ISSN: 2281-1346

UNIVERSITÀ DI PAVIA

Department of Economics and Management

DEM Working Paper Series

ICOs White Papers: identity card or lark mirror?

Paola Cerchiello
(Università di Pavia)

Anca Mirela Toma
(Università di Pavia)

Marco Caluzzi
(Università di Pavia)

 # 197 (01-21)

Via San Felice, 5
I-27100 Pavia

economiaweb.unipv.it

Working Paper manuscript No.
(will be inserted by the editor)

ICOs White Papers: identity card or lark mirror?

Paola Cerchiello, Anca Mirela Toma, Marco Caluzzi

Received: 27 January 2021

Abstract The Initial Coin Offering (ICO) is one of the operations based on DLT or blockchain technology that allows fundraising activities for an entrepreneurial project, by issuing utility tokens instead of a security or an equity token. ICOs are a new and promising tool to support innovative ideas with the potential, due to the underlying technology, to shape the future of the fundraising systems and architectures. The present paper is twofold: on one hand it offers a dataset of 760 ICOs containing several variables, completely checked, harmonized and validated through the comparison of alternative sources, that can be used as benchmark for further analysis. On the other hand, it investigates research hypothesis aimed at highlighting plausible success drivers that can be extracted from white papers taking into account also the team composition and the social media exposure. Our results show that the variables derived from the white papers, such as the existence of the appendix, the picture of the team, the sections and the nr. of pages are statistically significant with a differentiated impact on the probability of success or failure of an ICO.

Keywords ICO · white-paper · information disclosure · blockchain · social exposure

Paola Cerchiello
University of Pavia
E-mail: paola.cerchiello@unipv.it

Anca Mirela Toma
University of Pavia
E-mail: ancamirela.toma01@universitadeglistudidipavia.it

Marco Caluzzi
University of Pavia
E-mail: marco.caluzzi01@universitadipavia.it

1 Introduction

Initial Coin Offerings got famous as the finance model of cryptocurrencies. They are a digital way of public capital funding for entrepreneurial use through the issue of an own virtual token [Blaseg, 2018]. A token is a 'crypographically secured digital asset' [Howell et al., 2018]. For companies whose business model stands in connection with blockchain technology, ICOs have surpassed the traditional venture capital financing in the shortest of time[Hahn and Wilkens, 2019a]. This means that ICOs are a new way to raise capital for young and unestablished ventures.

The first ICO was held in July 2013 by Mastercoin, which is a digital currency built on the Bitcoin blockchain [Shin, 2017]. Since then, over a thousand ICOs have followed. CoinSchedule, a leading website monitoring current ICOs, reports that 366 ICOs took place in 2017, raising a combined amount of USD 6.2 bn. According to Fisch [Fisch, 2019a], the aggregated 2017 funding volume was surpassed in the first three months of 2018 alone. 254 ICOs raised USD 7.8 bn in this period. The premier crowdfunding platform Kickstarter in contrast has raised a total of USD 4.6 bn since its inception in 2009 [Kickstarter, 2019]. In the month of June 2018 alone, the ICO funds raised amounted to over USD 5 bn with 91 ICOs ending in this period. However since then, monthly funds raised remained more or less distinctly under the mark of USD 1 bn with the exception of May 2019. The number of token sales has also declined from the staggering number of 146 ended ICOs at the maximum in April 2018 to four ended token sales in September 2019 [Coinschedule, 2019].

It is to note that an officially recognized definition of ICOs does not exist [Blaseg, 2018]. The name initial coin offering is a reference to the well established concept of initial public offerings (IPOs). However, at first sight, ICOs have relatively few things in common with traditional public offerings. In an ICO, a new firm offers a token to a crowd of investors for the first time. In IPOs the company is most often already established and has had rather a successful past. In an IPO, shares of the company are sold. In an ICO, the sold token is created by the firm offering it using distributed ledger technology (DLT) and can be bought in exchange for fiat money or other cryptocurrencies. The functions of the token may equal classical shares but are manifold [Hahn and Wilkens, 2019b].

The spike in ocurrence of ICOs followed the development of the Blockchain by Nakamoto in 2008 and the subsequent development of cryptocurrencies such as ethereum (short: ether) [Catalini and Gans, 2018]. A blockchain enables the direct, secure transfer of value over the Internet between parties that do not trust each other [Howell et al., 2018]. It consists of a sequential list of transactions in a unit of value that is native to the blockchain. For example the Bitcoin blockchain uses the cryptocurrency bitcoin and the Ethereum blockchain uses ether. Additional text, such as for example contingent terms of contracts, can be appended to a transaction. Bitcoin permits simple and limited additional text, whilst other blockchains such as Ethereum essentially permit any code to be executed as part of a transaction. Blockchains are so called distributed ledgers, providing decentralized record-keeping that cannot be retroactively edited. Cryptography enables rapid verification and prevents hacking [Howell et al., 2018]. Distributed Ledger Technology (DLT) describes a decentralized database stored on a set of individual nodes. The records are syncronized through a consensus algorithm, which allows peer-to-peer transactions without the need for an intermediary. This is the technical basis of an ICO in various forms [Blaseg, 2018]. The token sales themselves are conducted via complex, self-enforcing and state contingent smart contracts, which are pieces of code embedded in a blockchain. This enables the exchange of money, property or other assets without an intermediate party. Smart contracts guarantee the fulfillment of the transaction and regulate for example the conditions of sale for the tokens [Blaseg, 2018]. Therefore, due to low transaction costs similar to crowdfunding, ICOs might become a significant driver for financial inclusion by democratizing access to investments and capital [Mollick and Robb, 2016].

Recently, there has been a growing literature studying the ICOs drivers aiming to predict their future outcome. [Blaseg, 2018] offers an exploratory empiric classification of ICOs and the dynamics of voluntary disclosures. It examines to what extent the availability and quality of the information disclosed can explain the characteristics of success and failure among ICOs and the corresponding projects. Another important research focuses on the effectiveness of signalling ventures and ICOs projects technological capabilities to attract higher amounts of funding [Fisch, 2019a]. To the discussion of information asymmetries also Boreiko and Risteski [Boreiko and Risteski, 2019] compare serial and large ICO investors with other ICO investors finding that the average serial ICO investor invests earlier than other ICO investors. However, serial investors are not more informed than other ICO investors, per se, and fail to choose better quality ICOs. The situation is different from large serial investors, who seem to have an information advantage over other ICO investors, as they are more likely to invest in more successful ICO campaigns.

Other streams of research concentrate on the impact of ICO managers quality and entrepreneurial decision making process. Those are specific aspects of a wider prospective that goes under the behavioural economics and entrepreneurial finance umbrella. [Momtaz, 2018] studies the impact of CEOs loyalty disposition and the magnitude of asymmetry of information between managers and investors on ICOs performance . Moreover, to remain in the management area, an interesting spark comes from a research specifically directed on CEOs emotions signalling role and their effects on ICOs results [Momtaz, 2020a]. Momtaz, in another paper aims at identifying the likelihood and possible timeframe of value creation for investors by combining several factors (financial return, amount of capital raised, listing and delisting alternatives, industry events study etc.) to analyse the ICOs success drivers [Momtaz, 2018]. In a recent study again [Momtaz, 2020b] focuses on information asymmetries and agency costs raising between entrepreneurs and investors. From the decision making process studies, he have found that entrepreneur's social identity in conjunction with the enabling mechanisms of the blockchain technology shape entrepreneurial pursuits and funding choice [Schückes and Gutmann, 2020], [Kher et al., 2020]. Finally, another area of studies focuses on the driving factors impacting the liquidity and trading volume of crypto tokens listed after the ICOs. Among those factors have been identified the quality level of disclosed documentation (source code public on Github, white paper published, an intended budget published for use of proceeds), the community engagement (measured by the number of Telegram group members), the level of preparation of the management (using as proxy the entrepreneurial professional background of the lead founder or CEO), and other outcomes of interest (i.e., the amount raised in the ICO, outright failure - delisting or disappearance, abnormal returns, and volatility) [Howell et al., 2018].

The aim of this paper is twofold: first, we make available a tailor made data set composed of 760 ICOs containing relative relevant variables, completely checked, harmonized and validated using several alternative sources. This constitutes a value per se, given the absence of a unique, reliable and complete dashboard of ICOs.

Second, we provide a full statistical analysis focused on highlighting the most relevant success drivers that take into account covariates extracted from White papers as an additional source of information. Indeed, we believe that, although not regulated or properly formatted, white papers can disclose interesting insights about the ICO issuers both in terms of reliability and trustworthiness. Our results show that most of the variables derived from the white papers are statistical significant with a differentiated impact on the probability of success or failure of an ICO. In particular our main findings are:

- The number of pages of white papers positively affects the probability of success. The longer the document, the higher the probability. This would suggest, as expected, that a more detailed and complete description of the business idea helps in reaching the target.
- The number of sections of white papers does not represent a relevant factor. The structure complexity of the documents is not an added value. Moreover, the number of sections cannot be considered a proxy of the length of the white papers.
- The presence of team's pictures is significant but with negative sign. This finding appears counterintuitive, but stable across several robustness checks. It is not straightforward to give an interpretation, but since it often happens that team members, especially advisors, participate in several ICOs, the clear visual identity more than names listing, may affect negatively the perception. A client is likely to think that the participation to several business venture reveals lack of specific commitment and background knowledge. A sort of mercenary behaviour of people gravitating around in ICOs world.
- The presence of an appendix, somehow technical, contributes positively to the success of an ICO. This is reasonable, if intended as a further level of disclosure and openness towards a full engagement pf prospect clients, in particular the more skilled ones.

