ICO Success and Post-ICO Performance *

Evgeny Lyandres† Berardino Palazzo‡ Daniel Rabetti§

July 2020

Abstract

We compile a comprehensive dataset of initial coin offerings (ICOs) from 19 data sources including 11 ICO aggregators. We alleviate severe limitations of available ICO data by performing the first systematic analysis of ICO data quality and use our dataset to study determinants of ICO funding success as well as post-ICO operating and financial performance. We highlight determinants of ICO success that are new to the literature and overturn some findings in existing studies. In addition, we provide evidence on some determinants of initial and longer-term ICO success on which existing literature has not reached a consensus. We also show that entrepreneurs' skin in the game is an important determinant of venture's post-ICO operating performance. Finally, we demonstrate that post-ICO operating success translates into financial one.

*We thank Yakov Amihud, Vasiliki Athanasakou, Ben Charoenwong, Deeksha Gupta, Bjorn Jorgensen, Paymon Khorrami, Steven Kou, Katya Malinova, Samuel Rosen, Ioanid Rosu, Gloria Yu, as well as seminar participants at Boston University, Sun Yat Sen University, Temple University, University of Arkansas, U.S. Securities and Exchange Commission, Federal Reserve Board and participants of the 2018 MJAC & CryptoCompare Blockchain Summit, 2019 Mont Tremblant Risk Management and Financial Innovation Conference, 2019 Adam Smith Asset Pricing Workshop, 2019 Financial Management Association Wine Country Conference, 2019 ABFER, CERP and CUHK Symposium in Financial Economics, 2019 Asian Bureau of Economic Research Conference, and 2019 Eagle Labs IDC Conference. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Board or the Federal Reserve System. All errors are our own.

†Questrom School of Business, Boston University, lyandres@bu.edu.

‡Federal Reserve Board, dino.palazzo@frb.gov.

§Coller School of Management, Tel Aviv University, rabetti@mail.tau.ac.il.

1 Introduction

In the past three years, a new form of financing–initial coin offering (ICO)–has emerged, fueled by developments in blockchain technology and its applications. An ICO enables an entrepreneurial venture to raise funds in exchange for cryptographically secured tokens intended to be the sole means of payment for the venture's future products or services. In 2016–2019, over 7,400 entrepreneurial ventures attempted an ICO, raising a staggering $US 35 billion.

Our paper contributes to the emerging empirical ICO literature along two dimensions. The first contribution is our comprehensive ICO dataset, compiled using data from 19 sources. Available ICO data suffer from serious limitations, which we alleviate by performing the first systematic analysis of ICO data quality. We propose and implement a procedure for identifying the most trustworthy components of data coming from various sources, which we use in the construction of our dataset, thereby substantially reducing measurement error.

Our second contribution is to the literature that studies determinants of ICO funding success and of post-ICO operating and financial performance. In analyzing determinants of post-ICO operating and financial success, we are guided by theoretical models focusing on entrepreneurs' post-ICO incentives and on investors' assessment of ventures' post-ICO operating performance. Many of our empirical findings are new to the ICO literature. We also overturn some of the findings in other ICO studies. In addition, we provide evidence on some determinants of initial and longer-term ICO success on which existing literature has not reached consensus.

The ICO process is mostly unregulated and decentralized. ICO information is scattered across a multitude of online sources, which aggregate various pieces of information regarding ICO characteristics, mostly by retrieving this information from ICO "white papers". Consequently, various data sources cover subsets of attempted ICOs, and the degree of pairwise overlap in coverage varies widely, resulting in large differences in sample sizes and compositions across existing studies. More importantly, even if two datasets cover the same ICO, they often disagree on the values of ICO characteristics. Discrepancies among data sources are often on such a scale that using data from different sources may lead to dramatically different estimates. Thus, before conducting any meaningful empirical investigation of ICOs, it is critical to build a dataset of ICOs that contains the most trustworthy pieces of data.

To mitigate data limitations, we obtain and compare data from no less than 11 ICO aggregator websites, which collectively likely cover almost the entire ICO population.[1] Our resulting initial sample covers over 7,500 attempted ICOs and over 5,300 completed ones, and has information on some ICO characteristics coming from 3 sources per ICO on average. In addition, we collect the most comprehensive auxiliary data on the time-series evolution of social media activity initiated by ventures performing an ICO across four most popular platforms (Twitter, Medium, Reddit, and Bitcointalk), on the evolution of project-related code updates on the world's leading open-source platform (GitHub), and on the evolution of the number of cryptographic wallets containing the token issued in an ICO as well as the number of on-chain transfers involving the token across wallets. These data allow us to study determinants of both the initial success in raising funds in an ICO as well as of post-ICO operating and financial performance.

While the size of our sample is one of the largest in the ICO literature, our main contribution is to data quality, not quantity. Bringing discipline to the data collection process and using the most trustworthy pieces of data are two crucial ingredients necessary for making reliable empirical inferences regarding determinants of ICO success. To overcome the large discrepancies in reported values of ICO characteristics across aggregators, we develop a procedure for ranking data quality both at the source-variable level and at the ICO level. Not only do we show that data sources vary in terms of their data quality–a hypothesis that has been proposed in existing papers, although never carefully examined empirically–but that there is also a substantial variation in the quality of data at the level of ICO, which leads us to examine subsets of ICOs with relatively high-quality data.

After compiling our dataset, we turn to using it to examine determinants of ICO success and post-ICO operating and financial performance. We begin by analyzing determinants of ICO funding success, measured by whether funds were raised in an ICO, by the (absolute and relative to funding objective) amount raised in an ICO, and by whether the venture's token is eventually listed on a cryptographic exchange. The determinants of ICO success that we focus on include those proposed and used in other empirical ICO studies (e.g., Adhami et al. (2018), Amsden and Schweizer (2018), Benedetti and Kostovetsky (2018), Bourveau et al. (2019), Davydiuk et al. (2019), Deng et al. (2018), Howell et al. (2020), Hu et al. (2018), and Lee et al. (2019)), as well as new ones, which have not been examined previously.

[1]The 11 aggregators were chosen as those with the highest historical access counts (Alexa ranks, see https://www.alexa.com/siteinfo) and Google search counts.

The novel results in our paper include the positive relation between ICO hardcap and the intensive margin of funding success and the positive effect of ICO white paper informativeness on ICO success. In addition, our paper is the first to provide an extensive analysis of the relation between pre-ICO venture-initiated social media activity and ICO funding success. The effects of social media activity on ICO success depend crucially on the social media platform, in ways consistent with costly signaling.

Some of our results overturn those in existing papers. For example, we find that the presence of bonus (typically a discount to early investors in tokens during an ICO) is positively associated with funding success at both the extensive and intensive margins. Existing papers report either a negative or an insignificant relation between bonus availability and funding success (e.g., Lee et al. (2019), Amsden and Schweizer (2018), and Bourveau et al. (2019)). Differently from some existing studies (e.g., Amsden and Schweizer (2018), Bourveau et al. (2019)), we also find that after controlling for white paper informativeness, a white paper's length has only a marginally significant impact on ICO funding success.

Additional results shed light on contrasting findings in the existing literature. We find that a presale is largely insignificant in explaining ICO funding success. The question of the effects of presale on ICO success has not been settled in the existing literature (e.g., Adhami et al. (2018), Lee et al. (2019), Bourveau et al. (2019), and Deng et al. (2018)). We also find that a measure of ICO transparency – an indicator equaling one for ICOs that include a know-your-customer (KYC) provision – is significantly positively related to both the minimal ICO funding success indicator and to the amount raised in an ICO. Current literature examining this relation is divided (e.g., Davydiuk et al. (2019), Deng et al. (2018), and Lee et al. (2019)).

Yet some other findings are consistent with those in existing studies. Examples include the negative relation between ICO funding success and the percentage of tokens offered for sale in an ICO and positive relations between ICO success on one hand and white paper availability, the size of venture team, and the cumulative code revision activity at the time of ICO on the other hand.

Following the analysis of ICO funding success, we proceed to examine determinants of ventures' post-ICO operating performance. In particular, we are interested in whether entrepreneurs' skin in the game–which emerges as one of the most important determinants of ICO funding success–also affects post-ICO operating performance. In this analysis, we are guided by the model of Gan et al. (2020), which shows that lower skin in the game leads to entrepreneurial shirking after ICO completion.

In our empirical tests of Gan et al. (2020), we first examine the association between the percentage of tokens for sale in an ICO and inputs to ventures' post-ICO production–code revisions (commits) on Github and venture-initiated social media activity. The percentage of tokens available for sale is an inverse proxy for entrepreneurs' skin in the game, which is an increasing function of the fraction of tokens retained by the venture. We find that post-ICO code revision activity is increasing in entrepreneurs' skin in the game. This result is highly economically and statistically significant and holds at all horizons, ranging from one month to one year after ICO. In addition, post-ICO social media activity on three out of four platforms–Twitter, Medium, and Bitcointalk–is increasing in entrepreneurs' skin in the game.

We also examine the relation between entrepreneurs' post-ICO skin in the game and venture's operating success, as measured by product/platform adoption by users, which increases the need in holding the venture's token and in transferring the token across cryptographic wallets. To this end, we estimate the associations between the percentage of tokens for sale in an ICO on one hand and post-ICO growth in the number of wallets holding the token and in the cumulative number of transfers of the token across wallets on the other hand. The only other paper that investigates the effects of skin in the game on post-ICO operational success, as measured by platform adoption by users, is Deng et al. (2018). We find that post-ICO product/platform adoption is increasing in the skin in the game, and our results tend to be highly economically and statistically significant at various horizons. This outcome is markedly different from that in Deng et al. (2018), who report an insignificant relation between skin in the game and platform adoption.

Overall, the negative relations between the percentage of tokens for sale in an ICO on one hand and inputs to the venture's production function and its product/platform adoption on the other hand are consistent with the positive impact of entrepreneurs' skin in the game on post-ICO operating performance, highlighted in Gan et al. (2020).

Lastly, we ask the following important question: Do the effects of entrepreneurs' skin in the game on venture's post-ICO operating performance have financial ramifications? In other words, is there a relation between the venture's post-ICO product/platform adoption by users and contemporaneous returns on the venture's token, as predicted by the model of Cong et al. (2020)? Obvious data limitations preclude us from examining long-term (3 to 5 years) token returns, thus we focus on post-ICO returns ranging from one month to one year from the time the token is first listed on a cryptographic exchange. We find that the contemporaneous relations between the two measures of product/platform adoption and post-ICO token returns are positive. These relations are especially significant economically and statistically at longer horizons, consistent with the positive relation between post-ICO operating performance and financial success, highlighted in Cong et al. (2020).

The remainder of the paper is organized as follows. Section 2 describes the sample construction, various data sources and their limitations, and the derivation of our ICO data quality measure. Section 3 reports summary statistics of the main variables of interest. Section 4 presents an empirical analysis of determinants of ICO funding success. Section 5 focuses on post-ICO operating and financial performance. Section 6 concludes.

2 Data

2.1 Data Acquisition and Sample Construction

Our objective is to construct the most comprehensive dataset of ICOs to examine the determinants of various measures of ICO success – initial fundraising, as well as post-ICO operating and financial performance. Information on ICO characteristics can be obtained from two non-mutually-exclusive types of sources. The first is ICO white papers, often available on project websites. The second is ICO aggregators, which contain information on large subsets of ICOs.

Scraping data from white papers and project websites, while intuitively appealing, has several limitations. First, not all ICOs have white papers and/or websites. Second, and more importantly, the information in white papers is updated frequently for some projects, with new versions of white papers replacing old ones on project websites.[2] Changes in the white paper contents often concern crucial pieces of information. The updating of white papers over time potentially introduces a problem of look-ahead bias. In other words, information available from a project's white paper after its ICO may be different from the information that was available to investors at the time of the

[2]For example, in our sample, the cover page of 37% of white papers currently appearing on project websites includes one of the following words or their variants: "revision", "update", "release", and "version". Notably, 37% is clearly a conservative estimate of the proportion of white papers whose contents have changed over time.

