisby itself admits "there are no proofs that Satoshi is Szabo".

Smart contracts might rely on the proof-of-work or on the proof-of-stake consensus mechanism. While the first has already been briefly discussed in 1.2.2., proof of stake was created as an alternative to proof of work, which was the original Blockchain consensus mechanism algorithm. A proof-of-stake protocol system cannot be joined by everyone. The number of tokens owned by a miner represents its "stake": the greater the stake, the higher are the probabilities to be selected as next block validator. In fact, the number of blocks that can be mined by a miner is proportional to its stake. There are also balancing methods in order to avoid the risk of rich validators always being selected and so get richer. PoS requires less energy and computing power than PoW and is less risky in terms of network attacks. (Seang & Torre, 2018).

Despite the great interest beyond this innovative instrument, confidence in Ethereum and in this type of application have been decreased after the DAO attack in 2016 (which is discussed in 1.5.1.) that led to a huge monetary loss and highlighted the security vulnerability of the system. Nowadays however it finds several applications in business areas such as financial services and derivatives, insurance premiums, credit law contracts and, more in general, in all those areas regarding business relationships.

## 1.4. Initial Coin Offerings

### 1.4.1. What is an Initial Coin Offering?

Initial Coin Offerings are emerging as new alternative method to raise capital, which is revamping the whole concepts of crowdfunding and seed funding[18], relying on Distributed Ledger Technology (Fisch, 2018). ICOs are a decentralized method of financing which lean on the issue of tokens to investors in exchange of capital investments for blockchain-based projects published by a startup or by a group of entrepreneurs. Tokens are digital units of value that can be issued as security tokens,

[18] Seed funding, also known as seed money or seed capital, represents an investment in a startup company, in exchange of equity stake, at its very early stages of development.

providing ownership rights with respect to the company issuing them, or guaranteeing utility functions such as access to the company's products or services. Security tokens derives their value from a tradable asset and are so subjected to the Security Law regulations. Cryptocurrencies can be considered as a subcategory of utility tokens and can be accepted by ventures as payment for other tokens they're issuing: the most common cases of cryptos used for this purpose are Bitcoin or Ether[19]. Utility tokens grants access to a service, without providing any holding rights. For utility tokens is easier to be exempt from complying security legislation, since they are not sold as an investment asset (Fisch, 2019; Fisch et al., 2019; Saameh, 2018). A third category visualizes payment tokens (that are sometimes considered as a subcategory of utility tokens) which are a decentralized tool to buy and sells goods and services without any intermediary[20].

ICOs represent an innovative crowdfunding alternative because of the technological structure where rely both its method and the tokens issued through it. Being ICOs a Blockchain-based crowdfunding instrument, they allow the reduction of the cost of capital raised because of the unnecessity of financial intermediaries such as agents or banks. Furthermore, ICOs' structure leads to better support decentralized businesses and open-source projects, enforcing the generation of a built-in customer base and constructing positive network effects towards the ICOs' environment (Adhami et al.,2018).

The token-mechanism upon where the whole system relies allows the creation and the development of a token secondary market. In most of the cases investors buy tokens so that they can exchange them among other investors on trade markets: as a matter of facts, many token owners expect to resell them in order to realize a financial gain (Li & Mann, 2018). In fact, a company may create a digital coin publicly offered in exchange

[19] Ether ETH is the cryptocurrency associated with the open-source platform Ethereum. Ether can both be traded as other cryptocurrencies or can be used inside the Ethereum environment in order to access its services and run applications.
[20] From https://www.planetcompliance.com/2019/09/04/what-is-the-difference-between-utility-security-and-payment-tokens/

of the investment, then these coins can be traded or converted into other currencies by the investor, rather than utilizing the services related to them.

Usually, related to an initial offering, there's also a technical white paper[21], which is a document providing to the public the first detailed information about the project for which the company is raising funds. Generally, a white paper contains information about IT protocols, public ledger adopted, the specific token issued, pricing methods and distribution mechanisms, plus others information related to the published scheme. As the white paper contains the relevant information for the investors, it can be compared to traditional projects' business plans. Effective signals of venture technological capabilities are good drivers for ICOs' funding. To publicly disclose the source code and to publish highly detailed technical information in the white paper became 'de facto' standards for organizations that wants to effectively raise capital through token offerings. Technical white papers and high quality source code lead to higher funds raised by the company, while there seems to be no correlation between the patents owned by the company and the amount of funds raised for the project (Ante et al., 2018; Fisch, 2019; Fisch et al., 2019; Huang et al., 2018).

Because of its Blockchain-based structure, ICOs represent an interesting crowdfunding method just for a restricted segment of high-tech organizations that intend to develop their projects through blockchain technology, which has to represent a crucial component for their structure. On the other side, investors that want to allocate their money through this mechanism should be aware of the elevated risks related to this type of investment, but especially should be aware that the sector is characterized by a highly technical environment: in order to adequately understand the technical details and applications of the proposed projects, they should rely on a consistent personal technological knowledge.

[21] Also known as "token sale term" (Adhami et al., 2018)

### 1.4.2. ICOs' market and recent trends

The first ICO has been issued in 2013 by Mastercoin [22], which was launching a homonymous digital currency based on Blockchain Technology. After Mastercoin's project, several ICOs have been released over years, and the phenomenon reached its peak of popularity in the last months of 2017 when, according to CoinSchedule.com[23], more than USD 5 bn were raised by the projects issued in that period. To make a comparison, Kickstarter [24], one of the most used and well-known crowdfunding platforms in the market, raised USD 4.74[25] bn since its creation in 2009 (Fisch, 2019).

The positive trend continued in 2018, when ICOs raised more than USD 21 bn, but of which USD 17.2 bn were collected on the first half of the year and almost USD 4 bn on the second half, emphasizing a big reduction in ICOs' funds raised over the year. The amount raised in the first half of 2019 almost reached USD 1 bn, confirming the negative trend observed during the previous year.

According to Bitcoinist, regulations were one of the reasons of the decrease of ICOs success and popularity, which started, after its peak in the end of 2017, after March 2018 and then fall over time. "One of the reasons that ICOs have gained popularity and been such fast and effective ways of raising funds is a perceived lack of regulation. Although, as they grew in popularity, there were some regulations placed on ICOs […]. Rather than having to jump through hoops, fill out paperwork, and see to it that all formalities are followed, and in some cases even ask permission, both those seeking funding and buyers alike have been able to get down to business and get deals done more quickly when working through a token sale"[26] (Bitcoinist).

[22] Nowadays known as Omni, Mastercoin was a platform intended for the creation and negotiation of digital personalized assets.
[23] https://www.coinschedule.com/stats
[19] Kickstarter is an American organization and is one of the most important crowdfunding platforms in the world. The platform is intended for projects based on innovation and creativity. Its utilization is widened over several part of the world.
[25] Information available at: https://www.kickstarter.com/help/stats
[26] https://bitcoinist.com/regulations-killed-icos-stos/

In their early stages in fact ICOs were nearly unregulated. The few regulations barriers provided entrepreneurs to access less costly external finance with respect to other financing approach. On the other side, the lack of existing securities regulations led a higher number of ICOs' projects to be fraud or scams: there were several cases of projects where, after having raised a big amount of money, the company and its team disappeared leaving the investors without any prize for their apparent capital contribution[27].

As public authorities started introducing regulations about the standards that organizations should maintain, and as people's awareness about the high-risk component of the projects increased, the deflection reflected by ICOs' market trend was an inevitable consequence. On one side the organizations had to face increased costs because of a stricter regulation system, having to comply pre-existing[28] or new ICO-specific regulations.

On the other side, ICOs performances in terms of quality of the published projects increased after the introduction of stricter regulations. The overall rating[29] of ICOs passed from 2.9 in the beginning of 2018 to 3.3[30] in the beginning of 2019. The most recent projects published should register a much lower failure rate with respect to the ones of previous years, as an effect of the stricter regulations that organizations must comply for the crowdfunding process through coin offerings. In such a way investors' trust towards this innovative seed funding solution should increase and ICOs' projects would gain a higher attractiveness from the investors' point of view. On the long term in fact, ICOs' more detailed and complete regulations' systems lead to a stronger diffusion of coin offerings. As it is emphasized by Huang, Meoli and Vismara, "ICOs occur

[27] Some examples of ICOs revealed to be fraud or scams can be found at https://www.finance-monthly.com/2018/10/the-10-biggest-ico-scams-swindled-687-4-million/

[28] ICOs, rather than having a regulation system specific and intended for this crowdfunding method, are often decided by some public authorities to be subjected to the Security Law regulation system, according to the type of tokens issued.

[29] Rate given on a scale from 1 to 5, the value 3.3 refers about projects published in the beginning of 2018 and 2019 (January and February). Ratings are given by Icobench.com and are based on a combination of an assessment algorithm that considers more than 20 criteria and the ratings given by independent experts (https://icobench.com/ratings)

[30] Data are taken from https://en.cryptonomist.ch/2019/03/09/ico-trends-2019/

more frequently in countries with more developed digital regulation environment" (Huang et al., 2019).

Because of the importance and the influence - for ICOs' development - of the external conditions represented by the regulations adopted by countries, the focus of this work will now shift towards this important theme.

## 1.5. ICOs' Regulations

As it has been earlier described, ICOs' absence of regulation - at least in the first period - has been highly debated over years. Most of world's financial markets are tightly regulated, leading frauds in these sectors to be very rare. The consequence is that scammers are turning into unregulated markets such as the one of coin offerings. Many ICOs' investors tend to be afraid of frauds and they share the opinion that its regulation should be stricter: in fact, unlike traditional capital market, ICOs' market is young and loosely regulated. Fraudulent ICOs[31] are a negative phenomenon which is discouraging investors, and that policy makers would like to better manage in order to guarantee a higher security of the investments. The diffusion of ICOs' fraudulent projects relied on the potentially high profits[32] that ICOs could lead to, on the massive hype around this seed funding method and on the time-constrained nature of ICOs, which makes investors committing their money quickly to the projects in order not to lose a potentially good deal.

Over the last years several warnings have been released about ICOs being a high-risk investment by public administrations and control authorities: in fact, in addition to the risk of fraudulent ICOs, very often the proposed projects are in their early development

[31] A fraudulent ICO, or exit scam, is when a venture team, after having raised funds for the project, steals the amount raised and disappears, making the investors losing the amount invested.
[32] Some ICOs, such as NXT or Ethereum, registered in 2017 absurdly high Return on Investments (ROI). On the other side, less than half of the total number of ICOs issued in that year were successful. (https://cointelegraph.com/ico-101/top-10-icos-with-the-biggest-roi)

stage, which represents a high risk of failure factor. To protect investors and enforce the position against issuer of fraudulent ICOs, some authorities imposed to ventures the publication of certain information about the technology used and of performance indicators, which could easily highlight which are the higher quality projects among which the investor can choose to allocate its money. (Fisch et al., 2019; Huang et al., 2018, Kean, 2018).

The high investment risks and the potentials of failures and frauds for several ICOs projects led some jurisdictions to take actions in order to contrast this phenomenon. Different governments faced the problem in very different ways each other, leading to an inhomogeneous ICOs' regulation system over the world. (Huang et al., 2018). A homogeneous and more complete regulation system would contrast moral hazard and free-riding behaviors that affects the ICOs' market, leading to a higher quality of the blockchain-based projects and improving the evolution of the market itself (Adhami et al. 2018). Furthermore, stronger regulations can enforce contractual certainty and encourage the development of financial technology firms, leading to a stronger investors' protections. (Huang et al., 2018).

### 1.5.1. The Dao Event

One very successful ICO launched was organized by the DAO (Decentralized Autonomous Organization). DAO was a decentralized venture capital vehicle created on Blockchain Ethereum by Christoph Jentzsch, where funds were allocated according to voting. In exchange of ETH, which were accepted by the organization in order to carry on their projects, the DAO created DAO tokens, issued in a proportional way with respect to the amount contributed. The tokens gave to the holder voting and ownership rights. DAO tokens owners were also allowed to sell them on the secondary market. The Blockchain project promoted by DAO raised in 2016 USD 150 million, which were intended for the development of a business working in the smart-contract environment. In the same year an apparently small bug on DAO's code was exploited and the system has been hacked by a group of thieves, carrying out more than ETH 3.6 million, which led to a financial loss of around USD 60 million, according to the exchange trading rate earlier the hacking episode (Adhami et al., 2018). After this event, on July 25, 2017, "the SEC issued a Report of Investigation under Section 21(a) of the Securities Exchange Act of 1934 describing a SEC investigation of The DAO, a virtual organization, and its use of distributed ledger or blockchain technology to facilitate the offer and sale of DAO Tokens to raise capital. The Commission applied existing U.S. federal securities laws to this new paradigm, determining that DAO Tokens were securities. The Commission stressed that those who offer and sell securities in the U.S. are required to comply with federal securities laws, regardless of whether those securities are purchased with virtual currencies or distributed with blockchain technology."[33] (SEC, 2017).

The SEC document considered DAO tokens as financial instruments and so applicable to the Security Law. More in general, the document highlighted that ICOs' promoters should be able to demonstrate that the token issued is not a financial instrument, in order not to be applied to stricter regulations. This was one the first decisions taken regarding the regulation of the ICOs and represented an important step in order to start facing this problem.

### 1.5.2. ICOs' Regulations over the world

After the "publication of SEC's document", many other countries started to introduce regulations regarding the new crowdfunding method.

There are three most common ways in order to regulate the token issuances adopted by governments, with further differences among countries:

- The most drastic decisions have been taken by China, South Korea and Colombia, where ICOs are forbidden by the governments. In those countries people can invest on them but may not issue a new one, except in few special cases;

[33] From https://www.sec.gov/oiea/investor-alerts-and-bulletins/ib_coinofferings

- Some jurisdictions decided to rely on an already existing framework for token issuance, such as the U.S. - which use the financial instruments regulation framework - or Japan and Singapore using the payment services framework. In the U.S. the SEC introduced a set of rules (known as the Howey test[34]) in order to establish if an Initial Coin Offering is an "investment contract", and so subjected to the Securities Law, or if not. However, among the different states of the U.S., the more specific decisions have been faced in different ways by the authorities;
- The last strategy is contemplating the creation of a new regulatory framework for ICOs and in general for token issuance. An example of this may be France, where the regulation system is much more advanced with respect to the rest of Europe because of the introduction of ICO Visa[35], through which French companies wanting to launch an ICO can send all the information about it to the Autorite Des Marchés Financiers (AMF) that will authorize and validate - or not - the issuance of the related tokens. (De Vauplane, 2018). In this way the investors' trust is increased, and it decreases the risk of investing in an ICO which is a fraudulent one or a scam.

In other countries, such as India, Thailand, Hong Kong and Philippines ICOs are allowed but not regulated by authorities and so companies can find in those environments favorable situations for issuing their token offerings.

In Europe ICOs are allowed, with different approaches in regulation decisions among countries. On November 2017 the European Security and Markets Authority (ESMA) highlighted through two warning documents about ICOs being a high-risk investment. Several European countries adopted ICO-friendly regulations, as it is emphasized by De Vauplane (2018):

[34] "Under the Howey Test, a transaction is an investment contract if: (i) it is an investment of money (or other assets, as subsequent cases are also included), (ii) profits are expected from the investment, (iii) the investment of money is in a common enterprise, and (iv) any profit comes from the efforts of a promoter or a third party." (Adhami et al. 2018)
[35] Information taken from https://thetokenist.io/france-announces-first-ever-ico-approval-via-amfs-ico-visa/

- Gibraltar, which introduced specific DLT laws and it's going to launch the Gibraltar Blockchain Exchange (GBX);
- Portugal, which introduced regulations that would permit cryptos' services to expand;
- Malta, that introduced a favorable taxation system for companies issuing a coin offering;
- Switzerland, where banks are encouraging the issuing of ICOs and few regulations have been adopted by authorities. However, the Swiss Financial Market Supervisory Authority (FINMA) published guidelines and, if according to these an ICO constitutes a security, it will be subjected to Security regulation;
- Ukraine[36] and Estonia, where no specific regulations have been set yet by governments and authorities, and so entrepreneurs can find favorable situations for their projects.

Other European countries such as UK, Poland or Germany are regulating ICOs with a less favorable framework. In the UK, the Financial Conduct Authority (FCA) proposes case-to-case analysis to establish if an ICO must be subjected to Security Law or not, according to the rights obtained by the coin holder. A very similar approach has been taken by Germany's Federal Financial Supervisory Authority (BaFin). In Poland, the Polish Financial Supervisory Authority ("KNF") issued warnings related to the risks in investing in ICOs and cryptocurrencies, highlighting the lack of legal protection if the currency is not subjected to financial markets regulations (Kaal, 2018). However, a stronger and complete regulation system has not been introduced yet.

Even Russia[37], which passed from banning cryptos over the country to legalize digital currencies in 2018, still must manage how to deal tax regulations with digital rights.

In a similar way to the one adopted by ESMA, also the Australian market authority emphasized the risks related to the trading of digital assets, even if an ICO-friendly

[36] In Ukraine the power to regulate cryptocurrencies is held by the National Bank of Ukraine (Kaal, 2018).
[37] The Central Bank of Russian Federation is the authority encharged to regulates ICOs and cryptocurrencies (Kaal, 2018)

environment has been developed in the country through the remove of cryptocurrencies from double taxation policies and the diffusion of several Bitcoin's ATMs over the national territory. The Australian Security and Investment Commission (ASIC) is the authority enrolled to manage cryptos' regulation over the country. On the other hand, New Zealand considers cryptos as a payment infrastructure rather than a currency, leading to a stricter regulation for organizations. (De Vauplane, 2018; Kaal, 2018)).

## 1.6. ICOs' Geography

In this section it is analyzed why ICOs occurs more frequently in some countries with respect to others. The country with the most developed ICOs' diffusion, both considering amount of funds raised and number of projects issued, is the U.S., where USD 7.3 bn have been raised over years. Considering funds raised, after the U.S. there is Singapore (USD 2.5 bn), British Virgin Islands (USD 2.4 bn), Switzerland (USD 1.8 bn) and UK (USD 1.5 bn)[38]. If we consider the amount of ICOs issued, the first two positions are still the same, while then there are, respectively, UK, Russia and Estonia.

As it is emphasized by Huang, Meoli and Vismara (2018), 4 factors can be distinguished in order to analyze the diffusion of digital entrepreneurial activities:

- The development of financial systems: ICOs occur less frequently in country with more developed debt markets, public and private equity;
- The information and communication technology (ICT) development: ICOs occur more frequently in countries where ICT is more advanced;
- ICOs' regulation system: ICOs occur more frequently in countries with more developed digital regulation environment;
- The growth of online crowdfunding platforms: ICOs and crowdfunding play complementary roles in the financing of ventures.

[38] Data are taken from https://icobench.com/stats#stats-raised

These four hypotheses were analyzed and statistically demonstrated by Huang, Meoli and Vismara (2018)[39].

An interesting outcome is that, among the variables taken in considerations in order to decide where to launch an ICO, the taxing scheme is not considered as a crucially determinant variable for the organization's decision.

Also, it was earlier discussed about PROs and CONs of a more developed regulation scheme, which could lead to increase transaction costs but it also helps to build a safer environment for investors: the third demonstrated hypothesis is another proof that a better developed regulation system would have positive effectives on the ICOs' market development and diffusion, despite the consequent increase of transaction costs.

[39] Huang, W., Meoli, M., Vismara, S. (2018) The geography of Initial Coin Offerings.

# Chapter II

## Scope of the chapter

The main purpose of this chapter is related to the second phase that has been carried on for the aim of this elaborate. During the previous chapter of the work, an initial and general overview about Initial Coin Offerings and Distributed Ledger Technologies has been given to the reader, whereas the content of this section will be related to the construction and population of an Initial Coin Offerings' database. In order to carry on empirical analysis about Initial Coin Offerings, a crucial step has been that of collecting information about the involved projects in order to build a trustworthy dataset. After having created a reliable dataset, the conducted analysis and the main obtained results and outcomes will be discussed and highlighted in the following chapters of this elaborate.