All the findings have been checked and validated by defining alternative models and by controlling for other relevant factors like the team composition and the social media exposure.

The rest of the paper is organized as follows: in section 2 we fully detail the ICOs characteristics and background. In section 3 we describe the statistical methodology used to prove the importance of the analyzed variables. In section 4 we describe the data gathering protocol and in section 5 we show main results. Finally in section 6 we draw conclusions and future steps.

2 The Initial Coin Offerings world

Although an international consensus on this topic cannot be found, the Swiss Financial Markets Supervisory Authority (FINMA) proposes a valid taxonomy of tokens based on their different economic functions [ESMA, 2018]:

1. Payment tokens
2. Utility tokens
3. Asset tokens

Payment tokens can be used to acquire goods or services. The most famous example for this is Bitcoin. Bitcoin is the worldwide leading decentralized cryptocurrency since its implementation in 2009, following its first description by Nakamoto in 2008 [Nakamoto, 2008]. The holder has no claim on the issuer and the currency is purely virtual [ESMA, 2018].

Utility tokens serve to provide access to a specific application or service which is provided by the issuer of the token. Contrary to payment tokens they do not work as payment for any other applications. An example for this would be Filecoin which held an ICO in September 2017 selling a utility token to enable the buyer to use its decentralized cooperative data storage solution [ESMA, 2018]. One can imagine the system of utility tokens to function in a similar way as crowdfunding presales on platforms like Kickstarter. A fitting example would be that one buys a designated seat in a stadium before it is even built [Howell et al., 2018].

Asset tokens have an economic function which is similar to traditional bonds or derivatives. They represent assets such as a debt or equity claim on the issuer and promise for example a share of future company earnings [ESMA, 2018].

In addition, there also exist hybrid tokens, representing a mixture of the types mentioned above.

By increasing competition in payment markets, payment tokens can inspire increased efficiency in cost, speed, user-friendliness and security of traditional payment methods. Furthermore, the DLT network approach empowers the individual as the businesses and services function without institutional barriers such as a bank. This means that payment tokens might also act in favour of financial inclusion of people lacking a bank account. Due to the anonymity of the users in the network there is however increased risk of new types of fraud as well as money laundering. Furthermore, investor protection problems arise from the fact that payment tokens do not represent a claim or entitlement but are speculative investments and thereby subject to price fluctuation and high volatility. Bitcoin for example regularly is in news media due to high volatility, making headlines like 'Bitcoin price roller coaster continues as value plunges by USD 1,000 in less than an hour' [Cuthbertson, 2019]. According to the SMSG, there is also a risk of market abuse, as 10 'whales' (large investors) own 50% of the largest ICOs and could therefore potentially engage in price manipulation [ESMA, 2018]. Also, due to the lack of centralization there is no control over the tokens' money mass, a task executed by central banks for legal tender.

Utility tokens allow the prefunding of a future business without diluting ownership. Therefore they are a source of early stage funding for innovative projects and an alternative to, for example, crowdfunding. By selling utility tokens, which enable the customer to use the product, issuers can also create a network of users before actually implementing the business, thus yielding business advantages. The advantages also directly translate into disadvantages as the issuer might not deliver the expected service (counterparty risk) or may go out of business (performance risk).

Asset tokens giving the holder a monetary claim on the issuer are similar in characteristics to securities and derivatives. Therefore, their benefits are also similar. They facilitate the (pre-)financing of a business and are a measure for risk-transfer. However, the risks also resemble the risks of securities. Namely counterparty risk as mentioned above, dilution risk if there is no issuance control and custody risk.

Table 1 depicts the results of an explanatory data analysis on a set of 4,564 token sales events listed on Icobench.com. It revealed the following geographical dispersion of ICOs. In total ICOs were taking place in 135 countries:

Table 1: Geographical dispersion of ICOs

Country	**Number of ICOs**
Estonia	230
Hong Kong	137
Russia	311
Singapore	441
Switzerland	226
UK	394
USA	661
Miscellaneous	1972

It can be seen that the vast majority of ICOs take place in the United States of America (USA), Singapore, United Kingdom (UK), Russia, Estonia and Switzerland. The choice of country under whose law one wants to hold an ICO is influenced by many factors. Initiators of ICOs are also associated to have high technological skills as they have to be able to create a business based on blockchain [Fisch, 2019a]. It is clearly observable that the countries with the highest numbers of ICOs are all somewhat considered as technologically and financially advanced as already pointed out by [Huang et al., 2020]. Also, according to [Howell et al., 2018], ICO issuers are concentrated in a set of countries that seem more related to technical expertise than to legal systems, with the highest number of ICOs in Russia, the USA, and Switzerland. ICOs may permit early stage, risky investment to circumvent well-functioning property rights and contract enforcement. The legal status of ICOs is not clearly defined at this point and the country approaches vary greatly. However, the subject is matter to great discussion.

According to [Howell et al., 2018], tokens are natural targets for speculation because they are usually exchangeable for fiat and cryptocurrency. This liquidity is an attractive feature compared to conventional VC securities. Furthermore, tokenized real assets, which are tokens tied to real-world assets such as the price of gold or USD, tend to have higher failure rates. These are essentially the opposite of utility tokens and more often appear to be scams .

2.1 Functionality of ICOs

When dealing with an ICO, a prospective buyer submits a purchase order for a token by sending a payment to the issuer. Payment is usually in cryptocurrency, and most commonly in ether, which prospective buyers can purchase for FIAT money on cryptocurrency exchanges. At the conclusion of the sale, the smart contract automatically sends the purchased tokens to the blockchain addresses of successful buyers. One reason that the number of ICOs has risen so quickly is that in their most basic form they impose essentially zero costs on the issuer.

Token sales events are typically preceded by the release of white paper disclosure documents that are similar in spirit to prospectuses for initial public offerings. As there is no official regulation on what white papers need to include, their content varies greatly. However most of the times they include a description of how the token will be used and its benefits as well as some information on the background and business idea of the project. Often missing is basic information about issuer and a contact address. Most of the time, no information about the legal address or the individuals behind the ICO can be found. In addition to a white paper, issuers promote their project before the ICO launch through social media channels such as Twitter, Facebook or Telegram communication channels. Sometimes, source code of the project or information about the token distribution mechanism is made public on code sharing platforms such as Github. It is also common that the issuing team links their online presence on websites such as LinkedIn. If there exists an advisory team for the ICO, their presence might also be reported.

Investors can get information about future and past ICOs, including links to all relevant information sources mentioned above on websites such as Coinschedule.com, Icobench.com, IcoWatchList.com and many more.

For the issuers, an initial coin offering can roughly be divided into two phases. Firstly, the Pre -Public-Engagement phase. The initiators of the project are developing their business idea and need to get to know whether there is significant surplus to holding an ICO for the project or not. There are many things to be taken into account. One question being whether an ICO is actually legal in the intended country for the ICO. In most countries, ICOs are either legal, regulated or subject to future regulation. Once a venture has made the decision to hold an ICO, it needs to create a coin. There are many different ways to hold an ICO, including different phases of token sale, pre sales, the actual ICO and general sale. The venture needs to decide on how many coins he/she wants to sell and how much remains within the ICO team. Also, a platform for the issue of coins needs to be selected. According to [Howell et al., 2018], most ICOs use ERC20 tokens, which are smart contracts hosted by the Ethereum blockchain. Anyone can create such a contract for free. If all basics are settled, the Public Engagement Phase with the goal of spreading the word about the ICO through above mentioned channels begins. After the launch of the ICO, the issuer has no control over the tokens beyond what was specified ex-ante in the contract. According to [Howell et al., 2018], when launching an ICO, the issuer has to decide on:

- Target proceeds
- Fraction of total token supply sold
- Pricing mechanism
- Distribution method
- Token rights
- Exchange listing

After the completion of an ICO, the newly created tokens can be traded online on certain websites, known as digital currency exchanges (DCEs), such as CoinBase or Kraken. These enable customers to exchange the new tokens for other assets, such as other digital currencies or legal tender [Stacher, 2018]. According to estimates by the European Securities and Markets Agency, around 200 of such trading platforms exist globally. However the most flows concentrate on a small subset of them. A token being traded on a secondary exchange therefore is a sign that an ICO has been completed successfully and there still exists demand for the token. Between a fourth and a third of tokens offered in ICOs are traded. Daily trading volumes normally amount to USD 10-15 bn. This number needs to be taken with caution though, as it is suspected that some platforms seemingly inflate the volumes that they trade [European Securities and Markets Authority, 2019]. However, as with traditional exchanges, the value of the token is subject to fluctuation and may change with the sentiment of the market. There is also, as with any exchange, a risk of insider trading[ESMA, 2018].

As we can see, there are different perspectives from which ICOs can be further understood through qualitative or empirical works. The perspectives from which we can gather deeper understanding are multiple: entrepreneurial, functionality, technicalities, regulatory issues, illicit behaviour, investor protection, etc. In the following section, we will focus on a technical aspect of the ICO process that has to be analyzed: the white paper informative power.

2.2 The White Paper

As mentioned above, the white paper can be seen as the prospectus of an ICO. It is important to note that it currently does not exist a specific obligation for ICOs to publish a white paper, nor any guidelines on what the white paper has to include or its format. As will be elaborated later in this section, white papers can signal high quality business to investors and therefore might be an important tool to attract funds and improve the success of an ICO. However, due to lack of guidelines, content varies greatly and one cannot assume that only the presence of a White Paper guarantees a trustworthy ICO. In our dataset, white papers have, on average, a length of 36.27 pages, the shortest being 1 page and the longest 127.