ICO (e.g., Lee et al. (2019)).[3] A potential (partial) solution to the look-ahead bias would be to use the information only from white papers available prior to (and close to) respective ICOs. However, an analysis of white papers in our dataset reveals that: 57% of white papers do not specify the date when the white paper was written or published; in additional 18% of white papers only the year (but not the month or day) is specified; out of the remaining 25% of white papers with full date available, the date of the most recent white paper that appears on the project website postdates ICO completion in 1/3 of the cases and precedes ICO completion by over 6 months in further 1/3 of the cases. As a result, data obtained from white papers directly undoubtedly represents information available to investors at the time of ICO in just about 8% of ICOs in which white papers are available, and in about 3% of ICOs overall.

Data from ICO aggregators are less likely to suffer from a look-ahead bias. The reason is that aggregators are incentivized to deliver detailed information to ICO investors at the time of ICO, as aggregators are compensated by ICO issuers for traffic to ICO websites that they generate. Aggregators typically do not have incentives to update this information after the completion of an ICO. As a result, it is more likely that data appearing on aggregator websites have been retrieved from the version of a white paper available at the time of ICO, and are less likely to be updated thereafter based on the evolution of the white paper.

In addition, in their attempts to generate as complete a description of ICO terms as possible, aggregators often do not limit themselves to retrieving data from white papers and project websites. They often examine social media channels and, on occasion, contact project teams directly. In addition, some aggregators require ICO ventures to fill out forms with essential data as a condition to listing an ICO on their website. Notably, aggregators' effort to obtain data beyond those available in white papers and project websites translates into larger data coverage and larger sample size in our analysis relative to many other ICO studies.[4]

However, data from aggregators have several drawbacks as well. First, each aggregator does not cover the entire universe of ICOs. For example `www.ICObench.com`, one of the most popular

[3] For example, in ICO of Aditus, the pre-ICO white paper indicated that the percentage of tokens for sale in the ICO – a proxy for entrepreneurs' post-ICO "skin in the game", the effects of which we analyze extensively, was 45%; in the post-ICO white paper this value was reported as 25%; and in the currently available version of the white paper the information on the percentage of tokens for sale is missing altogether.

[4] The advantages of using data from ICO aggregators notwithstanding, we verify that using data from white papers instead of those from aggregators in cases in which white paper data can be reliably used (in 8% of observations with white papers available) does not affect any qualitative or quantitative conclusions.

sources used in the literature so far (e.g., Huang et al. (2019) and Lee et al. (2019)), has about 50 percent coverage. Second, and more importantly, information regarding various ICO characteristics frequently contains errors, and some projects are duplicated within the same source with similar or even identical names.

To overcome data coverage issues, we use data from no fewer than 11 ICO aggregators, which we chose based on popularity, as measured using average historical Alexa Traffic Rank between January 2016 and January 2020: www.Etherscan.io, www.CoinDesk.com, www.CoinGecko.com, www.CryptoCompare.com, www.ICObench.com, www.ICOdrops.com, www.ICOrating.com, www.ICOmarks.io, www.ICOdata.io, www.FoundICO.com, and www.Tokendata.io. In what follows, we omit www and the website address extension when referring to various aggregators.

Critical issues arise when attempting to match data across sources. First, there is no unique identifier for each project. Second, listed ICOs are traded on multiple cryptographic exchanges (there are 119 exchanges in our sample). As a result, several different projects may have the same ticker, which may also coincide with IDs of non-listed projects. In many cases, matching by project name is not helpful because of variations in project names, misspellings, names unrelated to original projects, and outdated or incomplete names.

Etherscan is the most popular source of ICO data. A possible reason is that besides disclosing information on key ICO-related and project-related variables, it provides blockchain transaction data for Ethereum-based ("ERC") tokens, which represent a large portion of tokens issued in ICOs.[5] CoinDesk is the second most popular source of ICO-related data, but it contains information on fewer variables. Its popularity is mostly due to its role as a source of crypto news and ICO analysis. CoinGecko, CryptoCompare, and ICObench tend to provide relatively high-quality data, but the range of variables it covers is limited. ICOdrops provides some of the most reliable data on the amount raised in an ICO, with sparser coverage of some other variables. ICOrating and Tokendata mostly provides data on auxiliary ICO variables, discussed below, whereas ICOmarks has good coverage of the number of tokens issued and offered to investors in an ICO, but lacks coverage of other important variables. The most popular aggregator websites are located in the United States and Western Europe. However, we include ICOdata and FoundICO to

[5]Currently, Ethereum-based tokens are used in 90 percent of ICOs and are responsible for 75 percent of ICO proceeds.

improve the geographical diversity of our sample. Both these sources focus on Pacific Asian and Eastern European ICOs and, though less popular than other aggregators, provide information on a large number of listed ICOs.

Since many ICOs are covered by multiple ICO aggregators, we match data across various sources to generate a sample of uniquely identified ICOs. Project names and tickers cannot be reliably used for matching for aforementioned reasons, therefore we use projects' website addresses to resolve potential conflicts. A project may have several addresses that we can exploit for matching purposes, however not all reported addresses are accurate, up to date, or even related to the project. Thus, we adopt a website address validation criterion to reduce discrepancies in our matching process. In particular, we match ICOs according to the following order of preference: 1) the ICO website address as reported on www.CoinMarketCap.com, which is the source of post-ICO price and volume data for tokens traded on exchanges, 2) the project website address as reported on some of the aggregator websites, and 3) the addresses of accounts on the following social media sources: Twitter, Medium, Reddit, Bitcointalk, Linkedin, Slack, and Telegram. We validate each match ex-post using the name of the project and the ticker symbol.

Our final sample comprises 7,514 unique merged projects in 133 countries, carried out between 2013 and 2019. As evident from Figure 1, all but a few ICOs happened (that is, had an end date) in 2017, 2018, or 2019. ICO activity peaked between September 2017 and June 2018, with over 100 ICO-funded projects each month that were able to raise money from investors. During this 10-month period, close to $U.S. 20 billion was raised, with the peak at almost $U.S. 6 billion in June, 2018, in which the blockchain project EOS ended its ICO, raising a staggering $U.S. 4.2 billion. Possibly due to regulatory uncertainty, ICOs became less frequent in the end of 2018 and in 2019, being partially replaced by "Security Token Offerings" (STOs), which adhere to securities regulations, and "Initial Exchange Offerings" (IEOs), in which an issuer combines raising capital with listing the token on a crypto exchange.[6]

[6]An analysis of differences between ICOs on one hand and STOs and IEOs on the other hand would be very interesting. However, currently, the samples of STOs and IEOs are too small to conduct a meaningful empirical analysis (71 attempted STOs and 469 attempted IEOs as of January, 2020). In addition, the ICO type is largely driven by the timing of token issuance, as opposed to characteristics of the venture. In particular, the market for financing ventures via issuance of crypto tokens has been dominated by ICOs until the end of 2018, when STOs were created as a proposed solution to antitrust authorities' anticipated treatment as securities of tokens issued in ICOs. STOs have been quite popular until the middle/end of 2019. The beginning of 2019 saw a migration from ICOs and STOs to IEOs, which are listed on crypto exchanges immediately after ICO completion and have emerged as a solution to the the problem of non-tradable tokens, fraud, and exit schemes prior to listing.

Table 1 contains detailed descriptions of dependent and independent variables used in the empirical analysis. Table 2 summarizes the distribution of ICO data availability across various sources. Slightly over half of uniquely identified ICOs are covered by at most two aggregator websites, whereas 22 percent of ICOs are covered by five aggregators or more. In the empirical analysis, we further restrict our attention to a subsample of 5,376 ICOs for which we have data on the the number of tokens issued for sale, the amount raised in the ICO, or both. We do so in an attempt to eliminate incomplete ICOs, which are those that are halted before offering tokens to investors, as opposed to completed but unsuccessful ICOs (that is, those that fail to raise money), which we keep in the sample.

2.2 ICO Aggregators and Data Quality

Unfortunately, there are substantial inconsistencies in the reported values of main ICO characteristics–the amount raised, the hardcap, the number of tokens available for sale, and the overall number of tokens issued–across aggregators.[7] Table 3 reports the number of observations for each of the above variables across the 11 aggregators. All four variables are available for some ICOs in six to nine partially overlapping sources.

2.2.1 Data Quality at the Source-Variable Level

We begin by developing a procedure for identifying the value of a variable (x) that is most likely to be correct, in cases in which it is available across multiple sources and there is disagreement among them. Our procedure consists of the following steps. We begin by computing a measure of data quality at the source-variable level. The data quality measure is inversely related to the average (across all observations with available data) disagreement between the value of a variable reported by a given data source and the mean value for that variable reported by all data sources. The first step is computing for each observation (k) and source (i) with data available on variable x, the "relative distance" of the value reported in source i, $x_{i,k}$ from consensus (mean) value of this variable for observation k across all sources, $\bar{x}_k$, defined as $\left|\frac{x_{i,k}-\bar{x}_k}{x_{i,k}+\bar{x}_k}\right|$. If $x_{i,k}$ equals the average value, $\bar{x}_k$, then the relative distance of source i and observation k for variable x is zero.

[7]When variables concern a monetary amount (e.g., the amount raised in the ICO and the hardcap) – we convert these values to the $U.S. using exchange rates on the final day of the ICO.

If $x_{i,k}$ approaches zero or infinity, the relative distance approaches one. Consider as an example, the data available on the total amount raised by Blocklancer during its ICO. The reported values are \$300,000 (`ICObench`), \$5,475,789 (`CoinGecko`), \$4,420,000 (`CryptoCompare`), \$10,000,000 (`ICOrating`), and \$258,850 (`ICOdata`), with the mean value across sources being \$4,286,874. The relative distance of the amount raised reported by `ICObench` is 0.89, while the relative distance of the amount raised reported by `CoinGecko` is 0.12. This example, while extreme, illustrates quite a common occurrence in the data.

In the second step, we compute the average relative distance for variable x for each source across all observations in which data for a given variable are available across multiple sources, $\bar{x}_i$ for source i, i.e. we compute the averaged across all k of $\left|\frac{x_{i,k}-\bar{x}_k}{x_{i,k}+\bar{x}_k}\right|$. We refer to the resulting average relative distance from consensus value as the "mean deviation" for source i and variable x. This mean deviation for each of the 11 sources and four main variables in the analysis is reported in Table 3. As evident from the table, the values of the amount raised in an ICO, as reported in `CoinGecko`, tend to be the closest to consensus, with an average relative distance of 0.062. On the other end of the spectrum, data from `CryptoCompare` are the farthest from consensus, with an average relative distance of 0.111. The largest disagreement among the sources is regarding the number of tokens supplied in an ICO, with the average distance across all sources being 0.132.

The choice of the source of information about a particular ICO-related variable is crucial in cases in which there is disagreement among aggregators. To identify the most trustworthy pieces of information, we construct a measure of data quality at the source-variable level. To build such a measure, we first calculate for each available source and variable a measure of quality given by the inverse of that source's mean deviation reported in Table 3. For example, the inverse of mean deviation of the amount raised reported in `CoinGecko` (`CryptoCompare`) is $1/0.062 = 16.13$ ($1/0.111 = 9.01$). Then, we compute the relative quality of data coming from a given source for a given variable by dividing the inverse mean deviation by the highest inverse mean deviation across all sources reporting data for that variable. This normalization is important in light of variation in average quality of various ICO characteristics. The relative quality of the amount raised data from `Etherscan` is $16.13/16.13 = 1$, while the relative quality of `CryptoCompare` is $9.01/16.13 = 0.56$.