The primary focus of this section is that of examining the methodology that has been adopted for the construction and the population of the database: the main sources – together with their alternatives - that has been consulted by the author will be listed and outlined (namely websites or online platforms that collect and list information about Initial Coin Offerings); it will be highlighted how the information have been imported from the websites to the Excel database; how have been managed the different information taken from multiple diverse sources about the same parameters of the same projects; and, in conclusion, a broad overview about the database and its structure will be illustrated.

In order to give to the reader a clearer understanding, all the parameters displayed on the database will be described in detail, in doing so it will be easier to comprehend the structure upon which will rely the main analysis carried on in the next chapters. This chapter will also be emphasizing to the reader some data related to the availability of information for each of the fields included in the database, and for each of them respectively the source used by the author will be mentioned. A description of the assumptions that has been made and about projects that has been discarded will also be given to the reader. The results of this work are saved in an Excel file whose data will be converted so as to enable the drafting of the next steps of this work, which consists in carrying on empirical analysis through the utilization of the statistical software STATA 12.

## 2.1. Database Methodology: introduction

In order to carry on descriptive statistics and empirical analysis regarding the phenomenon of Initial Coin Offerings – that will be discussed in the next chapters –, one of the key goals of this work was to construct and populate a database with complete and reliable information regarding as many projects as possible issued by ICOs' entrepreneurs. With the intention to carry on future text mining analysis[40], there was also the need to dispose of a complete set of data about the projects analyzed, therefore the author downloaded the respective whitepapers (when available on the Internet) of the projects included in the database.

This phase of the elaborate has begun from a long and careful analysis and consultation by the author of the different sources that are available, with the aim to understand from which of them the collectable data were more accurate and complete, and from which ones there was the permission (or was not) to download and use their data for the development of this work. The selection of the web sources to be analyzed has been made according to the information available online, but the most important and reliable sources that have been analyzed are the ones that have been used by the existing literature[41] regarding this subject area; as a consequence for this research were taken

[40] These analyses will not be discussed in this work.
[41] (Fisch & Momtaz, 2019; Fisch et al., 2019; Fisch, 2019; Adhami et al., 2018; Meoli et al., 2019)

into account also the sources that have been utilized in already existing papers about ICOs, papers that provided a work developed on a dataset construction.

Therefore, several different sources were analyzed and compared, and a well-weighted selection of multiple of them was carried on according to the magnitude and reliability of data available on the online platforms, with the intention to use the designated platforms as the main sources for the dataset. The database has been populated with the information made available by the selected sources, concomitantly the downloading of the accessible projects' technical whitepapers has been carried out (all the whitepapers that have been found are saved in a dedicated folder).

## 2.2. Online sources

### 2.2.1. List of sources

This section of the chapter will highlight and describe the different sources examined, with the aim of obtaining the data required for the construction of the coin offerings' database. The website ICOBench.com, which originally represented the principal focus for the construction of this work's dataset, does not allow to download its data automatically from the website. As a consequence of this, in order to start building a reliable dataset, other alternatives have been taken into consideration by the author and several online sources have been considered for downloading the data.

Hereafter will be listed some of the alternatives that have been taken into examination as sources for the data needed for the work, briefly descripted, in order to assess their strengths and their weaknesses as resources:

- **Coinschedule.com**[42] [43]: the website offers statistics and general information about ICOs, IEOs and other financial instruments in the market, giving just very limited information for those users that are not subscribed to the website. The

[42] https://coinschedule.com/
[43] As of June 2020, the website used by the author is no longer in operation.

complete list of ICOs is not available for users which are not signed up, while for subscribers is possible to access a list of 1671[44] projects developed from the 1st January 2016 to the end of 2019[45]. Data about projects issued in 2020 are available as well but were not considered by the author for the purpose of this work. Regarding the information pertaining the projects, the majority of them provided data referred to the project's name, the business area in which it operates, the ending date of the crowdfunding phase, an information about the amount of money raised (data are available in USD for every ICO listed in the dataset) and the percentage of Hard Cap[46]. Coinschedule.com made possible to download automatically and import into an Excel file all the information related to the considered 1671 listed on the website;

- **ICOBench.com**[47]: this website is perhaps the internet source with the most abundant, complete and precise information regarding the Initial Coin Offerings that have been issued. ICOBench.com comprises a total of 5725[48] projects developed from 2015 to 2020. Information that can be encountered on the portal are usually more widened and deepened in comparison with the ones that can be found on other sources. On this website it is commonly possible to find: a description regarding the project, the beginning and ending dates of the crowdfunding phase (and sometimes also data about the pre-sale phase), information about the country issuing the project, the restricted areas (if any), the token issued by the company, information about the contract type and a multitude of other financial information such as price of the token (the price during sale phase and sometimes also during pre-sale phase), how many tokens have been issued and sold, currencies accepted, soft and hard cap. Furthermore, the website made accessible other information, such as the amount of money

[44] As of May 2020
[45] For the scope of this work have been taken into consideration just ICOs ending before January 1st, 2020.
[46] Later in this chapter will be given a short description about what is the percentage of Hard Cap and how it's calculated its value.
[47] https://icobench.com/
[48] As of May 15th, 2020

that have been raised by the company, a project rating (from 0, lowest, to 5, highest) given by ICOBench.com team[49], a link to the company's website and also a section where it is possible to have a view of the technical whitepaper of the selected project (but this section does not make available the possibility of downloading the whitepaper).

- **ICODrops.com**[50]: on this website around 650 projects are at the disposal of everyone, divided in active ICOs (5 published projects), upcoming ICOs (39) and ended ICOs (more than 600)[51]. Approximately half of the projects (the 53% of the published ICOs) have been given a rating amongst the three options, namely "LOW", "MEDIUM" or "HIGH". Information available about the projects are related to the amount raised, the previous goal, if the there is any, related to the amount to be raised set by the company (it can be associated to the parameter soft cap), the category in which the project will operate, the market where the token was ("is" for present ICOs) going to operate and the ending data of the crowdfunding period;

- **Tokendata.io**[52]: this platform lists approximately 2400 different projects, and presents – for nearly all of them – their status (e.g. "Completed ICO", "Active ICO", "Planned ICO"), the amount raised by the company and the month in which the ICO has been issued, as well as a link to the technical whitepaper – if existing. For some ICOs the platform published also information regarding the token issued, its sale price and its actual value. In general, for further information regarding a specific plan, the portal leads directly to the said project company's

[49] The rating is given as a combination of an assessment algorithm using more than 20 criteria (such as ICOs' team, information that have been published, product presentation etc.) and a set of ratings given by independent experts. Some of the projects may also be evaluated by legal experts, especially in cases of suspicious ICOs. As the ICOs are monitored on daily basis, rating is not permanent and so the actual rating of a project may differ from the one registered on the database (https://icobench.com/ratings). Rating information on the database have to be considered as of May 2020.
[50] https://icodrops.com/
[51] As of May, 2020
[52] https://www.tokendata.io/

website, and so Tokendata.io does not provide very much other information itself;

- **ICOData.io**[53]: this website makes accessible around 2000 different projects. Each of them is complete with information about the amount raised, the token sale and current price and the ending date of the ICO. Further information may be available for some records, including: a description of the project, its starting date, the Hard Cap, the market capitalization, the number of tokens issued and other general financial information;

- Other websites have been reviewed and analyzed, inter alia **ICOtokennews.com**[54], **listico.io**[55], **icohotlist.com**[56] or **icomarks.com**[57] whereas other online platforms like **icosbull.com**[58] and **neironix.io**[59] have been used to collect few others missing information.

### 2.2.2. Methodology

After having analyzed and evaluated the various sources available on the Internet, the next step is to start the database by including the records and information taken by the primary source. Coinschedule.com has been chosen as the principal and main resource. Then wherever there will be missing information, a secondary source shall be chosen to fill the dataset.

The choice to take data from Coinschedule.com permitted to download automatically information about the 1671 projects available – for the period of interest of this paper – directly from their website. The parameters downloaded for each ICO comprised the

[53] icodata.io
[54] https://icotokennews.com/it/
[55] https://www.listico.io/
[56] https://www.icohotlist.com/
[57] https://icomarks.com/
[58] https://icosbull.com/
[59] https://neironix.io/it

business area in which the project was going to operate, the ending date of the crowdfunding phase, the amount of money raised by the entrepreneurs and the percentage of Hard Cap.

Coinschedule.com has therefore been used as groundwork for the ICOs' list on the database. The website is a "comparatively comprehensive and established data source that has been featured in many reputable outlets"[60].

After having filled the database with the information from Coinschedule.com, in order to have a complete and more reliable database, other information about the selected ICOs have been manually analyzed and taken by ICObench.com in order to use them to populate the database, together with the information already downloaded from Coinschedule.com. In a limited number of cases, other information have been taken by other online sources in order to incorporate the data that were already collected, adding to them supplementary details.

From the original 1671 ICOs downloaded from Coinschedule.com, it was made the decision to not include in the final database those ICOs that were not available on ICObench.com, because of the lack of information about those projects. It must also be underlined that some of the ICOs taken from Coinschedule.com were replicated because they have been registered on multiple records, according to the different crowdfunding phases (i.e. two or more different records for pre-sale, private sale and final sale phases), so the records not directly referred to the final period of the ICO were removed. After having detected and eliminated ICOs according to these criteria, from the original 1671 projects downloaded, 1475 were still available in the database. Out of those 1475 left, 1354 projects disposed of the technical whitepaper, which have been downloaded in order to carry on future text mining analysis. The rest of those ICOs, for which the whitepaper was not available on the Internet, have however been included into the database.

[60] (Fisch, 2019; Economist, 2017; Roose, 2017)

The main sources used with the aim to find the technical whitepapers on the internet, and subsequently download them, were: the individual company's website, ICOsbull.com, Neironix.io and ICOmarks.com.

When populating the database with information from multiple sources, it was possible to find uncoherent information for the same parameters according to the two different platform consulted. This happened especially for the field referring to the ending date of the token sales period and for the one referring to the amount raised by the company. In order to clarify as much as possible in the database all these information, 2 parameters about each of these two voices have been inserted, specifying in the field's name the source from which it has been taken. In projects where the information from Coinschedule.com and those from ICObench.com coincided or were similar, the field corresponding to the second source has been left blank, otherwise both the parameters corresponding to the two sources have been completed by the author. It is important to specify that all the values extracted from Coinschedule.com have been inserted in the database, whether it was identical, similar or different to the same parameter taken from ICObench.com: this is due to the fact that it has been downloaded automatically from the web platform.

## 2.3. Database structure

As it was anticipated in the Section 2.2.2, the database is composed of 1475 issued ICOs, while the structure of the database comprehends 28 fields containing information about each of the projects. For all the fields in the database, a blank value was accepted because of the high unavailability of certain information for some of the ICOs. The only exception where a blank field was not authorized concerned the fields "ICO_ID" and "ICO's Name" (both of which will be described later), due to the fact that the information for these two parameters is required to be always available and univocal, especially for "ICO_ID", as they work as primary keys of the database.

It’s worth mentioning that the parameter “ICOBench rating”, which has been filled for every project, has always been possible to find, hence why this field is present in all the records collected in the dataset.

In this section of the chapter it will be described to the reader each of the parameters available in the database’s structure and how it has been managed its information. The fields will be divided according to the sources from which they have been taken, which may be Coinschedule.com, ICObench.com or other Internet sources that have been consulted by the author.

### 2.3.1. Information from Coinschedule.com

As previously anticipated, information from Coinschedule.com has been automatically downloaded from the Coinschedule’s web portal. The fields included in the final database regarding the data from Coinschedule.com were a total of five.

The information, regarding the 1671 ICOs that were initially taken from the website, were about:

- **Name:** the name of the issued project, which does not usually match with the name of the company issuing it. In some instances, when dealing with the data taken from Coinschedule.com, multiple ICOs with the same name were listed, specifying next to the name the selling phase for the crowdfund raising phase (private sale, pre-sale, final sale). Information about these ICOs have been collected together in a unique record in order not to have a duplicate of the same project;

- **Category:** category in which the projects that has been issued are going to operate when their completion will be achieved. For 120 of the 1475 final records included in the database, this information was not available from Coinschedule.com. The others 1355 ICOs are subdivided according to a total of

35 different categories. The category more represented by the projects in the database was “Finance” with 251 projects linked to it, while other categories such as “Trading & Investing” and “Infrastructure” had respectively 166 and 107 ICOs related to them, being the second and the third most represented categories in the constructed dataset;

- **Ending date (Coinschedule):** related to the ending of the token’s final sale phase. Inside the database have been included projects whose ending date is comprised between the beginning of 2016 and the end of 2019: it’s important to recall that no ICOs ending in 2020 have been taken into consideration. Since the information about this field may be different from the one from ICOBench.com, it has been inserted a field for each of the two sources used, filling the ICOBench website’s ending date section just when it was significantly different from the data that has been downloaded from Coinschedule.com;

- **Amount raised (Coinschedule):** which represents the total amount of money raised by the company during the crowdfunding periods (including all the different phases). The currency indicated for this field is the United States Dollar (USD) for all the ICOs. Since the information about this field may be different from the one provided by ICOBench.com, similarly to what has been done for the field about the ending date, it has been created a field for each of the two sources used, and again the ICOBench.com’s amount raised section has been filled with data exclusively when significantly different from the data already available, which have been downloaded from Coinschedule.com;

- **Percentage of Hard Cap:** this parameter refers to the total amount of money that has been raised by the company with respect to the Hard Cap. “The Hard Cap is defined as the maximum amount of money a cryptocurrency can receive from investors in its Initial Coin Offerings"[61]. The Hard Cap is usually estimated before launching the ICO, hence it's not rare that the amount raised at the end of the ICO becomes greater than the forecasted Hard Cap, and as a consequence in the end the percentage of Hard Cap obtained, in those cases, will be higher than 100%. New values of this parameter sometimes may also be set by the entrepreneurs during the crowdfunding period. From Coinschedule.com, the information about the percentage of Hard Cap for about 244 of the 1475 was not available, while for 49 ICOs the reported percentage was above 100%.

Another parameter, not directly taken from Coinschedule.com, that is included in the final database refers to the univocal code **ICO_ID** that has been used in order to indicate the projects. This field of the database can be considered as a primary key for the database development. The parameter, that has been set by the author, is a non-repeatable number that has been assigned to each project according to the order in which Coinschedule.com listed them, when downloading the information. To each of the 1671 ICOs that were originally downloaded was given a univocal code from 1 to 1671: eventually, as explained beforehand, some ICOs has been removed with the intention of preventing information from being duplicated, but the univocal codes for the ICOs still have been left as they were listed, going from 1 to 1671.

When downloading whitepapers, the files were saved in a dedicated folder, in which all the whitepapers were collected together and named according to their corresponding univocal code in the database.

### 2.3.2. Data from ICOBench.com

Data from ICOBench.com has been manually added to the database in order to have a more complete set of information about all the projects. During database's population of these information, when an ICO from Coinschedule.com was not found on

[61] https://decryptionary.com/dictionary/hard-cap/

ICOBench.com, the project was removed from the dataset: the result is that, as a consequence, there have been included in the database ICOs that were available both on Coinschedule.com and on ICOBench.com. The fields included in the final database that have been manually retrieved from ICOBench.com are represented by a total of 20 parameters.

Parameters taken from ICOBench.com are now listed and explained and consists of:

- **Description**: this field consists of a brief description about the company and the project that has been launched by the entrepreneurs. It deals especially with the project's main purpose and the methods used (or planned to be used), through which the entrepreneurs will achieve their goals, after having completed the crowdfunding phase. It can be considered as a description about how the money collected by the company will be used in order to realize the project. A description was found on ICOBench.com for about 1471 of the 1475 ICOs that populates the database, whereas for the other 4 ICOs no information have been found on the website;

- **Starting date**: since Coinschedule.com allowed to take information solely about the ending date of the ICO, from ICOBench.com was taken the information about the starting date of the project. In some instances, this field may be in contrast with the one about the ending date extracted from Coinschedule.com: however, both information has been collected and recorded in the database. No specifics about the starting date were given on ICOBench.com for around the 8% of the ICOs included in the dataset;

- **Ending date (ICOBench)**: this field of the database has been populated just in those situations in which the information from Coinschedule.com was uncoherent or different with the ones available on ICOBench.com. This phenomenon happened for around the 13% of the ICOs in the database. It has not been taken into consideration by the author a difference of less than three days regarding the dates;

- **Token**: this field refers to the name or the symbol used to indicate the token issued to investors. This information for an ICO is usually univocal with respect to the others in the database. The only exception is represented by ICOs for which was not specified any information about the token, which are around the 3% of the total number of ICOs in the final database;

- **Platform**: this field represents the platform on which the token was developed. For a significant part of the ICOs (1300, around the 88%) the platform used was Ethereum. Apart for the ICOs using Ethereum, as for the others there was a great heterogeneity about where the token has been developed. The platform Waves have been used for 27 of the tokens in the database, while the Bitcoin has been used just for few of the ICOs collected. A lot of other platforms have been used just for few ICOs each one. For just 4 ICOs collected it was not specified nothing about the platform used for the development of the token;

- **Type of token**: it refers to the fact that a token can be a utility or a payment token. No security tokens were available in the database. For most of the cases the tokens included were utility tokens (1450), 24 were payment tokens while for one project it was not specified. Tokens have previously been classified and differences between utility and payment tokens are available in the first chapter of this paper;

- **Type of contract**: it refers to the standard model of smart contract adopted by the company. Occasionally, it may happen to use more than one standard contract for each ICO, although it's something quite rare in the dataset built. In general, the most used type of smart contract was the ERC20, adopted by the 75% of companies. ERC20 has emerged as a new smart-contract standard for Ethereum-based token issued. The key advantages of ERC20 are represented by its compatibility with other Ethereum-based tokens, the easy way in which these tokens can be traded with each other, and the transparency standards set for token transfers[62]. The contract type provides a standard set of rules that the token itself should adhere to[63]. Among other standards for smart contracts, ERC223 and others ERC[64] contracts has been adopted by some of the ICOs available in the populated database;

- **Country**: it represents the country where the ICO has been issued. In this field, the most represented countries are Singapore and the U.S., where 187 projects for each one of them have been issued. In the UK there have been issued 138 of the projects populating the database, but this number increases even more when considering other British territories (such as British Virgin Islands, Isle of Man, Cayman Island and others). This field cannot be populated by more than one country. In the built dataset, information about the country where the project has been issued was not available for 35 ICOs;

- **Restricted areas**: areas in which is not possible for local investors to allocate their resources into the selected projects. This field is completed by inserting one or more countries as a restricted area. For 49 of the ICOs in the database was explicitly declared that there were not restricted areas, while for 932 was not made available any information. For all the other projects (around 500), at least one restricted area was indicated on ICOBench.com;

- **KYC/Whitelist**: usually ICOs' investors may face some controls before being authorized to allocate their money into one of the launched projects. In order to be allowed to invest its money the actor may have to register himself into a "Whitelist": for some projects in fact the distribution of coins will be allowed only to "Whitelisted" investors. In this way, the company's entrepreneurs are capable

[62] https://icospotters.com/blog/what-is-erc20-token-standard-and-why-is-it-useful-for-icos
[63] https://www.investopedia.com/news/what-erc20-and-what-does-it-mean-ethereum/
[64] ERC stand for Ethereum Request for Comments (https://icospotters.com/blog/what-is-erc20-token-standard-and-why-is-it-useful-for-icos);

to track who is receiving their tokens and investing in their ICO. A similar approach can be highlighted when referring to KYC. KYC, which stands for “Know Your Customer” is a professional financial protocol allowing to identify the other party through which exists a business relationship. These checks have the objective to prevent businesses to be used by criminals for money laundering and other illicit purposes. An ICO may have both of these checks, just one or may also have none of them. There are of course other types of checks that were developed and used for the same purpose, but this work will cover specifically only these two, which are among the most well-known and important. Talking about the ICOs that have been inserted in the database, 644 of the 1475 ICOs explicitly indicated that at least one of these two checks was used (and for 354 of them both controls have been adopted); 181 of them explicitly reported that none of them was used while for 650 projects was not possible to find any information about adopting KYC or Whitelist;

- **Price in ICO**: this field refers to the price of the token during the crowdfunding period. The price indicated in this section is not related to the price during private sales or pre-sales, but specifically to the one during the final sale. This information was given for most of the ICOs (1365 out of 1475) in USD, while for 87 of them this information was not available. For the other projects the information was available in ETH. On pre-sale the price of the token could have been smaller than the final sale price, being the token usually sold at discount in this phase, but however this data was not included because of the few projects having specified this information on ICOBench.com;

- **Tokens for sale**: this information is related to the number of tokens issued by the company for developing its project. It comprehends the entirety of tokens available in all the different phases of the ICO. For around the 22% of the ICOs in the database this information was not available on ICOBench.com;

- **Sold tokens**: tokens that have been actually sold by the company during all the phases of the ICO. Its value of course cannot be higher than the one for "Tokens for sale" for the same project. Information about this parameter was made available solely for a few ICOs in the database: in fact, at the beginning of the database's population, it was initially decided to not include this information: just in a second moment this field has been included in the final dataset. When completed the database's population, 114 ICOs had completed this field with an indication about the sold tokens;

- **Minimum investment**: the minimum amount of money that an investor could allocate when deciding to invest on a project. The minimum investment may be indicated in USD, in ETH, in other local currencies or even in the issued token. Even in this case, it was not available on the online platform any information for more than half of the projects;

- **Currency accepted**: for this field were indicated the currencies that the company was willing to accept in order to finance its project. This field may be filled with more than one currency, as most of the companies would accept several types of them. Among the most accepted currencies there are USD, BTC and ETH. For this field, for 193 ICOs was not available any information about the currencies that the company was willing to accept;

- **Distributed in ICO**: it's a percentage value that represents the amount of token that the company is going to distribute among investors. Being a percentage referred directly to the total issued tokens, its value can't be higher than 100%. This information was available for around the 71% of ICOs in the database. Sometimes this field has been filled not with a single percentage but with a range of percentages;

- **Soft Cap**: this field refers to the minimum amount of money that the project must raise in order to be carried on by the company. If the amount collected at the end of the ICO is lower than the Soft Cap that has been initially set, the company should return to the investors the money allocated in the project and stop its development, or it can decide to organize another crowdfunding phase. It's not mandatory to set a Soft Cap before launching an ICO, but it can be a good indicator for investors in order not to allocate their money in a project for which it's clear that its minimum goal will not be accomplished. For most of the projects in the database this information was available on ICOBench.com in USD or in ETH, but it was also possible to find it sometimes in other local currencies, such as EUR, GBP or CHF, or in other cryptocurrencies such as BTC. For around half of the projects in the database was made available an information about the Soft Cap that has been set by the entrepreneurs.