A white paper typically consists of numerous sections. Issuers might or might not disclose source code, give an insight into their business idea and the added value their project is supposed to generate. Often, a time frame of how the business is planned to develop is included. The white paper might also include photographs as well as a short biography of the team members in their individual functions. Issuers often employ advisors to their projects which might also be mentioned in the white paper. Although not too common, a white paper can also include a legal disclaimer or a warning about the speculative nature and connected risks of ICOs. Relating to the source code which might be included in the White Paper, the Securities and Markets Stakeholder Group notes that 'there is a 'code risk' because the instructions programmed into the token software may not always reflect the algorithm and features described in the whitepaper' ([ESMA, 2018], p.12).

Figure 1 shows three examples of white paper formats to illustrate the lack of specific guidelines. Often, it seems that a nice layout is almost, as important as, the information depicted in the document. The German Federal Financial Supervisory Authority BaFin also argues that white papers, in first line, are used as a PR measure and for communication. They often get changed during the lifetime of an ICO and the information included is often very imprecise. This results in not providing security or protection to investors [BaFin, 2018]. White papers can therefore not directly be compared to a prospectus regulated by law.

As already stated, the present paper deals with the understanding of success factors of ICOs. As ICOs are subject to few regulations, it is of high interest to examine how additional laws might need to be structured to prevent fraudulent ICOs and ensure consumer protection whilst also enabling innovation in the FinTech field. There is however a string of economic literature which suggests that the ventures themselves have a great interest in voluntarily disclosing important information which might help investors in making smart and informed choices. This, in turn, might imply that regulation on this issue is obsolete.

In his paper on 'Job market signaling' [Spence, 1973], Spence introduced the now well- known concept of economic signaling. Signaling is part of the discipline of information economics and tries to solve principal agent problems. The agent has more information than the principal and tries to signal its quality to the principal to persuade a conclusion of contract. Therefore, the goal of signaling is to avoid adverse selection. Spence examines this situation for the job market environment. Traditionally, there are high information asymmetries and information is distributed unevenly between employees and employers. The employer is at an information disadvantage. Before hiring someone, he/she cannot estimate exactly the abilities of the applicant. The model includes two types of applicants, one type has a higher productivity, the other is less productive. The problem, the more productive workers face, is that an employer cannot distinguish them from the other sort of workers without signaling. Without signaling, the employer will therefore pay an equilibrium salary in between the appropriate salary for both worker groups. This is so unsatisfactory for the more productive worker that he/she will try to signal his quality to the employer. In Spences' Model, he/she will do this by getting a higher education, which for the more productive type bears lesser costs than for the less productive type. This means he/she can actually signal his quality [Spence, 1973].

Spences theory has been adapted to many different fields of economics. For the field of finance, and in our case ICOs, signaling can be interpreted as high-quality ventures can attract higher amounts of funding by sending signals to potential investors. It is argued that, once high-quality ventures voluntarily begin to disclose information, other ventures will follow until all but the ventures of lowest quality fully disclose [Blaseg, 2018].

Because ICOs are mostly held by ventures in early stages, there is a considerable amount of information asymmetry. Additionally, the amount of speculation and hype around ICOs is very high and therefore the amount of objective information is normally quite low [Fisch, 2019a].

Fisch examines the concept of signaling in context related to ICOs [Fisch, 2019a]. In his study, he assesses the determinants of amount raised by 423 ICOs. He argues that the technological capabilities of the venture represent a crucial indicator for relative quality as ICOs are highly technical. From this conclusion, three indicators of a venture quality are derived. These can be understood as signals in the ICO context. The indicators are: technical white paper, high-quality source code and patents. Fisch finds that a technical white papers, as well as, high-quality source codes do lead to higher funding amounts. Patents however do not seem to have an effect on the amount raised. In addition, token supply, Ethereum-standard and Twitter activity seem to have a positive impact on amount raised. Fisch's study highlights that, under certain conditions,

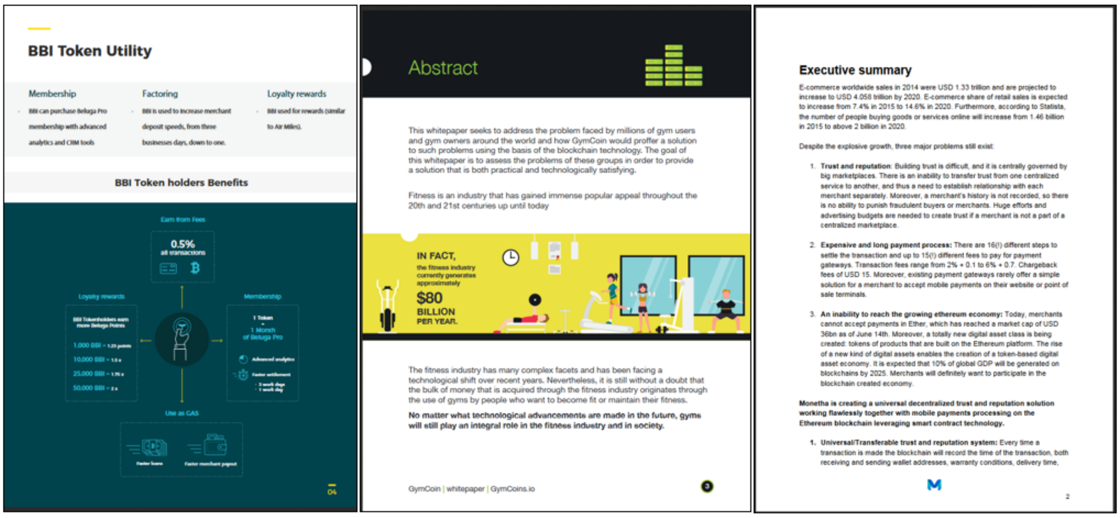

Fig. 1: Examples of white paper format
White Paper of BelugaPay (left), GymCoin (middle) and Monetha (right)

white papers can serve as signals for successful ventures. This intuition is important, as it shows that even absent of regulation, ICOs have an interest in signaling their quality and investors have focal points which may demonstrate the presence of a trustworthy venture. [Florysiak and Schandlbauer, 2019] stresses the importance of white papers as privileged communication channel to prospect clients, allowing the reduction of information asymmetry. In particular, by running a textual analysis, they found out that signaling is likely to be biased during the ICO process as investors make their decisions upon possibly misleading expert ratings. This can lead to inefficiencies in the financial market which allocates wrongly funds and produces high volatile, risky and prone to failure systems.

Therefore, if we focus on the entrepreneurial aspect of ICO, we see that there are many common aspects with crowdfunding financing tools. We can state that the ICOs lay at the interplay between crowdfunding and blockchain. Thus, investigating on the aspects that the two systems share and on the specific ones, is fundamental for investors and entrepreneurs in order to chose between the two tools[Block et al., 2020].

Although initially, Crowdfunding and ICOs appear to be very similar in their characteristics and mechanisms, there are meaningful differences between the two funding tools. Entrepreneurs and investors need to understand these differences to use these funding methods in the most efficient way. The same applies to policymakers who are asked to define an appropriate regulatory framework for the two financing instruments.

[Ante et al., 2018] confirm that ICOs exhibit several similarities with classical crowdfunding and venture capital markets. In particular, they found a match in the determinants of funding success specifically with regards to human capital characteristics, business model quality, project elaboration, and social media activity.

Generally, ICO issuers use three forms to voluntarily disclose information and thereby reduce information asymmetries and influence investors: (i) White paper (Cohney et al., 2019), [Cohney et al., 2019] (ii) Github and (iii) Online information tools (Bourveau et al., 2018) [Bourveau et al., 2019]. In line with the above literature and specifically focusing on white papers characteristics, we investigate the relative structure and the technical content controlling over other ICOs drivers such as team composition and social media exposure. Such two components have been widely investigated and showed their importance like in [Toma and Cerchiello, 202 and in [Cerchiello et al., 2019].

The literature on the analysis of white papers considers different relevant aspects. The length of the white paper has a positive association with the amount raised in the ICO. This confirms previous findings in the literature by [Amsden and Schweizer, 2018] and [Fisch, 2019b] Fisch (2019) and strengthens our hypothesis that the length of a white paper impacts the success of an ICO. Further, the file size does not have a significant relationship with the amount raised in the ICO. Insofar, we set our first two research hypotheis, that are:

H1: the length of white papers positively influences the probability of success.

The white paper length is measured as the natural logarithm of the number of pages (as in [Bourveau et al., 2019])

H1bis: the number of sections of white papers positively influences the probability of success.

We deem $H1bis$ as an alternative formulation to $H1$: the level of completeness of the white papers measured through the number of sections instead of the number of pages.

Again [Feng et al., 2019] puts emphasis on the quality of white papers. In particular, they found out that the more technical details are described in the document, the larger is the raised funded amount. Once again, it is stressed the importance of full disclosure and the white paper as a mean of communication more traditional (compared to Github or Telegram chats) and expression of the credibility of the ICO issuers, From such considerations we derive our two more hypothesis:

H_2: the presence of team pictures in white papers positively influences the probability of success.

H_3: the presence of an appendix in white papers positively influences the probability of success.

Such hypothesis are tested controlling over some ICOs characteristics, namely the composition of the team and the social web exposure. Those two components have been already investigated and tested in [Toma and Cerchiello, 2020].