2.2.2 Choosing the Most Trustworthy Pieces of Data

When choosing the most trustworthy pieces of data we incorporate a) the degree of (dis)agreement among sources reporting values for a given observation and b) the quality of these sources, when choosing the value of a variable. For each reported value of the variable for a given observation, we add the source-variable-level quality measures for each data source reporting that value, which results in the "quality-weighed" number of sources reporting a given value for the variable for a given observation. The value chosen is the one with the highest quality-weighted number of sources reporting it. As an example, consider all of the data available for the Bancor ICO, reported in Panel A of Table 4. Panel B reports the quality of each source for each of the four variables. The data for the amount raised are consistent across seven sources, thus we use $U.S. 153,000,000 as the true value. In contrast, the four sources that report token supply disagree on the values of that variable. Among these sources, CoinGecko is the highest-quality data source for token supply (0.854), hence the chosen value of token supply in Bancor ICO is the one reported by CoinGecko, $U.S. 75,783,855.

Importantly, we do not base the choice of the value for a variable on the relative distance between that value as reported by a given source for a given observation and the mean value across all sources for that observation. Instead, we use the value coming from a source (or multiple sources) whose (combined) quality measure for that variable is the highest, where quality measures at the source-variable level are computed using all observations, not just the observation whose value we are attempting to choose.

Our procedure for choosing values of variables is robust to various modifications. For example, in 90% of the cases, the variable chosen using the quality-weighed number of sources is the same as the one that would have been chosen using a simple count of the number of sources reporting the same value for the variable (and choosing the value from the source with the highest estimated data quality for that variable if all sources disagree). In addition, when we restrict our choice of the value of a variable by picking values from the three sources that have the highest quality measure (for that variable) and disregarding information for that variable coming from all other sources, in 87% of the cases, the value of the variable chosen is the same as the value chosen using data from all sources.

2.2.3 Data Quality at the Level of Observation

While our procedure is aimed at choosing the most trustworthy values of the main ICO characteristics, some ICOs still exhibit poor data quality. To account for differences in data quality across ICOs, we construct an ICO-level measure of data quality and verify the robustness of our results using subsamples of ICOs with the highest-quality data. In constructing our data quality measure, we are guided by the following considerations. First, data quality of an ICO should be increasing in its coverage–namely, the number of sources with available data on variables characterizing that ICO. Second, the measure should be increasing in the quality of available sources. Third, the measure should be decreasing in the amount of disagreement among the sources regarding the values of main ICO characteristics.

The first step in building a measure of data quality at the ICO level is to identify all sources that report a value for the main four ICO variables. We define the consistency of a variable for a given ICO as one minus the mean relative difference of this variable across all sources reporting it, which, in the case of the amount raised in the Bancor ICO, equals one. On the other hand, the consistency of values of hardcap is far from one, as there are two sources reporting values of this variable, which are inconsistent with each other. The relative distance for hardcap reported by `CryptoCompare` (`ICOdata`) is 0.333 (1), resulting in consistency of $1 - (0.333 + 1)/2 = 0.333$.[8]

Next, for each variable, we use the sum of qualities of all sources reporting information on that variable for a particular ICO to compute the total quality of that variable for that ICO. We take into account the consistency of available data for a given ICO by multiplying the total quality of a given variable by its consistency. For, example, the total quality of token supply data for the Bancor ICO (i.e. the sum of source-level qualities of token supply data for sources reporting the value of token supply for Bancor) is 2.464, and the average consistency of this variable across sources is 0.944, resulting in the adjusted quality measure of $2.464 \times 0.944 = 2.326$. The overall data quality for the Bancor ICO is given by the simple average of adjusted quality values across the four variables, equaling 2.323.[9]

Figure 2 describes the distribution of our ICO data quality measure and its association with the

[8]Interestingly, both values of the hardcap are much lower than the amount raised in the Bancor ICO. The reason is that Bancor ignored its stated hardcap and continued to accept funds even after it was reached.

[9]Various alterations of the procedure for estimating ICO-level data quality, such as using variable-source-level measure of consistency instead of averaging it across variables before computing consistency do not have a material impact on ICO-level data quality estimates.

total number of sources reporting data on each of the four variables for a given ICO and with the average consistency of these sources. Importantly, the figure shows that our data quality measure is clearly increasing in both the number of sources and in their consistency. In untabulated results, we find that our ICO-level data quality measure is also significantly inversely associated with measures of ICO opaqueness, as intuition would suggest.[10]

2.2.4 Other Variables Obtained from ICO Aggregators

In addition to the main ICO characteristics (amount raised, hardcap, and tokens issued and offered for sale in an ICO), we define a set of binary variables characterizing projects and ICOs: the occurrence of an attempted presale of tokens (that is, an attempt to sell tokens to large/institutional/VC investors before the offering of tokens to the general public); the requirement for investors to register in advance in order to participate in the ICO (known as "whitelist"); and the presence of a "know your customer" (KYC) requirement, which obliges token buyers to prove their identity by providing passport, national ID, or driver's license information. We also have information on the number of team members involved in ICO-backed projects, on the existence of bonus and bounty programs (discounts to ICO token price and rewards programs, respectively). We also collect information on the type ("industry") of ICO-funded projects. We aggregate industries into five sectors: entertainment, business services, blockchain, other software, and finance. Finally, we use information on ICO location and aggregate locations into five regions: Western Europe, Canada, and Australia; Eastern Europe; Asia; the United States; and the rest of the world.[11]

[10] Regressions of data quality on proxies of opaqueness–such as the availability of a white paper, the "know your customer" requirement, the cumulative social media activity at the start of ICO, and the cumulative code revision activity at ICO start–produce highly significantly negative coefficients on all opaqueness measures and an R squared of 23%. We also find that larger and (ex-post) more successful ICOs tend to feature higher-quality data: Regressing the data quality measure on (log) hardcap produces a statistically significant association, and so does regressing the data quality measure on (log) amount raised in the ICO. However, size-related and success-related variables do not explain a large portion of variation in the data quality measure – the R squared in the aforementioned univariate regressions are 1% and 7% respectively.

[11] We also have information on a project's legal form, on the availability of a "minimum viable product", on the presence of maximum and minimum token purchase requirements, on the intended use of ICO proceeds, on the possibility of receiving tokens by means of solving a computationally difficult puzzle (aka mining), and on the presence of an escrow account. These additional variables tend to not be significantly associated with outcome variables in our empirical analysis.

2.3 Other Data Sources

2.3.1 White Paper-Based Variables

For 2,000 ICOs with available white papers, we obtain additional information by examining their contents with the goal of measuring white paper informativeness, which is likely to be inversely related to project opaqueness and to the degree of information asymmetry between ICO issuers and potential investors. In this analysis, we attempt to use a pre-ICO version of white papers whenever possible. Our first measure of white paper informativeness is the number of unique words identified by natural language processing (NLP).[12] The second measure is the ratio of "technical" words out of all words appearing in a white paper, with the idea that more technical white papers are found in projects in more advanced stages of development.[13,14]

2.3.2 Post-ICO Token Prices and Returns

For post-ICO token price data we rely on `www.CoinMarketCap.com`, which has become the standard source for researchers interested in measuring token performance post-ICO (e.g., Benedetti and Kostovetsky (2018), Lee et al. (2019), and Howell et al. (2020)). We match market price data with our ICO sample using the ICO website address, as explained previously. This matching procedure allows us to identify 1,007 unique completed ICOs, which end up being listed on at least one crypto exchange, out of 2,442 ICOs that have successfully raised some funds.

One crucial variable that is used in construction of ICO initial returns is the average price paid by ICO investors for tokens sold in the ICO.[15] We compute the average ICO price as the ratio of the amount raised at the ICO and the number of tokens issued in the ICO, as reported by `CoinMarketCap`.[16]

[12] NLP is focused on identifying common roots, such as "buy" and "buying", while eliminating stop words, such as "a", "the", and "and".

[13] Technical words are common words used in the blockchain and computer science white papers–for instance "block", "node", and "ledger". We build a dictionary of the 144 most frequent technical words extracted from words frequency in ICO white papers and based on several blockchain and computer science glossaries extracted from various websites and tech forums. Details are available upon request. See Florysiak and Schandlbauer (2020) for a more detailed analysis of ICO white paper contents.

[14] In addition to these variables, we obtain such white paper characteristics as page count, word count, image count, and .pdf file size, which tend to have lower explanatory power than the number of NLP words and the ratio of technical words.

[15] Most ICOs employ an accelerated pricing schedule, in which early (and presale) investors pay lower-than-average prices for issued tokens.

[16] We measure the number of tokens seven days after the beginning of trading or after the appearance of the first

2.3.3 Social Media Data

To examine the time-series evolution of the coverage of ICOs in social media, we rely on four popular social media channels used by ICO projects: Twitter, Reddit, Medium, and Bitcointalk.[17] Time-series data are extracted for all projects when the associated social media account is available, not suspended, and is public from its inception to date. We exclude social media accounts that are not clearly related to the project based on information related to the project's name, ticker, website address, and team members.

Importantly, we examine only social media activity that is initiated by the venture. In particular, we focus on tweets using the firm's handle, i.e. those that are initiated by the firm's insiders (as opposed to replies and retweets), articles by ”editors” on Medium (as opposed to articles by ”team” and “writers”, and to claps and responses), discussions on Reddit (as opposed to thumbs and comments), and posts on Bitcointalk (as opposed to activity and merits).

Different from reactions to ventures' social media activity, whose contents may be positive or negative, venture-initiated activity is unlikely to signal negative information given its voluntary nature. In unreported analysis, we examine the average tone of venture-initiated activity by the time of ICO start, following an accepted method of measuring the tone of text (e.g., Kraaijeveld and De Smedt (2020) in the context of cryptocurrencies).[18] The results of this analysis strongly suggests that venture-initiated social media activity tends to convey positive signals. The credibility of this signal is an empirical question, however – a question that we examine in our analysis of short-term and long-term ICO performance.

---

observation of market cap (whichever happens later). The reason is that not all tokens reach exchanges immediately, and the number of tokens typically stabilizes within a week.

[17]These sources differ widely in content and format. For instance, Medium articles are often well-written, have hundreds or thousands of words, and usually focus on a project's description, solutions, milestones, achievements, and information useful for potential token buyers or token holders. Twitter tweets, on the other hand, are limited to 280 characters, are often written in an abbreviated language, and are used for quick press releases and for sharing videos, photos, and additional content from other social and news channels.

[18]We compute the tone as $\frac{\text{\# positive words} - \text{\# negative words}}{\text{\# positive words} + \text{\# negative words}}$ for each tweet (article, discussion, post) and then aggregate the tone for all tweets (articles, discussions, posts) by a venture as of the time of the start of its ICO. This measure of average tone ranges from -1 (most negative) to 1 (most positive). The tone of firms' tweets tends to be very positive (the average measure of tone is 0.54. Similarly, the mean tone of editor articles on Medium is 0.38, the mean tone of Reddit discussions is 0.35, and the mean tone of Bitcointalk posts is 0.29.

2.3.4 Code Production Data

Most ICO-backed projects are in very early stages of development, and their R&D output is typically not protected by patents. As a result, many of these projects rely on open source code development, and code production is a natural proxy for the project's maturity, the strength of the development team, and its ties with the international community of coders. To examine the time-series evolution of project development, we follow Amsden and Schweizer (2018), Bourveau et al. (2019), Davydiuk et al. (2019), and Fisch (2019) and focus on code revisions ("commits"), namely revisions to files in a project's repository, posted on the largest open source platform, GitHub. By the last ICO day, 1,806 projects have some commits on GitHub.[19]

2.3.5 Blockchain Transaction Data

For ERC-based tokens that are listed on crypto exchanges, we collect from `www.Ethplorer.io` information on the evolution of the number of distinct cryptographic wallets containing the tokens, and of the number of cumulative off-chain transactions involving the token. Each transaction contains information about the addresses of wallets sending and receiving the tokens, the amount of tokens transferred, and the transaction's hash and time stamp.[20] In computing the number of wallets, we exclude wallets belonging to crypto exchanges, which aggregate holdings of multiple investors, and "genesis wallets"–that is, wallets belonging to ICO issuers that are used to transfer tokens to ICO investors, contributors, and miners. After merging blockchain transaction data with ICO data, we are left with 774 ICOs with wallet information available as of the first trading day.