- **Hard Cap**: as it was earlier anticipated in the previous section of this chapter, Hard Cap is defined as the maximum amount that can be collected by a company during its Initial Coin Offerings. As for the Soft Cap, it is not mandatory to set it before launching an ICO, and as for the Soft Cap information are available for most of the projects in USD or ETH, but it's also possible to find it in other local currencies or in cryptocurrencies. For almost the 80% of the ICOs an information about Hard Cap was available. Hard Cap may be – in some rare cases – equal to the Soft Cap, but it cannot be lower;

- **Amount Raised (ICOBench)**: similarly, to the field "Ending Date (ICOBench)", this field is populated just in case the information from ICOBench.com differs from the one of Coinschedule.com. The information from the two websites differed for this field in 35% of the cases, but in many of them the order of magnitude of their differences was very little, being of some thousands of USD or less. However, there were some cases in which the differences for this field were relevant, having completely different total amounts raised. While for

Coinschedule.com data about the amount raised by the project were always available, for several ICOs on ICOBench.com was not specified any information about this field;

- **ICOBench Rating**: this last parameter that has been taken from ICOBench.com refers to an overall rating to the project's reliability given directly by the ICOBench.com website. The rating is given through an algorithm developed by the website's team and through some experts' rating. The information about the project's rating was accessible for all the 1475 ICOs that are contained in the database. The rating can be given from 0 to 5: the project with the lowest rating among the ones in the database has an overall score of 0.5, while the one with the highest rating has a total score of 4.9. A low rating should be an indicator for investors to be careful before allocating their money in the project, both because may be a scam or because it may simply be an unprofitable ICO.

### 2.3.3. Data from Web sources

Two other fields were included in the database: these parameters were not available in either Coinschedule.com or ICOBench.com and have been filled consulting other Internet sources. These two fields do not add any relevant information to the reader, but give the possibility for whoever is consulting the database to deepen its knowledge about the project, consisting in two URL links that redirect to other web pages that can present more information about the company issuing the ICO and the ICO itself.

The two additional fields added to the database were taken from web sources and consists of:

- **Whitepaper link**: this field contains a link that redirects to an online version of the technical whitepaper of the project. Technical whitepapers have been downloaded by the author, when available, and saved all together into the same folder, with each whitepaper named as the ICO ID corresponding to that project in the database. The link available on the database redirects from the same source from which the whitepaper has been downloaded. Usually, the sources used in order to download the papers were ICOsbull.com, Neironix.com, the website of the company issuing the project or other free online sources. It may be possible that for a record in the database it is available the link to the technical whitepaper, while the whitepaper has not been downloaded: this happened when the consulted source allows the visitor just to view the paper but not to download it. In the database is possible to access to 1362 links to their respective technical whitepapers, while the whitepapers that has been downloaded are 1354. For the remaining ICOs was not possible to find any information about technical whitepapers on the Internet;

- **Website URL**: this field contains a link to the website of the company issuing the project or to a dedicated website created by the ICOs' entrepreneurs specifically for the/their project(s). On ICOBench.com for some of the ICOs was available a link to the company's web page, therefore in some cases the information has been taken from ICOBench.com, but in most cases the website has been looked for through online searches. For many of the ICOs in the database a website has not been created by the company or was not anymore available: a link to the company or project's website has been inserted for the 974 ICOs for which a web page was available, out of the 1475 ICOs that were collected into the database. For the remaining ICOs, around 500, it was not possible to collect this type of information.

As previously stated, ICOBench.com and Coinschedule.com are the primary sources of information for this work, but in some instances, it was needed to consult other online platforms (as described beforehand in the section 2.2) on the web listing Initial Coin Offerings, to have a database as complete as possible. All the information provided on this paper, collected by the author, are to be referred as of May 2020.

## 2.4. Next steps

After having completed and populated the database with all the information taken from Coinschedule.com, ICOBench.com and the other occasional online sources, the next step of the elaborate will be referred to the possibility of carrying on descriptive statistics and empirical analysis from the data that have been collected in this phase of the thesis, using for this purpose the software STATA.

In order to carry on the next phases of this work, some of the data collected in the final database may not be included in any of the future analysis, since some of the inserted information have been collected just in order to give to the reader a complete and more proper overview about each of the included projects. Thus, some of the data have been collected anyway, even if they are not useful for the development of the next part of the work, which is represented by a more complete objective and statistical analysis.

Referring to the information about the amount of money raised by companies and entrepreneurs for the ICOs, the data that have been downloaded from Coinschedule.com will be integrated with the ones that have been collected from ICOBench.com in those cases where the two information differed from each other. The same approach will be used for the information referring to the ending date of the ICOs.

Information about the tokens that have been actually sold won't be included in the analysis because of the high unavailability of such data. Information about the description of the projects or about the links to ICO's technical whitepapers and company's website were included, with the aim to give to those consulting the database a quick access, if needed, to a more complete and detailed set of information about the project. However, information about whitepaper's availability will be used. Other parameters, such as "Minimum Investment" or "Currency Accepted" won't be directly used both for the little added value they could give to the analysis that will be carried on and because of the high inhomogeneity of the currencies that characterize these fields of the database, being expressed for some ICOs in USD, for others in ETH and for other else in other cryptocurrencies or local currencies. Their information may however be used through some data manipulations that will be later deepened.

In conclusion, the information about the rating given by ICOBench.com also won't be described in the next chapter but may be used for further empirical analysis when discussing about a linear regression. In general, this field has been inserted by the author just to provide an indicator about the most reliable ICOs among the ones that have been selected and collected into the dataset of this work.

# Chapter III

## Scope of the chapter

During the first part of this elaborate it has been given a general overview regarding Initial Coin Offering and Distributed Ledger Technologies, taking into consideration the past literature regarding these topics. Moving on to the second phase of the thesis it has been described the part of the work regarding the database construction and methodology, analyzing sources, variables taken in consideration and giving to the reader an overview about the whole set of data.

On the other hand, this part of the work will face the first analyses conducted through the database built and the collected data: the first general analysis, which are discussed on this chapter, refers to some descriptive statistics regarding some of the variables included in the database, analyzing the data by at first considering the whole set of Coin Offerings that are populating the database, and then dividing the ICOs in sub-samples according to some of the variables and their value.

Before deepening the analysis carried on in this section of the work, a further step is necessary in order to allow the correct import of the data to the statistical software STATA: some adjustments to the cells' content were necessary in order to allow STATA to assign to the variables their correct format, avoiding misinterpretation of similar data and having a more coherent and intuitive structure. The conversion of some currencies or in general of other variables was needed and has been conducted before the descriptive analysis.

After having allowed the software to correctly take the data from the database, the descriptive statistics were carried on through the support of some of the most common and popular statistical tools. The output from the software has been inserted inside this work and briefly discussed for all the variables and the subsamples analyzed in this section of the elaborate.

Once this phase will be completed, on the next chapter the analysis will deepen the relationship between the success of an ICO and other variables included in the database, carrying on this activity through a linear regression analysis, obtained through the same software.

## 3.1. Descriptive Statistics: introduction

In this section of the work for the first time the data that have been collected in the previous phases are going to be used in order to carry on statistical analysis. On this chapter, the focus will be on descriptive statistics on the whole dataset and on some sub-samples organized according to variables' values. Before this part of the work, it was necessary a data conversion for some of the variables' information collected: the currencies, for instance USD, ETH, BTC and other local currencies or cryptocurrencies, have all been converted into USD in order to have a coherent unit of measurement for the variables referring to an amount of money. Other little adjustments to some values have been made necessary, and some variables were not included in the final analysis because of the high inhomogeneity of data, for the high unavailability of information or for the too little added value they were bringing to the whole process.

After the conversion, it was possible to give to the software coherent data, and as a consequence the descriptive analysis have been carried on using some popular statistical tools for a probability distribution such as the arithmetic mean, the standard deviation, the kurtosis, the skewness and some others. These analysis have been carried on at first for the total set of data collected, and then dividing the data into some subsamples according to the country where the ICO has been launched, the platform used by the company to develop the token offered, the protocols used by the entrepreneurs and other criteria. Each of the STATA output for this section has been briefly described and analyzed. After having completed this type of analysis, on the next chapter of the work the focus will shift towards a regression analysis with the data collected.

## 3.2. Preliminary activities

### 3.2.1. Conversion of currencies

In order to carry on the empirical analysis through the software STATA, some adjustments were required in order to allow the data that have been collected to be used for the purpose for which had been inserted in the database – otherwise the software would have misread the inputs. As it was previously anticipated, some of the data have not been used due to:

- high inhomogeneity between the values included in one or more variables;
- the little value added that the information of the data could have brought to the analysis;
- the lack of information regarding the variable.

For certain other data, when possible, it was only necessary to convert the information collected in the database into a homogenized unit of measure. While for the field “Amount Raised” all the data - which were downloaded from Coinschedule.com - were registered in USD, for the parameters “Hard Cap” and “Soft Cap” – that has been taken from ICOBench.com - a lot of information were saved in different currencies, such as EUR, ETH, BTC, GBP and other local currencies or cryptocurrencies. Most of these data, among the ones not saved in USD, had their information saved in ether ETH. The solutions adopted to manage these inconsistencies among information were different according to the diverse type of data faced. The approach used by the author for this part of the elaborate was the following:

- **For data saved in ETH:** for information collected in ETH, which represented a consistent percentage of the dataset for variables “Hard Cap” and “Soft Cap”, a manual approach was not possible because of the large amount of time that the process would have required. In order to manage the conversion of the cryptocurrency, information about all monthly exchange rates between ETH and USD have been downloaded[65] and then saved into an Excel file. The exchange rates downloaded covered the period from 2016 to 2020. From these data, with Excel functions MID()[66] and VALUE()[67] it has been extracted the information of interest (exchange rates and dates) and converted from string into numerical values. At the same time and through the same functions, from the data saved in the database (which were saved as a string containing “*amount_soft_cap* ETH”, e.g. “300 ETH”) it has been carried on a necessary conversion into numerical values for the information already saved as string for the two variables. As last step of this phase, through the Excel functions IF()[68] and VLOOKUP()[69] were compared – just for the ICOs with information in ETH – the ending date of the ICO and the dates saved for monthly exchange rates, and when the two information matched, the data originally saved into ETH was multiplied by the corresponding exchange rate, converting the information in USD. This new information was then saved and used for the analysis;

- **For data saved in other currencies**: for currencies which were not USD nor ETH it was decided to manually convert the data into USD according to the exchange rate present during the ICOs’ crowdfunding period. This approach was made possible because just a few data didn’t have information collected in USD or ETH.

---

[65] https://it.investing.com/crypto/ethereum/eth-usd-historical-data
[66] The Excel function MID() allows the user to extract from data saved as string some specific information according to the input parameters.
[67] The Excel function VALUE() allows the user to convert a non-numerical value into a numerical one, when this is possible.
[68] The Excel function IF() allows to perform an activity conditioned to other factors that may influence the desired result.
[69] The Excel function VLOOKUP() returns a value from a table corresponding to the finding of another VLOOKUP value.

Information about currencies' exchange rates have been taken from the website money.it[70];

- **For data saved in USD:** these data required no currency conversion, however a simple conversion from a string format to a numerical format has been used, in order to allow the software to use these data as numbers. Through Excel functions MID() and VALUE() have been removed information about the currency being in USD, in a similar way to what has been done for information saved in ETH, and the remaining result was converted from a string into a numerical value.

### 3.2.2. Other database adjustments

Other parameters in the database needed – whenever possible – additional little conversions. Some of these conversions will be carried on in the next step of the work, while for this one their application was limited to some variables. The "description" field has not been used for the empirical analysis or the descriptive statistics because of the low applicability of the variable to such processes, whereas as far as the variables containing dates information are concerned, only the ending date may have been taken into account, and specifically were mainly considered the information extracted from Coinschedule.com due to higher availability of data from this web source in respect to ICOBench.com. Furthermore, information about restricted areas have not been used too in this specific phase (but its information will be manipulated in order to be used in the next chapter), while for the field "Price in ICO" was necessary another currency conversion, which was carried manually – from money.it - because of the low number of data of this variable that did not have their information collected in USD. Information about minimum possible investment for the ICO and the various currencies accepted were not used in these analysis due to the high inhomogeneity of the data and the little impact they could have on the success of an ICO, while for the field Amount Raised have

[70] https://www.money.it/

been used the information from Coinschedule.com as opposed to those from ICOBench.com because of both the higher availability of data and the consistency of currencies for different records, which were already collected in U.S. dollars.

### 3.2.3. Import of the database to STATA

After having completed the crucial adjustments activities on the database, in order to use the software was necessary to import the database to STATA[71]. The adjustments were made in order to make the software register the fields of each record in the correct format[72]. When the adjustments were completed, a copy of the populated database has been converted from an Excel (.xlsx) file into a text file (.txt) tab delimited[73]. After this conversion, the new text file with database's information was directly imported into STATA and the whole dataset was memorized by the software, in in the expectation of being analyzed.

The records were correctly saved and the integrity of data has been checked through basic STATA commands such as describe()[74], summarize()[75], codebook()[76] looking for

[71] The STATA version used for this phase of the work was STATA 12.

[72] As format is intended string (str), integers (int), floating (float), dates and many other more specific formats

[73] In order to allow the import of data to STATA 12, it was necessary to convert the Excel file into a tab delimited one, otherwise the conversion was not possible. Alternatively, to the tab delimited, it was possible also to use a comma-separated value file (.csv). For the purpose of this work it has been used the .txt file approach.

[74] STATA's describe() function provides data for the variables specified (if any variable is not specified, it provides a general overview) regarding the format, the labels, the numerosity of data and general statistics.

[75] STATA's summarize() function provides for the variables specified (if any variable is not specified, it provides a general overview) some more widened information with respect to the ones of the describe() function. It returns in fact some statistical information such as the number of observations, the arithmetic mean, the standard deviation and the range of values assumed by the variable, from the minimum to the maximum present in the database

[76] STATA's codebook() function provides for the variables specified (if any variable is not specified, it provides a general overview) more detailed information with respect to the other two already explained functions. This command, in addition to all the information that the other two functions were providing, highlights also the number of missing observations for the analyzed variables and their percentiles' subdivision, showing the 10% percentile, the 25$^{th}$, the 50$^{th}$, the 75$^{th}$ and the 90$^{th}$.

their output in order to compare it with the expected results, checking for eventual mistakes or data's misinterpretation.

## 3.3. Statistical tools

Descriptive statistics have been carried on through STATA, analyzing some of the variables in the dataset through the main popular statistical tools. Variables that have been included in this analysis are Soft Cap, Hard Cap and the most important Amount Raised, which represents the best indicator to determine the success of a token offering. The other two variables have been included with the aim to look for any evidence about a hypothetical relationship between them and the success of the project, and additionally to have a rough comparison among more variables.

It was made the choice to use 8 main statistical tools, amid the most popular ones, to analyze these variables. The statistical instruments that have been used were the arithmetic mean, the median, the standard deviation, the variance, the standard error of the mean, the coefficient of variation, the kurtosis and the skewness of the distribution. The tools just listed are now briefly described and explained one by one:

**Arithmetic Mean**: is the average value of the distribution, evaluated as the sum of all the recorded observation divided by the total number of observations. It represents a non-biased expected value of a distribution. The arithmetic mean is the most common and popular, but not the only one, estimator of the mean of a random variable. The formula used to evaluate the arithmetic mean of a distribution is $x = \frac{\sum_{i=1}^{N} x_i}{N}$.

**Median**: is described as the intermediate value between the extremes of a numerically ordered sequence, which subdivides the distribution in two equal parts. The median is another estimator of the mean, which in some cases, according to the distribution, may be preferred to the most common arithmetic mean. The median observation is the one which splits the sequence of records in 2 parts, where the 50% of observations have a lower (or equal) value than the median of the distribution and the other 50% have a higher (or equal) observed value with respect to the median. The formula used to evaluate the median can be explicated as $x = \frac{N+1}{2}$ when N is an odd number, whereas there are multiple criteria that can be used in case N is a pair number, however leading all of them to similar results each other's.

**Standard Deviation**: it is a measure used for the dispersion of values in a set of observations. The standard deviation is evaluated from the square root of the sum of the square distances between each observation from the estimated mean of the distribution, divided by the number N of observation minus 1[77]. It has the same unit of measurement of the mean and its value is also dependent on the order of magnitude of the observations recorded. The standard deviation it's the square root of the variance. Its formula is $s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - x)^2}{N-1}}$;

**Variance**: it's another unit of measure for the variability of values in the set. It's an indicator about how much the value of the observation quadratically differs from the mean of the analyzed population. It's strictly related to the standard deviation of the distribution, in fact it's evaluated as $s^2 = \frac{\sum_{i=1}^{N}(x_i - x)^2}{N-1}$ as it's the square of the standard deviation;

**Standard error of the mean**: it can be considered as an estimation of the standard deviation of the error of the sample mean with respect to its true value, which is unknown. It is evaluated as $s_x = \frac{s}{\sqrt{N}}$. As it's rational to think, increasing N will lead to a decrease of the magnitude of the mean error, as the sample is more representative of the population analyzed and with N→∞ $s_x$ tends to assume the same value of the standard deviation s. Its unit of measurement is the same with respect to the one of arithmetic mean, median and standard deviation;

**Coefficient of variation**: the coefficient of variation is another measure of dispersion of a distribution, but with respect to the standard deviation and variance it has the

[77] The subtraction by one is necessary because it has been made an estimation according to the observed data, which leads to the loss of a degree of freedom.

advantage of being standardized, and so not being affected by the order of magnitude of the data analyzed. In order to calculate the coefficient of variation, the standard deviation of the distribution is divided by the arithmetic mean, leading to have a non-dimensional coefficient as indicator for the dispersion. The formula used to evaluate the coefficient of variation is $c_v = \frac{s}{x}$.

