Do Digital Coins Have Fundamental Values? Evidence from Machine Learning

Yukun Liu Jinfei Sheng Wanyi Wang*

December 2019

Abstract

The rapid development of digital coins has raised questions about whether it is driven by technology innovations or investor speculations. Using machine learning techniques, we construct a novel technology index (Tech) of individual digital coin from their Initial Coin Offering (ICO) whitepapers. We find that the ICOs with high Tech Indexes are more likely to succeed and less likely to be delisted. Moreover, the Tech Index strongly and positively predicts ICOs' long-run performance. Overall, the results suggest that fundamentals are an important driving force for the valuation of ICOs.

Keywords: Initial coin offerings, Whitepaper, Machine learning, Textual analysis, FinTech, Blockchain.

*Yukun Liu is with Simon School of Business at University of Rochester. Jinfei Sheng (Corresponding Author) is with Merage School of Business at University of California, Irvine. Email: jinfei.sheng@uci.edu. Wanyi Wang is with Merage School of Business at University of California, Irvine. For helpful comments, we thank Chong Huang, Arthur Iinuma, Ye Li, Fangzhou Lu (discussant), Yushui Shi, Siew Hong Teoh, Aleh Tsyvinski, David Yang, and conference and seminar participants at University of California Irvine and 2019 Conference on Financial Economics and Accounting at New York University. All errors are our own.

1 Introduction

The rise of FinTech is one of the most critical developments in finance in the past decade. One important FinTech solution is initial coin offering (ICO), where investors can purchase blockchain-based digital tokens directly from entrepreneurs. 7.8 billion U.S. dollars were raised through ICOs in 2018 alone.[1] While ICO provides a new way of fundraising, there are growing concerns about whether speculations fuel the development of the market. For example, Satis—a security token advisory firm—claims that over 80 percent of ICOs in 2017 were scams.[2] Still, many digital coins, such as Bitcoin, are highly valued on the market. Despite the on-going discussions, little is known about how to measure the fundamentals of digital coins and the implications of them.

The theoretical literature of cryptocurrency suggest that investors participate in an ICO for two primary reasons (e.g., Cong, Li, and Wang, 2018; Sockin and Xiong, 2018; Prat, Danos, and Marcassa, 2019). First, they may prefer the underlying products and the convenience coming from transactions associated with the ICO. That is, the convenience yield of the coins. One example is utility tokens. Utility tokens give their holders the right to access products or services. Howell, Niessner, and Yermack (2019) document that 68 percent of ICOs have apparent utility values. Second, participants may purchase tokens for the expected capital gains. This reason of holding cryptocurrencies is closely tied to speculative motives of the investors. For example, Sockin and Xiong (2019) theoretically show that investor sentiments can drive cryptocurrency prices. In this paper, we study whether fundamentals play a role in the valuation of digital coins. If

[1] https://www.icodata.io/stats/2018.

[2] For the full report, please see: https://research.bloomberg.com/pub/res/d28giW28tf6G7T_Wr77aU0gDgFQ

cryptocurrency prices are purely driven by speculations, fundamentals should not matter when investors choose which ICOs to participate in.

What are the fundamentals of digital coins? Many investors of digital coins believe that the blockchain technology embodied in coins may become an important innovation and that at least some coins are assets that represent a stake in the future of this technology. Due to this nature of digital coins, we focus on the technology aspect of coin fundamentals. Several recent theoretical papers in digital coins also echo with this viewpoint and emphasize the importance of technology in determining the viability and valuation of digital coins. For example, Hinzen, John, and Saleh (2019) and Fanti, Kogan, and Viswanath (2019) show that the pricing implications of different technologies used in setting up the cryptocurrency platforms. Biais, Bisiere, Bouward, and Casamatta (2019) emphasize the platform security in determining the cryptocurrency valuations.

It is challenging to measure the technology component of a coin because there is limited information about it during the ICO. To overcome the challenge, we go to the disclosure of the coins—their whitepapers—to measure the technology components employed in the digital coins. In particular, we construct a technology index (Tech Index) from a comprehensive database of ICO whitepapers. We use a machine learning method—word embedding—to capture the importance of technology in an ICO. We validate the Tech Index using anecdotal evidence and data from GitHub. We show that the Tech Index positively associates with the measures indicating code quality from GitHub. The results show that the Tech Index offers a good proxy for the fundamentals of ICOs.

We study the relationship between the Tech index and ICO valuations. We first examine whether the Tech Index is related to ICO fundraising. If the entrepreneurs cannot raise any funding, the ICO is not likely to succeed. Therefore, the ability to raise funding is one of the most important steps in a successful ICO. If all ICOs are fully driven by speculations, investors would not care about the technology associated with the ICOs. Under this hypothesis, the technology index should not relate to ICO success. We find that ICOs with a high Tech Index is more likely to raise capital and more likely to be traded in the secondary market subsequently. The economic magnitude of the effect is significant. For instance, a one standard deviation increase in the Tech Index is associated with a 7.59 percent increase in the probability that the ICO raised capital, which is a 29 percent increase of the average. The results suggest that investors care about the underlying technology of the ICOs.

We also investigate whether the underlying technology of ICOs is associated with the ICO underpricing phenomenon as documented in Benedetti and Kostovetsky (2018). It is difficult to know whether ICOs with better technology would experience less or more underpricing ex ante. If investors are able to process and are attentive to the technology-related information, we would expect that the ICOs with high tech index experience similar underpricing compared to the ICOs with low tech index. However, if investors are not able to understand or process technology-related information, we may find that the ICOs with high tech index experience stronger underpricing compared to other ICOs. We find that the Tech Index is positively associated with ICO underpricing, indicating that the underpricing phenomenon is more severe for technologically advanced coins. This result suggests that investors have difficulty in fully incorporating technology-related information about ICOs, and thus undervalue the ICOs with strong underlying technology. This phenomenon may relate to the complexity of blockchain technology or come from investor inattention.

While ICOs with high Tech Indexes are not recognized by investors in the short run, they may eventually be valued by investors in the long run. The process to fully incorporate technology-related information may take months due to the complexity of blockchain technology. To test this conjecture, we examine the relationship between the Tech Index and the long-run performance of ICOs. We measure long-run performance using cumulative post ICO returns, abnormal returns, and liquidity measures. We find that the ICOs with higher Tech Indexes tend to have better performance in the long run compared to other ICOs.

We also investigate whether our indexes help understand ICO failure measured by delisting. We find that the ICOs with higher Tech Indexes are less likely to be delisted subsequently. The economic magnitude of the effect is also large. For instance, a one standard deviation increase in the Tech Index leads to a 29 percent decrease in delisting probability.

Overall, these results suggest that the underlying technology is an important determinant of digital coin prices, and support the argument that investors do take the technological components in the ICO whitepapers into their consideration. However, it takes time for investors to fully differentiate the fundamentally sound ICOs from the others. The delayed reaction from investors may be caused by investor inattention or the complex nature of ICOs, both of which necessitate more time to process related information.

Related Literature

This paper contributes to fast-growing literature on the economics of ICOs and digital assets in general. Harvey (2016) provides a general introduction for digital assets. Yermack (2017) is the first paper to explore the finance implications of blockchain. Liu and Tsyvinski (2018) provide one of the first comprehensive analyses of the risk-return tradeoff of cryptocurrencies. Liu, Tsyvinski, and Wu (2019) examine the cross-section of cryptocurrency and establish a cryptocurrency three-factor model. Recently, several theoretical papers examine the rationale and mechanisms of ICOs and cryptocurrencies (Cong and He 2018; Cong, He and Li, 2018; Cong, Li and Wang 2018; Chod and Lyandres 2018; Catalini and Gans 2018; Biais et al. 2019; Li and Mann 2018; Sockin and Xiong 2018). Our paper is closely related to Cong, Li and Wang (2018) and Sockin and Xiong (2019), which argue that the value of cryptocurrency is fundamentally anchored by the underlying utility value. In other words, their model predict that digital coins have fundamental values and the fundamental values are crucial for performance. However, there is little evidence showing the importance of the fundamental values of digital coins because it is hard to measure that empirically. A set of empirical papers study factors that contribute to ICO success, including Benedetti and Kostovetsky (2018), Howell, Niessner and Yermack (2019), Deng, Lee and Zhong (2018), and Lee, Li and Shin (2019). In general, they find social media and team play a significant role in ICO success and performance. Although some prior papers touch about whitepapers (e.g., Dittmar and Wu 2018, Florysiak and Schandlbauer 2019, Chen, Li, Wong and Zhang 2019, Lyandres, Palazzo and Rabetti 2019), our paper is the first paper that tries to measure the fundamental values of ICOs using a machine learning method and account for the relationship between whitepapers with ICO short and long-run performance. Our Tech Index appears to play a significant role in explaining ICO success, underpricing, and long-term performance, all of which are not well understood in the literature.

This paper provides support to the theoretical literature that links cryptocurrency fundamentals and valuations. Budish (2018), Abadi and Brunnermeier (2018), and Hinzen, John, and Saleh (2019) discuss the limitations of proof-of-work technologies and the pricing implications of them. Fanti, Kogan, and Viswanath (2019) show that the pricing implications of proof-of-stak. Biais, Bisiere, Bouward, and Casamatta (2019) emphasize the platform security in determining the cryptocurrency valuations. Consistent with the theoretical implications of the literature, our paper shows that the technology components affect the valuations of cryptocurrencies.

Our study also adds to the literature on textual analysis in finance (see Tetlock 2014 and Gentzkow et al., 2017 for reviews). Our paper is among the first set of papers that use the machine learning method “word embedding” to analyze financial text. Other papers using a word embedding model include Li et al. (2018) where they use this method to measure corporate culture from earnings call transcripts. One big advantage of word embedding is that it allows to identify synonyms from neighboring words. This is based on the linguistic concept that words tend to co-occur with neighboring words with similar meanings (Harris 1954). Other textual analysis methods, such as word count, treat all words as equal and split sentences into words (i.e., bag of words). This type of method is useful if researchers have enough knowledge about what keywords they are looking for (see e.g., Fisher, Martineau, and Sheng 2019, Liu and Matthies 2018). When the researchers do not have a prior about the keywords, machine learning techniques become useful. For example, Sheng (2019) uses Latent Dirichlet Allocation technique to study the text contents of Glassdoor employee outlook. ICO is a new phenomenon and most participants have limited knowledge about simple concepts in an ICO, let alone its technological components. Therefore, we employ an unsupervised machine learning method to detect latent topics from the text. Our results show that word embedding can be a powerful method in analyzing questions in finance and economics.

The rest of the paper is organized as follows. Section 2 explains the background of ICOs and the data we use. Section 3 introduces the construction and validation of the Tech Index. Section 4 describes our empirical results. We conclude and discuss implications for policy in Section 5.

2 Background and Data

2.1 Initial Coin Offering (ICO) basics

In a typical ICO, entrepreneurs issue digital assets (“tokens”) that are implemented on a blockchain or a contract to deliver such tokens in the future (e.g., a Simple Agreement for Future Tokens, or SAFT). Entrepreneurs then use the raised capital to create an online platform or ecosystem where the native token can be used.

In general, these tokens can be classified into three types based on their purposes. The first type is called “utility token” because its purpose is to redeem a product or service in the future. This is the largest group of tokens and it is the focus of our paper. The second type is called “security token”, which is similar to a conventional security but recorded and exchanged on a blockchain to reduce transaction costs and create a record of ownership. This type of token gives holders the rights for associated cash flows, such as dividends. The third type is called “asset token”, which serves as a general-purpose medium of exchange and store of value. These are often termed “coins”, such as Bitcoin.

Initial Coin Offering is appealing to both start-up companies and investors. The start-up companies that choose to issue ICOs are usually those that "conventionally finance themselves with angel or venture capital (VC) investment" (Howell, Niessner, and Yermack 2019). ICO are attractive to them because it allows them to avoid regulations from SEC and intermediaries such as venture capitalists and banks, leading to lower financing cost and easier access to capital. Investors invest in ICOs for various reasons. Some investors may believe in the intrinsic value of the project and are optimistic about the technological innovations embedded therein. Other investors are speculators who are attracted by the quick cash-out ability.

The first ICO was issued by Mastercoin in July 2013. In 2014, Ethereum also launched a token sale and raised over $15 million to support its development. Since 2017, ICO has become popular and 875 startups successfully raised capital using token sales during the year (see Figure 1). As of February 2019, ICOs have raised over 25 billion USD.[3]

As a new source for seed and early-stage funding, ICOs raise money from many small investors over the Internet. In that sense, the ICOs is similar to crowdfunding where investors get future rewards or deals on products and investor get securities for exchange. However, ICOs are different from crowdfunding in that they are blockchain-based and involve more advanced technology for their products and services. ICOs are also similar to initial public offerings (IPOs) in the sense that tokens can be listed on one or more cryptocurrency exchanges, so investors can benefit from the price appreciation of a listed token even before the project launches. This process is usually much faster than that of IPOs. It ranges from several days to several months, but there is no guarantee of listing.