[19]In unreported analysis, we examine additional proxies for code development activity – the number of contributors (outside of the venture), the number of closed issues, the number of pull requests – as suggested by Deng et al. (2018). While these are reasonable alternative measures of the code revision activity, they are quite sparsely populated in our data: only 10% (16%, 14%) of observations with non-zero commits have non-zero contributors (closed issues, pull requests). The results using these alternative measures of code production are generally statistically insignificant, likely due to insufficient statistical power. These results are available upon request.

[20]ERC protocol does not allow token transactions with non-integer values. As a result, information on token divisibility is reported using a variable called decimals. This variable represents the number of digits after the decimal place. e.g., if a transaction reports a number of tokens transferred equaling 150,000,000,000 and the value of decimals equaling 8, we adjust the reported value by subtracting 8 zeros, thus obtaining a value of 1,500 tokens transferred.

3 Summary Statistics

Table 5 presents summary statistics for key dimensions of ICOs and their outcomes. As evident from Panel A, the average ICO hardcap is $U.S. 63 million, while in more than 50 percent of ICOs it is larger than $U.S. 20 million. The percentage of tokens issued to the public in an ICO averages 56 percent of the total tokens outstanding, and in about 10 percent of ICOs all tokens outstanding are offered to investors. There are 2,789 ICOs (52 percent of the sample) in which ventures attempt to raise funds in a presale to large, institutional, or venture capital investors before the official ICO start. 40 percent of ICOs feature advanced investor registration ("whitelist"), whereas 50 percent of ICOs have a "know your customer" (KYC) requirement. The average (median) number of team members involved in an ICO is 11 (9). 37 percent of ICOs offer bonuses to early investors, while in 17 percent of ICOs bounties are offered for advertising ICOs on social media channels.

Industry affiliations are available for 61 percent of projects in our sample. Among projects with industry information available, the most frequent sector is finance, representing 39 percent of ICOs, while the least frequent sector is general blockchain (10 percent). Location information is available for 82 percent of ICOs. One-third of ICOs are performed in Western Europe, Canada, and Australia, 13 percent of ICOs are U.S.-based, and 42 percent of ICOs are performed in jurisdictions that have adopted crypto-friendly policies, such as Singapore, Hong Kong, Switzerland, Estonia, Malta, British Virgin Islands, and Gibraltar. White papers are available for 37 percent of attempted ICOs. A typical white paper has about 1,700 unique words filtered using NLP. The average ratio of unique technology-related words and the total number of unique words is 28 percent.

45 percent of ICOs are able to raise some funds. Conditional on raising money, the average (median) amount raised in an ICO is $U.S. 13 million ($U.S. 4 million). Ventures are able to reach 43 percent of ICO hardcap on average, and only 11 percent of ICOs reach or exceed their hardcap. Conditional on raising money, 41 percent of tokens end up being listed on at least one cryptographic exchange.

There is wide dispersion of our data quality measure across ICOs, which ranges between 0 and 4.34 with a standard deviation of 0.73, on the order of magnitude of average and median quality measures. This variation highlights the need to examine the robustness of empirical results using a subset of ICOs characterized by relatively high-quality data.

Panel B presents summary statistics of firm-initiated social media activity collected from four platforms: Twitter (tweets under firm's handle), Medium (articles by editors), Reddit (discussions), and Bitcointalk (posts) at the end of ICO. 3,625 (1,294, 935, 2,608) ICOs have non-zero cumulative tweets (discussions, editor articles, posts) on Twitter (Reddit, Medium, Bitcointalk) by ICO end. Panel C presents summary statistics of Github commits at the end of ICO. 1,806 projects have non-zero cumulative commits by ICO end.

In Panel D, we report summary statistics of the number of unique wallets containing the token issued in ICOs and the cumulative number of transactions involving the token by the end of the first day the token is traded on one of the crypto exchanges. 774 projects have information on wallets containing tokens issued in an ICO. The mean number of wallets on the first trading day is 1,836 and the mean cumulative number of transaction by the end of the first trading day is 7,588.

In Panel E, we present summary statistics of post-ICO returns of listed tokens over various horizons. We winsorize all returns at the top and bottom 5 percent to attenuate the influence of outliers. We calculate ICO "end-to-open" return using a token's opening price during the first day of trading on an exchange and the average ICO price computed as the ratio of the amount raised and the number of tokens in circulation. Conditional on a token being listed on an exchange, mean (median) ICO end-to-open return–that is, the adjustment of average ICO price from the ICO end day to its first trading day–is 384% (46%). These very large end-to-open returns are in line with results documented in other studies: Benedetti and Kostovetsky (2018) and Lee et al. (2019) report average ICO returns of 179% and 112%, respectively. As evident from the differences between mean and median returns, high mean end-to-open returns are driven by a few observations with extremely high returns, even after winsorization. These tend to be relatively small ICOs in terms of the amount raised. 59 percent of ICOs that successfully raise (some) funds from investors are never listed on a crypto exchange. In these cases, computing initial ICO returns is impossible. Thus, as an alternative treatment of these ICOs, instead of discarding them from the end-to-open return calculation, we assume the worst-case scenario, i.e. that investors lose 100% of their investments in tokens that are not listed on at least one exchange. Under this assumption, the mean ICO end-to-open return is 100%.

Mean first-day return, computed as the difference between the closing and opening prices of the first day of trading, is 10%, and more than 50 percent of ICOs have positive first-day returns. Median post-ICO cumulative returns – measured 30, 90, 180, and 365 days after the first trading day – are negative for all horizons, ranging from -30% to -85%. In addition, 69 percent of 30-day cumulative returns are negative, and this fraction increases to 84 percent for 365-day cumulative returns. Similar to the case of end-to-open returns, mean cumulative long-term returns are driven by a few ICOs with very high cumulative returns. As a result, mean post-ICO long-term returns are less negative (and equal zero at the 90-day horizon).

4 Determinants of ICO Success

In this section we analyze determinants of successful fund raising in an ICO and the likelihood of listing the issued token on a cryptographic exchange, while choosing the most trustworthy components of our data, as described in the previous section. The objectives of this analysis are: (1) to attempt to reconcile some of the disagreements in the existing literature regarding determinants of ICO success and (2) to uncover additional factors affecting ICO success not yet examined in the literature.

Table 6 presents results of estimating regressions of various measures of ICO success on ICO and project characteristics. In the first column of Table 6, we report results of a logistic regression in which the dependent variable is the extensive margin of funding success, i.e. an indicator equaling one if any funds were raised in an ICO.[21] In the second column, the dependent variable is the intensive margin of ICO success – the logarithm of the amount raised in the ICO plus one.[22] In the third column, the dependent variable is the relative funding success – the ratio of amount raised to ICO hardcap. Finally, in the last column, the dependent variable is an indicator equalling one if a token issued in an ICO eventually begins trading on at least one crypto exchange. To facilitate the interpretation of the results, in all logistic regressions here and below, we report marginal effects of each independent variable. Because of substantial time-series variation in average ICO characteristics (e.g., the proportion of technology-related words in ICO white papers has been declining over time), and because our sample's end point is relatively recent, reducing our ability to observe

[21] Raising any amount of funds may not necessarily quality as a success in an ICO. Thus, in Table A.3 in the Appendix, we examine the robustness of the main results discussed below to modifications of the ICO success indicator. In particular, we use three funding thresholds, reaching which is required for an ICO to qualify as a success – $U.S. 10,000, $U.S. 100,000, and $U.S. 1,000,000. The qualitative results are robust to the choice of the funding threshold and are generally more significant for higher thresholds.

[22] We add one to the amount raised (and some other variables discussed below) before taking logs in order to include in the sample observations with values of zero.

listings of some of the latest ICOs, all regressions include time (quarter) fixed effects. In addition, because of geography-driven differences in project types and industry-driven differences in ICO characteristics, we also include geographical region and industry fixed effects.

In constructing our sample, we generally do not fill in values of missing variables. One exception is that we follow Davydiuk et al. (2019) in assuming that whenever the amount raised is missing it equals zero. However, in Table A.1 in the Appendix, we examine the robustness of the results to more and less restrictive assumptions used in sample construction. First, in the most conservative specification, we do not fill in zeroes for missing values of amount raised. This reduces the sample size from 2,349 observations in the baseline specification in Table 6 to 1,346 observations. Second, we assume that hardcap is zero if reported missing, raising the number of observations to 2,798. Finally, in the least restrictive specification, we assume, in addition, that percentage or tokens for sale and the number of team members equal zero if reported missing, raising the number of observations to 5,138. All the qualitative results in these three samples are consistent with the results in the baseline sample in Table 6, discussed below.

The ability to raise funds in an ICO–at both the extensive and intensive margins, as well as the likelihood of listing the tokens issued in the ICO on a crypto exchange, are decreasing in the proportion of tokens available for sale in an ICO. A one-standard-deviation increase in the proportion of tokens for sale is associated with 7.5% reduction in the likelihood of raising funds and with 20% reduction in the likelihood of listing. In addition, a one-percentage-point increase in the proportion of tokens available for sale is associated with 1.5% reduction in the amount raised in an ICO. This finding is consistent with a negative signal conveyed by entrepreneurs attempting to sell a larger proportion of tokens to investors, resulting in lower remaining "skin in the game". This result is in line with Amsden and Schweizer (2018), Davydiuk et al. (2019), and Lee et al. (2019), and is reminiscent of a similar finding in the venture capital literature (e.g., Conti et al. (2013)).

Interestingly, an attempt to sell tokens to large/informed investors prior to an ICO does not significantly impact ICO funding success. This finding is consistent with Bourveau et al. (2019) and Deng et al. (2018), but is in contrast with the positive relation between presale and funding success, reported in Adhami et al. (2018), de Jong et al. (2018), Fisch (2019), and Lee et al. (2019). This result suggests that merely attempting a presale does not signal ICO quality to investors. In a robustness test, reported in Table A.2 in the Appendix, we replace the attempted presale indicator by the successful presale dummy, which equals one if the venture was successful in raising some funds from institutional/VC investors in a presale. Notably, while 52% of ICOs in our sample attempt a presale, in only 245 ICOs (less than 5% of the sample), the presale is successful. As shown in Table A.2, the effect of a successful presale on the amount raised in an ICO is significantly positive. This is not merely due to the mechanical effect of the amount raised in presale being part of the overall amount raised in the ICO. When we exclude the funds raised in presale from the overall amount raised, the coefficient on successful presale remains significantly positive, albeit four times smaller, suggesting that a successful presale, unlike a merely attempted one, does signal ICO quality to retail investors.

The likelihood of raising funds is independent of ICO hardcap. However, conditional on raising funds, ICOs attempting to raise more funds tend to end up doing so: A 1% increase in hardcap is associated with 0.35% increase in the amount raised in an ICO. However, ICOs with larger hardcap raise lower fraction of hardcap, as is evident from column 3. This is a novel finding in the ICO literature which is consistent with the theoretical argument that large offerings may send a negative signal to the market (e.g., Leland and Pyle (1977) and Miller and Rock (1985)) and with the idea that in the presence of downward-sloping demand, larger offerings reduce the likelihood of (relative) success (e.g., Scholes (1972)).

The likelihood of raising funds and of listing the token on an exchange is increasing in the KYC indicator: ICOs with KYC requirement are 15 percentage points more likely to be able to raise funds and 21 percentage points more likely to have the token listed on an exchange than ICOs without KYC requirement. This result is consistent with Davydiuk et al. (2019) and Deng et al. (2018), but stands in contrast with Lee et al. (2019), who find an insignificantly negative relation between ICO success and KYC requirement.