**Kurtosis**: it can be considered as a "tailedness" measure of the probability distribution of a random variable, i.e. it considers the information and amount of data contained on the tails of the distribution. It is a sort of distancing indicator from a normal distribution: the kurtosis of a normal distribution is equal to 3, when the kurtosis of a distribution is less than 3 the distribution is platykurtic, while when it's more than 3 it's leptokurtic. For a leptokurtic distribution will be easier to have outliers and so the tails are fatter, affecting the results, while for platykurtic distributions the tails are thinner, and outliers are something rare. The kurtosis has no unit of measurement as it is a non-dimensional indicator. The kurtosis is the fourth standardized moment[78] of a distribution, defined as $Kurt[X] = E[\left(\frac{X-\mu}{\sigma}\right)]^4 = \frac{\mu_4}{\sigma^4}$

**Skewness**: skewness can be referred as a measure for the asymmetry of a probability distribution of a random variable. The skewness may assume a positive or negative value: a negative skewness means that the left tail of the distribution is longer, while a positive skewness means that the right tail is the one longer. For a normal distribution, the skewness is always equal to 0 and the tails are the same, because there is perfect symmetry. When the skewness assumes a negative value, data are skewed to the left and the arithmetic mean leads to be always minor than the median, while in case of positive skewness data are skewed to the right and the arithmetic mean always results to be higher than the median. The skewness of a variable X is the third standardized

[78] The moment, in mathematics, is a specific measure of the shape of a function. It's evaluated as $\mu_n = \int_{-\infty}^{+\infty}(x-c)^n f(x)dx$. A standardized moment is a moment that has been normalized through expected value and standard deviation. In statistics, the standardized zeroth moment of a probability distribution is the probability function, the first is its expected value, the second is its variance, the third is the skewness and the fourth is the kurtosis.

moment of its distribution, and it’s mathematically defined as $SKEW[X] = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$

## 3.4. Descriptive statistics

All the statistical tools just described have been used with the intent to extract information from the dataset obtained in the previous steps. Through these tools have been carried out descriptive statistics to emphasize some characteristics of the ICOs’ database distribution. Descriptive statistics have been highlighted for the total ICOs listed in the database and for some subsamples, basing the creation of subsamples on the observed values of some of the variables.

The following Table 1, obtained by STATA’s output, highlights the statistics for the whole dataset imported on the software, for the variables that have been included in this part of the analysis. On the following subsections also results related to the other subsamples of the database will be emphasized, and the subsamples will be chosen according to the countries of ICO’s development, the categories in which the projects are going to operate, the protocols used by the entrepreneurs and some others relevant criteria.

### 3.4.1. Analysis on the entire dataset

*Table 1. Statistical analysis for the total sample*

| TOTAL | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 7348703 | 4.76E+07 | 1.84E+07 |
| Median | 2500000 | 2.00E+07 | 6900000 |
| std dev | 2.66E+07 | 1.19E+08 | 1.22E+08 |
| var | 7.08E+14 | 1.43E+16 | 1.49E+16 |
| std error | 972835.2 | 3510463 | 3179450 |
| cv | 3.62059 | 2.507799 | 6.647153 |
| kurtosis | 192.1392 | 64.81325 | 957.4691 |
| skewness | 12.31511 | 7.046789 | 29.23772 |

As it was reasonable to expect, analyzing the output related to the whole dataset allows to highlight the fact that the average amount raised by projects is between the Soft Cap (minimum amount that the company should raise in order to continue pursuing the project for the ICO has been launched) and Hard Cap (maximum goal for Amount Raised set by the company). It is possible to ascertain from the Table 1 that the arithmetic mean is for all the three variables more than twice the median, because of the very high amount of money collected (or set as Soft and Hard Cap) by particularly successful ICOs (e.g. Telegram Open Network[79]), which increases the mean but does not affect excessively the median.

Consequently, also the dispersion of the data clearly assumes a quite relevant value, as it is highlighted by the standard deviation and variance. The coefficient of variance highlights that for the variable Amount Raised, the standard deviation is more than 6 times the arithmetic mean of the distribution, emphasizing the high fluctuations that the variable value may assume. The very high kurtosis values show that all the three variables analyzed follow a leptokurtotic distribution, which are distributions with heavier tails and as a consequence a high probability to have extremes outlier values. This is also in line with the difference just found out between median and arithmetic mean, being the mean an order of magnitude higher than the median.

Finally, the skewness of the three variables have always positive values, highlighting asymmetries with a longer or fatter tail on the right side of the distribution. In case of positive skewness, the arithmetic mean results to be always higher than the median, as it was earlier explained, while in case of negative skewness the median gets always a higher value than the arithmetic mean. In the cases of Table 1, it is possible to see in fact that the positive skewness leads, for all the three variables, the arithmetic mean to assume a higher value than the median, while their relevant differences can be also emphasized and explained by the consistent kurtosis' value.

[79] An ICO referred to a "cryptocurrency based on multi-Blockchain Proof of Stake system […] designed to host a new generation of cryptocurrencies and decentralized applications". The project raised in 2018 around 1.7 Billion of USD. (https://icobench.com/ico/telegram-open-network)

### 3.4.2 Subdivision by countries

*Table 2. Statistical analysis for ICOs developed in the U.S.A.*

| USA | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 5427031 | 4.75E+07 | 4.23E+07 |
| Median | 2210749 | 2.50E+07 | 1.13E+07 |
| std dev | 8145178 | 1.01E+08 | 3.07E+08 |
| var | 6.63E+13 | 1.03E+16 | 9.43E+16 |
| std error | 1026196 | 8793595 | 2.25E+07 |
| cv | 1.500853 | 2.135258 | 7.253753 |
| kurtosis | 16.12816 | 38.54965 | 181.1497 |
| skewness | 3.226247 | 5.700887 | 13.35488 |

*Table 3. Statistical analysis for ICOs developed in U.K.*

| UK | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 4189969 | 4.07E+07 | 1.11E+07 |
| Median | 2000000 | 2E+07 | 7470000 |
| std dev | 6743192 | 8.50E+07 | 1.27E+07 |
| var | 4.55E+13 | 7.23E+15 | 1.61E+14 |
| std error | 794692.8 | 8108725 | 1080300 |
| cv | 1.609366 | 2.08792 | 1.140592 |
| kurtosis | 22.18867 | 48.08891 | 9.588779 |
| skewness | 3.974382 | 6.246241 | 2.258415 |

*Table 4. Statistical analysis for ICOs developed in Singapore*

| Singapore | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 1.29E+07 | 6.12E+07 | 1.25E+07 |
| Median | 3339490 | 2.00E+07 | 8562751 |
| std dev | 5.05E+07 | 1.77E+08 | 1.36E+07 |
| var | 2.56E+15 | 3.13E+16 | 1.85E+14 |
| std error | 4819744 | 1.42E+07 | 997197 |
| cv | 3.907309 | 2.889135 | 1.086253 |
| kurtosis | 80.4116 | 39.64076 | 15.18397 |
| skewness | 8.490713 | 5.764716 | 2.628943 |

In this section have been chosen three different countries where to focalize the attention for this subsample analysis. USA, Singapore and U.K. have been chosen because of being the most represented countries in the database developed. The variables and the statistical tools taken in consideration are the same of the previous case.

Looking at the values for the mean and median for the variable Amount Raised on Tables 2, 3 and 4, it seems that ICOs developed in the U.S. have been more effective in collecting money than others, having as an arithmetic mean more than 3 times the value of Singapore's and UK's and more than twice the value of the total database population. Despite this, the very high dispersion of data emphasized by standard deviations does not bring any statistical evidence about USA's ICOs collecting more money, i.e. we can't be sure that the differences highlighted by the analysis reflects an actual over-performance of USA's ICOs with respect to others. However, it's clear that the development country may play a major role in the ICOs' success.

It is possible to notice from Table 4 that the arithmetic mean for the Amount Raised variable is lower than the one for the Soft Cap, regarding ICOs developed in Singapore. This could allow to think that for many Singapore's projects the minimum amount to be raised has not been reached by the company: however, this information should be treated very carefully due to the fact that setting a Soft Cap is not mandatory and several ICOs may not have it, affecting the output of this analysis. In this case, it was verified that among the 188 ICOs developed in Singapore, for 112 of them a Soft Cap has been set by the entrepreneurs, and for 26 of these projects the Soft Cap has not been reached. Because of the high differences between Amount Raised and Soft Cap for these 26 cases, the Soft Cap results to be on average a little higher than Amount Raised, even if this may be a little counterintuitive. With this example it is possible to verify that the median, in some instances, may represent a better indicator for the average value of a probability distribution than the arithmetic mean, especially in these cases where the distributions analyzed are asymmetrical and the presence of many outliers are affecting the obtained results.

By the coefficient of variation, it's easy to verify that for the Amount Raised variable in the ICOs developed in the U.S.A. (Table 2) there is a relevant dispersion of observations, being the standard deviation more than 7 times the arithmetic mean. On the other side, for the U.K. (Table 3) and Singapore (Table 4) the values of mean and standard deviation for the same variable are more similar each other. A hypothesis test to verify if the US ICOs would collect more money would surely be rejected because of the high dispersion of the distributions.

The skewness of the three variables emphasize for all the subsamples an asymmetry with fatter right tails, in fact it's possible to verify that the median always assumes a lower value than the arithmetic mean. For the Amount Raised variable in the U.S.A. it is also possible to verify that there is a much higher value for the kurtosis and a higher value for the skewness with respect to the ones of other countries' subsamples: this is in line with the values just discussed at the beginning of this section, as the very high value for kurtosis emphasize the presence of outliers, and leads to register a higher[80] mean value and a higher dispersion of observations. Furthermore, the high value for skewness highlights an asymmetry on the right tails, leading to higher results for the arithmetic mean.

In conclusion, the higher value for the U.S.A for a successful variable like the Amount Raised may be explained by a set of few outliers represented by very successful ICOs that have collected a much bigger amount of money than almost the totality of other projects. The presence of very successful ICOs in the U.S. may be however discuss in a more deepened analysis, with a higher focus on the ICOs' geography (e.g. Huang et al., 2018).

### 3.4.3. Subdivision by categories

*Table 5. Statistical analysis for ICOs in financial market*

| Finance | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 7224791 | 5.48E+07 | 1.45E+07 |
| Median | 3000000 | 2.50E+07 | 7000000 |
| std dev | 1.41E+07 | 1.12E+08 | 2.13E+07 |
| var | 1.98E+14 | 1.25E+16 | 4.53E+14 |
| std error | 1243369 | 7897997 | 1343828 |
| cv | 1.947063 | 2.0418 | 1.469076 |
| kurtosis | 23.20083 | 32.66466 | 19.41022 |
| skewness | 4.221076 | 4.989571 | 3.549488 |

*Table 6. Statistical analysis for ICOs in Infrastructure*

| Infrastructure | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 1.65E+07 | 6.34E+07 | 5.88E+07 |
| Median | 3000000 | 2.01E+07 | 1.20E+07 |
| std dev | 5.07E+07 | 1.80E+08 | 4.05E+08 |
| var | 2.57E+15 | 3.23E+16 | 1.64E+17 |
| std error | 8220247 | 2.07E+07 | 3.91E+07 |
| cv | 3.078954 | 2.834121 | 6.88987 |
| kurtosis | 27.43076 | 56.16789 | 103.9457 |
| skewness | 4.915908 | 7.037384 | 10.12075 |

[80] The high kurtosis emphasizes a high value for the arithmetic mean because of the positive skewness, otherwise the high value for the kurtosis, in case of a negative skewness, would have led to a much lower mean value.

*Table 7. Statistical analysis for ICOs in Trading & Investment*

| Trading & Investment | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 1.38E+07 | 5.68E+07 | 1.36E+07 |
| Median | 2500000 | 2.00E+07 | 7240894 |
| std dev | 5.94E+07 | 1.44E+08 | 2.08E+07 |
| var | 3.53E+15 | 2.07E+16 | 4.33E+14 |
| std error | 6600645 | 1.26E+07 | 1614647 |
| cv | 4.307105 | 2.536111 | 1.52552 |
| kurtosis | 58.01636 | 26.9769 | 37.81915 |
| skewness | 7.281635 | 4.816863 | 4.896741 |

In this section ICOs have been divided in subsamples according to the market category on which the projects are going to operate once and if they will be accomplished. The categories included in this section, which are Finance (Table 5), Infrastructure (Table 6) and Trading & Investment (Table 7), have been chosen because of being the most represented categories present in the database's population. The variables and the statistical tools used are the same discussed in the previous cases.

For the Infrastructure sector, with respect to the other categories, it is possible to verify from tables 5, 6 and 7 that there is a higher average value for the variable Amount Raised both for arithmetic mean and for the median, but there is also a quite consistent dispersion of values represented by standard deviation and variance. Again, there is no statistical evidence that the ICOs operating on the Infrastructure sector may have on average a higher possibility to collect more investments than ICOs in other categories, as the great variability of observations affects any consideration that may be taken from these data.

As it was highlighted before for the ICOs from Singapore's subsample, for Trading & Investment is possible to visualize a higher average value for Soft Cap than for Amount Raised: for the 166 ICOs registered in Trading and Investing sector, 81 of them had a Soft Cap set by the company, and for 22 of them the minimum amount to be raised has not been achieved. In any case, the median highlights again a very different trend between Soft Cap and Amount Raised: this may be explained by the very high outlier values that Soft Cap may have due to fluctuations (coefficient of variation, standard deviation and variance), relevant kurtosis and positive skewness.

The higher dispersion of data and the more fluctuations of values in general seem to be more frequent for the subsample Infrastructure, especially for the Amount Raised variable. Another factor that may be explored with a qualitative analysis is that for Finance and Trading & Investing subsamples, which are two segments operating in similar fields, the distributions show very similar characteristics (tables 5 and 7), while for the Infrastructure subsample, which operates in a considerably different segment than the other two, very different values are emphasized from the data (table 6). These differences in the distributions may be so due to the segments where the projects are operating, but again the dispersion of data does not provide the possibility to highlight any clear and relevant statistical conclusion.

The high kurtosis of Infrastructure's Amount Raised variable shows the high probability for this variable to have outliers that may affect the result related to the average value evaluated from the arithmetic mean, specifying why for this subsample the value is so much bigger than for the others. Also, the skewness presents a high value emphasizing the relevancy of the probability distribution's right tail. In conclusion, even with all these subsamples it is possible to verify that in all the cases analyzed the distributions of the variables are leptokurtic with a relevant skewness on the right tail, enforcing the presence of outliers on the right side that may particularly increase the average value of the distribution.

### 3.4.4. Subdivision by protocols (KYC and Whitelist)

*Table 8. Statistical analysis Whitelisted and KYC ICOs.*

| KYC & Whitelist | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 6771469 | 4.23E+07 | 1.44E+07 |
| Median | 3000000 | 2.00E+07 | 8000000 |
| std dev | 1.98E+07 | 1.10E+08 | 3.38E+07 |
| var | 3.92E+14 | 1.21E+16 | 1.14E+15 |
| std error | 1285765 | 6136545 | 1797665 |
| cv | 2.923261 | 2.597498 | 2.342025 |
| kurtosis | 102.2165 | 66.20277 | 218.3627 |
| skewness | 9.078709 | 7.448058 | 13.45966 |

*Table 9. Statistical analysis for KYC ICOs.*

| KYC | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 6708666 | 4.72E+07 | 1.19E+07 |
| Median | 3000000 | 2.00E+07 | 5124000 |
| std dev | 2.38E+07 | 1.23E+08 | 3.16E+07 |
| var | 5.67E+14 | 1.52E+16 | 1.00E+15 |
| std error | 1837486 | 8207887 | 2028549 |
| cv | 3.550115 | 2.61537 | 2.649939 |
| kurtosis | 138.6421 | 89.68448 | 120.2384 |
| skewness | 11.28281 | 8.319995 | 9.900698 |

*Table 10. Statistical analysis for Whitelisted ICOs.*

| Whitelist | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 1981510 | 4.25E+07 | 9555103 |
| Median | 1000000 | 2.00E+07 | 3000000 |
| std dev | 2440146 | 1.25E+08 | 1.30E+07 |
| var | 5.95E+12 | 1.56E+16 | 1.68E+14 |
| std error | 431360.9 | 2.03E+07 | 1890355 |
| cv | 1.231458 | 2.939357 | 1.350364 |
| kurtosis | 7.514047 | 33.59647 | 6.080112 |
| skewness | 2.095083 | 5.61081 | 1.903279 |

*Tables 11. Statistical analysis for non-checked ICOs*

| None | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 7503961 | 5.95E+07 | 1.36E+07 |
| Median | 2084668 | 1.50E+07 | 4274600 |
| std dev | 2.38E+07 | 1.79E+08 | 5.56E+07 |
| var | 5.68E+14 | 3.21E+16 | 3.10E+15 |
| std error | 2360458 | 1.50E+07 | 4136289 |
| cv | 3.176916 | 3.014918 | 4.094942 |
| kurtosis | 46.31425 | 37.73022 | 157.8602 |
| skewness | 6.280036 | 5.563428 | 12.17982 |

In this section ICOs have been divided in subsamples according to the presence or not of KYC and/or Whitelist protocols when investing in the token offering. The projects have been divided in categories according to the following subdivision: projects which present both protocols (Table 8), projects which dispose just of the Know-Your-Customer (KYC) protocol (Table 9), the ones who have only been Whitelisted by the company (Table 10) and projects for which was found a sure information about not using both of the two just highlighted protocols (Table 11). The variables and the statistical tools that have been used in this section are the same used for the previous cases.

There are not very significant differences among the four subsamples looking at the arithmetic mean of the variable Amount Raised, having ICOs with both protocols raised on average a little amount of money more than the others. What can be noticed looking at tables 8, 9, 10 and 11 is that, comparing with table 1, the means of all these 4 subsamples are lower than the one for the total set of data. The high fluctuations of the observations do not allow to provide a clear statistical evidence about being inconvenient for entrepreneurs to provide these protocols (which is what someone could assume looking at the data). However, this result may be analyzed qualitatively rather than quantitatively: investors in fact may be discouraged by these instruments from investing in a coin offering, leading to lower values for the variable Amount Raised. On the other side, the presence of these protocols can give more trust to investors about the reliability of the company and the purpose that the company is pursuing, and on the long term their presence could become an important key factor in order to launch a successful ICO. In any case, looking at the median, values are more similar to the ones listed on Table 1 for the total distribution, and ICOs with both protocols have a higher median with respect to the total population, emphasizing how these differences are not so statistically relevant, giving different results with different statistical tools.

The fluctuations of the distributions seem to be relevant, but in any case, the values are quite in line with the ones of the samples previously highlighted. Furthermore, the kurtosis and the skewness offer results similar to the ones previously listed, presenting, like all the variables and subsamples already analyzed, an asymmetry with a heavier right tail and a leptokurtic characteristic of the distributions. It can be noticed the very high kurtosis value for the subsample of ICOs having both protocols: this information make appear less relevant the information about ICOs with both protocols having on average a higher amount of money collected at the end of the ICO phase than projects with just one of the protocols, since this value is affected by the presence of many outlier values. On the other side, the same relationships among the subsamples are however still supported by looking to the values of the median – not so much affected by outliers - instead of the ones of the arithmetic mean.

A set of ICOs with no protocols adopted has also been included in the dataset, but this subsample may not be represented by the correct number of project in the database: in fact, from the sources used, it was not possible for several ICOs to find out any information about the adoption (or not) of these protocols, not making clear if any protocol was used or not by the company. In other words, for some ICOs where no information about protocols have been listed in the database, it is not possible to state if any protocols was used, but just that the information was missing, and so the corresponding record has not been considered on this subsample. The subsample of ICOs not adopting protocols is populated by 181 Coin Offerings, but other ICOs not having any information about this parameter may also not using any of the two protocols earlier analyzed.