3 Methodology

As in the work of [Cerchiello and Toma, 2018], logistic regression is used to determine the influence of different factors on the success of ICOs. Using binomial logistic regression, the dependent variable is classified into two groups (1=success vs 0=failure/scam). The alternative classification into three classes (scam, success, fail) or into scam and no scam was thought of to be imprecise due to the rather small number of observations [Cerchiello and Toma, 2018]. The logistic regression model is as usual defined as follows [James et al., 2017]:

$$ln(\frac{p_i}{1-p_i}) = \alpha + \sum_j \beta_j x_{ij} \tag{1}$$

where p_i is the probability of the event of interest, $x_i = (x_{i1}, ..., x_{ij}, ..., x_{iJ})$ is a vector of each ICO_i's specific explanatory variables, α is the intercept parameter and β_j, for $j = 1, ..., J$ are the regression coefficients to be estimated from the available data. For the probability of success of an ICO it follows that:

$$p_i = \frac{1}{1 + exp(\alpha + \sum_j \beta_j x_{ij})} \tag{2}$$

The logistic function is a special case of the so-called sigmoid functions and maps into $(0, 1)$. This ensures that classification is possible and that $p_i \in (0, 1)$ [James et al., 2017].

Following the arguments of [Cerchiello and Toma, 2018] the approach of logistical regression is feasible as the target variable (success vs. no success) is distributed rather evenly.

In order to obtain a more robust and efficient model, we compare classical logistic regressions with the relative regularized versions: Lasso and Ridge.

Lasso logistic model is a shrinkage method that allows obtaining a subset of variables that are strongly associated with the dependent variable, through regularization of the coefficients bringing them to values very close or even exactly equal to zero. Since the L1 penalty is used, the variables with a coefficient equal to zero are excluded from the model [Hastie et al., 2009]. Mathematically:

$$L_{lasso}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} \left|\hat{\beta}_j\right|. \tag{3}$$

where y_i are the n-observations for the target variable (success-failure), x_i are the n-observations for the covariates, λ is the penalization parameter chosen by cross validation and β_j are the coefficient of the model.

Ridge regression appears to be very similar to Lasso with the exception of the penalization term that presents a different form, as in the following:

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} \hat{\beta}_j^2. \tag{4}$$

Formula 4) differs from 3) in imposing a different constraint on the parameters β_i. Ridge regression employs a L2 normalization ($\hat{\beta}_j^2$) which produces a different effect on parameters: they are shrunk toward zero without being exactly zero. From the employment and comparison of the three models, logistic-ridge-lasso regression, we can leverage the best model structure, understanding what are the key variables in predicting success or failure.

In multiple regressions, two or more explanatory variables might be correlated with each other. This situation is defined as collinearity. When collinearity exists between three or more variables, even if no pair of variables has a particularly high correlation, this situation is referred to as multicollinearity. The presence of multicollinearity leads the solution of a regression to become unstable. A multicollinearity check called the variance inflation factor (VIF) is employed. It measures how much the variance of a regression coefficient is inflated due to multicollinearity. The VIF is the ratio of the variance of $\hat{\beta}_j$ when fitting the full model divided by the variance of $\hat{\beta}_j$ if fit on its own [James et al., 2017]. The absence of multicollinearity is indicated by a VIF value of one. A rule of thumb used widely is that a VIF value exceeding 5 or 10 indicates problematic amounts of collinearity. The concerned variables should then be removed from the model since the occurrence of multicollinearity means that the information embedded in the concerned variable is abundant in the presence of the other explanatory variables of the model. The VIF for each variable can be computed using the formula [James et al., 2017]:

$$VIF(\hat{\beta}_j) = \frac{1}{1 - R^2_{X_j|X_{-j}}} \tag{5}$$

where $R^2_{X_j|X_{-j}}$ is the R^2 from a regression of X_j onto all of the other predictors. If $R^2_{X_j|X_{-j}}$ is close to one, then collinearity is present and the VIF will be large [James et al., 2017].

4 Data

4.1 Purpose and hypothesis of the analysis

The main purpose of this paper is to identify which characteristics of an ICO (based on the information disclosed and available to prospective investors) are meaningful to predict its outcome. Thus, the main hypothesis of this analysis is to find out if the outcome of the fundraising process is positively or negatively influenced by the features observed.

The total number of ICOs in the market (together with the total amount raised) increased extensively in 2017 and 2018. The expansion of the market was probably amplified by the increasing hype and consequent bubble that affected the bitcoin cryptocurrency towards the end of 2017 and the beginning of 2018 [Pilkington, 2018]. Given the quantity of information available in 2017 and 2018, it was decided to retrieve data from this time period since it provides a larger database increasing the probability to retrieve meaningful information on the outcome of ICO processes. The analysis was conducted investigating a dataset composed of 760 observations labelled as “successful”, “failed” or “scam”. During the analysis different models and techniques were used and compared to obtain the best possible classification performance.

The first step of the data gathering process was combining three initial datasets to obtain a list of observations that served as a basis to start collecting the information needed for the analysis. In particular:

1. The first dataset was retrieved from Icobench [1] one of the best ICO rating platforms.
This dataset was used to obtain a large number of ICO observations and then extract only their references, since it would be then enriched with features extracted from the second dataset and filled gathering information on Icobench. The file was initially composed of 13 features and 5552 observations. After dropping duplicates, missing values and unnecessary features, the number of observations decreased to 2465 and the number of variables decreased to 3 (“name”, “url”, “icoEnd”, which are used as reference of the ICO).
2. The second dataset was obtained from the work [Toma and Cerchiello, 2020] which provided 190 observations and an appropriate range of features for the purpose of the analysis.
Three main macro-categories of features were included in the dataset. The first is related to the presence of communication channels and social media (Website, Telegram, Twitter, Facebook, LinkedIn, Youtube, Github, Slack, Reddit, Bitcointalk, Medium). The second is related to information disclosed through rating platforms such as the number of team members, the number of advisors and the presence of their picture. The third set of features is instead related to whitepapers and their characteristics like the number of pages, the number of sections and the presence of the appendix. A full list of the features used in the analysis is provided in Table 2, while in Table 3 we report relative descriptive statistics.

Table 2: Explanatory variables

success	0=failure or scam 1= success
nat_fail	0=success or scam 1=failure
scam	0=success or failure 1=scam
Web_du	Website presence(dummy)
tw	Twitter (dummy)
fb	Facebook (dummy)
ln	Linkedin (dummy)
yt	Youtube (dummy)
gith	Github (dummy)
slack	Slack (dummy)
reddit	Reddit (dummy)
btalk	Bitcointalk (dummy)
mm	Medium (dummy)
social_du	There exits at least one social media (dummy)
nr_channels	number of activated social media channels (dummy)
nr_team	Number of Team members (quantitative)
nr_adv	Number of advisors (quantitative
picture_du	Presence of members pictures (Dummy)
Nr_pages	Number of White paper pages (quantitative)
sections	Number of White paper sections (quantitative)
appendix_du	Presence of an appendix in the white paper (dummy)

3. The third dataset was instead created by retrieving information from Deadcoins[2], the most popular cryptocurrency platform to report ICO scams, failures and phishing activities. The dataset, once filtered by the category “scam” and cleaned by duplicates or missing values, was composed by 641 observations and two main features, “name” and “cat-

[1] Retrieved from https://icobench.com/ on July 2019

[2] Retrieved from https://deadcoins.com/ on March 2020.

Table 3: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
success	760	0.478	0.49	0	0	1	1
nat_fail	760	0.48	0.50	0	0	1	1
scam	760	0.05	0.22	0	0	0	1
Oweb_dum	760	0.78	0.42	0	1	1	1
Nr_Telegram	760	4,6	10,257	0	0	4,176	108,871
tw	760	0.90	0.30	0	1	1	1
fb	760	0.81	0.39	0	1	1	1
ln	760	0.60	0.49	0	0	1	1
yt	760	0.63	0.48	0	0	1	1
gith	760	0.67	0.47	0	0	1	1
slack	760	0.21	0.41	0	0	0	1
reddit	760	0.74	0.44	0	0	1	1
btalk	760	0.80	0.34	0	1	1	1
mm	760	0.74	0.44	0	0	1	1
nr_team	760	9.05	6.44	0	5	12	52
nr_adv	760	4.29	4.20	0	0	7	24
Nr_pages	760	28.48	19.13	0	16	40	127
sections	760	9.53	6.54	0	6	13	50
foto_du	760	0.74	0.44	0	0	1	1
appendix_du	760	0.13	0.33	0	0	0	1

egory". This dataset was used to match it with the final dataset to determine which observations are affected by suspicious activities and therefore to simplify the scam identification.

After combining the datasets, 580 observations were collected and added to the 190 observations from the first dataset [Toma and Cerchiello, 2020], for a final total of 780 observations. The information collected was obtained from Icobench and, where missing, other rating and listing platforms were used, such as ICO Drops.com, CoinDesk.com, Tokendata.io, Icoholder.com. Furthermore, to fulfil missing data on Icobench regarding white papers and related information, other specialized websites were consulted, such as Allcryptowhitepapers.com and Whitepaperdatabase.com.

The collection process was guided taking as a reference the ICO market composition of 2017 and 2018. This means that the final dataset of 780 observations respects the structure of the market in both years. The data about the market structure were retrieved from the Report "ICO Market Analysis 2018" available on Icobench [BENCH, 2018]. As shown in figure 2, in 2017 58% (413 units) of ICOs in the market were successful while 42% (305 units) of them failed to raise funds. As reported in figure3 , the market scenario in 2018 is quite dissimilar, since the number of ICOs registered on Icobench increased dramatically and the composition as well, recording a success rate of 40% (1012 units) and a failure rate of 60% (1505 units). As regards the collection of scam ICOs, two studies were taken as a reference, "Cryptocurrencies Initial Coin Offerings: Are they Scams? - An Empirical Study" [Liebau and Schueffel, 2019] and "Initial Coin Offerings, Information Disclosure, and Fraud" [Hornuf et al., 2019] which based their empirical analysis on samples respectively composed by 6.7% and 12.6% of scams. Those are in contrast with findings from a recent study by Satis Group which finds in 2018 that the 80% were identified as scams. The article offers an alternative explanation for the allegedly poor performance of ICOs by relating them to studies from entrepreneurship literature.