The likelihood of ICO success is also higher for ICOs with white paper available–ICOs with white papers are 5 (7) percentage points more likely to raise funds (obtain an exchange listing). ICOs with white papers are also able to raise 86% more funds on average and 6 percentage points more as a fraction of hardcap. In addition, both the extensive and intensive margins of funding success as well as the likelihood of listing are increasing in the size of the team associated with the project. For example, a one-standard-deviation increase in the number of team members is associated with 7 (6) percentage points higher probability of raising funds (listing). This evidence is broadly in line with Amsden and Schweizer (2018) and Bourveau et al. (2019). In addition, consistent with Davydiuk et al. (2019), the likelihood of raising funds, the amount raised and the probability of listing are increasing in the number of commits by the start of ICO, which is a proxy for the project's technological maturity. All this evidence suggests that a reduction in opaqueness surrounding an ICO raises the likelihood of securing funds, the absolute and relative (to hardcap) amount of funding obtained, as well as the likelihood of listing on an exchange.

Existing ICO literature does not provide much evidence on the effects of social media on ICO success. Existing studies typically either examine a single social media platform (e.g., Twitter in the case of Benedetti and Kostovetsky (2018)) or use an indicator of social media activity presence at the time of ICO (e.g., Bourveau et al. (2019) and Howell et al. (2020)). We present a detailed analysis of the effects of the extent of venture-initiated social media activity on ICO funding and listing success. Our results reveal that the effects of social media activity are quite nuanced and that they depend heavily on the social media platform used. Ventures active on Medium and Bitcointalk by the beginning of their ICO enjoy larger success, along all dimensions. For example, a 1% increase in the number of Medium articles by editors (Bitcointalk posts) is associated with 0.2% (0.4%) increase in the amount raised. On the contrary, venture-initiated Reddit activity is negatively associated with measures of ICO success, whereas the number of Twitter tweets by the venture tends not to be significantly related to ICO success. This variation in the effects of social media activity on various platforms may be due to differences in target audience. Bitcointalk and Medium tend to contain more detailed/technical information and are targeted to audiences with deeper understanding of the crypto markets, whereas Twitter and Reddit usually provide shorter and more superficial content and are typically aimed at more casual users.[23] Taken together, this evidence may be consistent with the signaling via social media activity being effective only when it is costly – as impactful Medium articles and Bitcointalk posts are costlier to create than Twitter and Reddit content. Medium articles are by far the most lengthy, having on 1,084 words on average, followed by Bitcointalk with 186 words on average. Twitter and Reddit are markedly shorter with

[23]Medium articles are by far the most lengthy, having on 1,084 words on average, followed by Bitcointalk with 186 words on average. Twitter and Reddit are markedly shorter with 13 and 77 words on average, respectively. Medium articles tend to be written in a more formal (near journalistic/academic) language and contain wealth of details and explanations. While not as lengthy as Medium, Bitcointalk posts are targeted to a specialized audience. Created in 2009 by Satoshi Nakamoto, the platform aims exclusively to discuss cryptocurrencies and blockchain topics. Its content is more technical and nuanced and caters towards expert audience.

13 and 77 words on average, respectively. Medium articles tend to be written in a more formal (near journalistic/academic) language and contain wealth of details and explanations. While not as lengthy as Medium, BitcoinTalk posts are targeted to a specialized audience. Created in 2009 by Satoshi Nakamoto, the platform aims exclusively to discuss cryptocurrencies and blockchain topics. Its content is more technical and nuanced and caters towards expert audience.

Existence of bonus programs (discounts to early ICO investors) is positively associated with funding success at both the extensive and intensive margins. This evidence stands in contrast with existing papers, which report either a negative relation between bonus availability and funding success (e.g., Lee et al. (2019)) or an insignificant relation (e.g., Amsden and Schweizer (2018) and Bourveau et al. (2019)). Bounty programs, on the other hand, tend to be negatively associated with the amount raised in the ICO, potentially consistent with less promising ICOs employing bounty hunters to advertise token issuance on social media channels.

In Table 7, we examine the robustness of our main results to using various subsets of the data. First, we estimate a regression of log amount raised, using data from each of the four data sources, separately, which have non-missing information on amount raised, hardcap, token supply, and tokens for sale, simultaneously, for at least some ICOs – `CoinGecko`, `CryptoCompare`, `ICObench`, and `ICOdrops`.[24] Many of the results reported above disappear once data from a single source is used. For example, the effect of percentage of tokens for sale on ICO funding success tends to be insignificant; the presence of a white paper is not significantly related to ICO success in three out of four subsamples; the effects of bonus and bounty become insignificant in most subsamples, as do the effects of Github commits and various measures of social media activity by the start of ICO. This evidence highlights the importance of combining data from various ICO aggregators and using only the most trustworthy pieces of data, as we do in the regressions in Table 6.

In an additional robustness test, reported in Table 7 as well, we re-estimate the regression of raised indicator, while restricting the set of data sources used to pick the most trustworthy values of variables. For each of the four main variables – amount raised, hardcap, token supply, and tokens for sale – we include only top-three sources in terms of data quality for the respective variable. The results are qualitatively similar to those in the baseline specification, further highlighting the importance of picking data from top-quality sources. In addition, this finding suggests that our

[24] We also estimate regressions with other three measures of ICO success – raised dummy, raised-to-hardcup, and listing dummy – and find results consistent with the discussion below.

procedure is successful in weeding out data from questionable sources even if we do not discard data from such sources altogether. Finally, we re-estimate the ICO success regression for ICOs belonging to the top tercile of overall observation-level data quality.[25] Despite significant loss in the number of observations (898, compared to 2,349 in the baseline specification), most of the qualitative results are consistent with the baseline findings. The only exceptions are that: the coefficient on the white paper indicator becomes insignificant, likely because most top-data-quality ICOs have white papers available; and the effects of bonus and bounty on ICO success become insignificant. Overall, the fact that the results for observations with the best data quality are largely consistent with the results for the full sample suggests that our procedure for choosing the most trustworthy pieces of data results in removing significant part of the noise from an otherwise noisy data.

Table 8 narrows the analysis to the subsample of ICOs with available white papers, with the objective of examining the effects of white paper contents on measures of ICO success. Similar to Amsden and Schweizer (2018), Bourveau et al. (2019), and Howell et al. (2020), we examine the effects of white paper length on funding and listing success. Unlike existing studies, we also examine the effects on ICO success of white paper informativeness, as proxied by the proportion of technical words in the white paper, which we find to be significantly positively related to both the extensive and intensive margins of ICO success. This finding suggests a negative impact of the ICO white paper's opacity on the amount raised. Interestingly, differently from aforementioned papers, we find that after controlling for white paper informativeness, its length tends be only marginally associated with the extensive and intensive margins of funding success.

5 Post-ICO Operating and Financial Performance

In addition to analyzing factors associated with ICO funding success, our data enable us to examine determinants of ventures' post-ICO operating and financial performance. This analysis complements existing studies that examine determinants of success in raising funds via an ICO (e.g., Bourveau et al. (2019), Davydiuk et al. (2019), Howell et al. (2020), Hu et al. (2018), and Lee et al. (2019), as well as the analysis in Section 4 in this paper).

[25] Note that this is different from limiting the data to highest-quality data sources for each variable.

In particular, in what follows, we focus on the effect on ventures' post-ICO performance of entrepreneurs' "skin in the game". We examine the relations between the percentage of tokens for sale in the ICO (which is inversely related to insiders' skin in the game and a) post-ICO inputs into the venture's production and b) post-ICO product/platform adoption by users. This analysis is a test of a prediction of the model by Gan et al. (2020) that operational shirking in ICO-funded projects is decreasing in entrepreneurs' post-ICO skin in the game. Focusing on the effect of skin in the game on post-ICO operating performance complements the analysis in Davydiuk et al. (2019), who examine empirically the effect of the fraction of tokens offered in an ICO on initial success in raising funds.

In the following subsection, we examine whether a venture's operating performance–product/platform adoption by users–has financial implications. In other words, we are interested in the contemporaneous relation between post-ICO product/platform adoption and token returns. This analysis is a test of the hypothesis in Cong et al. (2020), whose model shows that token prices are increasing in platform adoption. An analysis of the contemporaneous relation between venture's operating performance and its token's long-term returns complements Lee et al. (2019), who examine other determinants of post-ICO long-term returns.

5.1 Skin in the Game, Inputs into Post-ICO Production, and Product/Platform Adoption

One of the important implications of the model in Gan et al. (2020) is that *"the firm is discouraged from pursuing production ex-post given it does not have enough 'skin' left in the game, [providing] a possible explanation for the loss of motivation or productivity post ICO of some well-funded startups in practice."* In other words, inputs into the venture's production and its resulting product/platform adoption are predicted to be decreasing in the percentage of tokens available for sale to outside investors in the venture's ICO.

Common inputs into ICO-funded ventures include code production, typically measured using code revisions on Github (e.g., Amsden and Schweizer (2018), Bourveau et al. (2019), and Davydiuk et al. (2019), and venture-initiated social media activity (e.g., Amsden and Schweizer (2018) and Bourveau et al. (2019)). Examining post-ICO code production and venture-initiated social media activity complements Howell et al. (2020), who examine determinants of post-ICO employment growth and of eventual project failure. Thus, the first prediction following from Gan et al. (2020) is:

Prediction 1: Post-ICO inputs into venture's production, measured by the number of code revisions on Github and by venture-initiated social media activity, are expected to be negatively related to the percentage of tokens for sale in the ICO.

Product/platform adoption by users raises the need to hold and transact in venture's token for reasons other than investing/trading. Thus, venture's operating performance can be proxied by the number of cryptographic wallets holding the venture's token, as well as by the number of on-chain transfers involving the token across wallets. Thus, a related prediction that follows from the model in Gan et al. (2020) is:

Prediction 2: Post-ICO product/platform adoption, measured by the number of wallets holding the venture's token and the number of on-chain transfers involving the token, is expected to be negatively related to the percentage of tokens for sale in the ICO.

In Table 9, we report results of regressions in which the dependent variable is the post-ICO growth in the number of cumulative commits, i.e. code revisions on Github, measured as the logarithm of the ratio of cumulative commits 30 (90, 180, 365 days) following ICO end and cumulative commits at ICO end. The main independent variable is the percentage of tokens for sale in the ICO. Since we are interested in the effects of entrepreneurs' post-ICO skin in the game on the change in post-ICO code production relative to pre-ICO production, we include lagged growth in commits, measured over 90 days prior to ICO end. In addition, since exponential growth in cumulative commits may be slower for high cumulative coding activity, we also control for the level of cumulative commits 90 days prior to ICO end. Finally, post-ICO code production is clearly related to success in obtaining funds in an ICO. To control for the effect of ICO funding success on post-ICO code development activity, we include an indicator variable equaling one if some funds were raised in the ICO.

Post-ICO code development is significantly negatively related to the percentage of tokens for sale across all horizons. This relation is also economically significant: A one-standard-deviation increase in the percentage of tokens for sale (0.24) is associated with 0.7 (2.8, 4.4, 7.8) percentage point reduction in the growth in cumulative commits at the 30-day (90-day, 180-day, 365-day) horizon, corresponding to 3% (6%, 7%, 9%) of the standard deviation of growth in commits over respective horizon.

In addition, the pace of post-ICO code production is strongly positively related to pre-ICO pace, suggesting persistence in code production activity. Finally, raising funds in an ICO has a pronounced positive effect on post-ICO code development – ventures that managed to raise funds in their ICO exhibit post-ICO growth in cumulative commits that is 5 (10, 16) percentage points higher than that of ventures that failed to raise funds in their ICO over 90-day (180-day, 365-day) horizon.

In Table 10, we analyze the relation between post-ICO growth over various horizons in venture-initiated social media activity on one hand and entrepreneurs' post-ICO skin in the game, as (inversely) proxied by the percentage of tokens for sale, on the other hand. Table 10 has four panels, devoted to four measures of firm insiders' social media activity – Twitter tweets by the firm (Panel A), Medium articles written by editors (Panel B), Reddit discussions (Panel C), and Bitcointalk posts (Panel D).