### 3.4.5. Subdivision by platforms

*Table 12. Statistical analysis for Ethereum-based ICOs*

| Ethereum | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 7496585 | 4.70E+07 | 1.71E+07 |
| Median | 2500000 | 2.00E+07 | 7038430 |
| std dev | 2.79E+07 | 1.21E+08 | 1.19E+08 |
| var | 7.79E+14 | 1.46E+16 | 1.42E+16 |
| std error | 1077943 | 3745243 | 3314332 |
| cv | 3.721943 | 2.572791 | 6.981396 |
| kurtosis | 177.0993 | 67.88017 | 1162.822 |
| skewness | 11.88141 | 7.308605 | 33.29867 |

*Table 13. Statistical analysis for Bitcoin-based ICOs*

| Bitcoin | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 7377269 | 5.35E+07 | 4.63E+07 |
| Median | 565950 | 1490413 | 1.20E+07 |
| std dev | 1.18E+07 | 1.04E+08 | 8.38E+07 |
| var | 1.39E+14 | 1.09E+16 | 7.02E+15 |
| std error | 6811365 | 5.22E+07 | 3.17E+07 |
| cv | 1.599187 | 1.951396 | 1.810241 |
| kurtosis | 1.5 | 2.333276 | 4.741813 |
| skewness | 0.7071068 | 1.154627 | 1.864291 |

In this section, the projects collected and listed in the database have been divided in two subsamples according to the platforms that have been used for the token issuance. For this analysis have been chosen the two most popular platforms for cryptocurrencies, i.e. Ethereum and Bitcoin: however, it should be specified that this rough comparison of the two platforms' performances may be highly affected by the numerosity of the ICOs for which it was used the platform. In fact, Ethereum platform has been used by around 1300 of the 1475 ICOs available on the built database, while just few entrepreneurs used the Bitcoin platform in order to issue their tokens. The results obtained for ICOs in the subsection related to the Bitcoin platform may so be affected by the low number of records populating the section, and so giving results which are not so reliable and representative of the entire population of projects whose token have been developed on Bitcoin. Again, the variables and the statistical tools that have been used are the same of the ones previously highlighted.

Looking at the Amount Raised variable (Tables 12 and 13), it looks like for the ICOs developed through the Bitcoin platform it is usually possible to raise more than twice the funds that are normally raised by Ethereum-based Coin Offerings, but besides that it should be taken in consideration that the high fluctuations related to the two variables do not allow to provide any certain result. Furthermore, the population available of ICOs which used Bitcoin platforms does not represent an adequate subsample in order to carry on any statistical evidences. Also looking at the median value for the two subsamples, Bitcoin performance seems to be better than the ones of the Ethereum platform. The variability is considerately high for both the subsections, being higher for the variables related to Ethereum's projects: but if we look at the standard error of the mean, which takes more into consideration also the numerosity of the sample analyzed, the Bitcoin's ICOs subsection seem to be more dispersive than the Ethereum-based ones.

The data of skewness and kurtosis for ICOs developed on Ethereum are quite in line with the ones that were earlier analyzed, while for Bitcoin the distribution seems to be platykurtic for variable Soft and Hard Cap, and very near to a normal[81] distribution's kurtosis for the variable Amount Raised. Also, the skewness can provide analogous results with respect to the ones found out earlier, especially for Ethereum-based ICOs, while for the Bitcoin ones there seems to be a little asymmetry on the right tail, but very limited with respect to the ones highlighted for the variables that were analyzed in the other sections. Again, the results of Bitcoin-based projects should be taken carefully by the reader because of the low ICOs which are populating the subsample analyzed.

### 3.4.6. Subdivision by other variables

*Table 14. Statistical analysis for ICOs using ERC20 standard.*

| ERC20 | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 7722375 | 4.63E+07 | 1.72E+07 |
| Median | 2426109 | 2.00E+07 | 6900000 |
| std dev | 2.92E+07 | 1.21E+08 | 1.29E+08 |
| var | 8.53E+14 | 1.47E+16 | 1.66E+16 |
| std error | 1183273 | 4007009 | 3890355 |
| cv | 3.781313 | 2.618416 | 7.514674 |
| kurtosis | 162.1128 | 71.10761 | 1004.305 |
| skewness | 11.38261 | 7.483794 | 31.0775 |

*Table 15. Analysis for ICOs issuing Utility tokens.*

| Utility | Soft Cap | Hard Cap | Amount Raised |
|---|---|---|---|
| Mean | 7373142 | 4.69E+07 | 1.53E+07 |
| Median | 25000000 | 2.00E+07 | 6772370 |
| std dev | 2.67E+07 | 1.18E+08 | 5.53E+07 |
| var | 7.12E+14 | 1.40E+16 | 3.06E+15 |
| std error | 979200.8 | 3497252 | 1453206 |
| cv | 3.620042 | 2.523664 | 3.628411 |
| kurtosis | 190.9545 | 67.43065 | 623.1878 |
| skewness | 12.27816 | 7.198176 | 22.23875 |

[81] The Gaussian distribution, where the probability density function is represented by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ where $\mu$ is the mean of the distribution and $\sigma$ is its standard deviation. For a normal distribution, the kurtosis is always equal to 3.

In this section ICOs have been subdivided in other 2 subcategories, chosen according to the variables' values which were more representative of the entire dataset. ERC20's information (on Table 14) refers to a standard for Ethereum contracts, and it's in general one of the most adopted standards for smart contracts. Among the projects in the database built, the Coin Offerings using an ERC20 standard contract were around 1000. Other Ethereum standards, such as ERC223, ERC721, ERC23 and ERC77 were "popular" among the projects available in the database, even if not as nearly as the ERC20. Example of non-Ethereum contract standards were NEP5, developed by platform NEO, and Waves, developed by the homonymous platform, but were not included in these subsamples because of the low numerosity of ICOs in the database adopting these types of smart contracts.

The utility sub-segment (Table 15) refers to the type of token issued by the company and the rights that the token can guarantee to its owner: utility token was the type of token more represented in the dataset, since more than 1000 companies were issuing this form of payment to the investors that were allocating their money in the ICO. Variables and statistical tools used are the ones that were analyzed and used for the other descriptive statistics already seen.

The values for the three variables for the arithmetic mean and the median are very similar to the ones of the total dataset (Table 1), which is something we expected since the two subsamples represent a large part of database's population. However, the fluctuations (even if are quite relevant) of the utility token's subsample seem to be a little lower (especially for the Amount Raised variable) with respect to the ones of the total number of records, and with respect to the ERC20 standard contract subsample. Also, the kurtosis and the skewness of the variables' distribution are in line with the ones of the whole dataset, having a considerably asymmetry on the right tail and a large presence of outliers highlighted by the leptokurtic characteristic of the probability distribution. The impact of this type of token and contract on the success of a Coin Offering, rather than being analyzed through descriptive statistics, may be better deepened and explained during the next phase of the work, faced in the next chapter, through a regression analysis.

## 3.5. Conclusions

The outcomes of the descriptive statistics carried on this chapter highlight for the total dataset a probability distribution which is affected by high fluctuations and in general dispersion of values. The variables analyzed do not follow a normal distribution, as it is clear from the asymmetry on the right tail emphasized by the skewness and from the strong probabilities of observations to be on the right tail of the distribution, represented by the leptokurtic characteristic of the probability density function.

The outputs for subsamples analysis provide results quite in line with the ones of the total dataset. The distributions, also for the subsamples, are always characterized by an asymmetry on the right tail and by a large presence of outliers. Estimates for the variables' expected values and fluctuations are also usually in line with the ones of the whole dataset. In some cases, such as for the ICOs developed in the U.S., or the ones operating on the Infrastructure market category, it was possible to speculate about a positive relationship with the Amount Raised variable, but the high dispersion of data usually does not allow to provide any quantitative and statistically relevant outcomes, but just some qualitative considerations.

On the next chapter, through the support of a regression analysis, carried on through the same software – STATA 12 – it will be deepened the impact of some of these variables on the success of an Initial Coin Offering in a different way than the one that has just been conducted on this chapter, i.e. through the development of a regression analysis.

# Chapter IV

## Scope of the Chapter

After having performed the database construction, and, from its construction, having imported the data collected into the software STATA in order to perform the first descriptive statistics about some of the variables included in the database, it is now going to be faced the core section of this work. In this chapter will be conducted a regression analysis through the statistical software STATA in order to check and highlights what are the key variables that may contribute for the success of an Initial Coin Offering.

The analysis will be carried on with the records that have been collected during the second part of this work, analyzed in chapter 2, and with some of the variables included in the original database. In addition, some other variables obtained from the data already collected will be introduced and used to carry on the analysis: in fact, for some variables, some adjustments and some data conversions were needed in order to carry on a more complete and objective analysis.

The chapter at first faces the various data conversions that were needed for the development of the analysis, dividing the information provided according to the different transformation that have been developed. Later, a brief introduction and description of the regression analysis will be given to the reader before deepening the results of the statistical analysis. Then it will be discussed the STATA output in order to highlight and emphasize correlations among variables, significance of the variables included in the regression and the impacts on the Coin Offerings' success. More than one regression has been developed in this phase of the work, modifying step by step the input provided for the regression analysis according to the results that were obtained in the previous steps. In the end of this final chapter of the work the main conclusions that were possible to extract from the analysis' output will be briefly highlighted and discussed.

## 4.1. Linear Regression Analysis: Introduction

In this section of the work is going to be faced the regression analysis, carried on through the software STATA, with the data that were collected in the previous phases of this thesis. The analysis has been developed not including all the variables collected, but just a selection of them, having carefully analyzed which may have been those satisfying the requisites in order to be included in a quantitative analysis.

Some adjustments to these variables were necessary, sometimes because their values were saved in a string format, while sometimes because the distribution of the variable was not allowing to perform a correct analysis. In the first part of the chapter have been highlighted the adjustments made to the variables and the transformations that were needed for the analysis. Generally, three types of data modifications have been done by the writer in this work phase:

- **introduction of dummy variables**: for those variables whose information was difficult to be included in the analysis, but whose data structure was allowing a conversion into a dummy variable according to the information contained by the variables and the records;
- **transformation into a categorical variable**: for those variables having their information available as string and where a finite set of values was possible to be applied to the variables. In these cases, when deciding to adopt such a conversion, the introduction of a dummy variable would have led to the loss of too much information: however, the working process for the categorical variables relies on the introduction of n-1 dummy variables. This approach has been performed for two variables in the dataset, country and category, and for an additional "Continent" variable obtained from Country variable's information,

where the information about the continent in which the project has been issued is saved.

- **Logarithmic conversion**: for the variables that have been included in the descriptive statistics of chapter 3 it has been done a natural logarithmic conversion of the values, in order to better manage their information despite the distribution (with many outliers and a relevant skewness on the right side) that has been highlighted in the previous phase of the work and which would have negatively affected the reliability of the results obtained in the regression analysis.

After this part of the chapter, the linear regression is finally introduced, briefly explaining and describing the whole process structure and the various output section that characterizes this analysis. Later, after having introduced the variables that have been included in the final analysis of the work, the results will be shown, described and discussed in the core section of this chapter. In order to deepen the implications of the obtained results and enforce their reliability, more than one linear regression has been carried on: in fact, changing some of the variables that were included in the process, more than one analysis was discussed by the writer.

This chapter will then conclude with a short discussion of the conclusions that may be highlighted specifically from the development of the linear regression analysis, and about the most important "takeaways" that the results at most can emphasize. After this chapter, at the end of this work, there will however be a dedicated section for the discussion of the results that have been extracted through this thesis, discussing the main outcomes of the whole activity that has been carried out for the development of the work.

## 4.2. Preliminary activities

### 4.2.1. Dummy variables

The regression analysis required multiple data adjustments with the intention to make the information more suitable, and more effectively manageable, for the purpose of the analysis. Indeed, several variables were saved as a string, which may be in some circumstances a crucial point to be handled in order to carry on an adequate regression. In those cases, multiple approaches were allowed in order to allow the information to be used in the analysis: one of these approaches was to introduce and use a set of dummy[82] variables.

To execute a reputable regression, the author decided to introduce a set of dummy variables, with the intention to manage that information which – if left as it were initially – would represent a problem for the developing of the analysis. The selection about which dummy variables to include in the regression have been made also consistently with respect to the past literature about this topic. At the end of the process, 9 dummy variables were introduced for further analysis, but not all of them have been used for the development of the work. The 9 dummy variables introduced were the following:

***Ethereum_dummy***: this dummy variable assumes a value equal to 1 in the case in which the development of the token's ICO has been made through the Ethereum platform. Alternatively, if the company had used any other platform different that Ethereum, the dummy variable assumes a value equal to 0. It is important to observe that for around the 88% of the projects in the database it has been used the Ethereum platform, and so for all of them this variable is equal to 1, whereas for the other projects it assumes a value of 0. The reason behind the inclusion of this variable is to search if there is a correlation between the success of a project and the platform used – especially if there

[82] A dummy variable is a numerical variable that can take just 2 values, specifically 0 and 1. Usually, a value of 1 indicates for the observation the presence of a certain characteristic, while 0 reflects its absence. Dummy variables can be used to statistically represent a categorical variable. To represent a categorical variable that may assume *n* different values, the introduction of *n-1* (the introduction of n variables would in fact lead to a multicollinearity problem) dummy variables is required in order to correctly manage the information.

is a difference in using Ethereum as opposed to an alternative platform in pursuing that success;

***Dummy_Utility***: this dummy variable assumes a value equal to 1 in case the token issued by the company is a utility token, otherwise it will correspond to 0 if any other type of token has been developed. Actually, for almost the totality of the ICOs in the database the developed tokens belonged to the category of utility token. This variable has been introduced in order to highlight a relationship between the utilization of a utility token as investors' reward and the success of a Coin Offering; however, because of the high percentage of records having the same observation, it has been decided not to include this variable in the conducted analysis;

***Restricted_area_dummy***: this variable adopts a value equal to 1 when no restricted area for the ICO was inserted by the token issuers, and equal to 0 when one or more countries were indicated as restricted areas. For 63% of the ICOs in the database have been assigned to the variable a value equal to 0, while for the others no restricted territories were highlighted by the company and so the variable assumed a value of 1. The inclusion of this variable has the goal to draw attention to the presence of the connection between the success of an ICO and the fact that restricted areas may have been indicated or not;

***Dummy_KYC***: this variable assumes a value equal to 1 when the Know Your Customer protocol has been used by the company, alternatively when both the KYC and the Whitelist protocols were used. A value of 0 has been assigned to those ICOs for which no information were available, to those for which just a Whitelist check was used or to those where it was clearly indicated that none of the two protocols were adopted. For the 40% of the ICOs present in the database a KYC protocol has been adopted. The inclusion of this variable has the aim to look for the presence of a relationship between the success of an ICO and the adoption of KYC protocol;

***Dummy_Whitelist***: in a similar way to the variable related to the KYC protocol, this dummy adopts a value of 1 in case investors were Whitelisted and a value of 0 in the opposite case. For around the 27% of the ICOs in the database, investors were

Whitelisted. This variable was included with the intention to look for any correlation between the success of a Coin Offering and the adoption of Whitelist protocols;

***Fiat_dummy***: this variable assumes a value equal to 1 in the case that fiat money has been accepted by the company as investments. In order to extract this information from the recorded data, the Excel functions IF() and SEARCH()[83] have been used, identifying if fiat money were listed between the multiple currencies mentioned in the file. For around the 11% of the ICOs in the database, fiat money was accepted by the company. The integration of this variable has been done in order to identify the presence of any potential relationship between the acceptance of fiat money as investments and the success of an Initial Coin Offering;

***Softcap_dummy***: this dummy adopts a value equal to 1 in case a Soft Cap has been set by the company, and a value equivalent to 0 otherwise. In the database, a Soft Cap has been set for the 51% of ICOs. This variable has been included in order to detect the eventual presence of an association between the success of an ICO and the fact that a Soft Cap was set, independently from its amount;

***Hardcap_dummy***: this dummy is equal to 1 in the case that a Hard Cap has been set by the company, whereas it equals 0 in case it has not been set. The author counted that for the 78% of the ICOs in the database a Hard Cap has been specified. This variable has been introduced in the regression analysis with the object to detect any involvement of the Hard Cap – regardless its amount – for the success of a Coin Offering.

### 4.2.2. Categorical variables

With regards to the other variables included in the database, for the sake of using the information they contain, a conversion was necessary. A linear regression in fact requires quantitative data in order to analyze the information entered. Among the variables in the data set, we are interested in evaluating the impact of the category of

[83] The Excel function SEARCH() allows to find a string inside a cell. With this function it's possible to specify from which point of the cell is better to start to look for the string.

the project and its relative country of development in the ICO's success, i.e. the amount raised. In order to use the information contained in these two variables, it was necessary to convert their information from a string format into a quantitative data in order to be taken into consideration by the software

The string variables have been converted into categorical variables: a categorical variable is considered as such when it can take one value among a fixed number of possible values. This may be obtained by dummy coding, i.e. the software creates a number of dummy variables according to the number of possible values that the variable can take, minus 1. The subtraction by 1 is a result of the multicollinearity effect, which would make part of the data redundant and not useful for the analysis. In practice, the process works in such a way that the dummy variable corresponding to the value of the observation is set to one, while all the others are forced to be 0. The last variable's observation is represented by the situation in which all the new dummies are set to 0[84]: this observation may be considered as a "reference observation", as in the regression output the coefficients of the other ones are related to it.

This approach has been used in order to include in the regression analysis the variables Country and Category of an Initial Coin Offering and evaluate their contribution for the success of a project. On STATA, the conversion of these variables has been done through the command encode()[85].

In addition to the two just highlighted variables, for the purpose of this analysis the author made this conversion to an additional parameter, which was obtained by the country variable, that is "Continents", with the intention to understand the impact of not just the single countries, but more generally of the continent where the ICO has been developed. From the variable "Country" – through the Excel functions IFS[86] and

[84] This is why the inclusion of n dummies instead of n-1 would lead to a multicollinearity problem: one of the n possible cases is already included in the analysis when all the dummies are set to 0. Adding another dummy variable would make the information redundant.
[85] The STATA command encode() allows to convert a string variable into a categorical one (or a factor one, which is a categorical variable allowing both strings and numbers), allowing the variable to be used for the regression
[86] The Excel command IFS() allows to manage a series of IF. It works in a similar way to the IF() function but allows to input a cleaner code

OR[87] – the information about the continent of origin of each of the ICO in the dataset has been included in the dataset. Subsequently, on STATA, the variable has been transformed into a categorical variable in order to include its information in the regression analysis.

### 4.2.3. Logarithm Transformation

In the previous analysis have been considered and examined some of the variables in the database, in accordance to the total list of projects in the dataset and to some other subsamples. The three inspected variables i.e. Amount Raised, Hard Cap and Soft Cap, in all the samples analyzed always showed a leptokurtic characteristic with a skewness on the right, i.e. the analysis emphasized for the variables a large presence of outliers and a consistent asymmetry on the right tail.

With such distributions, the results of the regression analysis that is going to be approached may be highly affected by these "anomalies", going often to consider data with consistent and different orders of magnitude. In such cases it is often approached, in order to better manage these data, a logarithmic transformation of the variables. Transforming the dependent and some of the independent variables through a natural logarithmic conversion leads to two different cases of comparison:

- Logarithmic dependent variable vs independent variable: the comparison among the logarithmic dependent variable and the non-logarithmic independent variable leads to a result where the regression coefficient of the independent variables represents the percentage change of the dependent variable for a unitary variation of the independent one.
- Logarithmic dependent variable vs logarithmic independent variable: the comparison among the logarithmic dependent variable and the non-logarithmic independent variable leads to a result where the regression coefficient of the

[87] The Excel command OR() allows to include a logical test where one among the n conditions given as input have to be satisfied in order the statement to be true

independent variable represents the elasticity among the two variables, i.e. a unitary percentage change of the independent variable leads to a percentage change of the dependent variable equal to the independent variable's regression coefficient.