The volume of Initial Coin Offerings (ICOs) has risen steeply with an all-time high market capitalisation of close to 1 trillion USD in December 2017. Since then the digital asset market has slumped, retreating to approximately 200 billion USD by mid-2018. Stakeholders of the crypto industry have pondered the reasons for this retrenchment and are increasingly focusing on the notion that many ICOs could be scams. A recent industry study even goes as far to claim that 80% of all ICOs are indeed scams. In this paper, we investigate the question whether these scams are as common as claimed. We do so by first defining what a scam is and secondly, by drawing on empirical data to assess the number of cases fitting such a definition. Building on Principal Agent Theory and based on the statistical analysis of our empirical data set we attempt to establish the current state of affairs with regards to scams in the crypto-currency world. The results of our study divert from salient belief Hornuf et al. [Hornuf et al., 2019] study based their empirical analysis on samples respectively composed by 6.7% and 12.6% of scams. They look at the extent of fraud in initial coin offerings (ICOs), and whether information disclosure prior to the issuance predicts fraud. Issuers that disclose their code on GitHub are more likely to be targeted by phishing and hacker activities, which suggests that there are risks related to disclosing the code. Generally, we find extremely difficult to predict fraud with the information available at the time of issuance. This calls for the need to install a third-party that certifies the quality of the issuers, such as specialized platforms, or the engagement of institutional investors and venture capital funds that can perform a due diligence and thus verify the quality of the project.

To reach the goal of reproducing the real market composition three new variables were created: "Success", "Nat_fail"[3] , "Scam". The classification of each observation into these three classes was made following as close as possible to the criteria used in the work of [Toma and Cerchiello, 2020] as follows:

1. *Success*: the observed ICO successfully collected the predefined cap within the time horizon of the campaign;
2. *Failure*: the observed ICO failed to collect the predefined cap within the time horizon of the campaign;

[3] Nat_fail stands for natural failure

Fig. 2: ICO market structure in 2017 (Icobench.com, 2018)

Fig. 3: ICO market structure in 2018 (Icobench.com, 2018)

3. *Scam*: the observed ICO is discovered to be a fraudulent activity and it is described as such by different sources (e.g. SEC announcements, blog articles, forums).

The most challenging part of the analysis was the data collection. The issues encountered during the data collection process mainly regards the poor quality or even the complete lack of information on rating websites. Most of the time missing information could be found by consulting the website, the white paper or the details page on rating platforms. However, as experienced on multiple occasions, this solution can be quite confusing or ineffective as the same information might result different from website to website or, in the worst case, might be completely missing. The underlying problem is represented by the complete absence of regulation regarding information disclosure of ICO projects. In fact, information published by ICO ventures is unsupervised. This condition also creates perfect circumstances for information asymmetry and moral hazard [Momtaz, 2018] [Momtaz, 2019]. Similar obstacles were experienced in determining ICO scams. Indeed, the aim of the third dataset was to simplify the identification process of ICO scams in the final dataset. The dataset was retrieved from Deadcoins.com, an online platform where the crypto community can freely report tokens characterized by natural failure, phishing, scam or suspicious activity. The issue lies on the fact that reports are not properly checked. Therefore, many observations collected were not really scam or fraud, but they were just the result of bad management or other reasons that adversely affected the project causing its premature failure. For these reasons, the collection of scams does not completely rely on the raw scam dataset retrieved from Deadcoins.com. After an initial match between the two datasets, a robustness check was made on every singular indicted observation to have a more reliable base for classifying them as scam ICOs (e.g. SEC announcements, repeated unofficial announcements in blogs and other medias). The dataset was cleaned by dropping duplicated values and fixing missing values. Table 2 offers a general overview of the features included in the final dataset used for implementing the analysis.

Whilst, we have carefully collected ICOs, according to the 3 labels above described (Success, Failure, Scam), we focused the analysis on disentangling what differs a successful ICO from the rest. This because the amount of reliable information regarding scam is too scarce and would lead to non significant or consistent statistical results. The analysis was conducted using Python 3.8 and R[4] language.

5 Empirical Analysis

Before starting the data modelling, we run some descriptive analysis. Different techniques are used to summarize the main characteristics of the data collected and to improve the general quality of the analysis. A fist general insight into the distribution of quantitative variables is given by using box plots. These charts are particularly effective tools to get a first idea of the distribution of variables (divided by quartiles) and to enable the identification of outliers. Note that variables were standardized for better comparison and visualization. As shown in figure 4, the sample is not well balanced due to the high presence of outliers, especially in the first variable representing the number of members in Telegram groups.

Rather informative can be also comparing the distribution of quantitative features related to the target variable "Success". For visualizing this relationship, box plots were combined with distribution plots distinguishing for both categories of the target variable, 1 for

[4] Python 3.8 retrieved at https://www.python.org/downloads/release/python-380/ - https://www.r-project.org/

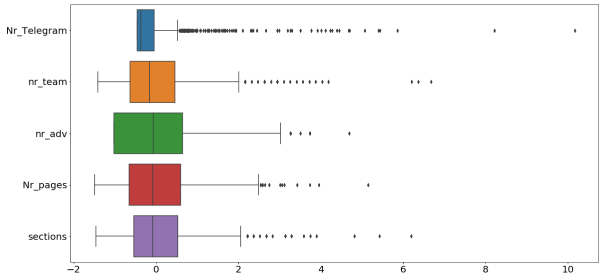

Fig. 4: Box plots of numerical variables

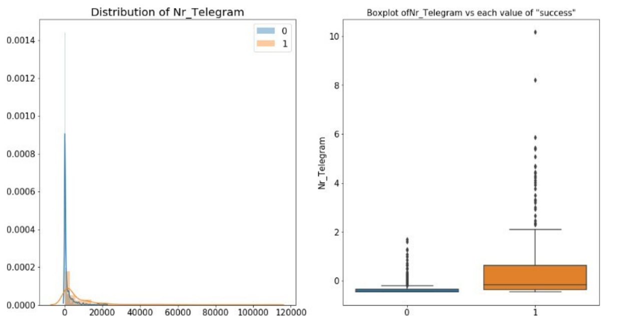

Fig. 5: Distribution of the number of Telegram groups of successful and failed ICOs

success and 0 for failure. As shown in Fig. 5, the observations that do not have a Telegram group, or just a few members, seem to concentrate in the category of failed ICOs while ICOs classified as successful are more likely to have larger a Telegram community (despite the presence of outliers in both categories).

Although the distribution of the number of team members assumes a slightly similar distribution in both cases of success and failure, figure 6 shows a higher concentration of failed ICOs with absence of information of the team or with fewer members involved in the project. A similar observation can be also applied to the number of advisors, since figure 6 shows a weak signal that failed ICOs are more likely to have a lower presence of advisors.

Finally, as regards the variables related to the white papers, both the number of pages and the number of sections seem to have a quite similar distribution for successful and failed ICOs, see figure 7.

Afterwards, the attention of the exploratory data analysis shifted to the distribution of the target variables "success" and "scam". The distribution of the variable "success" is balanced while the latter revels to be excessively unbalanced making it hardly usable for prediction purposes. As shown in Fig. 8, observations in the minority class "scams" are not sufficient to correctly train the algorithm. Therefore, to avoid misleading accuracy and inability to predict rare events as

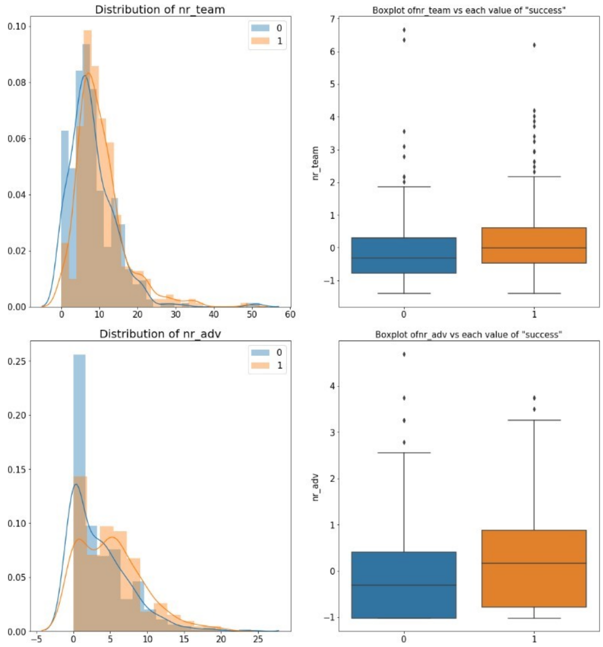

Fig. 6: Distribution of the number of team members and advisors of successful and failed ICOs

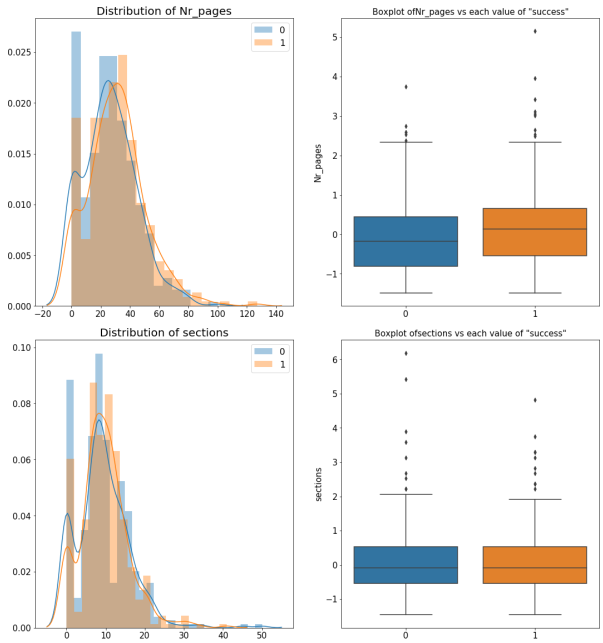

Fig. 7: Distribution of the number of pages and sections of successful and failed ICOs' white papers.

scams, it was chosen to focus the rest of the analysis on just the target variable "success", which contains two classes, 1 for successful ICOs and 0 for failed and scam ICOs.