The percentage of tokens for sale in an ICO is negatively related to post-ICO growth in cumulative social media activity. This relation is significant statistically and economically over all horizons for three out of four social media platforms – Twitter, Medium, and Bitcointalk. A one-standard-deviation increase in the percentage of tokens for sale is associated with 0.8 (3.7, 5.7, 8.8) percentage point decrease in cumulative growth in Twitter tweets over 30-day (90-day, 180-day, 365-day) horizon, corresponding to 2% (6%, 8%, 10%) of the standard deviation of growth in the number of tweets over the relevant horizon. When social media activity is proxied by the number of Medium editor articles, a one-standard-deviation increase in the percentage of tokens for sale is associated, depending on the horizon, with a reduction of 9%–14% of the standard deviation of the growth in cumulative number of articles. For Bitcointalk posts, the same increase in the percentage of tokens for sale is associated with 3%–7% standard-deviation reduction in the cumulative number of posts.

The association between the percentage of tokens for sale and the post-ICO growth of venture-initiated activity on Reddit is significant only at longer horizons, and is less meaningful economically. This finding is consistent with the lack of positive relation between pre-ICO Reddit activity and ICO funding success, discussed in the previous section. This result may be traced to the type of content typically displayed on Reddit–less detailed and more superficial than on other platforms, suggesting that ventures realize that potential product/platform adopters as well as token investors are less likely to treat social media activity as an input to the venture's production if engaging in such activity is nearly costless.

Similar to the results in Table 9, the growth in measures of cumulative social media activity exhibits persistence, as evidenced by the positive and significant coefficients on pre-ICO growth in measures of cumulative social media activity in three panels out of four. In addition, exponential growth in cumulative social media activity slows down as higher cumulative levels of it are achieved, as follows from the negative coefficients on the pre-ICO level of social media activity in all four panels. Finally, raising funds in an ICO has a clearly positive impact on the growth in cumulative social media activity: ventures that raised funds in their ICO exhibit up to 4.2 (14.3, 23.9, 33.6) percentage point higher growth in measures of cumulative social media activity at 30-day (90-day, 180-day, and 365-day) horizon.

Overall, the results in Tables 9 and 10 are strongly supportive of Prediction 1, that post-ICO code development and social media activity are positively associated with entrepreneurs' post-ICO skin in the game, as follows from the model in Gan et al. (2020).

In Table 11, we examine whether, in addition to affecting inputs into post-ICO production, entrepreneurs' skin in the game impacts ventures' post-ICO operational success, i.e. product/platform adoption. The first proxy for product/platform adoption, used in Panel A of Table 11, is the post-ICO change in the number of wallets containing the venture's token. The second, used in Panel B, is the post-ICO growth in cumulative number of on-chain transfers of the token across wallets.

The relation between the percentage of tokens for sale in an ICO and post-ICO growth in the number of wallets holding the venture's token is significantly negative at all horizons.[26] The economic magnitude of this relation is large: A one-standard-deviation increase in the percentage of tokens for sale is associated with a 5.6 (8.2, 9.1) percentage point reduction in the post-ICO

[26]Note that the number of wallets containing the token and the cumulative number of token transfers at ICO end date tend to be small. Thus, we compute the growth in the number of wallets and in the cumulative number of transfers at different points in time relative to day 30 post ICO end.

growth in the number of wallets containing the token at the 90-day (180-day, 365-day) horizon, or to 10% (9%, 7%) standard-deviation decrease in post-ICO wallet growth rate. Similar results are reported in Panel B. A one-standard-deviation increase in the percentage of tokens for sale is associated, depending on the horizon, with 8%-9% standard-deviation reduction in the growth rate of cumulative number of wallet transfers involving the token.

Notably, the negative associations between the percentage of tokens for sale on one hand and post-ICO evolution of the number of wallets containing the token and cumulative number of on-chain transfers of the token on the other hand stand in contrast with the evidence in the only other paper that examines this relation – Deng et al. (2018), who report an insignificant association between entrepreneurs' skin in the game and platform adoption. This discrepancy further highlights the importance of using multiple data sources and attempting to use only the most trustworthy pieces of data when examining ICO outcomes.

Similar to the evidence in Tables 9 and 10, post-ICO growth in the number of wallets and in the cumulative number of wallet transfers is positively related to the growth in these measures during the first 30 days following the ICO and is negatively related to the number of wallets and cumulative number of on-chain transaction by ICO end. The growth in the number of wallets containing the token is not significantly related to whether funds were raised in the ICO, likely because the vast majority (91%) of ICOs with wallet data available were successful in raising funds.

The robust negative relation between the percentage of tokens for sale and the two measures of post-ICO product/platform adoption supports Prediction 2, that entrepreneurs' post-ICO skin in the game is negatively related to post-ICO operating performance, and is consistent with entrepreneurs' lower skin in the game leading to shirking, as highlighted in Gan et al. (2020).

5.2 Product/Platform Adoption and Post-ICO Financial Performance

In this subsection, we focus on the relation between the adoption of a venture's product/platform and the financial performance of the venture's token. This analysis allows answering the following question: Do the effects of skin in the game on inputs into post-ICO production and on post-ICO operating performance have implications for ventures' long-term financial performance and valuation? One of the results in Cong et al. (2020) is that "token price increases quickly with adoption in the early stage, changes gradually in the intermediate stage, and speeds up again once the user base reaches a sufficiently high level.” This result translates into the following empirical prediction:

Prediction 3: Post-ICO long-term token returns are expected to be positively associated with contemporaneous product/platform adoption.

To test this prediction, we estimate regressions of longer-term returns following a token’s first listing on one of the cryptographic exchanges at horizons ranging from 30 days to 365 days post-listing. The estimates of these regressions are reported in Table 12. The main independent variables are the contemporaneous growth in the number of wallets containing the token (in odd columns) or, alternatively, the contemporaneous growth in the cumulative number of transfers involving the token (in even columns). Since post-ICO returns may be related to the visibility of the token at the time of listing, we control for the number of wallets containing the token or for cumulative number of wallet transactions involving it at the time of first trading. As post-ICO returns may potentially exhibit momentum and/or reversal patterns, we control for the relative change in token price between ICO end date and the open of the first day the token is traded on an exchange (“end-to-open return”) as well as for return on the first trading day (“first-day return”). We also include the amount raised in the ICO in the set of explanatory variables, since excessive success in fund raising may be a signal of ICO overvaluation. Finally, since returns on the majority of crypto assets tend to be correlated with returns on Bitcoin and Ethereum (e.g., Adhami et al. (2018)), we control for contemporaneous returns of these two cryptocurrencies.

Consistent with Prediction 3, the relation between post-listing return and contemporaneous growth in the number of wallets containing the token is positive. This relation is highly statistically significant at most horizons and is economically sizable: A one-standard-deviation increase in the growth in the number of wallets – 1.0 (1.2, 1.3, 1.5) at 30-day (90-day, 189-day, 365-day horizon) is associated with 8 (8, 15, 27) percentage point increase in contemporaneous token return. The fact that this relation is stronger at the short (30-day) horizon and long (365-day) horizon than at the medium horizons (90-day and 180-day) is consistent with a more refined interpretation of Cong et al. (2020) that the sensitivity of token price to the speed of product/platform adoption is the largest in the early and late stages of adoption than in the intermediate stage.

The association between post-listing return and contemporaneous growth in the cumulative number of on-chain transfers involving the token is also positive and highly statistically and economically significant only at all horizons. A one-standard-deviation increase in the growth in the post-listing growth in cumulative number of on-chain transfers across various horizons is associated with 13-34 percentage point increase in contemporaneous post-listing return.

Consistent with our conjectures, there is a positive relation between the number of wallets and cumulative transfers at ICO date on one hand and post-ICO longer-term return on the other hand. The relations between ICO end-to-open return and first-day return on one hand and longer-term post-ICO returns on the other hand are negative, suggesting that post-ICO returns exhibit reversal. Consistent with potential overvaluation of ICOs on average, those raising large amounts tend to have lower post-listing returns. Finally, token returns tend to load positively on contemporaneous Bitcoin and Ether returns (except for the 365-day horizon for the latter).

Overall, the results in Table 12 support Prediction 3, following from the model in Cong et al. (2020), that post-ICO returns are positively related to contemporaneous product/platform adoption. In other words, operational success post-ICO does translate into financial success, as measured by the return on the venture's token.

6 Conclusions

We provide one of the most comprehensive empirical analyses of initial coin offerings. ICO data is generally of low quality, therefore a significant portion of our paper deals with ways to characterize data quality. We propose a data quality measure both at the level of data source-variable pair and at the level of a particular ICO. This measure allows us to (i) identify the most trustworthy pieces of data for each ICO and (ii) verify the robustness of our results using subsamples of ICOs characterized by the highest data quality. Employing the most reliable data at the ICO level and focusing on subsamples of ICOs with the highest data quality mitigates concerns about wrong inference due to measurement error.

The first part of our empirical analysis focuses on factors influencing ICO funding success. We discover some novel determinants of ICO success, such as ICO hardcap, ICO white paper informativeness, and pre-ICO social media activity on various platforms. Some of our findings–e.g., the positive relation between ICO bonus programs and ICO funding success–overturn results in existing papers. Some other results–e.g., the positive relation between the know-your-customer requirement and ICO funding success and an insignificant effect on ICO success of attempted presale–shed light on inconsistent findings in existing literature.

In the second part of the empirical analysis, we focus on post-ICO operating and financial success. In particular, we test the prediction of the model in Gan et al. (2020) that ventures' operating success is impacted by entrepreneurs' skin in the game. We find that the relations between the skin in the game and post-ICO inputs into ventures' production–code revision activity and social media activity–are indeed positive and significant both statistically and economically. In addition, we find that entrepreneurs' skin in the game has significant impact on product/platform adoption by users.

Finally, we report a strong relation between product/platform adoption and contemporaneous medium-term token returns, consistent with the prediction of the model of Cong et al. (2020). This finding suggests that the effects of entrepreneurs' post-ICO skin in the game on venture's operating success have financial ramifications.

References

Adhami, S., G. Giudici, and S. Martinazzi (2018). Why do businesses go crypto? an empirical analysis of initial coin offerings. Journal of Economics and Business 100, 64–75.

Amsden, R. and D. Schweizer (2018). Are blockchain crowdsales the new "gold rush"? success determinants of initial coin offerings. McGill University Working Paper.

Bakos, Y. and H. Halaburda (2019). The role of cryptographic tokens and icos in fostering platform adoption. New York University Working Paper.

Benedetti, H. and L. Kostovetsky (2018). Digital tulips? returns to investors in initial coin offerings. Boston College Working Paper.

Bourveau, T., E. T. De George, A. Ellahie, and D. Macciocchi (2019). Information intermediaries in the crypto-tokens market. Columbia University Working Paper.

Chod, J. and E. Lyandres (2020). A theory of icos: Diversification, agency, and information asymmetry. Management Science forthcoming.

Chod, J., N. Trichakis, and S. A. Yang (2020). Platform tokenization: Financing, governance, and moral hazard. Massachusetts Institute of Technology Working Paper.

Cong, L., Y. Li, and N. Wang (2020). Tokenomics: Dynamic adoption and valuation. Review of Financial Studies Forthcoming.

Conti, A., M. Thursby, and F. T. Rothaermel (2013). Show me the right stuff: Signals for high-tech startups. Journal of Economics and Management Strategy 22(2), 341–364.

Davydiuk, T., D. Gupta, and S. Rosen (2019). De-crypto-ing signals in initial coin offerings: Evidence of rational token retention. Carnegie Mellon University Working Paper.

de Jong, A., P. Roosenboom, and T. van der Kolk (2018). What determines success in initial coin offerings? Venture Capital, 1–23.

Deng, X., Y. T. Lee, and Z. Zhong (2018). Decrypting coin winners: Disclosure quality, governance mechanism and team networks. Shanghai Univesity of Finance and Economics Working Paper.

Fisch, C. (2019). Initial coin offerings (icos) to finance new ventures. Journal of Business Venturing (34), 1–22.

Florysiak, D. and A. Schandlbauer (2020). The information content of ico white papers. University of Southern Denmark Working Paper.