As it was earlier anticipated, the variables that have been transformed through a natural logarithm conversion were "Soft Cap", "Hard Cap" and "Raised". The conversion has been done through the Excel Functions IF() and LN()[88]

## 4.3. Linear Regression

### 4.3.1. Linear Regression Introduction

The linear regression consists of an analysis where the primary aim is to emphasize the impact or the relationship between a dependent variable and a set of independent variables. The linear regression's principle relies on the minimization of the sum of the squared distances between the observed values of the variables included in the investigation and the values predicted by the regression function representing the analysis' output. Through the observation collected, the regression analysis evaluates a regression function for the dependent variable, where each of the independent variables is included in the function and has a coefficient reflecting the impact of that variable in the dependent one's result.

Before deepening the results of the regression analysis, in order to give to the reader a better comprehension of the results, some of the parameters included in the STATA regression's output section are now going to be described.

| Source | SS | df | MS |
|---|---|---|---|
| Model | 524.877493 | 51 | 10.2917155 |
| Residual | 734.879251 | 544 | 1.35088098 |
| Total | 1259.75674 | 595 | 2.11723822 |

| | |
|---|---|
| Number of obs = | 596 |
| F( 51, 544) = | 7.62 |
| Prob > F = | 0.0000 |
| R-squared = | 0.4166 |
| Adj R-squared = | 0.3620 |
| Root MSE = | 1.1623 |

*Figure 1. First section of STATA's output*

[88] The Excel function LN() allows to convert data saved as number into their natural logarithm.

The output includes a first subdivision between "Model" and "Residual": the Model section represents the distances between the regression function and the observed values that can be explained by the regression function obtained from the analysis, while the Residual section represents those distances which can't be explained by the model. For these 2 output sections there are information about the Sum of Squares (SS), the degrees of freedom (df) and the Mean Squared Errors (MS). The Sum of Squares in the Model section represents the part of variations that can be explained by the regression model, and it's evaluated from the distances between the model's predicted values and the average expected value of the observation. The Sum of Squares for the Residual section represents the part of variation that cannot be explained by the model and it's evaluated from the squared distances between the observed values and the model predicted values. The total Sum of Squares is the sum of these two parameters and can also be evaluated from the squared distances between the observed values and the average expected values of the observations.

The degrees of freedom for the Model is evaluated by the number of regressors that were included in the investigation, while the degrees of freedom for the Residuals are calculated from the total number of observations in the model minus the number of regressors minus 1 (one degree of freedom must be eliminated because of having made estimations from the observed values). The total degrees of freedom voice can be evaluated from the sum of these two numbers and can also be measured from the total number of observations minus 1.

The Mean Squared Errors represents the mean of the sum of squares and it's evaluated as the sum of squares divided by the correspondent degrees of freedom: this stands for the Model, for the Residuals and also for the Total.

Next to the Model and the Residual table, some parameters resulting from the analysis carried out are listed. The F-value represents the F-test, testing the null hypothesis that there is no correspondence between the dependent and the independent variables, i.e. all the coefficients are equal to 0. The F-value is evaluated from the division between the Model MS and the Residual MS or can be found out from the F-distribution tables[89] referring to the degrees of freedom of the Model and the Residual. From this value can be evaluated the probability, which is available immediately under the F-value, that there is no relationship between the dependent variable and the regressors included in the analysis.

The $R^2$ value represents the amount (specifically, it may be considered as the percentage) of variation in the dependent variable that can be explained by the independent variables included in the regression: this value is in fact evaluated as the Model Sum of Squares divided by the total Sum of Squares, and its value is always comprehended from 0 to 1. In the output section there is also an indication about the adjusted $R^2$, which is another form of $R^2$ for regressions having more than one regressor and which takes into account this availability of multiple regressors. It is evaluated as $1 - \frac{n-1}{n-k-1}\frac{Residual\ SS}{Total\ SS}$, where n is the total number of observations in the model and k is the number of regressors. It can be noticed that, in case of a model with a single regressor, the adjusted $R^2$ and the $R^2$ values leads to be the same.

The root MSE value is evaluated from the square root of the Residual's Mean Squared Error.

| lnamount_raised | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| ethereum_dummy | .2768134 | .1653597 | 1.67 | 0.095 | -.0480084 | .6016352 |
| restricted_areadummy | -.198185 | .111192 | -1.78 | 0.075 | -.4166032 | .0202332 |
| dummykyc | -.3697606 | .1254361 | -2.95 | 0.003 | -.616159 | -.1233623 |
| dummywhitelist | .2812255 | .116883 | 2.41 | 0.016 | .0516283 | .5108228 |
| icoprice | .0020549 | .0011476 | 1.79 | 0.074 | -.0001994 | .0043091 |

*Figure 2.Second section of STATA output*

STATA' regression output consists also of a section where for each of the regressors are highlighted its analysis' results. For each of the independent variables, some parameters are available on a table, which are:

- **Coefficient**: which represent the regressor coefficient in the built model, i.e. it reflects the impact of that variable on the analyzed dependent variable. A

[89] The F table shows standardized values related to the F distribution. The F-distribution is a continuous probability distribution used as the null distribution of a statistics test. It finds applications in the analysis of variance (ANOVA) (Mood et al., 1974).

positive coefficient implies that the variable positively affects the independent variable and vice versa. The higher the coefficient's absolute value, the higher the impact of that variable on the dependent one: if the coefficient is near to 0, the variable is not significant for the model;

- **Standard error**: which represents the standard error of the independent variable's coefficient and it is used to evaluate the t-value, dividing the coefficient by the standard error. This value is also used to evaluate the confidence interval;
- **t**: the t-value is obtained by dividing the variable's coefficient by its standard error and it is used to provide the p-value in testing the null hypothesis that the coefficient is equal to 0;
- **P>|t|**: it represents the p-value referred to the null hypothesis that the independent variable's coefficient is equal to 0. The smaller is this value, the higher probabilities there are that the variable's coefficient is different than 0 and so that the variable in the model is significant;
- **[95% Coefficient Interval]**: it represents the right and the left bounds of the 95% coefficient interval, i.e. the interval for which with a 95% confidence level it is possible to find the true value of the coefficient. This value is evaluated from the expected value of the coefficient (the first column of this table) and its standard error. The bigger is the standard error, the wider is the confidence interval and so the higher is the dispersion of data. If 0 is included in the confidence interval, p>|t| would always be greater than 0.05 in a two-sided test.

On the coefficient table is also available a row denominated "_cons" which represents the constant term of the regression function. This parameter represents the intercept of the function, i.e. the value that the dependent variable would assume when all the other variables in the model are set to 0. Of course, the regression function does not provide statistically relevant values at its extreme, and so this information usually is not too much relevant, but its contribute in the correctness of the function is however crucial. For this parameter, as it happens for all the other model's regressors, are available its coefficient, the standard error, the t value and its related probability and the 95% coefficient interval.

### 4.3.2. Regression: variables included

In the final regression analysis carried out it has been included a dependent variable representing the success of an ICO, i.e. the "Amount Raised" variable, converted into a logarithmic scale because of its distribution, and a set of independent variables that are now going to be listed.

The **Ethereum_dummy** variable has been included in the analysis, in order to check the dependency of the ICOs' success on the utilization of Ethereum platform. The **Utility_dummy** variable has been **discarded**, since almost every record in the dataset was issuing a utility token, and so its result would have not been statistically relevant in order to find any relationship evidence. The information about **Country** and **Continents** have been used separately in order to check the geographical impact on ICOs' success through the categorical variable conversion. Also the **category** of the project has been included through a categorical transformation of the variable. The **Restricted area dummy** variable has been included in order to check if the presence of at least one restricted area for the token sale would negatively impact on its success. Dummy variables about **KYC** and **Whitelist** have been included in order to check if their presence would impact positively or negatively on the success of the project. The **ICOprice** variable has also been included to find any long-term relationship with the Coin Offering performances. The **fiat dummy** variable and the **distributed in ICO** variable has also been included in the analysis. For **Soft Cap** and **Hard Cap** it has not be used their original values, nor the dummies related to the presence or not of a Soft or Hard Cap (at least in the first analysis), but it has been used the natural logarithmic transformation of their original value, in order to manage the skewed distribution of the two parameters. Finally, the dummy variable referred to the availability of a technical **Whitepaper** and the variable related to the ICOBench.com's **Rating** of the Initial Coin Offering have been included in the first regression analysis.

### 4.3.3. Regression analysis: first approach

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 524.877493 | 51 | 10.2917155 | Number of obs = | 596 |
| Residual | 734.879251 | 544 | 1.35088098 | F( 51, 544) = | 7.62 |
| | | | | Prob > F = | 0.0000 |
| | | | | R-squared = | 0.4166 |
| | | | | Adj R-squared = | 0.3620 |
| Total | 1259.75674 | 595 | 2.11723822 | Root MSE = | 1.1623 |

| lnamount_raised | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| ethereum_dummy | .2768134 | .1653597 | 1.67 | 0.095 | -.0480084 | .6016352 |
| restricted_areadummy | -.198185 | .111192 | -1.78 | 0.075 | -.4166032 | .0202332 |
| dummykyc | -.3697606 | .1254361 | -2.95 | 0.003 | -.616159 | -.1233623 |
| dummywhitelist | .2812255 | .116883 | 2.41 | 0.016 | .0516283 | .5108228 |
| icoprice | .0020549 | .0011476 | 1.79 | 0.074 | -.0001994 | .0043091 |
| fiat_dummysz | -.0365191 | .1500455 | -0.24 | 0.808 | -.3312587 | .2582204 |
| distributedinico | -.1477518 | .2713633 | -0.54 | 0.586 | -.6808001 | .3852966 |
| lnsoftcap | .3668187 | .0464386 | 7.90 | 0.000 | .2755978 | .4580396 |
| lnhardcap | .3087928 | .0528877 | 5.84 | 0.000 | .2049037 | .4126819 |
| dummy_whitepaper | .156502 | .1968316 | 0.80 | 0.427 | -.230141 | .5431451 |
| rating | .1746671 | .0871072 | 2.01 | 0.045 | .0035594 | .3457748 |
| category2 | | | | | | |
| 1 | .8589456 | .608135 | 1.41 | 0.158 | -.3356349 | 2.053526 |
| 2 | -.403997 | 1.230744 | -0.33 | 0.743 | -2.82159 | 2.013595 |
| 3 | .0956978 | .2637042 | 0.36 | 0.717 | -.4223055 | .6137011 |
| 4 | -.7754745 | .4933141 | -1.57 | 0.117 | -1.744508 | .1935594 |
| 5 | .2174303 | .3453019 | 0.63 | 0.529 | -.460858 | .8957187 |
| 6 | .0023671 | .4978904 | 0.00 | 0.996 | -.9756561 | .9803902 |
| 7 | -.1595901 | .4565658 | -0.35 | 0.727 | -1.056438 | .7372577 |
| 8 | .199307 | .495934 | 0.40 | 0.688 | -.7748731 | 1.173487 |
| 9 | .6108584 | .3174804 | 1.92 | 0.055 | -.0127791 | 1.234496 |
| 10 | .1983626 | .2869343 | 0.69 | 0.490 | -.3652722 | .7619975 |
| 11 | .334895 | .5377123 | 0.62 | 0.534 | -.7213518 | 1.391142 |
| 12 | .0053901 | .3601801 | 0.01 | 0.988 | -.7021239 | .7129042 |
| 13 | -.0510396 | .3451664 | -0.15 | 0.883 | -.7290617 | .6269826 |
| 15 | -.9051338 | .5389158 | -1.68 | 0.094 | -1.963745 | .1534769 |
| 16 | -.2645478 | .303409 | -0.87 | 0.384 | -.8605445 | .3314489 |
| 17 | -.3738445 | .2597817 | -1.44 | 0.151 | -.8841426 | .1364537 |
| 18 | 1.339855 | 1.172324 | 1.14 | 0.254 | -.9629818 | 3.642692 |
| 19 | .0526427 | .4663233 | 0.11 | 0.910 | -.8633723 | .9686576 |
| 20 | -.0629939 | .262385 | -0.24 | 0.810 | -.5784058 | .4524179 |
| 21 | .2571439 | .8380867 | 0.31 | 0.759 | -1.389139 | 1.903426 |
| 22 | .5883966 | 1.182221 | 0.50 | 0.619 | -1.733881 | 2.910674 |
| 23 | -.1650492 | .2672289 | -0.62 | 0.537 | -.6899761 | .3598776 |
| 24 | -.4995077 | .2420149 | -2.06 | 0.039 | -.9749058 | -.0241096 |
| 25 | -.3659109 | .4246915 | -0.86 | 0.389 | -1.200147 | .4683252 |
| 26 | -.0217124 | .2322045 | -0.09 | 0.926 | -.4778396 | .4344148 |
| 27 | -.2116139 | .350744 | -0.60 | 0.547 | -.9005924 | .4773645 |
| 29 | -.2610492 | .3578475 | -0.73 | 0.466 | -.9639814 | .4418829 |
| 30 | -.931803 | .6916196 | -1.35 | 0.178 | -2.290375 | .4267691 |
| 31 | -.1135046 | .2556642 | -0.44 | 0.657 | -.6157145 | .3887054 |
| 32 | .5001158 | .3784254 | 1.32 | 0.187 | -.2432383 | 1.24347 |
| 33 | -.210309 | .1877913 | -1.12 | 0.263 | -.5791938 | .1585759 |
| 34 | -.1231986 | .3761559 | -0.33 | 0.743 | -.8620946 | .6156974 |
| 35 | -.1433076 | .360192 | -0.40 | 0.691 | -.8508451 | .5642299 |
| 36 | -.0184752 | .2908675 | -0.06 | 0.949 | -.5898361 | .5528857 |
| continent2 | | | | | | |
| 1 | -.2962109 | .3656416 | -0.81 | 0.418 | -1.014453 | .4220315 |
| 2 | -.3059015 | .186273 | -1.64 | 0.101 | -.671804 | .0600009 |
| 3 | -.0741666 | .2261526 | -0.33 | 0.743 | -.5184059 | .3700727 |
| 4 | -.3826072 | .1746309 | -2.19 | 0.029 | -.7256408 | -.0395737 |
| 6 | -.3061285 | .3861405 | -0.79 | 0.428 | -1.064638 | .4523806 |
| 7 | -.2748422 | .6294101 | -0.44 | 0.663 | -1.511214 | .9615296 |
| _cons | 4.517423 | .8223574 | 5.49 | 0.000 | 2.902038 | 6.132808 |

*Figure 3. Screenshot from STATA regression output*

The previous screenshot in figure 3 has been taken from STATA's output. The analysis has been carried out through the STATA command regress()[90]. For this analysis it has been used the variable Continents instead of the variable country because the information of the variable Country was too imprecise since there were many countries and most of them with just few observations. For the categorical variables, i.e. Continents and Categories, it has been used, as reference observations, respectively North America and Finance: the first one has been chosen because, in addition to being one of the most represented observations of the variable, it allows to compare the rest of the observations with the U.S.'s one, while Finance has been used because of being the most representative observation of the variable. The two reference observations have been chosen through the command *ib_observationnumber.variablename* (e.g. in the table*, ib5.continent2* and *ib14.category2).*

As it is possible to verify from the first table of STATA's output, the high F-value highlights an elevate probability that there is a general correlation between the dependent and the independent variables, of almost 100%. The number of observations used is 596 because some of the observations in the database had a blank value for the variables that were taken into consideration and so were not considered for the linear regression analysis.

The $R^2$ value highlights the fact that around the 42% of variance was explained by the regression model, while the rest of it it's due to unexplained fluctuations. The adjusted $R^2$, which takes into consideration also the number of regressors that were used in the model, assumes a value which is a little smaller than the $R^2$ value, as it is normal to expect.

Focusing on the regressors and looking to the non-categorical ones (later will be discussed also Continent and Category variables), it's immediately clear that 3 regressors seems to be not in line with the model, because of the high value on the p>|t| column. These regressors are 2 dummy variables, i.e. the one referred to the availability of a

[90] The STATA command regress() allows to carry on a linear regression function with the variables of the saved dataset that have been included in the command.

whitepaper and the one referred to the acceptance of fiat money as investment, and one variable referred to the percentage of issued tokens distributed among investors for the specific project's investment. There are other three variables which highlights a possible, weak relationship with the dependent variable: the dummy variable about the development of the token on the Ethereum platform, the dummy variable related to the presence of any restricted area for the token offering and the variable including the information about the price of the token. As it's possible to see from the table, the variable ICOprice has a coefficient very near to 0, i.e. even if there's any relationship, it does not impact too much on the amount raised and the variable is not significant. However, the coefficient' s sign is positive and so if there is any correlation, this is positive, but the widen of the 95% confidence interval does not allow to make such a statement, at least if considering just p-values which are equal or smaller than the 5%. The other 2 dummy variables also have a 95% confidence interval which goes from a negative minimum to a positive maximum, hiding any statistical evidence about the impact of the two parameters on the ICO's success.

On the other side, there are five variables for which a null hypothesis about the relationship between the independents and the dependent variable would not be rejected with a 95% confidence interval: the variables are the two dummy variables related to the protocols adopted by the company (KYC and Whitelist), the amounts of Hard Cap and Soft Cap (in the analysis their natural logarithm transformations are used) that were set by the company and the rating that has been given by ICOBench.com to the project.

It can be noticed that the impacts of the two protocols are different: the KYC dummy variable has a $|t|=2.97$, which highlights a high probability referred to the relationship with the dependent variable, and exhibits a negative coefficient, i.e. a negative relationship between the success of an ICO and the adoption of KYC protocols is emphasized. On the other side, the Whitelist variable, with a $|t|=2.41$ shows a positive regression coefficient and so a positive relationship with the success of an ICO. The result seems to be that for a company issuing an ICO it may be profitable to whitelist their investors but not to adopt a Know Your Customer protocol. Since the two significant variables are dummies, the coefficient shown in the output section highlights the percentage change that the adoption of the protocol would lead to: the formula $(e^{\beta} - 1) * 100$, where β is the regression coefficient of the variable inspected, represents the percentage change in the independent variable following the activation of the dummy variable.

Considering the two variables related to Soft and Hard Cap, |t| and p highlights the correlation among these two variables and the dependent one, while the coefficient signs emphasize a positive relationship with the success of an ICO. The coefficient is higher for the Soft Cap variable, but however the two values are very near each other and it might be challenging to state which one impacts more on the ICO's success. However, from the output it seems clear that the increase of the Soft Cap and Hard Cap leads to higher amounts raised. This result might be discussed also qualitatively, in fact a company that know that the target investors window is quite large may set larger Caps, while it wouldn't be profitable to set a large Soft Cap if there are no guarantees that the amount would be reached. In conclusion, it might not be that the Soft Cap and the Hard Cap make the amount raised increasing, but it may be that the companies set them according to the amount expected to be collected. Considering the impact of these variables according to the model, since the two variables are considered through two logarithmic transformations, the two regression coefficients represents the elasticity of the dependent variable with respect to the independent, i.e. considering a β coefficient, a 1% increase in the independent variable would lead to a β% increase in the dependent variable.

In conclusion, considering the rating variable, it is emphasized a positive relationship among the success of an ICO and the rating given by ICOBench.com, highlighting how the judgement by the platform's algorithm and experts is reliable. However, since this rating can change over time, the result can be explained qualitatively also in another way, i.e. that it is the success of a project which leads to have a high ICOBench.com rating, and not vice-versa.

Talking about the two categorical variables, i.e. continent and category, the two reference observations that were taken are the North America for the first one and Finance for the second one. The STATA output shows, for each of the variables, the relationship of each observation with respect to the reference one, with the coefficient emphasizing if it would be more (or less) profitable to develop the ICO with that observation value with respect to the reference one. For the continents variable it can be shown that it would be less profitable to develop in any other continent than North America, but a clear statistical evidence can be found just for the Europe observation (fourth observation), while a weaker result that may however be considered relevant is with respect to the Asia continent (second observation), having a p-value which is around 5%.

For the Category variable also, there are not too much statistical evidences: coefficients are usually negative, and so in most of the cases other categories would be less profitable than the reference one, i.e. Finance, but the 95% confidence interval are usually too wide to emphasize such a statement. However, for some variables, some conclusions may be obtained: the data storage category (ninth observation) seems to be more profitable with respect to Finance, while Food & Beverage has a good probability of being less profitable than the reference variable. Finally, there is an evident less profitability for the variable Marketplace (represented by the twenty-fourth observation) highlighted from the negative coefficient and the strict 95% confidence interval shown in the STATA output. It can be noticed that, for the observation Unknown (last observation, thirty-sixth), where projects where no categories were indicated by the company are collected, any statistically evident result cannot be found.