2.2 Data on ICOs

Our dataset consists of three different components: ICO characteristics from *trackico.com*, daily trading data from *coinmarketcap.com,* and textual measures from ICO whitepapers. There are over 4,100 ICOs on *trackico.com*, with 2,452 closed, 575 trading, 264 ongoing, 82 pre-sale, 307 upcoming and 422 unknown. We focus on ICOs between January 2017 and December 2018. The final sample consists of 2,916 ICOs which raised more than $17 billion in total. For each ICO, we collect the following information: ICO start and end date, ICO price, total capital raised, trading

[3] Source: https://icobench.com/.

status, pre-ICO, bonus, platform, accepted currency, the founder team, country, industry, links of whitepaper, official website, GitHub and Twitter.

We define two measures of ICO success. The first one is *Trading*, a self-reported dummy variable by fundraiser to *trackico.com*, indicating whether the token is trading on cryptocurrency exchanges. The second one is *Success*, which equals to 1 if an ICO successfully raised any capital (Benedetti and Kostovetsky, 2018). Other ICO characteristics serve as control variables. *ICO length* is the number of days between the start and end of an ICO. *ICO price* is the cost per token in US dollars. *Total Raised* is the amount of money raised in millions of US dollars. *Pre ICO*, *Bonus*, *Ethereum Based* and *Accept BTC* are indicator variables about whether the ICO has a pre-ICO, offers bonus to investors, is built on Ethereum platform and accepts Bitcoin as a payment currency. *Team size* is calculated as the number of team members. We define *Has GitHub* and *Has Twitter* to be dummy variables of whether the fundraiser has a GitHub or a Twitter homepage. We further control for Bitcoin price on the ICO start date or the coin's listing day as a proxy for the market sentiment. Finally, we control for quarterly, categorical and geographical (continent-level) fixed effects.

Next, we merge ICO data with information from *coinmarketcap.com*, the leading information source of cryptocurrency trading data. It is also a primary choice in the ICO literature. By the end of 2018, *coinmarketcap.com* has provided data for over 3,600 cryptocurrencies, among which 2,070 are active while 1,583 are delisted. We collect daily opening price and 24h dollar trading volume on all coins from August 2013 to December 2018. We then use token names, ticker symbols and website slugs to merge these variables with our ICO data. Since many coins on *coinmarketcap.com* were not issued through ICO, and many ICOs do not list their coins on any exchange, we get a merged sample of 765 observations.

With the merged sample, we first define a third ICO success measure, *CMC Trading*, which equals to one if the coin has ever appeared on *coinmarketcap.com*. Note that this measure also aims at characterizing the same fact (i.e. whether the coin is traded on an exchange) as the self-reported measure *Trading*, but is more comprehensive.[4] Therefore, we apply *CMC Trading* in our main analysis and consider the other measures in the robustness tests. We define *First Open/ICO Price* to measure the premium on the listing day and *Delist* to characterize whether the coin is delisted from cryptocurrency exchanges. We also calculate the cumulative rate of return, Bitcoin-adjusted rate of return and 24h trading volume after the coin has been listed for 7 days, 30 days, 90 days, 180 days, 240 days and 300 days. These measures capture the short- and long-term performance and liquidity of cryptocurrencies.

The last set of variables comes from textual analysis of ICO whitepapers, which are downloaded from *trackico.com*. We obtained 1,629 valid whitepapers in PDF format. In Table A1, we list all other variations of whitepaper status. Next, we convert PDF files into TXT format, which can be used as raw inputs of textual analysis.

Using this whitepaper corpus, we first construct our main measure, *Tech Index*, which will be explained in detail in Section 3. Moreover, we consider three well-known textual measures as control variables: *Readability*, *Tone* and *Uncertainty*. *Readability* is characterized by *Fog Index*, a widely adopted measure in finance and accounting literature. Developed by Robert Gunning in 1952, *Fog Index* is a linear combination of the percentage of complex words and the average number of words per sentence.[5] *Tone* is the difference between positive and negative words divided by the total number of words, while *Uncertainty* is the percentage of uncertainty words

[4] The correlation between CMC Trading and Trading is 75.8%. Trading is highly accurate if it equals to 1, but is not comprehensive, as we identified approximately 200 more trading tokens on *coinmarketcap.com*.

[5] The complete formula of Fog Index is: $\text{Fog Index} = 0.4[(\frac{\text{words}}{\text{sentences}}) + 100(\frac{\text{complex words}}{\text{words}})]$. "Complex words" are words consisting of three or more syllables.

among all words used in a whitepaper. All lexical categories are defined in Loughran and McDonald (2011).

2.3 Summary Statistics

Table 1 panel A presents summary statistics on variables related to ICO characteristics. On average, it takes 51 days to complete an ICO with a team of 11 people. 18% of ICOs are reported as trading, 26% are initiated in 2017 and 38% had non-zero values of capital raised. Moreover, 60% have a GitHub homepage for their project and over 90% have set up their Twitter accounts.

Table 1 Panel B presents summary statistics on the merged sample. Consistent with the literature, we identify 26% of ICOs that ever list tokens on an exchange. Among them, only 10% are delisted while the remaining 90% are still active. On average, investing in a cryptocurrency during an ICO can earn a premium of 120% on the first trading day, indicating a large amount of money left on the table. Moreover, the return of cryptocurrency investment increases sharply as time goes by, from 19% during a 7-day holding period to 151% during a 300-day holding period. The 24h trading volume fluctuates with different time spans, varying from $1.5 million to $2.78 million. ICO characteristics with respect to the merged subsample are also reported in this panel.

3 Measuring Fundamental: Tech Index

In this section we discuss how we measure the fundamental of digital coins based on their whitepapers. We first present how we construct the Tech Index using a machine learning method, and then we validate the measure by several ways.

3.1 Measure Construction

We now introduce how we use word embedding to measure the technology component of ICOs. Word embedding is an advanced technique in natural language processing (NLP). It maps a word to a vector of real numbers representing the frequency distribution of its neighboring words. The intuitive understanding comes from the famous quotation of Firth, J. R. (1957): "*You shall know a word by the company it keeps.*" To put it in another way, the meaning of a word can be inferred from the distribution of words around it, and words appearing in similar contexts tend to have similar meanings. Compared to traditional approaches to NLP (e.g. bag-of-words model), word embedding can capture semantic relationships between words, which can help us acquire more information from a given text, such as, in our case, the technological component of ICO whitepapers.

We start with the vector representations of whitepaper vocabulary. The word embedding toolkit we use is "word2vec", a two-layer neural network developed by Google in 2013. For every whitepaper, it accepts sentences in whitepapers as input and outputs the vector representation of each word. Since semantically similar words tend to appear in similar contexts, they should be mapped to adjacent vectors in a geometric space.[6] Therefore, we use K-means clustering method to classify all words into five topics and try to find one that best describes technology. Table 2 presents the 50 most frequently used words of each topic.

To identify the most suitable topic, we constructed a test set *ex-ante* by searching "blockchain terminology"/"blockchain dictionary"/"blockchain glossary" online.[7] We find that

[6] To reduce computational burden during classification, we exclude words that appear less than 10 times in over 2,000 whitepapers. Words with such low frequency are basically proper nouns providing no information to the technological content of whitepapers. This simplification largely reduces the vocabulary by 2/3 and greatly speeds up computation.
[7] Here are some examples: https://blockgeeks.com/guides/blockchain-glossary-from-a-z/; https://medium.com/my-blockchain-bible/101-blockchain-terminology-874f007c0270; https://www.blockchaintechnologies.com/glossary/.

Topic 4 is most closely related to technology, because 58% of words in the test set are classified as this topic. We enumerate the test set and the topics they belong to in Table A3. The Tech Index, our main measure, is the percentage of words in a whitepaper that belong to Topic 4. Figure 2 shows the word cloud of the 50 most frequent words of the technology topic. We report the summary statistics of the Tech Index in both Panel A and Panel B of Table 1. An average whitepaper has a Tech Index of 17.6, meaning that 17.6% of the words in whitepapers are used to discuss technology-related topics.

Despite its widespread use in computer science, word embedding is a relatively new approach in finance. A more common topic modeling approach in the financial literature is the Latent Dirichlet Allocation (LDA) model. However, there are two main benefits of using word embedding: first, each word can only be classified into one topic, which reduces ambiguity; second, this further allows us to use a test set to choose the right topic, which minimizes subjective interventions.

3.2 Measure Validation

One concern about the Tech Index is that whitepaper may be an advertising tool, reflecting the *strategic disclosure* of ICO issuers rather than the *fundamental value* of ICO tokens. That is, the Tech Index may capture how technological the fundraisers want the ICO to be looked like, instead of the fundamental of the ICO.

To mitigate these concerns, we validate the Tech Index with data from GitHub. GitHub is an open-source online platform which provides repository hosting service for developers. Many ICO projects voluntarily disclose code on GitHub to increase their transparency and credibility (Howell, Niessner and Yermack, 2019). Since developing code is much more costly than writing a whitepaper, and all codes on GitHub are open to public for inspection and improvement, GitHub data is less likely to suffer from the strategic disclosure and manipulation concern. If the Tech Index measures the technology aspect of ICOs, we would expect a positive correlation between the Tech Index and the measures indicating code quality.

Using the API provided by GitHub, we obtain the number of (1) users subscribing updates of the repository (*watch*), (2) "likes" received by the repository (*star*), (3) copies made by other developers (*fork*), (4) code revisions (*commit*), (5) pointers to specific versions (*branch*) and (6) developers who have contributed to the source code (*contributor*). These measures are often used by researchers to proxy for product quality and post-ICO technological development (Deng, Lee and Zhong, 2018; Dittmar and Wu, 2018; Lyandresy, Palazzoz and Rabetti, 2019). We aggregate these metrics from the repository level to the project level and convert them to logarithmic form.

Table 3 documents the results that relate the Tech Index to GitHub measures. Panel A shows that the Tech Index is positively correlated with these measures from GitHub, with all correlation coefficients ranging from 34 percent to 38 percent. Moreover, this strong positive correlation only exists for the technology topic identified in the previous section. Panel B shows the result of univariate regression between the Tech Index and GitHub measures. The positive relationship is highly significant at 1 percent level for all specifications. Figure 3 provides graphical evidence and confirms this positive relationship.

We provide two additional validation tests in the appendix. First, we use two examples to illustrate that the Tech Index does provide a sensible measure of technology relevance of an ICO. The intuition is that, technologically sound projects often use many technology-related words in the whitepaper, so that they can provide a concrete and precise description about the details of the project. However, for ICO projects without a real technological foundation, the description is usually vague and lacks specific technological content. Our Tech Index successfully captures the nuances between the two types of projects. Second, we compare the Tech Index across different industries. *Trackico.com* divides all ICO projects into 20 industries (categories). We calculate the average Tech Index of each industry and sort them from high to low in Figure A1. The top three industries according to the Tech Index are connectivity, software and Internet, while the lowest three industries are real estate, production and ecology. In general, industries that are more relevant to technology have higher Tech Index.

Taken together, results in this section show that whitepapers are not just an advertising channel and that the Tech Index constructed from whitepapers is a good proxy for the fundamental values of ICO projects.

4 Empirical Results

In this section, we examine whether the technology component of ICOs are associated with ICO success, short-run, and long-run performances. We evaluate an ICO using both its fund-raising stage information and its subsequent performance data.

4.1 ICO Success

First, we study the set of characteristics in ICO whitepapers that are most related to ICO success. We use two ways to measure ICO success. The first measure is based on whether the cryptocurrency is listed on the *Coinmarketcap.com* (CMC trading) and the second measure is based on whether the ICO successfully raised capital. If the entrepreneur cannot raise any funding, the ICO is not likley to succeed. Therefore, the ability to raise funding is one of the most important steps in a successful ICO. If investors care about the fundamentals, especially the technology components, of ICOs, we should expect that it is easier for ICOs with more sophisticated technologies to raise funding. Companies voluntarily disclose whitepapers to communicate with investors in the fund-raising stage, and one of the primary ways that investors evaluate coins is through whitepapers. If whitepapers indeed inform investors about the different aspects of the ICOs, we would be able to extract information from the whitepapers. We use Tech Index to summarize the characteristics of ICO whitepapers. The index is constructed using the word embedding method described above.

Table 4 documents the results that relate ICO whitepapers' characteristics to ICO successes. Panel A of Table 4 presents results based on *CMC trading* and Panel B presents results based on whether the cryptocurrency successfully raised capital. We report coefficient estimates for each of the two whitepaper indexes as well as the control variables. Time, categorical and geographic fixed effects are included for some specifications when indicated.