Gan, J., G. Tsoukalas, and S. Netessine (2020). Initial coin offerings, speculators, and asset tokenization. Management Science forthcoming.

Goldstein, I., D. Gupta, and R. Sverchkov (2020). Initial coin offerings as a commitment to competition. University of Pennsylvania Working Paper.

Howell, S. T., M. Neissner, and D. Yermack (2020). Initial coin offerings: Financing growth with cryptocurrency token sales. Review of Financial Studies forthcoming.

Hu, A., C. Parlour, and U. Rajan (2018). Cryptocurrencies: Stylized facts on a new investible instrument. Financial Management 48(4), 1049–1068.

Huang, W., M. Meoli, and S. Vismara (2019). The geography of initial coin offerings. Small Business Economics, 1–26.

Kraaijeveld, O. and J. De Smedt (2020). The predictive power of public twitter sentiment for forecasting cryptocurrency prices. Journal of International Financial Markets, Institutions and Money 65, 43–76.

Lee, J., T. Li, and D. Shin (2019). The wisdom of crowds in fintech: Evidence from initial coin offerings. University of Florida Working Paper.

Leland, H. E. and D. H. Pyle (1977). Informational asymetries, financial structure, and financial intermediation. Journal of Finance 32(2), 371–387.

Li, J. and W. Mann (2018). Digital tokens and platform building. UCLA Working Paper.

Lyandres, E. (2020). Product market competition with crypto tokens and smart contracts. Boston University Working Paper.

Miller, M. H. and K. Rock (1985). Dividend policy under asymmetric information. Journal of Finance 40(4), 1031–1051.

Scholes, M. S. (1972). The market for securities: Substitution versus price pressure and the effects of information on share prices. Journal of Business 45(2), 179–211.

Sockin, M. and W. Xiong (2020). A model of cryptocurrencies. Princeton University Working Paper.

Figure 1. ICOs over time. This figure reports monthly values of the number of ICOs that are able to raise funds (dashed blue line, left axis) and the total amount raised across all ICOs each month (billions of dollars, right axis). The solid red line excludes the EOS ICO in June 2018, while the dotted red line includes it. Monthly observations go from August 2015 to August 2019. The observations reported for the month of August 2015 group all ICOs up to August 2015.

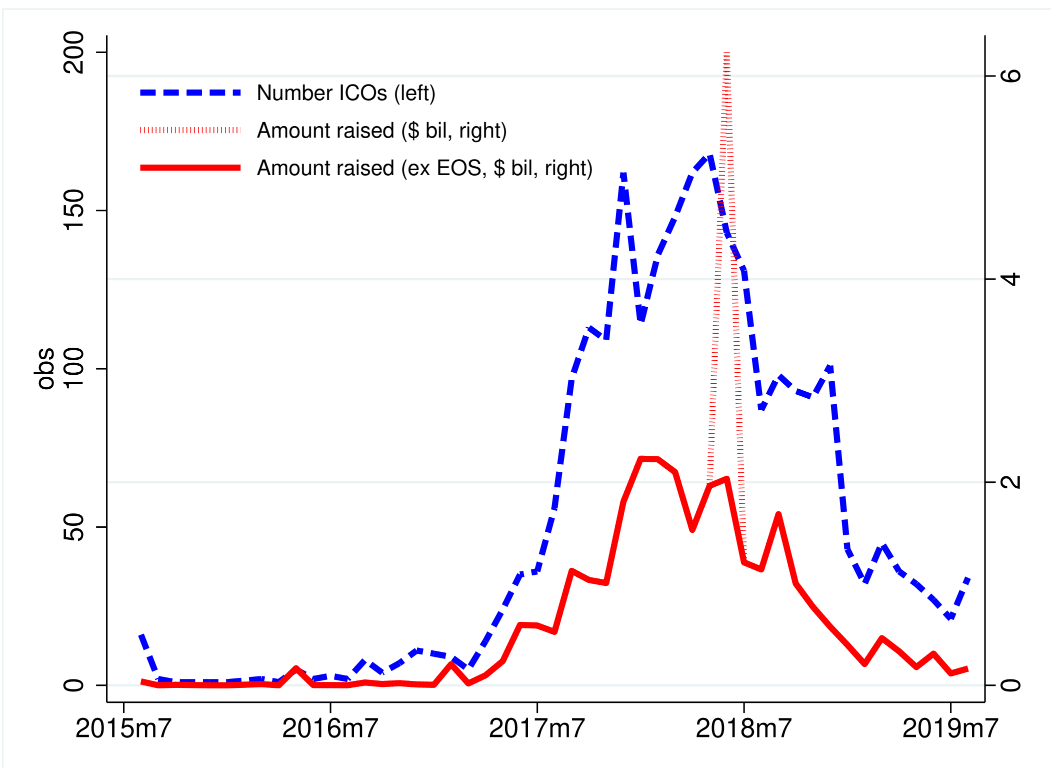

Figure 2. ICO data quality. This figure reports the overall ICO data quality in our dataset as a function of the total number of sources and average consistency across sources for each ICO. Darker points refer to low ICO data quality, while lighter points refer to high ICO data quality.

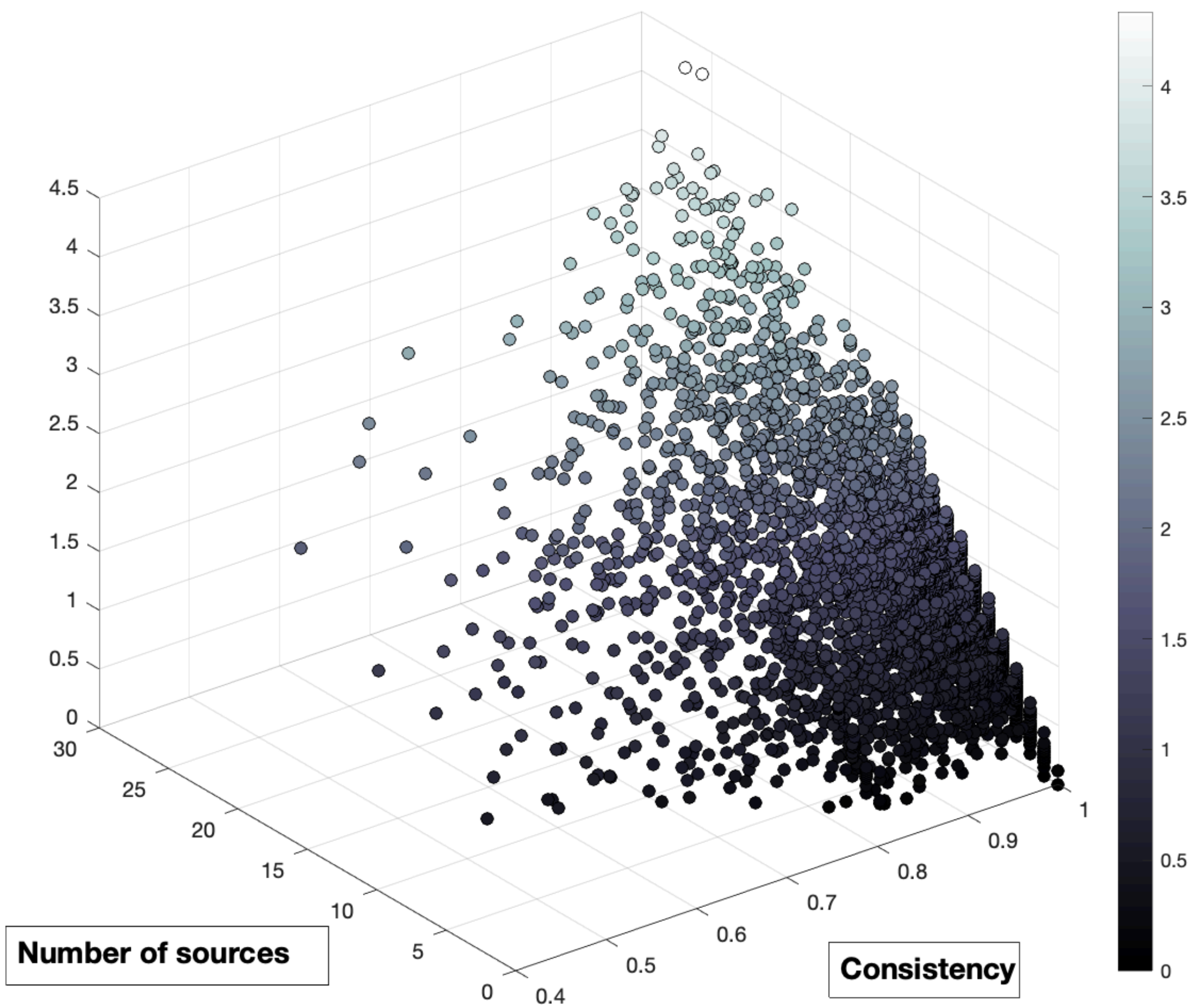

Table 1. Variable definitions. This table lists the variables used in the empirical analysis.

Name	Type	Description
amount raised	continuous	total amount raised in an ICO (in all currencies, converted to $U.S., including presale)
hardcap	continuous	maximum amount allowed to be raised (in all currencies, converted to $U.S.)
token supply	continuous	total amount of tokens that can be issued according to the smart contract
tokens for sale	continuous	total amount for tokens released for the crowd sale
percent for sale	continuous	ratio of token supply to tokens for sale
presale	indicator	whether the project attempted a presale prior to ICO (either successfully or not)
presale raised	indicator	whether the project has information on the amount raised in a presale
whitelist	indicator	whether the project offers whitelist to early investors
KYC	indicator	whether the project complies with the "know your customer" requirement
team members	discrete	total number of team members verified using LinkedIn
bonus	indicator	whether the project has information on bonus provided to token buyers
bounty	indicator	whether the project has information on bounty programs
white paper	indicator	whether the project has a white paper associated with it
industry	indicator	whether the project has information about the industry
entertainment	indicator	whether the project is in the entertainment sector
business services	indicator	whether the project is in the business services sector
general blockchain	indicator	whether the project is in the general blockchain sector
other software	indicator	whether the project is in the other software sector
finance	indicator	whether the project is in the finance sector
location	indicator	whether the project has information about location
West. Europe, Can., Austr.	indicator	whether the project is located in Western Europe, Canada, or Australia
Eastern Europe	indicator	whether the project is located in Eastern Europe
Asia	indicator	whether the project is located in Asia
USA	indicator	whether the project is located in the USA
other location	indicator	whether the project is in located in a different location from the ones listed previously
crypto friendly	indicator	whether the project is in a crypto friendly country
#NLP words	discrete	word count in the white paper after natural language treatment
tech ratio	continuous	ratio of unique tech words in the white paper and the total unique words in it
raised dummy	indicator	whether any funds were raised in an ICO
raised-to-hardcap	continuous	ratio of the amount raised in the ICO to its hardcap

Table 1. Variable definitions. (Continued)

Name	Type	Description
listing	indicator	whether the project is listed on at least one cryptographic exchange
ICO quality	continuous	ICO data quality measure
Twitter	continuous	cumulative Twitter activity at time t. Only tweets from the ICO project official account are considered.
Reddit	continuous	cumulative Reddit activity at time t. Only discussions from the ICO project official account are considered.
Medium	continuous	cumulative Medium activity at time t. Only articles by editors from the ICO project official account are considered.
Bitcointalk	continuous	cumulative BitcoinTalk activity at time t. Only posts from the ICO project official account are considered.
Commits	continuous	cumulative Github commit activity at time t.
Wallets	continuous	total number of cryptographic wallets containing at least one token at time t
Wallet transfers	continuous	cumulative number of on-chain transfers of token at time t
t-day return (log)	continuous	percentage difference between token price at closing after t days and token price at closing in the first trading day
ICO end-to-open return (conditional)	continuous	percentage difference between token price at opening during the first trading day and token price at the end of ICO (if available), computed for ICOs that are eventually listed on an exchange
ICO end-to-open return (unconditional)	continuous	percentage difference between token price at opening during the first trading day and token value at the end of ICO (if available), computed for all ICOs that raised nonzero funds, while assuming -100% return for ICOs not listed on an exchange
ICO first day return	continuous	percentage difference between token price at closing and token price at opening during the first trading day

Table 2. Summary of data sources. This table provides a summary of the data sources we employ. In Panel A we list each source and the number of projects covered by it. An ICO is considered to be covered by a source if information is available on the value of at least one of four key variables: total amount raised in an ICO, maximum amount allowed to be raised (hardcap), total supply of tokens that may be issued (token supply), and the number of tokens released for crowd sale (tokens for sale). We also provide a description of the type of data and main variables that we extract from each source. In Panel B, we report the distribution of projects across the 11 ICO aggregators. Number of matches is the number of data sources reporting (some) data on a given ICO.