The final part of the exploratory data analysis is focused on the identification of correlations among variables see figure 9. We take into account high level of cor-

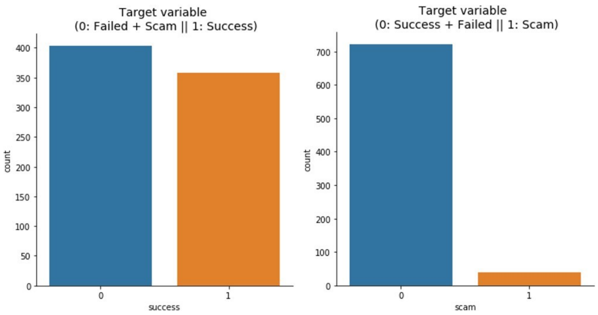

Fig. 8: Comparing the composition of target variables ”success” and ”scam”

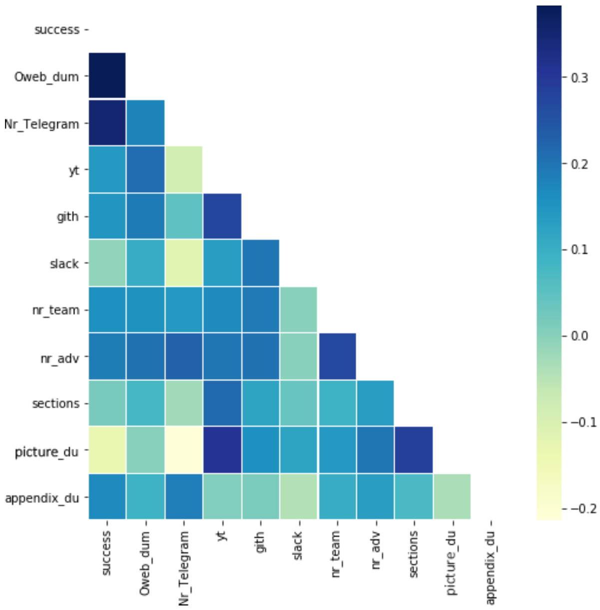

Fig. 9: Correlation heatmap

relations during the estimation of the inferential models to so avoid any multicollinearity issue. Nevertheless, the highest values do not exceed 30% of linear correlation.

As introduced in section 3 the core of the analysis focuses on Logistic Regression, deepening some important aspects of the model and allowing to test it accurately. The model was trained according to several configuration with and without controls.

In table 6 and 7 we report results of $HP1$ controlled over the social media exposure. As it clearly appears the length of the white paper is a driver of success of an ICO: the longer the document the better is the informative inclusion. This conclusion holds even controlling over the different communication channels. Whether we account for the usage of at least one channel ($social\_du$) or we count how many are used ($nr\_channels$), we still have significance. The same applies, if we focus on each specific social media platform (model from 3 through 11 in table 6 and 7).

For sake of robustness, we also check whether the length of the white papers matters, taking into account the composition of the team as well. We investigate if a team, large enough, can compensate a detailed document. In table 8, we can clearly see that is not the case, the length of the white papers still remains positive and significant, confirming our $HP1$.

The validity of our analysis is also corroborated by the confirmed importance of social media exposure and team characteristics variables as in [Toma and Cerchiello, 2020] and in [Cerchiello et al., 2019].

Moving to $HP1bis$, we study an alternative in the measurement of the completeness of white papers: instead of considering the length, we look at the number of sections. As clearly appears from table 9, such hypothesis does not hold at all. The number of sections cannot be considered a proxy of the amount of information disclosure from a given ICO.

We then consider our $HP2$, in particular we investigate the role of a visual representation of an ICO team. Thus, we test the importance of team pictures in the white paper. Do prospect clients put more trust in ICOs that fully disclose their composition and visual identity? Table 10 and 11 report results. We see that the picture variable is significant across all the configuration and surprisingly with a negative sign. In table 12 we also control for the team composition and again we find that the visual identity matters but with a negative sign.

We finally test the last hypothesis $HP3$ which investigates the role of the appendix. Once again, it clearly emerges from Table 13,14 and 15that the appendix plays a relevant and positive role with regards to the success of an ICO. The appendix, which often contains a technical and more detailed description, helps in increasing the prospect clients knowledge and awareness about the trustworthiness of the ICO. At the same time, both the social media exposure and the team composition keep on being relevant and statistically significant.

Thus for, all but one hypothesis result to be confirmed by our analysis; the length of white papers, the presence of team pictures and the appendix clearly play a role in determining the outcome of an ICO launch. Insofar suggesting that the stronger is the level of information disclosure, the higher is the chance of being funded.

6 Further Robustness check

To further test our results, we compared logistic regression with 5 well-known machine learning classifiers:

1. K-nearest Neighbors

2. Support Vector Machine
3. Naive Bayes
4. Decision Trees
5. Random Forest

After splitting the dataset into train and test set and dividing the dependent variable from independent variables, the target subsets were stratified in order to guarantee the same proportions of class labels in the train and test set. Each model was tested analyzing its accuracy (table 4) and obtaining its confusion matrix (figure 10).

	Accuracy
Logistic regression	0.76
Ridge regression	0.75
Lasso regression	0.75
Random forest	0.75
Naive Bayes	0.73
Linear Svm	0.71
KNN	0.68
Decision Tree	0.67

Table 4: Accuracy score for each model

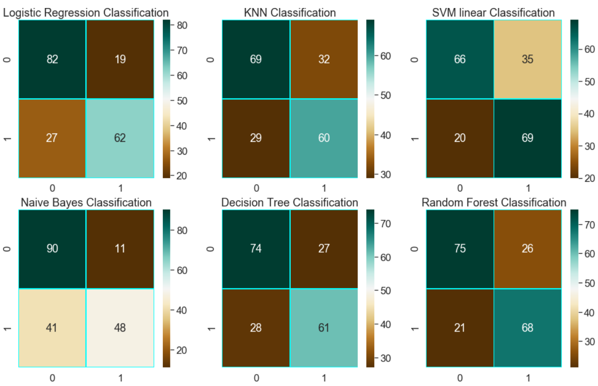

Fig. 10: Confusion matrices for each model

As already seen in figure 8, the target variable “success” seems to be quite balanced. However, it is a good practice to use cross-validation to train the models in each and every fold of the dataset and use the average of all the recorded accuracies for each model. Despite the significant decrease in accuracy, a 5-fold cross-validation was used to avoid overfitting. Figure 11 shows the distribution of cross-validated accuracy scores for each model and table 5 gives a rank of their cross-validated mean scores.

To analyze the performance of the models, ROC curves were also used. With 5-fold cross- validation, random forest seems to obtain the best accuracy score

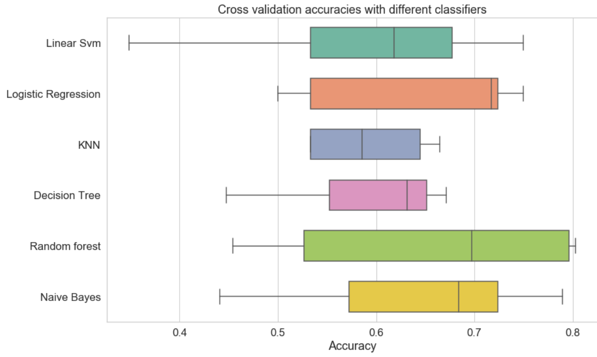

Fig. 11: Distribution of cross-validated accuracy scores

	CV Mean
Random forest	0.66
Logistic regression	0.64
Naive Byes	0.64
Decision Tree	0.60
KNN	0.59
Linear Svm	0.59

Table 5: Mean of cross-validated scores

but looking at the ROC and AUC, Logistic regression seems to perform better since it has better True Positive and False Positive rates (figure 12).

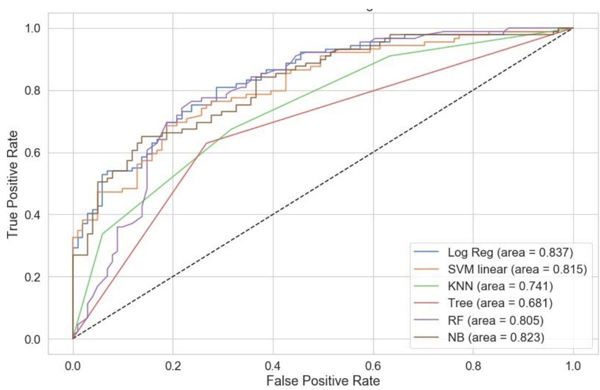

Fig. 12: ROC curves with AUC

7 Conclusions

The recent evolution of the ICO phenomenon tremendously enlarged the availability of data and consequently, its academic literature experienced significant developments. The rapid growth of the market, combined with the stunning results obtained by different ventures, attracted more and more investors which tried to better understand the dynamics of this relatively new approach to raise funds and leverage on the blockchain technology novelty for creating new projects.