Panel A shows that Tech Index is positively associated with the *CMC trading* dummy, suggesting that when the tech index is high, the cryptocurrency is more likely to be listed on *Coinmarketcap.com*. The positive relationship is highly significant at the 1 percent level. The economic magnitude is large. The coefficient estimate on the Tech Index is 0.019 and the standard deviation of the tech index is 5.84. In other words, a one standard deviation increase in the tech index leads to an increase of the listed probability by 11.1 percent—a 43 percent increase of the average of the listed probability.

Panel B measures ICO success based on whether the ICO raised capital (*Success* dummy). The coefficient estimates are largely consistent to those in Panel A. Tech index remains highly statistically and positively significant. The coefficient estimate on the tech index is 0.013, which suggests that a one standard deviation increase in the tech index is associated with a 7.59 percent increase of the probability that the ICO raised capital—a 20 percent increase of the average.

Further evidence that Tech Index serves as an important factor for ICO success is the R^2 . For example, Column (1) in Panel A of Table 4 shows that the single variable of Tech Index explains 5% of the variation of CMC trading, while other 12 variables' marginal contribution to the R^2 is about 10% (Column (2)). Interestingly, judging from the R^2 , Quarterly Fixed Effects, seem to be the most important factor because it explains 16% more variation of the *CMC Trading* variable (Column (3)). In other words, the timing of the ICOs is important in determining whether they can successfully raise capital. Nevertheless, Tech Index is still one of the most important factors that contribute to the success of an ICO.

Taken together, the results show that when the ICO whitepaper contains more discussion on technology related topics, the ICO is more likely to be successful.

4.2 ICO Underpricing

Extensive research has shown that there is substantial underpricing in initial public offerings in the equity market. Recently, Benedetti and Kostovetsky (2018) document a similar underpricing phenomenon in the initial coin offering market. In this section, we study whether the Tech Index helps to explain the underpricing phenomenon of ICOs. If investors can process the information of whitepapers and understand the technological components of the ICOs, we would expect that the ICOs with high tech index experience similar underpricing compared to the ICOs with low tech index. However, if investors fail to completely incorporate the technology of ICOs, then we would expect that the ICOs with high tech index experience stronger underpricing compared to other ICOs.

Our measure of ICO underpricing is defined as the natural logarithm of the ratio between first opening price and the ICO offer price. By definition, the sample only includes coins with trading records. Table 5 reports the results for ICO underpricing. Quarterly, categorical and geographic fixed effects are included for some of the specifications.

We find that the Tech Index is positively associated with the ICO underpricing phenomenon. In other words, the underpricing phenomenon is more severe for coins with more technically advanced whitepapers. The economic magnitude of the coefficient estimate is large. The coefficient estimate on the Tech Index is 0.068 and the standard deviation of the Tech Index is 5.84. A one standard deviation increase in the Tech Index leads to an increase of the listed probability by 0.39–an 18 percent increase of the average of the underpricing measure.

The Tech Index is strongly and positively associated with both the ICO success measure and the ICO underpricing measure. These two results suggest that, although coin market investors take the technical aspects of coins into consideration, they fail to fully incorporate the information.

4.3 Long-term performance

In this section, we investigate whether the Tech Index helps to understand the medium- to long-horizon ICO returns. While ICOs with higher Tech Index are not recognized by investors in the short run, they will eventually be valued by investors in the long run. In the equity market, initial public offerings tend to underperform in the long run (see Ritter, 1991). In sharp contrast, initial coin offerings perform well in the medium to long horizon (see Benedetti and Kostovetsky, 2018).

In order to study the speed of information acquisition of the investors, we ask whether overperformance in the ICO markets is related to the ICO Tech Index.

We track the subsequent returns of the ICOs over different horizons—from 7 days to 300 days. In each specification, we exclude coins that are delisted and hence the samples across different horizons are not constant. In the robust section, we examine the effect of delisting on the results.

First, we look at how the Tech Index affects the subsequent performance of initial coin offerings. The results are documented in Table 6. We find that the Tech Index is positively associated with the ICO's subsequent performance. The point estimates are positive across all horizons. The point estimates are insignificant at the 7-day and 30-day horizons but start to become significant in longer horizons. At the 90-day horizon, the point estimate increases to 0.029, indicating a 16.9 percent increase in cumulative returns at this horizon for one standard deviation increase in the tech index. At the 300-day horizon, a one standard increase in the tech index leads to a statistically significant 40.9 percent increase in cumulative returns. Figure 4 plots the point estimate of the coefficient on Tech Index with confidence intervals. The pattern from the plot shows that the impact of Tech Index increases with horizons.

Our return measures are not adjusted to other factors so far. A common factor that is important for the ICO market is Bitcoin returns. Thus, we also conduct a similar exercise with abnormal returns that are adjusted to Bitcoin returns. Table 7 reports the results of this test and shows similar results in terms of statistical significance and economic magnitude as in Table 6.

Overall, the medium and long horizon results are consistent with the short horizon ICO underpricing results we documented in the previous section. Although coins with high tech scores have high probability to raise funds, investors undervalue these high-tech coins on average. Interestingly, it takes a long time for investors to fully incorporate the information.

4.4 Other Measures of Performance

In this section, we use two additional measures to evaluate ICO performances. The first one is the liquidity measure and the second one is the delisting probability measure.

We measure coins' liquidity as the log transformation of the 24-hour trading volume. On average, we find that liquidities are higher for older coins, consistent with Howell, Niessner and Yermack (2019). We examine the relationships between characteristics of whitepapers and coins' liquidity measure. We report the results in Table 8. In our model specifications, we include quarterly, categorical, and geographic fixed effects. We find that the tech index is positively associated with coin liquidity. These results are always statistically significant across the different horizons since inception.

We then investigate the relationships between coins' delisting probability and the characteristics of the whitepapers. We define *Delist* as an indicator variable which is equal to 1 if a token is delisted from CMC. The results are reported in Table 9. The results show that coins with high tech scores are less likely to be delisted subsequently. The economic magnitude of such effect is large. For instance, a one standard deviation increase in the Tech Index leads to a 29 percent decrease in delisting probability.

The results in this section highlight that coins with high tech scores are intrinsically superior. The results provide supports to our argument that the investors in the coin market take technical aspects of the ICOs into consideration. However, as we have shown above, it takes a considerable amount of time for investors to eventually reach proper pricing of the ICOs.

4.5 Tech Index vs. GitHub Measures

Although there is a concern that other measures of technology of ICOs such as source code on GitHub are already good proxies, we show that our results are not affected after controlling these measures. In this section, we provide further evidence that our results are driven by Tech Index rather than other measures by using the orthogonal Tech Index. Specifically, we regress Tech Index on a GitHub measure (commit) and define the residuals as the Tech Index$^{\text{Orthogonal}}$ because this is orthogonal to the GitHub measure. In other words, Tech Index$^{\text{Orthogonal}}$ captures the variation from Tech Index that is not related to GitHub measures. We then re-estimate several tests using this Tech Index$^{\text{Orthogonal}}$.

Table 10 reports the results. Column (1) shows that Tech Index$^{\text{Orthogonal}}$ is positively related to ICO success. The economic magnitude is similar to the main result (i.e., Table 4). Other columns report results for underpricing, long-term returns, and liquidity and the results are consistent with the main results.

4.6 Robustness

In this section, we conduct several robustness tests. First, we use an alternative measure of success, *Trading*, which indicates whether the token is traded on a cryptocurrency exchange. We examine whether Tech Index affects ICO success under this measure and run a similar regression as in

Table 4. Table 11 Panel A reports the result. The coefficient on Tech Index is positive and significant and support the same conclusion as in Table 4.

Second, we use linear regression in Table 4 where the dependent variable is a binary variable. Alternatively, we can use a Logit or Probit model. Table 11 Panel B reports the results from a Logit regression and finds similar results as in Table 4. The results from Probit (untabulated) are robust as well.

Third, it is well-documented that we have to impute delisted returns for equity to avoid survivor bias in the data (Shumway 1997). The equity return data from CRSP automatically contain imputed returns for delisted stocks. For the same reason, we may need to consider impute returns for delisted ICOs. We set a large negative value -99% as their returns after listed for all delisted ICOs. We then redo the tests on whether Tech Index affect short-run and long-run returns with and without adjusting Bitcoin returns as in Table 6 and 7. Table 11 Panels C and D report the results. Similar to the results in Tables 6 and 7, ICOs with higher Tech Index tend to outperform in the long-run. The economic magnitudes are also close. Interestingly, we find that Tech Index has impacts on relatively short-run (7 and 30 days) returns after adjusting for delisted returns. This is not the case for abnormal returns adjusted by Bitcoin returns.

5 Conclusion

There are two views about cryptocurrency and blockchain technology. The first view is that the cryptocurrency market represents bubbles and fraud. The second one believes that the value of the cryptocurrency market comes from the innovative technologies and that a stake in cryptocurrencies is an investment in the future of the technology. This study contributes to this debate by providing a measure of fundamental of ICOs from textual analysis of ICO whitepapers. We construct a text-based Technology Index (Tech Index) from a comprehensive sample of ICOs' whitepapers. We find that the ICOs with higher Tech-Index are more likely to succeed and less likely to be delisted subsequently. Although the Tech-index does not statistically significantly affect the short-run returns of ICOs, it has positive impact on their long-run performance. In short, our findings suggest that fundamental is an important driving force for the growing and performance of ICOs.

Our findings have important policy implications. Although SEC has launched several initiatives on regulating ICOs, there are no clear disclosure requirements. Our results show that the disclosures such as whitepaper are potentially important for the long-term development of the cryptocurrency market. Thus, it might be useful to set up a requirement or guideline for formats or necessary components in the whitepaper. There is a natural analogy for disclosure requirement for public firms (e.g., 10K) and financial firms (e.g., 497K for mutual funds).

References

Abadi, Joseph and Markus Brunnermeier, 2018, Blockchain economics, Working paper, Princeton University.

Benedetti, H. and L. Kostovetsky, 2018, Digital tulips? Returns to investors in initial coin offerings. Working paper, Boston College.

Biais, Bruno, Christophe Bisiere, Matthieu Bouvard, and Catherine Casamatta, 2019, Equilibrium bitcoin pricing, Working paper, Toulouse School of Economics.

Borri, Nicola, 2019, Conditional tail-risk in cryptocurrency market, *Journal of Empirical Finance*.

Budish, Eric, 2018, The economic limits of bitcoin and the blockchain, Working paper, University of Chicago.

Catalini, Christian, and Joshua S. Gans, 2018, Initial coin offerings and the value of crypto tokens, Working paper, National Bureau of Economic Research.

Chen, Feng, Nan Li, Barry Lu, M. H. Wong, and Mingyue Zhang, 2019, Initial coin offerings, blockchain technology, and white paper disclosures, Working paper, University of British Columbia and University of Toronto.

Chod, Jiri, and Evgeny Lyandres, 2018, A theory of ICOs: Diversification, agency, and information asymmetry, Working paper, Boston College and Boston University.

Cong, Lin W., and Zhiguo He, 2019, Blockchain disruption and smart contracts, *Review of Financial Studies*.

Cong, Lin W., Zhiguo He, and Jiasun Li, 2018, Decentralized mining in centralized pools, Working paper, University of Chicago and George Mason University.

Cong, Lin W., Ye Li, and Neng Wang, 2018, Tokenomics: Dynamic adoption and valuation, Working paper, University of Chicago, Ohio State University, and Columbia University.

Dittmar, Robert, and Di Wu, 2018, Returns to initial coin offerings: An empirical examination, Working paper, University of Michigan.

Deng, Xin, Yen Teik Lee, and Zhengting Zhong, 2018, Decrypting coin winners: Disclosure quality, governance mechanism, and team networks. Working Paper, Shanghai University of Finance and Economics.

Fanti, Giulia, Leonid Kogan, Pramod Vismanath, 2019, Economics of proof-of-stake payment systems, Working paper, MIT.

Firth, John R. 1957, A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis* (1957).

Fisher, Adlai, Charles Martineau, and Jinfei Sheng, 2019, Macroeconomic attention and the stock market, Working paper, University of British Columbia.

Florysiak, David, and Alexander Schandlbauer, 2019, The information content of ICO white papers, Working paper, University of Southern Denmark.

Gunning, R., 1952. The technique of clear writing.

Harris, Z. S., 1954. Distributional structure, *Word* 10, 146–162.

Harvey, Campbell R., 2016, Cryptofinance, Working paper, Duke University.

Hinzen, Franz, Kose John, and Fahad Saleh, 2019, Bitcoin's fatal flaw: The limited adoption problem, Working paper, New York University.

Howell, Sabrina, Marina Niessner, and David Yermack, 2019, Initial coin offerings: Financing growth with cryptocurrency token sales, Working paper, New York University.

Lee, Jongsub, Tao Li, and Donghwa Shin, 2019, The wisdom of crowds in FinTech: Evidence from initial coin offerings, Working paper, University of Florida.