From this regression output have been made some adjustments in order to improve the results and develop another one. In fact, with these variables, the observation included in the analysis are much less with respect to the total ones that have been added to the database (596 record included in the analysis, with respect to the 1475 observations that have been included in the final database), especially because Soft and Hard Cap are parameters not mandatory to be specified for the company when launching an ICO, and so for many projects there is no information regarding these variables. This leads to have less records included in the analysis because the projects where those parameters were not set by the entrepreneurs were excluded from the analysis. The following step will so be to not include the variable related to the amount of Soft and Hard Cap set by the company, but just the 2 dummy variables that are going to assume a value of 1 when, respectively, the Hard and Soft Cap were set by the company and 0 when those parameters were not considered by the companies.

Also the variable “Distributed in ICO” will be excluded from the next regression analysis because of two reason: the first one is that even this variable is available for not too much ICOs in the database, while the second one is that the regression analysis just carried out emphasized that there is no relationship between this parameter and the success of a Coin Offering.

Finally, the rating variable will also be excluded because of the reasons that were earlier emphasized about this variable not being a good indicator for the success of an ICO even if the regression analysis highlights a positive relationship between the rating and the ICOs’ success.

## 4.3.4. Regression analysis: second output

| Source | SS | df | MS |
|---|---|---|---|
| Model | 332.905153 | 50 | 6.65810307 |
| Residual | 3211.73322 | 1337 | 2.40219388 |
| Total | 3544.63838 | 1387 | 2.55561527 |

| | | |
|---|---|---|
| Number of obs | = | 1388 |
| F( 50, 1337) | = | 2.77 |
| Prob > F | = | 0.0000 |
| R-squared | = | 0.0939 |
| Adj R-squared | = | 0.0600 |
| Root MSE | = | 1.5499 |

| lnamount_raised | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| ethereum_dummy | .2093519 | .1346759 | 1.55 | 0.120 | -.0548473 | .473551 |
| restricted_areadummy | .0262695 | .1078176 | 0.24 | 0.808 | -.1852406 | .2377795 |
| dummykyc | -.0699817 | .1205056 | -0.58 | 0.562 | -.3063823 | .1664188 |
| dummywhitelist | .2118311 | .1182598 | 1.79 | 0.073 | -.0201639 | .4438262 |
| icoprice | .0002397 | .000167 | 1.44 | 0.151 | -.0000879 | .0005674 |
| fiat_dummysz | .2612644 | .1365853 | 1.91 | 0.056 | -.0066805 | .5292094 |
| softcap_dummy | -.2504355 | .0972765 | -2.57 | 0.010 | -.4412667 | -.0596044 |
| hardcap_dummy | -.0487076 | .122505 | -0.40 | 0.691 | -.2890305 | .1916153 |
| dummy_whitepaper | .5216068 | .1608419 | 3.24 | 0.001 | .2060768 | .8371368 |
| category2 | | | | | | |
| 1 | -.4088101 | .4600377 | -0.89 | 0.374 | -1.311284 | .4936643 |
| 2 | .2912545 | .9071592 | 0.32 | 0.748 | -1.488356 | 2.070865 |
| 3 | .1214754 | .2302784 | 0.53 | 0.598 | -.330271 | .5732218 |
| 4 | -.483266 | .4441565 | -1.09 | 0.277 | -1.354586 | .3880536 |
| 5 | .5865772 | .2824384 | 2.08 | 0.038 | .0325066 | 1.140648 |
| 6 | .0102984 | .443405 | 0.02 | 0.981 | -.8595468 | .8801436 |
| 7 | -.1372008 | .3915264 | -0.35 | 0.726 | -.9052737 | .6308721 |
| 8 | -.0115581 | .4799973 | -0.02 | 0.981 | -.953188 | .9300718 |
| 9 | .5768344 | .3476565 | 1.66 | 0.097 | -.1051772 | 1.258846 |
| 10 | .1205659 | .2975728 | 0.41 | 0.685 | -.4631945 | .7043264 |
| 11 | -.3353306 | .4609287 | -0.73 | 0.467 | -1.239553 | .5688916 |
| 12 | .2841335 | .3714521 | 0.76 | 0.444 | -.4445588 | 1.012826 |
| 13 | -.3971757 | .306583 | -1.30 | 0.195 | -.9986118 | .2042604 |
| 15 | -.4623032 | .5593941 | -0.83 | 0.409 | -1.559689 | .6350826 |
| 16 | -.6418924 | .2536427 | -2.53 | 0.011 | -1.139473 | -.1443114 |
| 17 | -.3869881 | .2284987 | -1.69 | 0.091 | -.8352432 | .061267 |
| 18 | 1.256235 | .9050857 | 1.39 | 0.165 | -.5193074 | 3.031778 |
| 19 | -.3136143 | .4336535 | -0.72 | 0.470 | -1.16433 | .537101 |
| 20 | .4381586 | .1878951 | 2.33 | 0.020 | .0695572 | .8067599 |
| 21 | -.0041131 | .7872421 | -0.01 | 0.996 | -1.548477 | 1.540251 |
| 22 | .3796654 | .9030272 | 0.42 | 0.674 | -1.391839 | 2.15117 |
| 23 | .047725 | .2699786 | 0.18 | 0.860 | -.4819028 | .5773528 |
| 24 | -.2470225 | .2236187 | -1.10 | 0.270 | -.6857043 | .1916592 |
| 25 | -.1552269 | .3963784 | -0.39 | 0.695 | -.9328183 | .6223645 |
| 26 | -.0994692 | .1961485 | -0.51 | 0.612 | -.4842615 | .2853232 |
| 27 | .2317634 | .3288608 | 0.70 | 0.481 | -.4133759 | .8769027 |
| 28 | -.6154988 | 1.102994 | -0.56 | 0.577 | -2.779287 | 1.548289 |
| 29 | -.7275605 | .3439225 | -2.12 | 0.035 | -1.402247 | -.0528739 |
| 30 | -1.092133 | .528909 | -2.06 | 0.039 | -2.129715 | -.0545508 |
| 31 | -.2149141 | .2514151 | -0.85 | 0.393 | -.7081252 | .278297 |
| 32 | .2461914 | .3402486 | 0.72 | 0.469 | -.4212879 | .9136707 |
| 33 | .027087 | .1594375 | 0.17 | 0.865 | -.285688 | .339862 |
| 34 | -.3986842 | .3630138 | -1.10 | 0.272 | -1.110823 | .3134544 |
| 35 | -.3687693 | .3912929 | -0.94 | 0.346 | -1.136384 | .3988456 |
| 36 | -.7160797 | .1795366 | -3.99 | 0.000 | -1.068284 | -.3638755 |
| continent2 | | | | | | |
| 1 | -.6773164 | .3044244 | -2.22 | 0.026 | -1.274518 | -.0801149 |
| 2 | -.2139626 | .1426154 | -1.50 | 0.134 | -.493737 | .0658118 |
| 3 | .0806804 | .1914188 | 0.42 | 0.673 | -.2948335 | .4561943 |
| 4 | -.4939125 | .1299566 | -3.80 | 0.000 | -.7488535 | -.2389714 |
| 6 | -.3485935 | .3075281 | -1.13 | 0.257 | -.9518837 | .2546967 |
| 7 | -.8807789 | .3716229 | -2.37 | 0.018 | -1.609806 | -.1517514 |
| _cons | 15.35044 | .2734032 | 56.15 | 0.000 | 14.81409 | 15.88678 |

*Figure 4 STATA regression output*

The screenshot shown in figure 4 represents the STATA output for the linear regression carried out according to the criteria earlier specified. With respect to the other STATA output discussed, in this case the number of observations is way bigger (1388 observations included, with respect to the 596 included in the first analysis and the 1475 available in the database). It is clear from the F-value that there is a dependency between the dependent variable and the independent ones. What can be also noticed unfortunately is that the $R^2$ and the Adjusted $R^2$ are just, respectively, around the 9% and 6%: this may be a consequence to the fact that some continuous variables have been replaced for a set of dummy variables, and now most of the variables included in the analysis are dummies. Dummy variables may be helpful in many circumstances when having to manage information saved in string formats but, like in this case, they may not be effective in explaining a large part of the variance of the dataset since the variables can assume just values of 0 and 1.

With respect to the analysis carried out earlier, the Ethereum dummy variable has a coefficient a little lower than before, but its p-value does still get a similar value, while for the restricted area dummy variable the results are completely different: while earlier the parameter had a negative coefficient and its p-value assumed a value which was relatively small (around 7.5%), now the variable has a positive coefficient (which is also counterintuitive, since it means that the presence of any restricted area would increase the ICOs' success) and its confidence interval is very large, emphasizing the fact that the variable does not have any impact on ICOs' success or, better to say, that the model's output does not allow to state that the variable has any involvement in determining the ICO's success.

Talking about protocols adopted by the companies, the results highlighted by the output for Whitelisted ICOs are still in line with the ones obtained previously, while for Know Your Customer protocol, even if the coefficient still assumes a negative value, its variability drastically increases, hiding any evidence about the variable being correlated with respect to the projects' success.

The ICO price's output seems also to be similar to the previous one if looking just at the coefficient, but its variability increased, leading to a widener confidence interval. There are two variables that passed from being not significant in the previous model to have a much higher impact in this one: one of them is the parameter related to the acceptance of fiat money, which became quite significant, positively affecting the Coin Offerings' success, even if a small part of its 95% confidence interval also comprise negative values, and so the 95% hypothesis test about the variable being correlated with the ICO's success would not be accepted. The other variable which passed from being not significant to highlight an impact on the amount raised by the company is the one related to the availability of a whitepaper, which leads to be the variable which contributes more to the projects' success, as it is emphasized by the high coefficient value that this parameter assumes. The whitepaper availability variable also presents a very low variability of its coefficient, leading to a strict confidence interval and a significant result regarding the relationship with the ICO's success. This results may seem too much distant with the ones earlier obtained, but what should be considered is that in this analysis more than twice the observation have been considered, and that two significant variables such as the amounts of Soft Cap and Hard Cap have been removed: the consequence of those activities are that for some variables it is normal to find results very different with respect to the ones that were shown in the previous subchapter.

Talking about the Hard and the Soft Cap, the model leads to a result where the presence of both these two parameters negatively impact the success of a project, but while for the Soft Cap the variability of the coefficient is small, and so the regressor is significant, for the Hard Cap the coefficient's variability is way higher and the 95% confidence interval may have both positive and negative values, leading the variable not to be significant for the model's purposes. Looking at the value assumed by the Soft Cap, an interpretation may be that, when setting a soft the ICO's profitability decreases, but once the Soft Cap has been set, the higher it is this parameter the more amount of money the project is going to raise. Analogous implications may not be pointed out for the Hard Cap because of not being the parameter significative.

Shifting the focus on the categorical variables, the trends of the previous analysis are quite confirmed, but the variability of the regressors in many cases decreased because of the more availability of observations. For the Category variable, Communications (the fifth observation) and Infrastructure (the twentieth observation) emphasized a positive trend with respect to the reference observation (which still is Finance) and provides a coefficient variability which makes the information significant. On the other side, looking at the observations Gambling & Betting (the sixteenth observation), Real Estate (the twenty-ninth observation) and Recruitment (the thirtieth observation) there is a negative trend with respect to the reference observation, having a coefficient variability which makes the information about these records significant. Other observations have a p-value which is between the 5% and 10% and may be taken into consideration as significant records. Lastly, the thirty-sixth observation represent the "Unknown" observations, i.e. those records for which there was no information about the category of the project: for this observation, it is emphasized a strong negative trend and a p-value equal to 0, supporting the reliability of the information. This last outcome does make sense also in a qualitative reasoning, as it is reasonable to think that those projects for which the company did not make very clear in which market category they are going to operate will be able to raise an amount of money usually lower than those projects for which this statement has been clearly made.

Considering Continents, the Center and South America (both collected together in the third observation) seems to be more profitable than the reference observation (which still is North America) but the coefficient's variance is too high to evaluate any statistical evidence. All the other observations have a negative trend with respect to North America, reflected by the negative coefficient, and among these for Africa (first observation) and Europe (fourth observation) the results obtained are significant. The last observation represents those records for which there was no information about the country of ICO's development: even for this set of projects is emphasized the strong negative profitability with respect to the ICOs developed in North America, highlighted by high coefficient value and the low p-value which characterizes the record.

Because of the low $R^2$ value assumed by the model and other of its characteristics, a last regression analysis will be carried on making some further modifications to the variables included in the model.

## 4.3.5. Regression Analysis: Final Results

| Source | SS | df | MS |
|---|---|---|---|
| Model | 656.360713 | 50 | 13.1272143 |
| Residual | 1864.19563 | 1075 | 1.73413547 |
| Total | 2520.55635 | 1125 | 2.24049453 |

| | |
|---|---|
| Number of obs = | 1126 |
| F( 50, 1075) = | 7.57 |
| Prob > F = | 0.0000 |
| R-squared = | 0.2604 |
| Adj R-squared = | 0.2260 |
| Root MSE = | 1.3169 |

| lnamount_raised | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| ethereum_dummy | .031065 | .1371881 | 0.23 | 0.821 | -.2381219 | .3002518 |
| restricted_areadummy | -.0495935 | .0984763 | -0.50 | 0.615 | -.2428211 | .1436341 |
| dummykyc | -.0973161 | .108881 | -0.89 | 0.372 | -.3109595 | .1163273 |
| dummywhitelist | .2212207 | .1066909 | 2.07 | 0.038 | .0118747 | .4305667 |
| icoprice | .0002399 | .0001653 | 1.45 | 0.147 | -.0000845 | .0005643 |
| fiat_dummysz | .0513127 | .1269547 | 0.40 | 0.686 | -.1977943 | .3004197 |
| softcap_dummy | -.1576965 | .085946 | -1.83 | 0.067 | -.3263373 | .0109444 |
| lnhardcap | .4859215 | .0309662 | 15.69 | 0.000 | .4251605 | .5466824 |
| dummy_whitepaper | .4555465 | .1572701 | 2.90 | 0.004 | .1469553 | .7641377 |
| category2 | | | | | | |
| 1 | .3562955 | .4111078 | 0.87 | 0.386 | -.4503691 | 1.16296 |
| 2 | -.2473418 | .7736099 | -0.32 | 0.749 | -1.765298 | 1.270615 |
| 3 | .2146326 | .2159733 | 0.99 | 0.321 | -.2091444 | .6384095 |
| 4 | -.5473963 | .4534311 | -1.21 | 0.228 | -1.437107 | .342314 |
| 5 | .1237808 | .2719808 | 0.46 | 0.649 | -.4098928 | .6574543 |
| 6 | .0684267 | .4105062 | 0.17 | 0.868 | -.7370576 | .8739111 |
| 7 | .0260151 | .3788933 | 0.07 | 0.945 | -.7174392 | .7694694 |
| 8 | .261965 | .4292946 | 0.61 | 0.542 | -.5803854 | 1.104315 |
| 9 | .4588726 | .3112948 | 1.47 | 0.141 | -.1519416 | 1.069687 |
| 10 | .1942718 | .2765256 | 0.70 | 0.482 | -.3483194 | .7368629 |
| 11 | -.1473799 | .4100912 | -0.36 | 0.719 | -.9520498 | .65729 |
| 12 | -.2330712 | .345383 | -0.67 | 0.500 | -.9107725 | .4446301 |
| 13 | -.4064077 | .2919682 | -1.39 | 0.164 | -.9792998 | .1664845 |
| 15 | -.6636122 | .5484521 | -1.21 | 0.227 | -1.73977 | .4125459 |
| 16 | -.4672626 | .2429688 | -1.92 | 0.055 | -.9440094 | .0094843 |
| 17 | -.3454949 | .2135874 | -1.62 | 0.106 | -.7645905 | .0736006 |
| 18 | -.6174139 | .9407341 | -0.66 | 0.512 | -2.463297 | 1.228469 |
| 19 | -.2615817 | .4149374 | -0.63 | 0.529 | -1.075761 | .5525973 |
| 20 | .415175 | .1835613 | 2.26 | 0.024 | .0549959 | .7753541 |
| 21 | .4209297 | .7714013 | 0.55 | 0.585 | -1.092693 | 1.934553 |
| 22 | .2278159 | .7685633 | 0.30 | 0.767 | -1.280238 | 1.73587 |
| 23 | -.0333424 | .2467154 | -0.14 | 0.893 | -.5174407 | .450756 |
| 24 | -.262493 | .199244 | -1.32 | 0.188 | -.6534442 | .1284583 |
| 25 | -.224923 | .3699733 | -0.61 | 0.543 | -.9508747 | .5010287 |
| 26 | -.114083 | .1827972 | -0.62 | 0.533 | -.4727628 | .2445967 |
| 27 | .0402771 | .320494 | 0.13 | 0.900 | -.5885876 | .6691419 |
| 28 | -.5413531 | .9382399 | -0.58 | 0.564 | -2.382342 | 1.299636 |
| 29 | -.5823905 | .3319794 | -1.75 | 0.080 | -1.233791 | .0690105 |
| 30 | -1.037494 | .6008958 | -1.73 | 0.085 | -2.216556 | .1415674 |
| 31 | -.1691934 | .228188 | -0.74 | 0.459 | -.6169377 | .2785509 |
| 32 | .0996824 | .3112883 | 0.32 | 0.749 | -.5111192 | .710484 |
| 33 | .0628617 | .1508053 | 0.42 | 0.677 | -.2330444 | .3587679 |
| 34 | -.4532765 | .3111383 | -1.46 | 0.145 | -1.063784 | .1572308 |
| 35 | -.1000264 | .3681938 | -0.27 | 0.786 | -.8224864 | .6224335 |
| 36 | -.4024091 | .1784568 | -2.25 | 0.024 | -.7525724 | -.0522459 |
| continent2 | | | | | | |
| 1 | -.4404061 | .2852515 | -1.54 | 0.123 | -1.000119 | .1193067 |
| 2 | -.0314716 | .1375597 | -0.23 | 0.819 | -.3013876 | .2384444 |
| 3 | -.0468223 | .1809244 | -0.26 | 0.796 | -.4018274 | .3081828 |
| 4 | -.3891985 | .1259551 | -3.09 | 0.002 | -.6363442 | -.1420528 |
| 6 | -.3568309 | .2946166 | -1.21 | 0.226 | -.9349197 | .2212578 |
| 7 | .104363 | .4194473 | 0.25 | 0.804 | -.7186652 | .9273912 |
| cons | 7.279593 | .5886246 | 12.37 | 0.000 | 6.124609 | 8.434576 |

*Figure 5. Final regression analysis*

In the previous table (Figure 5) the regression analysis' variables included in the model are the same of the previous one, except for the variable considering the amount set by the company as Hard Cap, in his logarithmic transformation. This substitution has been made since the dummy variable considering the presence of hard cap was not significant, but from the first analysis this variable highlighted a relationship between the amount of Hard Cap set and the ICO's success. For the variable Soft Cap was not applied this modification because the Soft Cap dummy variable was already considered in the first analysis and because of the few projects for which a Soft Cap was set by the company. In addition, Soft Cap's impact on the ICOs' success was already verified also by the dummy variable introduced in the previous analysis.

Looking at the STATA output, it can be noticed that the number of observations decreased, since setting a Hard Cap is not mandatory for the company and the information of projects not having set a Hard Cap so has been not used for this analysis. The $R^2$ and the adjusted $R^2$ increased, since a dummy variable was substituted with a continuous variable, better explaining the model.