The main objective of our analysis is to test if the information disclosed by ICOs is relevant to effectively predict the result of the fundraising process. The disclosed information, on which we rely, takes into account different perspectives such as the entrepreneurial and strategic, the engagement strategy, and the information quality strategy. Some important aspects, highly correlated with the information disclosure process and the presence of criminal activity, are related to the novelty of this phenomenon and the lack of regulation that affects its market. This, in fact, is also reflected in the disclosure of information, which in many cases appears inconsistent, uncertain, or completely missing. The data collecting process took into account this by crosschecking, both manually and automatically, the information disclosed under structured and unstructured data types. The work presented is in line with the relevant empirical works in literature considering a representative sample of the ICO phenomenon [Adhami et al., 2018] [Boreiko and Risteski, 2020]. To investigate rare events, as scams, for instance, the process is even more challenging because ventures that attempt to perform frauds accurately erase their online presence once the ICO is completed. Despite these difficulties, some interesting insights were obtained based on our analysis. As reported in the section 5, the impact of the white paper is assessed as well through the features summarising the technical characteristics of the paper such as the number of pages, the number of sections and the presence of an appendix (containing legal statements) and the presence of pictures in the biography section as a proxy of the reliability of the member's identity. At the same time, we control the impact of those variables by adding the classical success drivers when considering ICOs like: the presence of a working website, the presence of the business on the social media in general and a specific drill down on which channels. From the significance of both the website presence and social media coverage, in particular, it can be deduced that building a consistent online presence and creating a broad community can considerably increase the probability of reaching the fundraising goals.

Our findings let us consider that an all-round marketing strategy leveraging on online presence, community engagements and quality informative white paper is the key for successful projects.

Further improvement of our findings would consider a larger dataset encompassing more ICOs and consideration of additional information. Moreover, similarly to what proposed in [Toma and Cerchiello, 2020], sentiment analysis based on Telegram chats would help even more in defining the ICOs behaviour and success of failure drivers.

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest.

References

[Adhami et al., 2018] Adhami, S., Giudici, G., and Martinazzi, S. (2018). Why do businesses go crypto? An empirical analysis of initial coin offerings. *J. Econ. Bus.*, 100:64–75.

[Amsden and Schweizer, 2018] Amsden, R. and Schweizer, D. (2018). Are blockchain crowdsales the new 'gold rush'? success determinants of initial coin offerings.

[Ante et al., 2018] Ante, L., Sandner, P., and Fiedler, I. (2018). Blockchain-based icos: Pure hype or the dawn of a new era of startup financing? *Journal of Risk and Financial Management*, 11(4):80.

[BaFin, 2018] BaFin (2018). Initial coin offerings: Advisory letter on the classification of tokens as financial instruments.

[BENCH, 2018] BENCH, I. (2018). Ico$_{m}arket_{a}nalysis_{2}018.pdf$.

[Blaseg, 2018] Blaseg, D. (2018). Dynamics of voluntary disclosure in the unregulated market for initial coin offerings. *Draft*.

[Block et al., 2020] Block, J. H., Groh, A., Hornuf, L., Vanacker, T., and Vismara, S. (2020). The entrepreneurial finance markets of the future: a comparison of crowdfunding and initial coin offerings. *Small Bus. Econ.*, pages 1–18.

[Boreiko and Risteski, 2019] Boreiko, D. and Risteski, D. (2019). Serial and Large Investors in Initial Coin Offerings. [Online; accessed 9. Nov. 2020].

[Boreiko and Risteski, 2020] Boreiko, D. and Risteski, D. (2020). Serial and large investors in initial coin offerings. *Small Bus. Econ.*, pages 1–19.

[Bourveau et al., 2019] Bourveau, T., De George, E. T., Ellahie, A., and Macciocchi, D. (2019). Information Intermediaries in the Crypto-Tokens Market. [Online; accessed 9. Nov. 2020].

[Catalini and Gans, 2018] Catalini, C. and Gans, J. S. (2018). Initial coin offerings and the value of crypto tokens. *Working paper*.

[Cerchiello et al., 2019] Cerchiello, P., Tasca, P., and Toma, A. M. (2019). ICO Success Drivers: A Textual and Statistical Analysis. *Journal of Alternative Investments*, 21(4):13–25.

[Cerchiello and Toma, 2018] Cerchiello, P. and Toma, A. (2018). Icos success drivers: a textual and statistical analysis. *Draft*.

[Cohney et al., 2019] Cohney, S., Hoffman, D., Sklaroff, J., and Wishnick, D. (2019). Coin-operated capitalism. *Columbia Law Rev.*, 119(3):591–676.

[Coinschedule, 2019] Coinschedule (2019). Crypto token sales market statistics. *Coinschedule.com*.

[Cuthbertson, 2019] Cuthbertson, A. (17.05.2019). Bitcoin prices rollercoaster continues as value plunges by usd 1000 in less than an hour. *The Independent (Online)*.

[ESMA, 2018] ESMA (2018). Advice to esma. *Own Initiative Report on Initial Coin Offerings and Crypto Assets*.

[European Securities and Markets Authority, 2019] European Securities and Markets Authority (2019). Advice on initial coin offerings and crypto-assets. Technical report, European Securities and Markets Authority.

[Feng et al., 2019] Feng, C., Li, N., Wong, M. H. F., and Zhang, M. (2019). Initial Coin Offerings, Blockchain Technology, and White Paper Disclosures. [Online; accessed 9. Nov. 2020].

[Fisch, 2019a] Fisch, C. (2019a). Initial coin offerings (icos) to finance new ventures. *Journal of Business Venturing*, 34(1):1.

[Fisch, 2019b] Fisch, C. (2019b). Initial coin offerings (ICOs) to finance new ventures. *Journal of Business Venturing*, 34(1):1–22.

[Florysiak and Schandlbauer, 2019] Florysiak, D. and Schandlbauer, A. (2019). The information content of ico white papers. *International Corporate Finance eJournal*.

[Hahn and Wilkens, 2019a] Hahn, C. and Wilkens, R. (2019a). Ico vs. ipo - prospektrechtliche anforderungen bei equity token offerings. *Zeitschrift fuer Bankrecht und Bankwirtschaft*.

[Hahn and Wilkens, 2019b] Hahn, C. and Wilkens, R. (2019b). Ico vs. ipo - prospektrechtliche anforderungen bei equity token offerings. *Zeitschrift fuer Bankrecht und Bankwirtschaft*, (1).

[Hastie et al., 2009] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

[Hornuf et al., 2019] Hornuf, L., Kück, T., and Schwienbacher, A. (2019). Initial Coin Offerings, Information Disclosure, and Fraud. [Online; accessed 20. Oct. 2020].

[Howell et al., 2018] Howell, S. T., Niessner, M., and Yermack, D. (2018). Initial coin offerings: Financing growth with cryptocurrency token sales. *Working paper*.

[Huang et al., 2020] Huang, W., Meoli, M., and Vismara, S. (2020). The geography of initial coin offerings. *Small Bus. Econ.*, 55(1):77–102.

[James et al., 2017] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017). *An Introduction to Statistical Learning: with Applications in R*. Springer.

[Kher et al., 2020] Kher, R., Terjesen, S., and Liu, C. (2020). Blockchain, Bitcoin, and ICOs: a review and research agenda. *Small Bus. Econ.*, pages 1–22.

[Kickstarter, 2019] Kickstarter (2019). Kickstarter statistics. *Kickstarter (Online)*.

[Liebau and Schueffel, 2019] Liebau, D. and Schueffel, P. (2019). Cryptocurrencies initial coin offerings: Are they scams? - an empirical study. *The Journal of the British Blockchain Association*, 2:1–7.

[Mollick and Robb, 2016] Mollick, E. and Robb, A. (2016). Democratizing innovation and capital access: The role of crowdfunding. *California Management Review*, 58(2):72–87.

[Momtaz, 2020a] Momtaz, P. (2020a). Ceo emotions and underpricing in initial coin offerings. *SSRN Electronic Journal*.

[Momtaz, 2018] Momtaz, P. P. (2018). Initial Coin Offerings. [Online; accessed 16. Oct. 2020].

[Momtaz, 2019] Momtaz, P. P. (2019). CEO Emotions and Underpricing in Initial Coin Offerings. [Online; accessed 16. Oct. 2020].

[Momtaz, 2020b] Momtaz, P. P. (2020b). Initial coin offerings, asymmetric information, and loyal CEOs. *Small Bus. Econ.*, pages 1–23.

[Nakamoto, 2008] Nakamoto, S. (October 31, 2008). Bitcoin: A peer-to-peer electronic cash system. Technical report, www.cryptovest.co.uk.

[Pilkington, 2018] Pilkington, M. (2018). The Emerging ICO Landscape - Some Financial and Regulatory Standpoints. [Online; accessed 13. Oct. 2020].

[Schückes and Gutmann, 2020] Schückes, M. and Gutmann, T. (2020). Why do startups pursue initial coin offerings (ICOs)? The role of economic drivers and social identity on funding choice. *Small Bus. Econ.*, pages 1–26.

[Shin, 2017] Shin, L. (2017). Here's the man who created icos and this is the new token he's backing. *Forbes Online*.

[Spence, 1973] Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*.

[Stacher, 2018] Stacher, D. (2018). Regulation of initial coin offering: An international comparison with focus on switzerland. Master's thesis, Universitaet Basel.

[Toma and Cerchiello, 2020] Toma, A. M. and Cerchiello, P. (2020). Initial coin offerings: risk or opportunity? *Frontiers in Artificial Intelligence*, 3:18.