Li, Jiasun, and William Mann, 2018, Initial coin offering and platform building, Working paper, George Mason University and UCLA.

Li, Kai, Feng Mai, Rui Shen and Xinyan Yan, 2018, Measuring corporate culture using machine learning, Working paper, University of British Columbia.

Liu, Yukun and Ben Matthies, 2018. Long run risk: Is it there? Working paper, University of Rochester.

Liu, Yukun and Aleh Tsyvinski, 2018. Risks and returns of cryptocurrency, Working paper, University of Rochester and Yale University.

Liu, Yukun, Aleh Tsyvinski, and Xi Wu, 2019. Common risk factors in cryptocurrency, Working Paper, University of Rochester and Yale University.

Loughran, T. and McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), pp.35-65.

Lyandres, Evgeny, Berardino Palazzo, and Daniel Rabetti, 2019, Do tokens behave like securities? An anatomy of initial coin offerings, Working paper, Boston University, Federal Reserve Board and Tel Aviv University.

Prat, Julien, Vincent Danos and Stefania Marcassa, 2019, Fundamental pricing of utility tokens, Working paper, Ecole Polytechnique.

Sheng, Jinfei. Asset pricing in the information age: Employee expectations and stock returns, Working Paper, University of California Irvine.

Shumway, T., 1997. The delisting bias in CRSP data. *Journal of Finance*, 52(1), pp.327-340.

Sockin, Michael and Wei Xiong, 2019, A model of cryptocurrencies, Working paper, Princeton University.

Tetlock, P.C., 2014. Information transmission in finance. *Annual Review of Financial Economics*, 6(1), 365-384.

Yermack, David, 2017, Corporate governance and blockchains, *Review of Finance* 21(1), 7-31.

Appendix

Variable definition

This table provides a detailed description of all variables used in our analysis, including ICO success measures, trading variables, whitepaper measures and ICO characteristics.

Variable	Definition
ICO Success Measures:	
CMC Trading	A dummy variable that equals to one if a cryptocurrency is listed on *coinmarketcap.com* after ICO.
Trading	A self-reported dummy variable of fundraiser to *trackico.com*, indicating whether the token is trading on an exchange.
Success	A dummy variable that equals to one for ICOs that raised capital.
Trading Variables:	
First Open/ICO Price	The ratio between the first day's opening price and the ICO price.
Delist	An indicator about whether a token is delisted from CMC.
Rate of Return	We calculate rate of return by assuming investors buy tokens at the opening price on the first listing day and sell them after at the opening price after the following holding periods: 7 days, 30 days, 90 days, 180 days, 240 days and 300 days.
Trading Volume	We calculate trading volume as the 24h trading volume in millions of USD on the listing day and after they have been listed on *coinmarketcap.com for* 7 days, 30 days, 90 days, 180 days, 240 days and 300 days. Note that these are measures of flows instead of stocks.
Whitepaper Measures:	
Tech Index	The percentage of words in a whitepaper belonging to "technology topic".
Fog Index	A readability measure developed in 1952 by Robert Gunning and widely adopted in finance and accounting literature. It is defined as the linear combination of the percentage of complex words and the average number of words per sentence, where "complex words" are words consisting of three or more syllables.
Tone	The difference between number of positive and negative words defined in Loughran and McDonald (2011) divided by the total number of words in a whitepaper.
Uncertainty	The number of uncertainty words defined in Loughran and McDonald (2011) divided by the total number of words in a whitepaper.

Variable	Definition
ICO Characteristics:	
Has GitHub	A dummy variable that equals to 1 if the ICO project has a GitHub homepage.
Has Twitter	A dummy variable that equals to 1 if the ICO project has a Twitter account.
ICO Length	Number of days from the start to the end of an ICO.
Team Size	Number of team members.
Pre ICO	A dummy variable that equals to 1 if a pre-ICO exists.
Bonus	A dummy variable that equals to 1 if the fundraiser offers bonus (?) to investors.
Ethereum Based	A dummy variable that equals to 1 if the ICO project is built on Ethereum platform.
Accept BTC	A dummy variable that equals to 1 if the ICO accepts Bitcoin as a payment currency.
BTC Price (ICO)	The price of Bitcoin in thousands of US dollars on the day an ICO initiates. We use it to proxy for the market sentiment on that day.
BTC Price (List)	The price of Bitcoin in thousands of US dollars on the day an ICO lists on *coinmarketcap.com*. Also used to proxy for the market sentiment.

Figure 1. ICO Market Overview

This figure plots the number of ICOs and the amount of fund raised in each year from 2014 to 2018. The grey bar represents total money raised in millions of US dollars. The red dashed line represents number of ICOs. Data source: *icodata.io*.

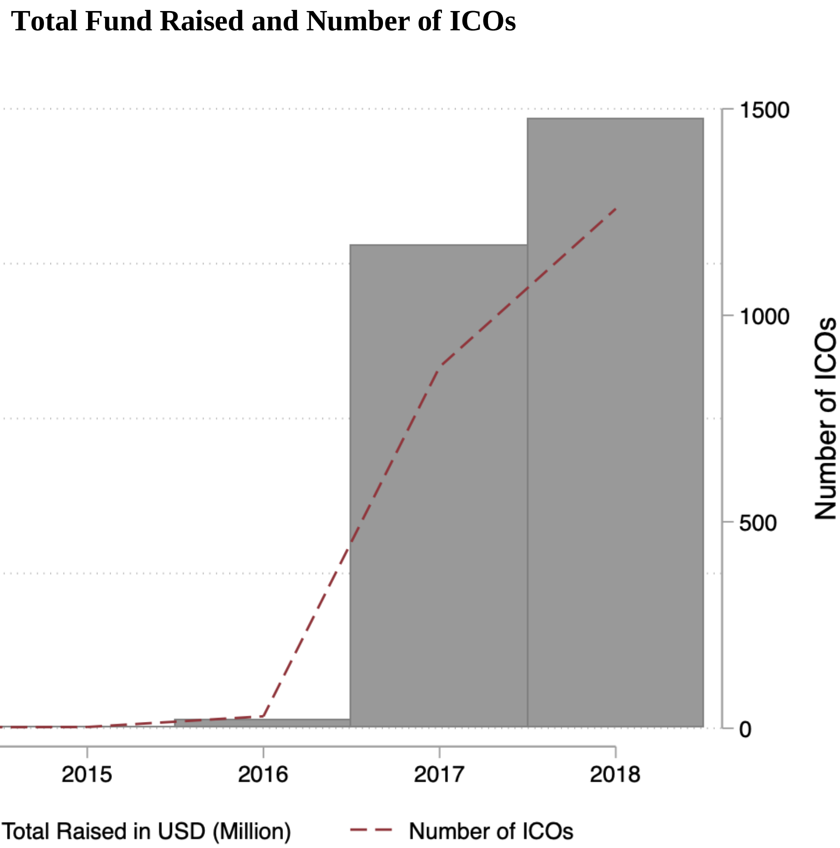


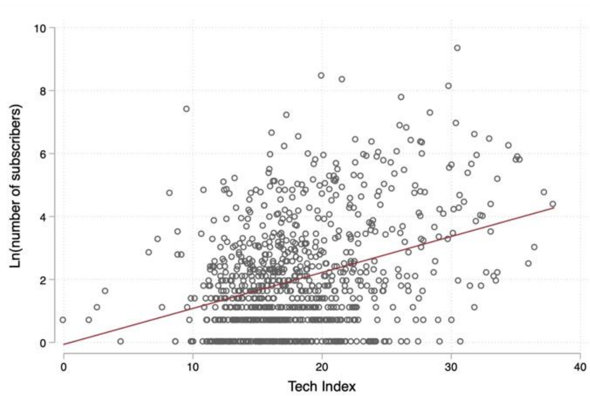
Figure 2. Word Cloud

This figure displays the 50 most frequent words of the "technology topic".

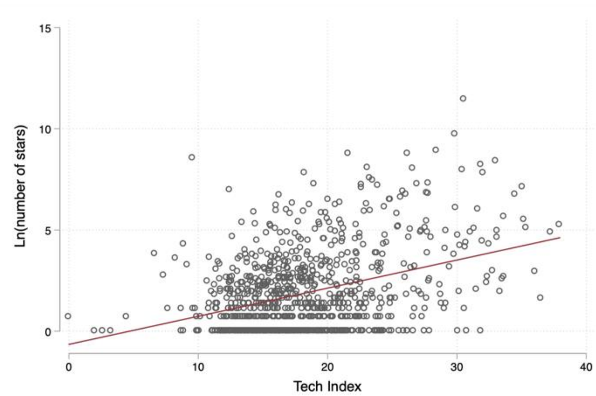
Figure 3. Tech Index validation

This figure illustrates the relationship between *Tech Index* and GitHub measures. In panel A, the variable of interest is *subscriber*, which measures the number of users subscribing updates of the repository; in panel B, *star* indicates the number of "likes" received by the repository; in panel C, *fork* proxies for copies made by other developers; in panel D, *commit* represents how many times the code has been revised; in panel E, *branch* is the amount of pointers to specific versions; and in panel F *contributor* reflects how many developers have contributed to the source. The red solid line represents the linear fitting of GitHub measures on *Tech Index*. Regression results are shown in Table 3.

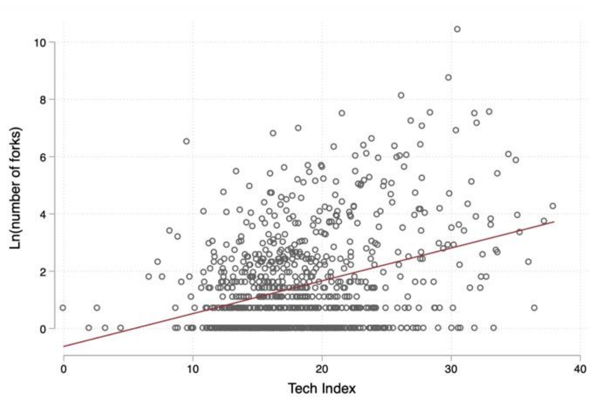
Panel A: subscriber



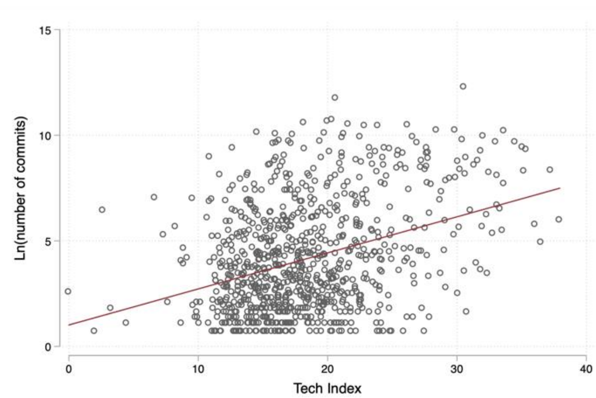
Panel B: star



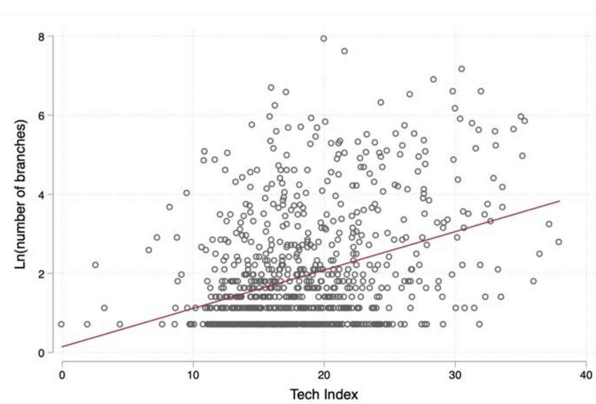
Panel C: fork



Panel D: commit



Panel E: branch



Panel F: contributor

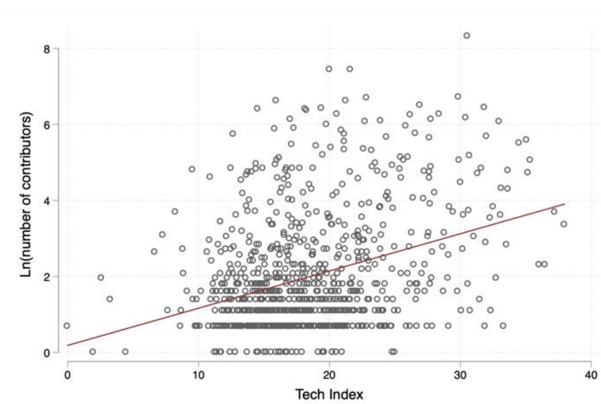
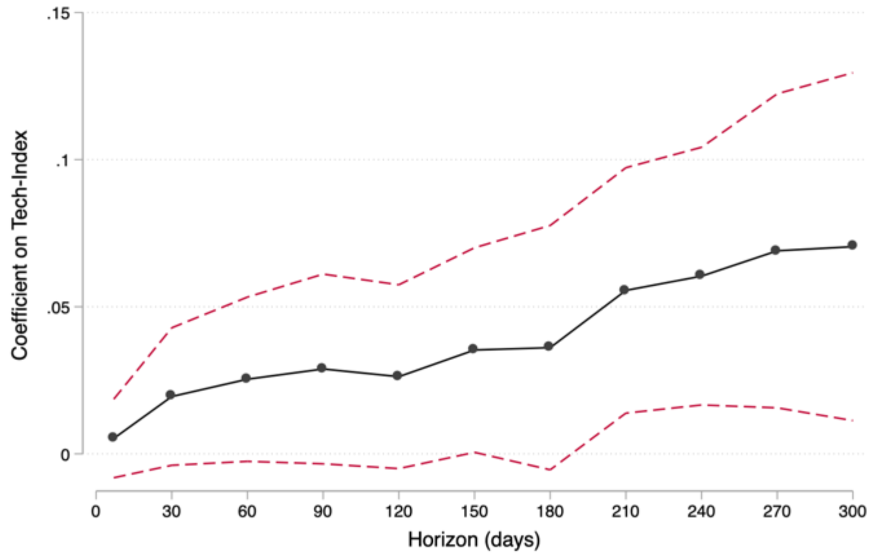


Figure 4. The effect of technology on performance over different horizons

This figure presents the effect of technology on trading performance over different horizons. Panel A and panel B show the coefficient on *Tech Index* for raw rate of returns and Bitcoin-adjusted rate of returns. For each panel, the x-axis represents horizons from 7 days to 300 days and the y-axis is the regression coefficient. The red dashed line represents the 95% confidence interval.