Focusing on the regressors, the dummy variables related to the utilization of the Ethereum platform, the presence of any restricted area and the adoption of KYC protocol were not significant in the earlier analysis and still appears not to be in this one, while the Whitelist dummy variable is now significant with a p-value which is around 4%, and with a positive relationship with the ICOs' success. The ICOs' price variable still in this analysis results not to be significant, while the fiat dummy variable passed from a significant dependency to a non-significant one. For the Soft Cap dummy variable there seems to be a negative relationship with respect to the ICOs' success, but the hypothesis can't be accepted with a 95% probability as the p-value is equal to 6.7%. On the other side, for the Hard Cap parameter there is a clear and positive relationship with the ICOs' success, and the confidence interval for the variable's coefficient is relatively small: being the Hard Cap variable in its logarithmic transformation, and being in its logarithmic transformation also the Amount Raised variable, the coefficient represents the Amount Raised's elasticity with respect to the Hard Cap. The whitepaper presence's dummy variable is again significative with similar parameters to the one of the previous analysis, and it can’t be rejected the null hypothesis which considers a relationship between the variable and the ICO’s success.

For the categorical variables cannot be highlighted too much relationships among the observations: most of the variables highlights p-values greater than 5%, and so a 95% confidence interval with both positive and negative values. For the Category variable the results are quite similar with the ones that have been found earlier, with the observations without a category indication highlighting a negative relationship with the reference one, and the other significant relationships which more or less still assume the same values. For the continent variable, as in the previous analysis, almost all the relationship with respect to the reference observation (North America) are negative. The coefficient related to the seventh observation, which represent projects for which the country was not specified, should not be taken into consideration too much, since the highlighted coefficient is a rough estimation of the observation’s impact on ICO’s success because of the wide confidence interval. The only relationship that appears to be statistically significant is the one with the Europe (fourth observation) for which a negative correspondence is again confirmed.

## 4.4. Conclusions

The analysis performed in this chapter has been carried out in order to detect which may have been the most important success factors for a company which is launching a project and it is collecting money through an Initial Coin Offering. The data that were collected during the database construction allowed to involve in the investigation several parameters concerning the development of a token offering. Among these attributes, according to the analysis developed in this part of the work, some of them resulted to be key factors for the Coin Offerings success while others appeared not to be impacting the ICOs’ success too much.

The utilization of the Ethereum platform in general did not seem to be an ICOs’ success factors from the results obtained: this may be discussed also qualitatively since many projects and many ICOs develop their token on the Ethereum platform, and so, even if the Ethereum-based characteristic of the ICO would have been important, the fact that most of the projects are developed in such a way provides less importance to this parameter. Also, the presence of at least one restricted area, according to the analysis, never seemed not to impact too much on the success of the crowdfunding process.

Regarding protocols, from the three regressions analyzed it seems clear that Whitelisting a project leads to a higher Amount Raised, while for the Know Your Customer protocol is not clear if there is any dependence, but, if any exists, the adoption of this protocol would probably lead to a decrease on the Amount Raised by the company. In other words, the KYC protocol's adoption seems to be unprofitable for the company, while the adoption of a Whitelist one is a profitable factor.

The acceptance of fiat money and the price of the token during the sale phase do not seem to have an impact on the ICOs' success, and so it is for the percentage of issued tokens distributed among investors for their contribution.

Talking about Hard and Soft Cap, what was emphasized by the regression outputs is that, increasing the two parameters, also the Amount Raised by the company will increase. On the other side, at least for the Soft Cap, it seemed that setting a Soft Cap leads to decrease the ICOs' success, while it's not possible to find such evidence for the Hard Cap (which is something we may have expected, since an Hard Cap usually is set and so it is available more often than a Soft Cap). The quantitative results obtained for these two parameters are in line with the qualitative predictions that may have been done.

About the availability of a Whitepaper, even if in the first regression the variable was not significative, usually the results showed the relevant impact of this factor towards the success of a Coin Offering. Even in this case the results are rational with respect to a qualitative approach, since the Whitepaper is a document describing key factors of the projects, the Coin issuing mechanism and many other important details, and its availability and contents are crucial for an investor when deciding if allocating its money in the project or not.

As already discussed, the ICOBench.com’s rating is quite in line with the ICO’s success, as reported by the STATA’s output, but the relationship among the two variables may be inverse, i.e. the ex post assignment of the rating is heavily dependent on ICOs’ success or not, and so the rating information collected is not useful for this analysis. The information would have been more useful if the ex-ante rating from ICOBench.com was available for the analysis, in order to establish if the rating specified by the web site before the starting of the crowdfunding phase was a good indicator or not.

Lastly, speaking about the categorical variables, the impacts of market category and continent of development were not clear for all the observations collected and analyzed, but just for some of them. What it seemed from the regression output is that usually the North America is the continent for which a higher profitability during the crowdfunding phase may be reached. Instead, for the category variable, such a constant trend among the observations was not emphasized by the analysis, in fact just few relevant relationships among the observation were found out to be significative. What may have impacted the results for the Category variable are the high numbers of category available, which led to have few observations for most of the categories and so less significative results.

# Final Conclusions and Possible Future Developments

The principal aim of this work was to build an Initial Coin Offerings' database in order to use it to carry on some descriptive statistics and empirical analysis regarding the identification of the most important ICOs' success factors. The database construction represented one of the most challenging part of the work: among the several online sources available on the Internet, it was necessary to assess which ones were the most reliable and complete of useful information that may have been involved in an empirical analysis. The choice fell on two main sources, which are ICOBench.com and Coinschedule.com. The information from both the online sources have been used to construct the ICOs' dataset, with ICOBench.com's information going to fill the database with those data that were not included on the initial Coinschedule.com's database population. In some cases, the information available from the two platforms about the same field were different, and the information from Coinschedule.com has been privileged because of the accuracy of its data and the more availability of them. For further analysis regarding this subject the author would suggest a database enlargement, going to consider all the ICOs available on ICOBench.com (more than 5000) instead of just the ones available on both the two platforms. The database's expansion would allow to carry on a more precise analysis, with a greater availability of data and information, and with a sample that represents more effectively the total Coin Offerings population.

The variables that were collected for each observation were a total of 26, but some of them were not used because of the low availability of some information: an useful and interesting deepening of the work done may be to include in the regression analysis also the data regarding the amount of tokens issued and sold by the company, in order to check for any correlation among the two variables – and with other variables – or to verify their impact on the ICO's success. The two variables were included in the database, but the few availabilities of their information did not allow to use them effectively for the analysis. The results may have emphasized if it is more profitable to issue many tokens at a low price rather than less tokens with a higher price, or vice-versa.

Other further analysis that may be interesting to carry on may be related to general post-ICO parameters, e.g. the market value of the issued tokens over time, in order to assess the success of an ICO not only at the end of the crowdfunding phase, according to the total amount of money that has been raised, but also during the final development of the project over time.

Regarding the regression analysis, the results emphasized a clear dependency of some of the variables collected on the success of the Initial Coin Offering: those variables were mainly related to: financial parameters set by the company, such as Hard and Soft Cap, to the accessibility of information which may be represented by the availability of a public technical white paper, to the adoption of protocols with the aim to protect company's entrepreneurs and investors and to some other parameters. These results highlighted especially the importance, for the success of an ICO, of setting a Hard and Soft Cap, illustrating also the trend of the amount raised with respect to those parameters, and also the importance on issuing a technical whitepaper or to adopt a Whitelist protocol rather than a Know Your Customer one. Other variables in the analysis led to be non-significative according to the results, such as the acceptance of fiat money as investment or the development of the token on the Ethereum platform. According to this last point, many records used in the regression were developed on an Ethereum-based environment, leading inevitably to a less significance of this factor: an interesting future development may be related to the inclusion of more projects whose token has not been developed on the Ethereum platform, to better assess the impact of its utilization with respect to the token offering's success.

Another further development of the analysis, that may strengthen the results obtained, is to carry on a non-linear regression where the following are included: the dummy variables related to either setting or not a Soft and a Hard Cap, and the variables related to their respective amounts that were set by the company for these two parameters.

The aforementioned factors would be multiplied together (i.e. the Soft Cap dummy multiplied by the amount of Soft Cap that was set, and the same applies to the Hard Cap), and as a consequence the resulting variables would assume a value of 0 when the parameter was omitted and a value equal to its total amount (or a logarithm transformation) when it was set. The non-linear model would allow on the same analysis to assess both the impact of the presence of the parameter and its amount: the analysis just conducted in fact emphasized for the Soft Cap that its presence is a negative factor for the ICOs' success, but when the parameter has been already set, increasing its value would lead to an increase in the amount of money collected by the company. However, in a linear model as the one developed in this thesis, the impact of the two aspects has been analyzed separately, while a multiple regression, examining the two factors in the same model, would lead to have more reliable and complete results about this topic.

The investigations regarding the impact of the projects' market category and development country did not emphasize too many evidences: for the development country it was very often explicated that developing the ICO in the North America would lead to a more profitable fund raising, but even if this trend was highlighted by the STATA output, few times the parameters allowed to give significance to these results. The explanation of this phenomenon may be that even if there may be a more profitability in developing the ICO in the U.S., this increase on the amount raised by the company is not too much evident and so it does not affect too much the projects' success in most of the cases. Regarding the Category variable, the outputs did emphasize few significant results, but quite consistent (changing the regression carried on, the results did not change too much). One of the reasons of the low significance of many observations may have been the presence of many categories that were related with few records among the whole dataset, leading to have non-significant statistical relationships among the observations analyzed. A future improvement regarding this analysis may be to collect some of these categories in few classes of categories, each of them having more observations than the single categories. It is a process very similar to the one used for the development countries, which have been grouped according to the countries' continent. In this way, it would be possible to have more significant information regarding some specifically selected macro-areas, rather than collecting many non-significant results about a large number of categories.

In conclusion, the work that has been developed led to the construction of a reliable ICOs' dataset, complete of information, and the analysis carried on through the collected data emphasized quantitative results which were in line with considerations based on a qualitative approach. It was possible to detect some of the main success factors for Initial Coin Offerings and their impact over the amount raised by the company. Further deepening and more detailed analysis may be possible including more observations (i.e. more Coin Offerings) in the dataset used for the analysis, more variables if there is a large availability of some information (e.g. data about issued tokens, pre-sale and post-sale parameters) and carrying on a more complete non-linear regression in order to better detect the impact of multiple factors, which are affecting each other's, together.

# References


Adhami, S., Giudici, G., Martinazzi, S. (2018). Why do businesses go crypto? An empirical analysis of Initial Coin Offerings. Journal of Economics & Business.

Amsden R., Schweizer, D. (2019). Are Blockchain Crowdsales the new “Gold Rush”? Success determinants of Initial Coin Offerings. Working paper.

Ante, L., Sandner, P., Fiedler, I. (2018). Blockchain-based ICOs: Pure Hype or the Dawn of a New Era of Startup Financing. Article from Journal of Risk and Financial Management.

Baliga, A. (2017). Understanding Blockchain Consensus Models. Persistent Whitepaper

Bashir, I. (2018). Mastering Blockchain: Distributed Ledger Technology, decentralization and smart contracts explained.

Bertani, B. (2014). La crittografia nel sistema di moneta digitale Bitcoin. Bachelor thesis.

Block, J.H., Colombo, M.G., Cumming, D.J., Vismara, S. (2018). New players in entrepreneurial finance and why they are there. Article from Small Business Economy.

Block, J.H., Fisch C.O., Obschonka, M., Sandner, P.G. (2018); A personality perspective on business angel syndication. Article from Journal of Banking and Finance.

Block, J.H., Groh, A., Hornuf, L., Vanacker, T., Vismara, S. (2020). The entrepreneurial finance market of the future: a comparison of crowdfunding and Initial Coin Offerings. Article from Small Business Economy.

Boreiko, D., Risteski, D. (2020). Serial and large investors in Initial Coin Offerings. Article from Small Business Economy.

Buterin, V. (2013). A next generation smart contract & decentralized application platform. Ethereum Whitepaper.

Chohan, U.W. (2017). The Double Spending Problem and Cryptocurrencies. Working paper.

Colombo, M.G., Franzoni, C., Rossi-Lamastra, C. (2015). Internal social capital and the attraction of early contributors in crowdfunding. From Entrepreneurship Theory and Practice.

D’Anconia, F. (2017). Future of Digital Currencies may not Involve Blockchain. Article from Cointelegraph.

De Vauplane, H. (2018). Overview of ICO regulations around the world. Kramer Levin presentation.

Domingo, R.S., Piñeiro-Chousa, J., López-Cabarcos, M.A. (2020). What factors drive returns on Initial Coin Offerings? Article from Technological Forecasting & Social Change journal.

Ducas, E., Wilner, A. (2017). The security and financial implications of blockchain technologies: Regulating emerging technologies in Canada. Article from International Journal

Economist, (2017). Regulators Begin to Tackle the Craze for Initial Coin Offerings. From: https://www.economist.com/news/finance-and-economics/21731157-they-raise-difficult-legal-questions-regulators-begin-tackle-craze

Financial Action Task Force (FATF) Report (2014). Virtual Currencies. Key definitions and Potential AML/CFT Risks.

Feng, C., Li, N., Wong, M.H.F., Zhabg, M. (2020). Initial Coin Offerings, Blockchain Technology, and White Paper Disclosures. Working paper.

Fisch, C. (2019). Initial coin offerings (ICOs) to finance new ventures. Article from Journal of Business Venturing

Fisch, C., Masiak, C., Vismara, S., Block, J. (2019). Motives to invest in initial coin offerings (ICOs). Working paper.

Fisch, C., Masiak, C., Vismara, S., Block, J. (2020). Motives and profiles of ICO investors. From Journal of Business Research.

Fisch, C., Momtaz, P. (2019). Venture Capital and the Performance of Blockchain Technology-Based Firms: Evidence from Initial Coin Offerings (ICOs). Working paper

Florysiak, D., Schandlbauer, A. (2019). The information Content of ICO White Papers. Working paper.

Galuba, W., Girdzijauskas, S. (2009). Encyclopedia of Database Systems. pp 2081-2082.

Greeshma, R. N., Shoney, S., (2017). Blockchain Technology – Centralised Ledger to Distributed Ledger. Paper from International Research Journal of Engineering and Technology (IRJET).

Howell, S., Niessner, M., & Yermack, D. (2018). Initial coin offerings: Financing growth with cryptocurrency token sales. European Corporate Governance Institute (ECGI) - Finance Working Paper

Huang, W., Meoli, M., Vismara, S. (2018) The geography of Initial Coin Offerings. Working paper.

Kaal, W.A. (2018). Initial Coin Offerings: The Top 25 Jurisdictions and their Comparative Regulatory Responses (as of May 2018).

Kaal, W.A., Dell'Erba, M. (2019). Initial Coin Offerings: emerging practices, risk factors, and red flags.

Kean, B., 2018. Don't Believe the Hype. Five Largest ICO “Exit Scams”: Expert Take. From: https://cointelegraph.com/news/dont-believe-the-hype-the-fivelargest-ico-exit-scams-expert-take

Kher, R., Terjesen, S., Liu, C. (2020). Blockchain, Bitcoin and ICOs: a review and research agenda. Article from Small Business Economy

Lee, J., Li, T., Shin, D. (2018). The wisdom of crowds and information cascades in FinTech: Evidence on Initial Coin Offerings. SSRN working paper.

Li, J., Mann, W. (2018) Initial Coin Offerings and Platform Building. Working paper

Malladi, S., Alves-Foss, J., Heckendorn, R.B. (2002). On Preventing Replay Attacks on SecurityProtocols.

Masiak, C., Block, J.H., Masiak, T., Neuenkirch, M., Pielen, K.N. (2020). Initial Coin Offerings (ICOs): market cycles and relationship with Bitcoin and Ether. Article from Small Business Economy.

Momtaz, P.P. (2020). Entrepreneurial Finance and Moral Hazard: Evidence from Token Offerings. From Journal of Business Venturing.

Mood, A., Graybill, F.A., Boes, D.C. (1974). Introduction to the Theory of Statistics. Third Edition

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Working Paper.

Natarajan, H., Krause, S., & Gradstein, H. (2017). Distributed ledger technology (DLT) and blockchain. World Bank FinTech note No. 1, working paper.

Ølnes, S., Janssen, M., Ubacht, J. (2017). Blockchain in government: benefits and implications of distributed ledger technology for information sharing. Article in Government Information Quarterly

Perez, C., Sokolova, K., Konata, M. (2020). Digital social capital and performance of Initial Coin Offerings. Article from Technological Forecasting & Social Change journal.

Peterson, A. (2014). Hal Finney received the first Bitcoin transaction. Here’s how he describes it. Article from The Washington Post.

Roose, K., (2017). Is There a Cryptocurrency Bubble? Just Ask Doge. From: https://www.nytimes.com/2017/09/15/business/cryptocurrency-bubble-doge.html

Saamer, T. (2018). Ico basics: the difference between security tokens and utility tokens.

Saiedi, E., Broström, A., Ruiz, F. (2020). Global drivers of cryptocurrency infrastructure adoption. Article from Small Business Economy.

Samiefar, S., Baur, D.G. (2020). Read me if you can! An analysis of ICO white papers. Article from Finance Research Letters.

Seang, S., Torre, D. (2018). Proof of Work and Proof of Stake consensus protocols: a blockchain application for local complementary currencies. Working paper.

SEC, 2017. Investor bulletin: initial coin offerings. Retrieved from: https://www.sec.gov/oiea/investor-alerts-and-bulletins/ib_coinofferings.

Schückes, M., Gutmann, T. (2020). Why startups pursue Initial Coin Offerings (ICOs)? The role of economic drivers and social identity on funding choice. Article from Small Business Economy.

Stock J.H., Watson M.W. (2002). Introduction on Econometrics

Szabo, N. (1997). Formalizing and Securing Relationships on Public Networks. BOH

Swan, M. (2015). Blockchain: blueprint for a new economy. O'Reilly, Sebastpol (CA).

Tanenbaum, A. S., Wetherall, D. J. (2011). Computer Networks. Fifth Edition. p. 368-369

Vismara, S. (2016). Equity retention and social network theory in equity crowdfunding. Article from Small Business Economy.

Yermack, D. (2017). Corporate Governance and Blockchain. From Review of Finance.

# Sitography


https://www.investopedia.com/terms/l/legal-tender.asp

https://en.cryptonomist.ch/2019/03/09/ico-trends-2019/

www.khanacademy.org/computing/computer-science/cryptography/crypt/v/random-vs-pseudorandom-number-generators

https://icobench.com/ratings

https://icobench.com/stats#stats-raised

https://www.coinschedule.com/stats

https://bitcoinist.com/regulations-killed-icos-stos/

https://www.finance-monthly.com/2018/10/the-10-biggest-ico-scams-swindled-687-4-million/

https://www.kickstarter.com/help/stats

https://www.finance-monthly.com/2018/10/the-10-biggest-ico-scams-swindled-687-4-million/

https://cointelegraph.com/ico-101/top-10-icos-with-the-biggest-roi

https://cointelegraph.com/news/dont-believe-the-hype-the-fivelargest-ico-exit-scams-expert-take

https://www.sec.gov/oiea/investor-alerts-and-bulletins/ib_coinofferings

https://thetokenist.io/france-announces-first-ever-ico-approval-via-amfs-ico-visa/

https://www.washingtonpost.com/news/the-switch/wp/2014/01/03/hal-finney-received-the-first-bitcoin-transaction-heres-how-he-describes-it/?arc404=true

https://neironix.io/it

https://icosbull.com/

https://icomarks.com/

https://www.icohotlist.com/

https://www.listico.io/

https://icotokennews.com/it/

icodata.io

https://www.tokendata.io/

https://icodrops.com/

https://decryptionary.com/dictionary/hard-cap/

https://icospotters.com/blog/what-is-erc20-token-standard-and-why-is-it-useful-for-icos

https://www.economist.com/news/finance-and-economics/21731157-they-raise-difficult-legal-questions-regulators-begin-tackle-craze

https://www.nytimes.com/2017/09/15/business/cryptocurrency-bubble-doge.html

https://www.money.it/

https://www.investopedia.com/terms/l/leptokurtic.asp

https://it.investing.com/crypto/ethereum/eth-usd-historical-data

https://icobench.com/ico/telegram-open-network

http://web.mit.edu/course/17/17.846/OldFiles/www/Readout.html

https://stats.idre.ucla.edu/stata/output/regression-analysis/

https://www.planetcompliance.com/2019/09/04/what-is-the-difference-between-utility-security-and-payment-tokens/

https://www.investopedia.com/news/what-erc20-and-what-does-it-mean-ethereum/