Table 6: Results of logistic regression for HP1 about the relevance of the length of the white papers, controlled withe the social media exposure

	Dependent variable:					
	success					
	(1)	(2)	(3)	(4)	(5)	(6)
Nr_pages (scaled)	0.255***	0.227***	0.148*	0.214***	0.182**	0.194**
	(0.075)	(0.075)	(0.078)	(0.077)	(0.077)	(0.077)
social_du		2.419**				
		(1.039)				
nr_channels			0.175***			
			(0.036)			
tw				1.981***		
				(0.365)		
fb					1.050***	
					(0.208)	
ln						0.547***
						(0.156)
Constant	-0.122*	-2.510**	-1.201***	-1.953***	-0.985***	-0.456***
	(0.073)	(1.036)	(0.236)	(0.356)	(0.191)	(0.121)
Observations	760	760	760	760	760	760
Log Likelihood	-519.410	-514.121	-506.944	-497.680	-505.538	-513.173
Akaike Inf. Crit.	1,042.820	1,034.242	1,019.887	1,001.360	1,017.076	1,032.347

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: Results continuation from table 6

	Dependent variable:						
	success						
	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Nr_pages (scaled)	0.201***	0.211***	0.255***	0.212***	0.275***	0.210***	0.153*
	(0.076)	(0.076)	(0.075)	(0.076)	(0.076)	(0.076)	(0.081)
yt	0.519***						
	(0.157)						
gith		0.570***					
		(0.161)					
slack			-0.040				
			(0.180)				
reddit				0.500***			
				(0.173)			
btalk					-0.322*		
					(0.188)		
mm						0.578***	
						(0.175)	
Web_dum							2.307***
							(0.258)
Constant	-0.452***	-0.509***	-0.114	-0.493***	0.137	-0.555***	-2.029***
	(0.125)	(0.133)	(0.082)	(0.149)	(0.168)	(0.152)	(0.244)
Observations	760	760	760	760	760	760	760
Log Likelihood	-513.918	-513.053	-519.385	-515.174	-517.934	-513.813	-461.102
Akaike Inf. Crit.	1,033.836	1,032.105	1,044.770	1,036.348	1,041.868	1,033.625	928.205
Note:							*p<0.1; **p<0.05; ***p<0.01

Table 8: Results of logistic regression for $HP1$ controlled over the composition of ICO team

	Dependent variable:	
	success	
	(1)	(2)
Nr_pages (scaled)	0.180**	0.144*
	(0.078)	(0.079)
nr_team	0.045***	
	(0.013)	
nr_adv		0.082***
		(0.019)
Constant	-0.528***	-0.472***
	(0.136)	(0.110)
Observations	760	760
Log Likelihood	-512.758	-510.045
Akaike Inf. Crit.	1,031.515	1,026.089
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 9: Results of logistic regression for $HP1bis$ controlled over the social media exposure

	Dependent variable:					
	success					
	(1)	(2)	(3)	(4)	(5)	(6)
sections	0.006	-0.016	0.001	-0.011	0.001	-0.003
	(0.011)	(0.013)	(0.011)	(0.012)	(0.011)	(0.011)
scale(Nr_pages)		0.307***				
		(0.085)				
social_du			2.615**			
			(1.037)			
nr_channel				0.202***		
				(0.036)		
nr_team					0.053***	
					(0.012)	
nr_adv						0.094***
						(0.019)
Constant	-0.174	0.035	-2.714***	-1.255***	-0.608***	-0.498***
	(0.129)	(0.143)	(1.034)	(0.236)	(0.165)	(0.145)
Observations	760	760	760	760	760	760
Log Likelihood	-525.273	-518.579	-518.747	-508.295	-515.448	-511.670
Akaike Inf. Crit.	1,054.546	1,043.158	1,043.494	1,022.589	1,036.897	1,029.340
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 10: Results from logistic regression for $HP2$ regarding the importance of pictures in white papers controlled over the social media exposure

	Dependent variable:						
	success						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Picture_du	-0.608***	-0.719***	-0.719***	-1.255***	-0.738***	-0.896***	-0.933***
	(0.168)	(0.173)	(0.173)	(0.201)	(0.177)	(0.183)	(0.184)
social_du		2.983***	2.983***				
		(1.043)	(1.043)				
nr_channel				0.302***			
				(0.041)			
tw					2.164***		
					(0.368)		
fb						1.380***	
						(0.218)	
ln							0.907***
							(0.166)
Constant	0.329**	-2.528**	-2.528**	-1.053***	-1.570***	-0.589***	0.015
	(0.145)	(1.036)	(1.036)	(0.240)	(0.370)	(0.210)	(0.157)
Observations	760	760	760	760	760	760	760
Log Likelihood	-518.757	-509.904	-509.904	-487.539	-492.723	-495.843	-503.027
Akaike Inf. Crit.	1,041.514	1,025.808	1,025.808	981.078	991.446	997.685	1,012.054
Note:					*p<0.1; **p<0.05; ***p<0.01		

Table 11: Results: continuation of results from table

	Dependent variable:						
	success						
	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Picture_du	-0.951***	-0.762***	-0.613***	-0.755***	-0.620***	-0.874***	-0.736***
	(0.186)	(0.175)	(0.169)	(0.175)	(0.181)	(0.182)	(0.187)
yt	0.907***						
	(0.170)						
gith		0.778***					
		(0.165)					
slack			0.042				
			(0.181)				
reddit				0.743***			
				(0.176)			
btalk					0.035		
					(0.200)		
mm						0.921***	
						(0.185)	
Web_dum							2.411***
							(0.261)
Constant	0.006	-0.083	0.324**	-0.113	0.310*	-0.162	-1.564***
	(0.158)	(0.170)	(0.147)	(0.180)	(0.182)	(0.177)	(0.271)
Observations	760	760	760	760	760	760	760
Log Likelihood	-503.720	-507.206	-518.729	-509.524	-518.741	-505.644	-454.887
Akaike Inf. Crit.	1,013.440	1,020.411	1,043.459	1,025.047	1,043.483	1,017.287	915.774

Note: *p<0.1; **p<0.05; ***p<0.01

Table 12: Results for logistic regression on $HP2$ controlled over the composition of the ICO team

	Dependent variable:		
	success		
	(1)	(2)	(3)
Picture_du	-0.608***	-0.777***	-0.860***
	(0.168)	(0.176)	(0.178)
nr_team		0.064***	
		(0.013)	
nr_adv			0.115***
			(0.020)
Constant	0.329**	-0.121	0.023
	(0.145)	(0.171)	(0.155)
Observations	760	760	760
Log Likelihood	-518.757	-505.442	-499.649
Akaike Inf. Crit.	1,041.514	1,016.884	1,005.297
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 13: Results for logistic regression for $HP3$ on the importance of the appendix in a white paper controlled over the social media exposure

	Dependent variable:						
	success						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
appendix_du	1.055***	1.077***	1.073***	0.985***	1.067***	1.030***	1.071***
	(0.233)	(0.237)	(0.238)	(0.238)	(0.239)	(0.235)	(0.236)
social_du		2.691***					
		(1.043)					
nr_channel			0.196***				
			(0.035)				
tw				1.988***			
				(0.365)			
fb					1.143***		
					(0.208)		
ln						0.617***	
						(0.153)	
yt							0.620***
							(0.156)
Constant	-0.252***	-2.909***	-1.460***	-2.080***	-1.192***	-0.625***	-0.648***
	(0.078)	(1.042)	(0.235)	(0.357)	(0.194)	(0.123)	(0.129)
Observations	760	760	760	760	760	760	760
Log Likelihood	-514.344	-507.515	-497.793	-492.425	-497.616	-506.138	-506.270
Akaike Inf. Crit.	1,032.688	1,021.029	1,001.586	990.850	1,001.231	1,018.275	1,018.541
Note:					*p<0.1; **p<0.05; ***p<0.01		

Table 14: Results: continuation results table 13 with a control over the the social media exposure

	Dependent variable:					
	success					
	(8)	(9)	(10)	(11)	(12)	(13)
appendix_du	1.064***	1.055***	1.061***	1.044***	1.051***	0.963***
	(0.236)	(0.233)	(0.235)	(0.234)	(0.235)	(0.251)
gith	0.652***					
	(0.161)					
slack		-0.001				
		(0.181)				
reddit			0.600***			
			(0.172)			
btalk				-0.172		
				(0.187)		
mm					0.660***	
					(0.174)	
Oweb_dum						2.326***
						(0.259)
Constant	-0.695***	-0.251***	-0.697***	-0.112	-0.745***	-2.161***
	(0.137)	(0.088)	(0.152)	(0.171)	(0.154)	(0.247)
Observations	760	760	760	760	760	760
Log Likelihood	-505.961	-514.344	-508.107	-513.920	-506.919	-454.965
Akaike Inf. Crit.	1,017.922	1,034.688	1,022.215	1,033.840	1,019.838	915.931
Note:				*p<0.1; **p<0.05; ***p<0.01		

Table 15: Results from logistic on the relevance of the appendix in white papers controlled for the characteristics of the ICO team

	Dependent variable:		
	success		
	(1)	(2)	(3)
appendix_du	1.055***	0.992***	0.936***
	(0.233)	(0.236)	(0.237)
nr_team		0.049***	
		(0.013)	
nr_adv			0.084***
			(0.019)
Constant	-0.252***	-0.689***	-0.599***
	(0.078)	(0.137)	(0.110)
Observations	760	760	760
Log Likelihood	-514.344	-506.018	-503.396
Akaike Inf. Crit.	1,032.688	1,018.036	1,012.793
Note:	*p<0.1; **p<0.05; ***p<0.01		