Panel A: Raw rate of return



Panel B: Bitcoin-adjusted rate of return

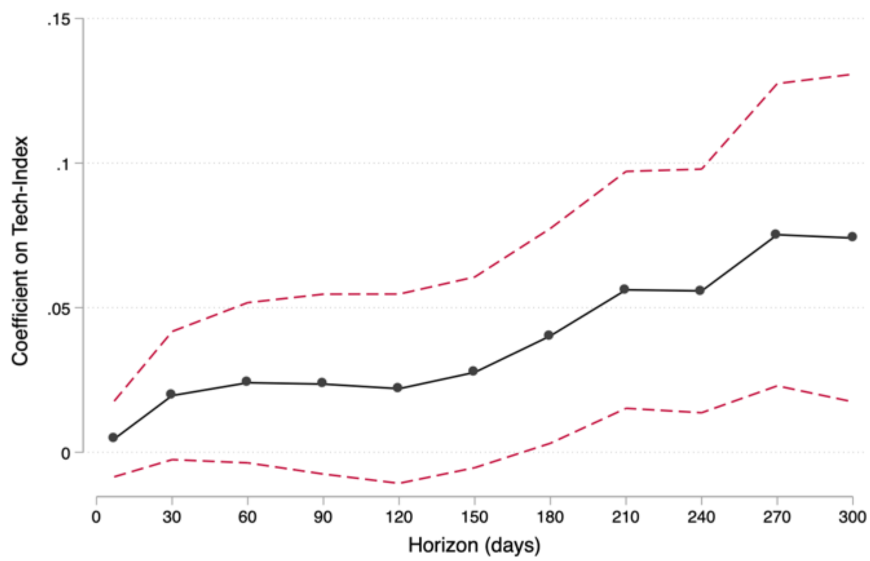


Table 1. Summary statistics

This table presents summary statistics on variables related to ICO characteristics, outcomes and whitepapers. Panel A shows descriptive statistics for 2,916 ICOs that were completed before 12/31/2018. Panel B summarizes a subsample of 765 ICOs listed on *coinmarketcap.com*. For each variable, we provide the number of non-missing observations, mean, standard deviation and the 10^{th}, 50^{th} and 90^{th} percentile values. Please refer to "variable definition" for a detailed definition of each variable.

Panel A: Full Sample						
	Count	**Mean**	**SD**	**P10**	**P50**	**P90**
ICO Success Measures						
CMC Trading	2916	0.26	0.44	0	0	1
Trading	2916	0.18	0.39	0	0	1
Success	2916	0.38	0.49	0	0	1
Whitepaper Measures						
Tech Index	1629	17.6	5.19	12.1	16.9	24.4
Fog Index	1629	16.7	12.6	13.2	15.7	18.5
Tone	1629	0.28	0.73	-0.58	0.29	1.10
Uncertainty	1629	0.75	0.39	0.35	0.67	1.25
ICO Characteristics						
Has GitHub	2916	0.60	0.49	0	1	1
Has Twitter	2916	0.91	0.29	1	1	1
ICO Price ($)	1684	1.57	17.8	0.01	0.10	1
ICO Length	2683	50.7	45.8	14	32	100
Team Size	2916	11.0	7.05	3	10	20
Pre ICO	2916	0.51	0.50	0	1	1
Bonus	2916	0.20	0.40	0	0	1
Ethereum Based	2916	0.83	0.37	0	1	1
Accept BTC	2916	0.40	0.49	0	0	1
BTC Price (ICO, $1000)	2669	7.80	3.07	4.23	7.28	11.3

(Table 1 continued)

Panel B: Listed Sample	**Count**	**Mean**	**SD**	**P10**	**P50**	**P90**
Trading Variables:						
First Open/ICO Price	413	2.20	4.66	0.16	0.97	3.79
Delist	765	0.10	0.30	0	0	1
Rate of Return:						
7 Days	741	0.19	0.83	-0.45	-0.041	1.03
30 Days	730	0.30	1.84	-0.71	-0.28	1.60
90 Days	686	0.65	3.11	-0.87	-0.43	3.21
180 Days	566	1.00	5.14	-0.95	-0.64	4.00
240 Days	486	0.69	3.89	-0.96	-0.70	3.20
300 Days	397	1.51	8.91	-0.97	-0.74	3.85
Bitcoin-adjusted ROR:						
7 Days	700	0.17	0.76	-0.42	-0.04	1.06
30 Days	687	0.23	1.57	-0.69	-0.27	1.51
90 Days	645	0.16	1.77	-0.87	-0.43	1.70
180 Days	525	-0.05	1.68	-0.93	-0.68	1.59
240 Days	448	-0.22	1.36	-0.94	-0.70	0.98
300 Days	358	-0.33	1.21	-0.95	-0.73	0.61
Trading Volume ($ MIL):						
Listing Days	751	2.40	11.9	0.0023	0.12	3.90
7 Days	739	1.63	5.58	0.0015	0.083	3.60
30 Days	725	1.50	5.62	0.0011	0.066	2.53
90 Days	680	1.60	5.58	0.00045	0.11	3.22
180 Days	564	2.78	13.5	0.00039	0.069	3.99
240 Days	482	2.60	12.4	0.00023	0.048	3.31
300 Days	393	2.60	11.5	0.00025	0.058	3.50
Whitepaper Measures:						
Tech Index	422	19.6	5.84	13.1	18.7	27.8
Fog Index	422	17.2	18.9	13.3	15.5	18.3
Tone	422	0.20	0.72	-0.70	0.23	1.03
Uncertainty	422	0.79	0.40	0.35	0.71	1.30
ICO Characteristics:						
Has GitHub	765	0.70	0.46	0	1	1
Has Twitter	765	0.96	0.19	1	1	1
ICO Price ($)	420	2.37	19.9	0.01	0.12	1.22
ICO Length	656	34.9	42.0	2	30	63
Team Size	765	12.1	8.02	3	11	22
Pre ICO	765	0.26	0.44	0	0	1
Bonus	765	0.075	0.26	0	0	0
Ethereum Based	765	0.80	0.40	0	1	1
Accept BTC	765	0.30	0.46	0	0	1
BTC Price (List, $1000)	710	7.59	3.70	2.73	7.03	13.5

Table 2. Most Frequent Words by Topic

This table presents the most frequent 50 words of each topic after removing common stop-words (i.e. meaningless words such as the, an, and). Topic 4 is the "technology topic" we use.

Topic1	Topic2	Topic3	Topic4	Topic5
tokens	development	market	platform	world
token	team	business	blockchain	marketing
use	project	new	data	management
sale	digital	technology	users	global
exchange	ico	ecosystem	network	experience
value	first	crypto	system	years
company	bitcoin	cryptocurrency	user	companies
contract	pre	community	information	online
used	revenue	provide	smart	million
number	main	industry	time	projects
access	launch	most	services	sales
payment	page	trading	based	year
price	start	between	service	research
investment	program	future	transaction	media
funds	stage	well	using	billion
paper	back	financial	transactions	partners
amount	strategy	make	content	capital
whitepaper	day	high	ethereum	group
available	next	about	security	currently
able	plan	create	process	bank
currency	successful	people	public	international
purchase	roadmap	work	wallet	founder
legal	advisors	real	contracts	top
white	generation	way	decentralized	working
own	campaign	need	chain	countries
coin	reserved	products	order	games
investors	volume	social	product	insurance
distribution	compliance	support	key	developer
set	funding	while	protocol	university
assets	live	solution	model	local
risk	last	being	application	professional
without	introduction	growth	private	advisor
distributed	contents	like	app	partner
total	allocation	current	software	manager
account	overview	cost	mobile	medical
part	copyright	get	different	united
money	table	customers	end	member
offer	report	costs	open	production
fees	portfolio	possible	block	commerce
following	net	mining	systems	ceo
fund	here	potential	game	investments
receive	plans	solutions	example	finance
same	gold	however	applications	leading
asset	please	large	level	others
case	dnd	various	nodes	country
risks	daily	platforms	storage	tech
third	summary	due	customer	banking
participants	post	existing	secure	worldwide
made	founders	increase	proof	expert
terms	disclaimer	supply	node	estate

Table 3. Tech Index validation

This table compares *Tech Index* with measures from GitHub. Panel A displays the correlation matrix. Panel B presents the univariate regression results. *Watch* measures the number of users subscribing updates of the repository; *star* indicates the number of "likes" received by the repository; *fork* proxies for copies made by other developers; *commit* represents how many times the code has been revised; *branch* is the amount of pointers to specific versions; and *contributor* reflects how many developers have contributed to the source code. All GitHub measures are in logarithmic form. The reported t-statistics are based on robust standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively.

Panel A: Correlation matrix

	Ln(watch)	Ln(star)	Ln(fork)	Ln(commit)	Ln(branch)	Ln(contributor)	Tech Index
Ln(watch)	1.0000						
Ln(star)	0.8998	1.0000					
Ln(fork)	0.8712	0.9282	1.0000				
Ln(commits)	0.8359	0.7630	0.7640	1.0000			
Ln(branch)	0.8774	0.7952	0.8063	0.8804	1.0000		
Ln(contributor)	0.8639	0.7833	0.8055	0.9130	0.9212	1.0000	
Tech Index	0.3490	0.3740	0.3544	0.3551	0.3581	0.3472	1.0000
Other topics:							
Topic 1	-0.0846	-0.0864	-0.0585	-0.0772	-0.0664	-0.0549	-0.2100
Topic 2	-0.1003	-0.1050	-0.1231	-0.1074	-0.1097	-0.1122	-0.3982
Topic 3	-0.0500	-0.0451	-0.0275	-0.0685	-0.0671	-0.0607	0.0578
Topic 5	-0.1858	-0.2094	-0.2289	-0.1901	-0.2158	-0.2021	-0.3842

Panel B: Univariate regression of GitHub measures and Tech Index

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln(watch)	Ln(star)	Ln(fork)	Ln(commit)	Ln(branch)	Ln(contributor)
Tech Index	0.114***	0.139***	0.115***	0.171***	0.097***	0.098***
	(9.95)	(10.18)	(9.13)	(11.06)	(10.01)	(10.05)
Constant	-0.061	-0.652***	-0.626***	1.015***	0.144	0.186
	(-0.30)	(-2.75)	(-2.92)	(3.58)	(0.84)	(1.08)
Observations	861	861	861	861	861	861
Adjusted R^2	0.121	0.139	0.125	0.125	0.127	0.120

Table 4. ICO success

This table presents OLS estimates of the relationship between *Tech Index* and ICO success. The dependent variable is *CMC Trading* in Panel A and *Success* in Panel B. Column (1) displays univariate result; column (2) includes control variables for ICO characteristics; column (3) also takes several textual measures of whitepapers into account; and column (4) further considers quarterly, categorical and geographical fixed effects. The reported t-statistics are based on robust standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively.

Panel A: CMC Trading

	(1)	(2)	(3)	(4)
Tech Index	0.019***	0.010***	0.011***	0.009***
	(0.002)	(0.002)	(0.002)	(0.002)
ICO Length		-0.002***	-0.002***	-0.001***
		(0.000)	(0.000)	(0.000)
Team Size		0.008***	0.008***	0.009***
		(0.002)	(0.002)	(0.001)
Has GitHub		0.063***	0.061***	0.044**
		(0.022)	(0.022)	(0.021)
Has Twitter		0.129***	0.129***	0.170***
		(0.034)	(0.034)	(0.035)
BTC Price (ICO)		0.011**	0.011**	0.010
		(0.005)	(0.005)	(0.006)
Pre ICO		-0.182***	-0.182***	-0.032
		(0.022)	(0.022)	(0.023)
Bonus		-0.080***	-0.078***	0.010
		(0.023)	(0.023)	(0.022)
Accept BTC		-0.038*	-0.039*	-0.011
		(0.022)	(0.022)	(0.021)
Ethereum Based		-0.021	-0.022	-0.013
		(0.030)	(0.030)	(0.030)
Fog Index			0.000	0.000
			(0.001)	(0.001)
Tone			-0.001	-0.000
			(0.014)	(0.014)
Uncertainty			0.045	0.031
			(0.029)	(0.028)
Constant	-0.078**	-0.045	-0.086	0.409***
	(0.037)	(0.069)	(0.075)	(0.110)
Quarterly FE	N	N	N	Y
Categorical FE	N	N	N	Y
Geographical FE	N	N	N	Y
R^2	0.051	0.160	0.162	0.324
Observations	1629	1483	1483	1382

(**Table 4 continued)**

Panel B: Capital Raised > 0

	(1)	(2)	(3)	(4)
Tech Index	0.013***	0.009***	0.010***	0.007***
	(0.002)	(0.002)	(0.003)	(0.003)
ICO Length		-0.001***	-0.001***	-0.001***
		(0.000)	(0.000)	(0.000)
Team Size		0.008***	0.008***	0.009***
		(0.002)	(0.002)	(0.002)
Has GitHub		0.112***	0.109***	0.082***
		(0.026)	(0.026)	(0.025)
Has Twitter		0.043	0.041	0.091*
		(0.052)	(0.052)	(0.054)
BTC Price (ICO)		-0.001	-0.002	-0.005
		(0.005)	(0.005)	(0.007)
Pre ICO		-0.172***	-0.172***	-0.009
		(0.026)	(0.026)	(0.028)
Bonus		0.001	0.003	0.104***
		(0.031)	(0.031)	(0.031)
Accept BTC		0.032	0.030	0.069***
		(0.025)	(0.025)	(0.024)
Ethereum Based		0.003	0.000	-0.011
		(0.034)	(0.034)	(0.034)
Fog Index			0.000	-0.000
			(0.001)	(0.001)
Tone			0.016	0.011
			(0.017)	(0.018)
Uncertainty			0.084**	0.080**
			(0.034)	(0.033)
Constant	0.145***	0.171**	0.089	0.468***
	(0.042)	(0.082)	(0.090)	(0.135)
Quarterly FE	N	N	N	Y
Categorical FE	N	N	N	Y
Geographical FE	N	N	N	Y
R^2	0.020	0.085	0.089	0.255
Observations	1629	1483	1483	1382

Table 5. ICO underpricing

This table presents OLS estimates of the relationship between *Tech Index* and ICO underpricing. The dependent variable is the log transformation of the ratio between the first day's opening price and ICO price. Column (1) displays univariate result; column (2) includes control variables for ICO characteristics; column (3) also takes several textual measures of whitepapers into account; and column (4) further considers quarterly, categorical and geographical fixed effects. The reported t-statistics are based on robust standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively.

Panel A: Ln(First Opening Price/ICO Price)

	(1)	(2)	(3)	(4)
Tech Index	0.068***	0.057***	0.061***	0.056***
	(0.016)	(0.016)	(0.017)	(0.020)
ICO Length		-0.004*	-0.003*	-0.001
		(0.002)	(0.002)	(0.002)
Team Size		0.015	0.015	0.008
		(0.011)	(0.012)	(0.012)
Has GitHub		-0.219	-0.265	-0.108
		(0.234)	(0.238)	(0.254)
Has Twitter		0.302	0.397	-0.218
		(0.474)	(0.459)	(0.396)
BTC Price (ICO)		-0.041	-0.042	0.002
		(0.029)	(0.029)	(0.048)
Pre ICO		-0.411	-0.401	-0.235
		(0.255)	(0.259)	(0.347)
Bonus		-0.204	-0.242	0.104
		(0.373)	(0.376)	(0.405)
Accept BTC		0.055	0.057	-0.017
		(0.210)	(0.212)	(0.234)
Ethereum Based		-0.049	-0.061	-0.264
		(0.310)	(0.314)	(0.379)
Fog Index			-0.010	-0.006
			(0.007)	(0.007)
Tone			0.228*	0.098
			(0.131)	(0.147)
Uncertainty			0.281	0.090
			(0.311)	(0.322)
Constant	-1.556***	-1.172	-1.394*	-3.882***
	(0.356)	(0.710)	(0.824)	(1.027)
Quarterly FE	N	N	N	Y
Categorical FE	N	N	N	Y
Geographical FE	N	N	N	Y
R^2	0.075	0.122	0.138	0.304
Observations	238	212	212	199

Table 6. Rate of return

This table presents the effects of *Tech Index* on the rate of returns of listed ICOs. The dependent variable is the raw rate of return, calculated by taking log transformation of one plus the rate of returns. Column (1) - (6) display results for six different horizons: 7 days, 30 days, 90 days, 180 days, 240 days and 300 days. We include control variables concerning ICO characteristics and whitepapers in all columns. Quarterly, categorical and geographical fixed effects are considered under all circumstances. The reported t-statistics are based on robust standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	7 Days	30 Days	90 Days	180 Days	240 Days	300 Days
Tech Index	0.005	0.019	0.029*	0.036*	0.060***	0.070**
	(0.007)	(0.012)	(0.016)	(0.021)	(0.022)	(0.030)
ICO Length	0.000	0.000	-0.001	-0.003	-0.003	0.004
	(0.001)	(0.001)	(0.001)	(0.004)	(0.005)	(0.007)
Team Size	-0.001	0.003	0.004	-0.002	0.001	-0.002
	(0.004)	(0.007)	(0.010)	(0.013)	(0.015)	(0.018)
Has GitHub	0.026	-0.012	0.121	0.197	0.203	0.339
	(0.075)	(0.134)	(0.171)	(0.231)	(0.249)	(0.375)
Has Twitter	0.063	0.018	0.028	-0.022	0.045	0.157
	(0.149)	(0.569)	(0.723)	(0.848)	(0.795)	(0.819)
BTC Price (ICO)	0.000	-0.027	-0.056**	-0.090***	-0.071**	-0.082*
	(0.013)	(0.019)	(0.026)	(0.029)	(0.034)	(0.047)
Pre ICO	-0.001	-0.128	0.038	0.236	0.035	0.808
	(0.078)	(0.133)	(0.177)	(0.289)	(0.369)	(0.595)
Bonus	-0.002	0.130	-0.459	-0.178	0.391	-2.336*
	(0.122)	(0.179)	(0.285)	(0.780)	(0.563)	(1.314)
Accept BTC	-0.063	0.018	0.057	-0.302	-0.414	-0.251
	(0.082)	(0.132)	(0.185)	(0.235)	(0.261)	(0.340)
Ethereum Based	0.024	0.104	0.518**	0.364	0.460	0.741
	(0.093)	(0.158)	(0.245)	(0.357)	(0.419)	(0.508)
Fog Index	0.002	0.015***	0.020***	-0.017	-0.006	-0.006
	(0.003)	(0.006)	(0.006)	(0.012)	(0.011)	(0.014)
Tone	0.030	0.051	-0.027	-0.018	-0.038	0.058
	(0.045)	(0.080)	(0.105)	(0.131)	(0.152)	(0.206)
Uncertainty	0.090	0.219	-0.095	-0.268	-0.117	0.228
	(0.105)	(0.181)	(0.224)	(0.274)	(0.312)	(0.488)
Constant	0.000	-0.778	-1.972**	-1.173	-1.434	-2.212
	(0.685)	(1.304)	(0.952)	(1.298)	(1.216)	(1.437)
Quarterly FE	Y	Y	Y	Y	Y	Y
Categorical FE	Y	Y	Y	Y	Y	Y
Geographical FE	Y	Y	Y	Y	Y	Y
R^2	0.077	0.170	0.249	0.362	0.424	0.397
Observations	316	310	286	218	184	140

Table 7. Bitcoin-adjusted rate of return

This table presents the effects of *Tech Index* on the rate of returns of listed ICOs. The dependent variable is Bitcoin-adjusted rate of return, calculated by taking log transformation of one plus the rate of returns. Column (1) - (6) display results for six different horizons: 7 days, 30 days, 90 days, 180 days, 240 days and 300 days. We include control variables concerning ICO characteristics and whitepapers in all columns. Quarterly, categorical and geographical fixed effects are considered under all circumstances. The reported t-statistics are based on robust standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively.

Panel A: Tech Index

	(1)	(2)	(3)	(4)	(5)	(7)
	7 Days	30 Days	90 Days	180 Days	240 Days	300 Days
Tech Index	0.005	0.020*	0.024	0.040**	0.056**	0.074**
	(0.007)	(0.011)	(0.016)	(0.019)	(0.021)	(0.029)
ICO Length	0.000	0.000	-0.000	-0.002	-0.002	0.005
	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.006)
Team Size	-0.000	0.002	0.007	0.004	0.007	0.002
	(0.004)	(0.006)	(0.009)	(0.011)	(0.014)	(0.016)
Has GitHub	0.046	-0.065	0.067	0.139	0.121	0.154
	(0.067)	(0.125)	(0.159)	(0.195)	(0.233)	(0.342)
Has Twitter	0.191	-0.003	0.162	0.534	0.360	0.422
	(0.178)	(0.673)	(0.935)	(0.922)	(0.861)	(0.859)
BTC Price (ICO)	0.009	-0.006	-0.031	-0.071***	-0.049	-0.047
	(0.012)	(0.019)	(0.024)	(0.027)	(0.032)	(0.049)
Pre ICO	-0.008	-0.122	0.050	0.183	0.050	0.643
	(0.073)	(0.115)	(0.175)	(0.249)	(0.326)	(0.516)
Bonus	-0.073	0.146	-0.367	-0.518	0.534	0.000
	(0.112)	(0.177)	(0.283)	(0.628)	(0.742)	(.)
Accept BTC	-0.052	0.007	0.127	-0.111	-0.282	0.001
	(0.080)	(0.125)	(0.172)	(0.210)	(0.242)	(0.302)
Ethereum Based	-0.032	0.033	0.459*	0.184	0.294	0.619
	(0.090)	(0.155)	(0.235)	(0.298)	(0.377)	(0.441)
Fog Index	0.003	0.014***	0.017***	-0.021**	-0.010	-0.012
	(0.003)	(0.005)	(0.006)	(0.010)	(0.012)	(0.014)
Tone	0.028	0.069	0.027	0.015	-0.010	0.125
	(0.044)	(0.079)	(0.097)	(0.115)	(0.135)	(0.185)
Uncertainty	0.100	0.269	-0.053	-0.257	-0.045	0.302
	(0.099)	(0.169)	(0.214)	(0.259)	(0.304)	(0.440)
Constant	-0.158	-0.931	-2.435**	-3.777***	-4.459***	-4.470***
	(0.653)	(1.221)	(1.076)	(1.297)	(1.182)	(1.373)
Quarterly FE	Y	Y	Y	Y	Y	Y
Categorical FE	Y	Y	Y	Y	Y	Y
Geographical FE	Y	Y	Y	Y	Y	Y
R^2	0.098	0.165	0.179	0.335	0.316	0.314
Observations	311	305	281	213	180	137

Table 8. Liquidity

This table presents OLS estimates of the relationship between *Tech Index* and liquidity of listed ICOs. The dependent variable is the log transformation of 24-hour trading volume in US dollars. Column (1) displays results on the listing day. Column (2) to (7) display results for six different horizons: 7 days, 30 days, 90 days, 180 days, 240 days and 300 days. We include control variables for ICO characteristics and whitepapers in all columns. Quarterly, categorical and geographical fixed effects are considered under all circumstances. The reported t-statistics are based on robust standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Listing	7 Days	30 Days	90 Days	180 Days	240 Days	300 Days
Tech Index	0.092***	0.073**	0.075**	0.130***	0.095*	0.133**	0.145**
	(0.032)	(0.033)	(0.036)	(0.037)	(0.055)	(0.055)	(0.068)
ICO Length	-0.004	-0.005	-0.003	-0.005	-0.005	-0.001	-0.001
	(0.004)	(0.006)	(0.005)	(0.007)	(0.009)	(0.008)	(0.017)
Team Size	0.020	0.024	0.034*	0.032	0.088***	0.060**	0.096**
	(0.020)	(0.018)	(0.020)	(0.019)	(0.027)	(0.027)	(0.037)
Has GitHub	0.140	0.429	0.367	0.610	0.769	1.020*	1.082
	(0.366)	(0.391)	(0.469)	(0.511)	(0.560)	(0.580)	(0.774)
Has Twitter	-0.554	-0.117	-0.757	-1.161	0.598	-0.546	-1.242
	(1.263)	(1.200)	(1.797)	(1.824)	(2.200)	(1.649)	(1.527)
BTC Price (ICO)	0.084	0.105	-0.029	0.026	-0.030	-0.022	0.061
	(0.069)	(0.066)	(0.073)	(0.073)	(0.089)	(0.090)	(0.114)
Pre ICO	-0.227	-0.341	-0.299	-0.743	-1.096	-1.741**	-0.005
	(0.426)	(0.493)	(0.525)	(0.593)	(0.759)	(0.810)	(1.197)
Bonus	0.608	0.816	0.640	0.640	-2.127	-0.303	-8.568***
	(0.521)	(0.600)	(0.630)	(0.920)	(1.485)	(1.469)	(2.497)
Accept BTC	-0.135	-0.250	0.441	0.408	0.836	0.526	0.187
	(0.322)	(0.377)	(0.399)	(0.428)	(0.526)	(0.515)	(0.664)
Ethereum Based	0.109	0.061	0.220	0.913	0.105	-0.561	-0.864
	(0.510)	(0.481)	(0.600)	(0.744)	(0.845)	(0.914)	(1.029)
Fog Index	0.007	0.002	0.011	0.015	0.032	0.062*	0.075**
	(0.011)	(0.009)	(0.009)	(0.010)	(0.031)	(0.036)	(0.032)
Tone	-0.170	-0.193	-0.301	-0.297	-0.024	-0.352	-0.098
	(0.259)	(0.249)	(0.293)	(0.330)	(0.389)	(0.420)	(0.519)
Uncertainty	-0.036	-0.086	-0.330	-0.634	-0.431	-0.926	-0.390
	(0.451)	(0.497)	(0.536)	(0.593)	(0.730)	(0.721)	(1.021)
Constant	10.089***	12.549***	9.864***	7.247***	3.096	4.648*	6.489**
	(1.854)	(1.869)	(2.465)	(2.241)	(3.413)	(2.755)	(3.230)
Quarterly FE	Y	Y	Y	Y	Y	Y	Y
Categorical FE	Y	Y	Y	Y	Y	Y	Y
Geographical FE	Y	Y	Y	Y	Y	Y	Y
R^2	0.207	0.223	0.175	0.226	0.342	0.446	0.451
Observations	323	316	308	283	217	183	139

Table 9. Delisting probability

This table presents OLS estimates of the relationship between Tech Index and ICO delisting probabilities. The dependent variable is *Delist*, a dummy variable that equals to 1 if a token was shown as "inactive" in CoinMarketCap by the end of 2018. Column (1) displays univariate result; column (2) includes control variables for ICO characteristics; column (3) also takes several textual measures of whitepapers into account; and column (4) further considers quarterly fixed effects. The reported t-statistics are based on robust standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively.

	(1)	(2)	(3)	(4)
Tech Index	-0.006***	-0.005*	-0.005**	-0.005*
	(0.002)	(0.002)	(0.003)	(0.003)
ICO Length		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)
Team Size		0.002	0.002	0.001
		(0.002)	(0.002)	(0.002)
Has GitHub		-0.038	-0.035	-0.036
		(0.033)	(0.034)	(0.035)
Has Twitter		-0.165	-0.165	-0.166
		(0.154)	(0.153)	(0.156)
BTC Price (ICO)		-0.009***	-0.009***	-0.010**
		(0.003)	(0.003)	(0.004)
Pre ICO		0.017	0.015	-0.003
		(0.030)	(0.030)	(0.040)
Bonus		0.037	0.038	0.019
		(0.053)	(0.053)	(0.055)
Accept BTC		-0.006	-0.005	0.002
		(0.029)	(0.029)	(0.028)
Ethereum Based		-0.018	-0.016	-0.009
		(0.041)	(0.041)	(0.043)
Fog Index			-0.001**	-0.001**
			(0.000)	(0.001)
Tone			-0.034*	-0.031*
			(0.017)	(0.018)
Uncertainty			-0.004	-0.005
			(0.049)	(0.052)
Constant	0.183***	0.391**	0.427**	0.335*
	(0.048)	(0.165)	(0.170)	(0.186)
Quarterly FE	N	N	N	Y
R^2	0.017	0.052	0.061	0.084
Observations	422	358	358	358

Table 10. Tech Index vs. GitHub Measures

This table presents the effects of Tech Index not captured by GitHub measures. Tech Index^{Orthogonal} is defined as the residual from regressing Tech Index on *Has GitHub* and *ln(commit).* The dependent variable is *CMC Trading* in column (1), *Success* in column (2), and *ln(First Opening Price/ICO Price)* in column (3). Column (4)-(6) show results for raw rate of return, Bitcoin-adjusted return and trading volume after listing for 300 days. We include control variables concerning ICO characteristics and whitepapers in all columns. Quarterly, categorical and geographical fixed effects are considered under all circumstances. The reported t-statistics are based on robust standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	CMC Trading	Capital Raised > 0	Underpricing	Ln(ROR) 300 Days	Ln(ROR_{adj}) 300 Days	Ln(volume) 300 Days
Tech Index^{Orthogonal}	0.008***	0.006**	0.057***	0.054*	0.061**	0.137**
	(0.002)	(0.003)	(0.022)	(0.030)	(0.029)	(0.066)
Has GitHub	-0.005	0.013	-0.093	-0.448	-0.458	0.653
	(0.026)	(0.031)	(0.331)	(0.516)	(0.475)	(1.079)
Ln(commit)	0.019***	0.025***	0.024	0.171***	0.143***	0.147
	(0.005)	(0.005)	(0.042)	(0.058)	(0.053)	(0.133)
ICO Length	-0.001***	-0.001***	-0.001	0.004	0.005	-0.001
	(0.000)	(0.000)	(0.002)	(0.006)	(0.005)	(0.016)
Team Size	0.008***	0.008***	0.009	-0.006	-0.001	0.094**
	(0.001)	(0.002)	(0.012)	(0.018)	(0.016)	(0.038)
Has Twitter	0.166***	0.085	-0.223	0.216	0.443	-1.211
	(0.035)	(0.054)	(0.401)	(0.837)	(0.872)	(1.553)
BTC Price (ICO)	0.010	-0.005	0.002	-0.109**	-0.068	0.048
	(0.006)	(0.007)	(0.049)	(0.049)	(0.050)	(0.114)
Pre ICO	-0.031	-0.008	-0.231	0.665	0.541	-0.067
	(0.023)	(0.028)	(0.348)	(0.560)	(0.502)	(1.255)
Bonus	0.012	0.107***	0.102	-3.197**	0.000	-8.994***
	(0.022)	(0.031)	(0.408)	(1.325)	(.)	(2.463)
Accept BTC	-0.008	0.073***	-0.022	-0.240	0.011	0.191
	(0.021)	(0.024)	(0.237)	(0.327)	(0.298)	(0.672)
Ethereum Based	-0.010	-0.007	-0.270	0.667	0.561	-0.899
	(0.030)	(0.034)	(0.383)	(0.494)	(0.435)	(1.029)
Fog Index	0.000	-0.000	-0.006	-0.011	-0.015	0.073**
	(0.001)	(0.001)	(0.007)	(0.014)	(0.014)	(0.032)
Tone	-0.001	0.010	0.101	0.072	0.135	-0.091
	(0.014)	(0.018)	(0.149)	(0.197)	(0.178)	(0.521)
Uncertainty	0.028	0.076**	0.097	0.251	0.320	-0.379
	(0.028)	(0.033)	(0.332)	(0.477)	(0.435)	(1.037)
Constant	0.486***	0.488***	-2.926***	-0.990	-3.167**	8.922***
	(0.100)	(0.124)	(0.960)	(1.298)	(1.257)	(2.925)
Quarterly FE	Y	Y	Y	Y	Y	Y
Categorical FE	Y	Y	Y	Y	Y	Y
Geographical FE	Y	Y	Y	Y	Y	Y
R^2	0.328	0.263	0.304	0.435	0.345	0.453
Observations	1382	1382	199	140	137	139

Table 11. Robustness tests

This table displays several robustness tests. Panel A redoes Table 4 using *Trading* as the dependent variable. Panel B shows another version of Table 4 with Logit regression. Besides, to mitigate the concern of survivorship bias, we impute -99% to the rate of returns and Bitcoin-adjusted returns of delisted ICOs and redo the first panel in table 6 and table 7. Results are presented in Panel C and panel D respectively, with the dependent variable calculated by taking log transformation of one plus the imputed rate of returns or adjusted rate of returns. The reported t-statistics are based on robust standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively.

Panel A: Trading

	(1)	(2)	(3)	(4)
Tech Index	0.017***	0.010***	0.011***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
ICO Length		-0.001***	-0.001***	-0.001***
		(0.000)	(0.000)	(0.000)
Team Size		0.004***	0.004***	0.005***
		(0.001)	(0.001)	(0.001)
Has GitHub		0.012	0.010	-0.004
		(0.018)	(0.018)	(0.016)
Has Twitter		0.090***	0.090***	0.108***
		(0.028)	(0.028)	(0.028)
BTC Price (ICO)		-0.001	-0.001	0.007
		(0.005)	(0.005)	(0.006)
Pre ICO		-0.181***	-0.181***	-0.027
		(0.018)	(0.018)	(0.017)
Bonus		-0.072***	-0.070***	0.022*
		(0.015)	(0.015)	(0.013)
Accept BTC		0.001	-0.000	0.015
		(0.017)	(0.017)	(0.015)
Ethereum Based		0.001	0.001	0.011
		(0.024)	(0.024)	(0.023)
Fog Index			-0.000	-0.000
			(0.001)	(0.001)
Tone			-0.008	-0.004
			(0.012)	(0.011)
Uncertainty			0.042*	0.026
			(0.024)	(0.021)
Constant	-0.136***	0.013	-0.017	0.585***
	(0.033)	(0.061)	(0.066)	(0.078)
Quarterly FE	N	N	N	Y
Categorical FE	N	N	N	Y
Geographical FE	N	N	N	Y
R^2	0.057	0.166	0.169	0.386
Observations	1629	1483	1483	1382

(**Table 11 continued)**

Panel B: Logit Regression

	(1)	(2)	(3)	(4)
Tech Index	0.097***	0.054***	0.058***	0.067***
	(0.011)	(0.013)	(0.013)	(0.017)
ICO Length		-0.013***	-0.013***	-0.012***
		(0.003)	(0.003)	(0.003)
Team Size		0.048***	0.048***	0.070***
		(0.009)	(0.009)	(0.012)
Has GitHub		0.435***	0.421***	0.361**
		(0.149)	(0.150)	(0.175)
Has Twitter		1.084***	1.089***	1.477***
		(0.370)	(0.368)	(0.433)
BTC Price (ICO)		0.064***	0.064***	0.051
		(0.024)	(0.024)	(0.033)
Pre ICO		-1.059***	-1.060***	-0.225
		(0.140)	(0.140)	(0.186)
Bonus		-0.656***	-0.653***	0.127
		(0.199)	(0.201)	(0.245)
Accept BTC		-0.227	-0.231	-0.176
		(0.146)	(0.146)	(0.174)
Ethereum Based		-0.131	-0.144	-0.112
		(0.190)	(0.190)	(0.257)
Fog Index			-0.001	0.000
			(0.007)	(0.008)
Tone			0.021	0.005
			(0.094)	(0.117)
Uncertainty			0.308*	0.224
			(0.186)	(0.238)
Constant	-2.817***	-3.064***	-3.336***	-7.764***
	(0.214)	(0.525)	(0.558)	(1.081)
Quarterly FE	N	N	N	Y
Categorical FE	N	N	N	Y
Geographical FE	N	N	N	Y
Pseudo R^2	0.043	0.159	0.161	0.325
Observations	1629	1483	1483	1351

(**Table 11 continued)**

Panel C: Rate of return

	(1)	(2)	(3)	(4)	(5)	(7)
	7 Days	30 Days	90 Days	180 Days	240 Days	300 Days
Tech Index	0.019*	0.032**	0.039**	0.045**	0.067***	0.070**
	(0.010)	(0.014)	(0.017)	(0.021)	(0.022)	(0.028)
ICO Length	-0.000	0.000	-0.001	-0.004	-0.003	0.002
	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.008)
Team Size	-0.014**	-0.008	-0.006	-0.007	-0.002	0.005
	(0.007)	(0.008)	(0.010)	(0.012)	(0.013)	(0.016)
Has GitHub	0.058	-0.010	0.087	0.129	0.141	0.275
	(0.111)	(0.158)	(0.182)	(0.235)	(0.247)	(0.341)
Has Twitter	0.733	0.593	0.517	0.152	-0.149	-0.191
	(0.635)	(0.730)	(0.717)	(0.770)	(0.664)	(0.681)
BTC Price (ICO)	-0.002	-0.014	-0.056**	-0.069**	-0.043	-0.029
	(0.014)	(0.022)	(0.026)	(0.030)	(0.034)	(0.047)
Pre ICO	0.123	0.021	0.091	0.238	-0.227	0.165
	(0.148)	(0.172)	(0.190)	(0.313)	(0.388)	(0.636)
Bonus	0.026	0.184	-0.443	0.031	0.688	-0.474
	(0.220)	(0.252)	(0.321)	(0.719)	(0.559)	(0.912)
Accept BTC	-0.004	0.030	0.123	-0.356	-0.420	-0.286
	(0.114)	(0.161)	(0.192)	(0.252)	(0.264)	(0.319)
Ethereum Based	0.124	0.164	0.581**	0.309	0.407	0.495
	(0.170)	(0.205)	(0.270)	(0.346)	(0.380)	(0.425)
Fog Index	0.004	0.016***	0.023***	-0.010	-0.000	0.003
	(0.003)	(0.006)	(0.006)	(0.012)	(0.012)	(0.014)
Tone	0.091	0.119	-0.009	0.023	-0.027	0.142
	(0.065)	(0.095)	(0.109)	(0.133)	(0.151)	(0.218)
Uncertainty	0.236*	0.426*	0.069	0.009	0.174	0.511
	(0.136)	(0.228)	(0.239)	(0.293)	(0.314)	(0.436)
Constant	-2.265*	-3.473**	-3.198***	-1.336	-1.262	-1.460
	(1.277)	(1.483)	(1.065)	(1.223)	(1.333)	(1.534)
Quarterly FE	Y	Y	Y	Y	Y	Y
Categorical FE	Y	Y	Y	Y	Y	Y
Geographical FE	Y	Y	Y	Y	Y	Y
R^2	0.160	0.216	0.271	0.387	0.466	0.455
Observations	323	319	293	228	198	157

(**Table 11 continued)**

Panel D: Bitcoin-adjusted rate of return

	(1)	(2)	(3)	(4)	(5)	(7)
	7 Days	30 Days	90 Days	180 Days	240 Days	300 Days
Tech Index	0.020	0.033**	0.036**	0.055***	0.070***	0.081***
	(0.012)	(0.015)	(0.017)	(0.019)	(0.021)	(0.028)
ICO Length	0.001	0.001	-0.000	-0.003	-0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.007)
Team Size	-0.009	-0.006	0.001	-0.000	0.004	0.009
	(0.007)	(0.008)	(0.010)	(0.011)	(0.012)	(0.015)
Has GitHub	0.256*	0.125	0.191	0.178	0.194	0.187
	(0.151)	(0.183)	(0.188)	(0.201)	(0.235)	(0.323)
Has Twitter	1.523*	1.231	1.250	1.114	0.623	0.396
	(0.819)	(0.898)	(0.907)	(0.754)	(0.720)	(0.711)
BTC Price (ICO)	0.008	0.009	-0.032	-0.047*	-0.019	0.010
	(0.015)	(0.022)	(0.024)	(0.028)	(0.032)	(0.050)
Pre ICO	0.062	-0.000	0.122	0.228	-0.133	0.109
	(0.181)	(0.196)	(0.213)	(0.260)	(0.364)	(0.621)
Bonus	-0.307	-0.092	-0.667*	-0.824	0.054	-0.500
	(0.279)	(0.304)	(0.348)	(0.562)	(0.677)	(0.927)
Accept BTC	-0.036	-0.025	0.145	-0.191	-0.317	-0.113
	(0.124)	(0.163)	(0.186)	(0.227)	(0.250)	(0.301)
Ethereum Based	0.087	0.083	0.490*	0.049	0.110	0.273
	(0.201)	(0.233)	(0.261)	(0.285)	(0.340)	(0.387)
Fog Index	0.004	0.015***	0.019***	-0.014	-0.002	-0.003
	(0.003)	(0.005)	(0.006)	(0.012)	(0.014)	(0.017)
Tone	0.102	0.145	0.056	0.055	-0.019	0.203
	(0.066)	(0.094)	(0.103)	(0.122)	(0.139)	(0.216)
Uncertainty	0.226	0.470**	0.114	0.021	0.264	0.408
	(0.159)	(0.238)	(0.241)	(0.277)	(0.301)	(0.405)
Constant	-3.268**	-4.530***	-4.613***	-4.336***	-5.288***	-4.422***
	(1.356)	(1.583)	(1.217)	(1.204)	(1.374)	(1.532)
Quarterly FE	Y	Y	Y	Y	Y	Y
Categorical FE	Y	Y	Y	Y	Y	Y
Geographical FE	Y	Y	Y	Y	Y	Y
R^2	0.221	0.223	0.253	0.429	0.440	0.430
Observations	323	319	293	228	198	157

Online Appendix for

“Do Digital Coins Have Fundamental Values? Evidence from Machine Learning”

Appendix A: Tech Index validation with anecdotal evidence

In this appendix, we use two whitepaper examples to illustrate that the Tech Index does provide a sensible measure of technology relevance of an ICO.

We consider two contrasting ICO projects: *Filecoin* and *Neogame*. *Filecoin* is a successful ICO which raised more than $200 million in a month in 2017. Although the token has not been distributed to investors, the corresponding futures are being traded on cryptocurrency exchanges. The market value is three times the amount raised in the ICO, generating large returns to early investors. Howell, Sabrina and Yermack (2019) conducts a detailed case study of *Filecoin* in the appendix. On the contrary, *Neogame* raised no money during the ICO process (or did not report any amount raised to the public). The official website is also inaccessible. From their whitepapers, we can see clear differences between the two ICO projects.

The whitepaper of *Filecoin* starts with the following introduction:

> *“Filecoin is a protocol token whose blockchain runs on a novel proof, called Proof*-of-*Spacetime, where blocks are created by miners that are storing data. Filecoin protocol provides a data storage and retrieval service via a network of independent storage providers that does not rely on a single coordinator, where: (1) clients pay to store and retrieve data, (2) Storage Miners earn tokens by offering storage (3) Retrieval Miners earn tokens by serving data.”*

On the contrary, *Neogame* introduces itself as:

> *"This project seeks to change gambling principles throughout the industry. The gambling business has always been based on 'house edge' – the principle that the casino or organizer always has the advantage. With the advent of blockchain technology, there is no longer any need for organizers – people can gamble without intermediaries and claim their winnings without paying commission...The Neogame project stands in contrast to traditional lotteries, and we want to show the world how blockchain can become a game changer."*

We can see that *Filecoin* provides a concrete and precise description about its project. It uses many technology-related words to make the details of the project clear. Its Tech Index (37.01%) ranks the 3rd among all 1,629 available whitepapers. On the contrary, for *Neogame*, despite the hard effort to describe the project as promising and technically advanced, due to the lack of a real technological foundation, the description is vague and lacks technological content. Specifically, its Tech Index (4.43%) ranks 5th in the bottom.

Therefore, this evidence suggests that our Tech Index is able to filter out technologically sound projects.

Figure A1. Average Tech Index by industry

This figure shows the mean of Tech Index by industry. The x-axis is the average Tech Index in percentage. The y-axis displays 20 industries (categories) in *trackico.com.*

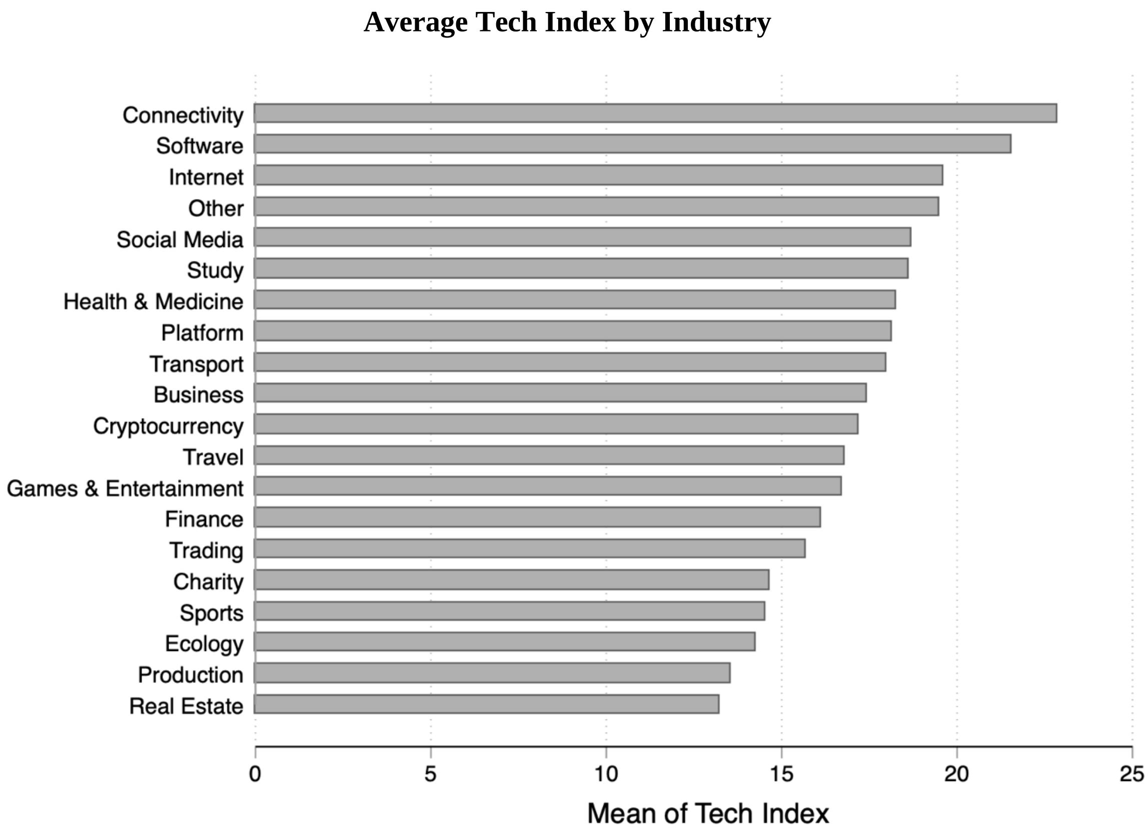


Table A1. Summary statistics on whitepaper status

This table lists all possible whitepaper status and their frequencies.

	Frequency	Percent (%)
Downloaded.	1629	55.90
URL response: client error.	535	18.36
URL response: server error.	104	3.57
Unable to get URL response.	403	13.83
Invalid PDF files.	155	5.32
Whitepaper not found.	54	1.85
Whitepaper is accessible but not downloadable.	27	0.93
Permission is required to access.	7	0.24
Total	**2914**	**100.00**

Table A2. Summary statistics on ICO industries

This table presents summary statistics on 20 ICO industries.

Category	Frequency	Percent
Business	212	7.27
Charity	15	0.51
Connectivity	40	1.37
Cryptocurrency	384	13.17
Ecology	30	1.03
Finance	219	7.51
Games & Entertainment	174	5.97
Health & Medicine	83	2.85
Internet	56	1.92
Other	223	7.65
Platform	977	33.50
Production	30	1.03
Real Estate	65	2.23
Social Media	45	1.54
Software	94	3.22
Sports	17	0.58
Study	17	0.58
Trading	158	5.42
Transport	40	1.37
Travel	37	1.27
Total	**2916**	**100.00**

Table A3. Test set classification result

This table presents the classification results of the test set. Panel A lists the words in the test set and the topic each word belongs to. Panel B presents the distribution of each topic.

Panel A.

Word	Topic	Word	Topic	Word	Topic
accenture	5	decryption	4	nft	2
address	1	encryption	4	oracle	4
altcoin	2	ethereum	4	platform	4
api	4	ether	1	protocol	4
asic	4	eth	1	provably	4
authentication	4	evm	4	proof	4
bitcoin	2	exchange	1	reproduction	2
btc	1	fiat	1	robustness	3
block	4	fork	4	satoshi	2
blockchain	4	gas	4	scrypt	4
chain	4	gigabyte	2	server	4
cipher	4	gartner	5	service	4
client	4	hash	4	system	4
coin	1	hashrate	2	nakamoto	2
collective	2	hashcash	2	hardfork	2
confirmation	4	halving	2	solidity	4
consensus	4	ibm	5	testnet	4
cryptocurrency	3	immutable	4	transaction	4
cryptography	4	ipfs	4	token	1
dapp	4	ledger	4	timestamp	4
data	4	liquidity	3	user	4
dao	2	mining	3	wallet	4
difficulty	3	node	4	workflow	4

Panel B.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 4	Total
Percentage	11.6%	18.8%	7.2%	**58.0%**	4.4%	100%