Panel A: ICO data sources

Source	Type	Projects	Main variables
www.Etherscan.io	ico	901	listed exchanges, circulating supply.
www.CoinDesk.com	ico	829	raised, project name, ICO end date, cumulative funding.
www.CoinGecko.com	ico	2,831	raised, hardcap, token supply, location.
www.CryptoCompare.com	ico	988	raised, hardcap, location, tokens for sale.
www.ICObench.com	ico	3,705	raised, hardcap, location, ratings.
www.ICOdrops.com	ico	591	raised, hardcap, location, token supply.
www.ICOrating.com	ico	3,628	raised, hardcap, location, ratings,
www.ICOmarks.io	ico	3,253	raised, hardcap, location, token supply.
www.ICOdata.io	ico	1,842	raised, hardcap, location, token supply.
www.FoundICO.com	ico	2,202	raised, hardcap, location, industry, rating.
www.Tokendata.io	ico	2,075	raised
www.CoinMarketCap.com	price	4,047	open, high, low, close, volume, market cap.
www.Ethplorer.io	transactions	1,352	wallet address, transaction value.
www.GitHub.com	source code	3,157	commits, source commits, feature commits.
www.twitter.com	social media	5,927	tweets, replies, retweets, likes.
www.reddit.com	social media	3,055	posts, thumbs, comments.
www.medium.com	social media	3,607	articles, claps, comments.
www.bitcointalk.org	social media	3,685	post, activity, merits.
white papers	white paper contents	2,962	NLP word count, tech ratio.

Panel B: Distribution of projects across ICO data sources

Number of matches	1	2	3	4	5	6	7	8	9	10	11
Number of projects	2,399	1,295	1,096	856	577	340	312	274	130	59	5
Percent of projects	33%	18%	15%	12%	8%	5%	4%	4%	2%	1%	0%

Table 3. Distribution of ICOs across data sources. For each of the 11 ICO aggregators, this table reports the number of observations for four key variables: total amount raised in an ICO, maximum amount allowed to be raised (hardcap), total supply of tokens that may be issued (token supply), and the number of tokens released for crowd sale (tokens for sale). For each of the four variables, we also report the mean deviation from the consensus (average) value at the source-variable level. For variable x, source i, and observation k, the deviation from the average value is measured as the absolute value of the difference between the variable value $x_{i,k}$ and the average value across all sources reporting values for this variable, $\bar{x}_i$, divided by the sum of the two values, namely $\left|\frac{x_{i,k}-\bar{x}_k}{x_{i,k}+\bar{x}_k}\right|$. The mean deviation is the mean of $\left|\frac{x_{i,k}-\bar{x}_k}{x_{i,k}+\bar{x}_k}\right|$for each source and variable, $\bar{x}_i$.

Source	Amount raised		Hardcap		Token supply		Tokens for sale	
	Obs	Mean deviation	Obs	Mean deviation	Obs	Mean deviation	Obs	Mean deviation
www.Etherscan.io	161	0.097	0	NA	809	0.106	0	NA
www.CoinDesk.com	827	0.074	0	NA	0	NA	0	NA
www.CoinGecko.com	1,095	0.062	1,814	0.063	1,289	0.089	1,383	0.107
www.CryptoCompare.com	519	0.111	706	0.071	769	0.12	748	0.124
www.ICObench.com	1,722	0.064	1,654	0.060	2,644	0.084	3,133	0.07
www.ICOdrops.com	605	0.071	534	0.068	536	0.099	463	0.092
www.ICOrating.com	1,118	0.090	1,863	0.061	2,334	0.189	0	NA
www.ICOmarks.io	0	NA	1,251	0.040	2,298	0.076	2,658	0.061
www.ICOdata.io	1,078	0.100	1,166	0.082	363	0.293	0	NA
www.FoundICO.com	0	NA	137	0.041	0	NA	1,977	0.099
www.Tokendata.io	921	0.083	0	NA	0	NA	0	NA

Table 4. Bancor data quality. This table provides an example of the calculation of our data quality measure for the ICO of Bancor. Panel A reports available data on amount raised, hardcap, token supply, and tokens for sale, across the 11 ICO aggregators. For each variable, we report the average value of the variable across sources and its consistency, defined as 1 minus the mean deviation from the average value of this variable across data sources, where mean deviation is defined in Table 3. Panel B reports for each available data source the quality of that source for that variable, computed as the inverse of that data source's mean deviation, reported in Table 3, divided by the largest value across data sources. For each variable we report the sum of the source quality values across all four variables above. The adjusted quality value for each variable is given by the product of the sum of source quality values and corresponding consistency value for Bancor, reported in Panel A. The ICO data quality is the simple average of the adjusted quality values.

Panel A: Available Data (Bancor Example)

Source	Amount raised	Hardcap	Token supply	Tokens for sale
www.Etherscan.io	153,000,000		77,566,371	
www.CoinDesk.com	153,000,000			
www.CoinGecko.com			75,783,855	
www.CryptoCompare.com	153,000,000	36,000,000	79,320,000	39,660,000
www.ICObench.com	153,000,000			
www.ICODdrops.com	153,000,000			
www.ICOrating.com	153,000,000			
www.ICOmarks.io				
www.ICOdata.io	153,000,000	0	56,889,807	
www.FoundICO.com				
www.Tokendata.io	153,000,000			
Average	153,000,000	18,000,000	72,390,008	39,660,000
Consistency	1.000	0.333	0.944	1.000

Panel B: Source Quality (Bancor Example)

Source	Amount raised	Hardcap	Token supply	Tokens for sale
www.Etherscan.io	0.660		0.717	
www.CoinDesk.com	0.865			
www.CoinGecko.com			0.854	
www.CryptoCompare.com	0.577	0.563	0.633	0.492
www.ICObench.com	1.000			
www.ICOdrops.com	0.901			
www.ICOrating.com	0.711			
www.ICOmarks.io				
www.ICOdata.io	0.640	0.488	0.259	
www.FoundICO.com				
www.Tokendata.io	0.771			
Sum of source qualities	6.125	1.051	2.464	0.492
Adjusted quality	6.125	0.350	2.326	0.492
ICO data quality	2.323			

Table 5. Summary statistics. This table reports the mean, standard deviation, minimum value, median value, maximum value, and number of observations for the variables used in the empirical analysis. Variables are described in Table 1.

Panel A: ICO Variables

	Mean	Std. Dev.	Min	Median	Max	Obs.
ICO characteristics						
hardcap	62.99	3,788.89	0.00	20.00	229,000	3,653
% for sale	0.56	0.24	0.00	0.57	1.00	3,883
presale	0.52	0.50	0.00	1.00	1.00	5,376
presale raised	0.04	0.19	0.00	0.00	1.00	5,376
whitelist	0.40	0.49	0.00	0.00	1.00	5,376
kyc	0.50	0.50	0.00	1.00	1.00	5,376
number of team members	10.64	8.03	1.00	9.00	74.00	3,452
bonus	0.37	0.48	0.00	0.00	1.00	5,376
bounty	0.17	0.38	0.00	0.00	1.00	5,376
Industry						
industry	0.61	0.49	0.00	1.00	1.00	5,376
finance	0.39	0.49	0.00	0.00	1.00	3,289
other software	0.12	0.32	0.00	0.00	1.00	3,289
business services	0.21	0.40	0.00	0.00	1.00	3,289
entertainment	0.18	0.39	0.00	0.00	1.00	3,289
general blockchain	0.10	0.30	0.00	0.00	1.00	3,289
Location						
location	0.82	0.38	0.00	1.00	1.00	5,376
West. Europe, Can., Austr.	0.32	0.47	0.00	0.00	1.00	4,428
Eastern Europe	0.20	0.40	0.00	0.00	1.00	4,428
Asia	0.19	0.39	0.00	0.00	1.00	4,428
USA	0.13	0.34	0.00	0.00	1.00	4,428
other location	0.11	0.32	0.00	0.00	1.00	4,428
crypto friendly	0.42	0.49	0.00	0.00	1.00	4,428
White paper characteristics						
white paper	0.37	0.48	0.00	0.00	1.00	5,376
# NLP words	1,728.89	792.47	139.00	1,618.00	7,776.00	2,000
# NLP words (log)	7.35	0.49	4.94	7.39	8.96	2,000
tech ratio	0.28	0.06	0.03	0.28	0.65	2,000
ICO Outcomes						
raised dummy	0.45	0.50	0.00	0.00	1.00	5,376
amount raised, conditional on raising	13.24	90.28	0.00	3.82	4,197.96	2,442
raised-to-hardcap, conditional on raising	0.43	0.39	0.00	0.29	1.00	1,927
listing	0.41	0.49	0.00	0.00	1.00	2,442
ICO data quality						
ICO data quality	1.07	0.73	0.00	0.95	4.34	5,376

Table 5. Summary statistics. (Continued)

Panel B: Social Media

	Mean	Std. Dev.	Min	Median	Max	Obs.
Total						
Twitter at ICO end (>0)	137.85	182.56	1.00	65.00	1,265.00	3,625
Medium at ICO end (>0)	29.06	33.90	1.00	19.00	364.00	935
Reddit at ICO end (>0)	1,945.24	3,800.02	1.00	518.00	30,223.00	1,294
Bitcointalk at ICO end (>0)	560.53	970.13	1.00	195.00	11,957.00	2,608

Panel C: GitHub Commits

	Mean	Std. Dev.	Min	Median	Max	Obs.
Commits at ICO end (>0)	1,250.24	5,756.30	1.00	35.00	116,666.00	1,806

Panel D: Wallets

	Mean	Std. Dev.	Min	Median	Max	Obs.
Wallets number at first trading day (>0)	1,836.24	5,163.06	1.00	884.5	126,445	774
Wallets transfers at first trading day (>0)	7,588.07	16,868.91	5.00	3,651	331,210	774

Panel E: Returns

	Mean	Std. Dev.	Min	Median	Max	Obs.
ICO end-to-open returns						
ICO return, conditional on listing (%)	384.39	936.82	2.96	46.25	3,870.72	1,007
ICO return, unconditional (%)	99.75	646.98	-100.00	-100.00	3,870.72	2,442
ICO first-day returns						
First day return (%)	9.88	22.98	-19.73	1.69	76.15	1,153
Longer-term cumulative post-ICO returns						
30-day return (%)	-2.58	80.44	-78.60	-29.58	233.05	1,142
90-day return (%)	0.07	127.55	-94.33	-47.52	415.91	1,119
180-day return (%)	-6.87	150.95	-97.69	-70.36	494.56	1,068
365-day return (%)	-37.97	110.53	-99.28	-85.07	337.34	891

Table 6. ICO success. This table reports estimates of regressions of determinants of ICO success. We use four measures of ICO success: (1) dummy variable that takes the value of one if some funds were raised in an ICO; (2) logarithm of total amount raised plus one; (3) ratio of total amount raised and ICO hardcap; (4) dummy variable that takes the value of one if the token ends up being traded on at least one cryptographic exchange. See Table 1 for definitions of all independent variables. In cases in which the amount raised is missing, it is assumed to equal zero. We estimate OLS regressions in columns (2) and (3) and logit regressions in columns (1) and (4). In column (4) the sample includes only observations with positive amount raised. In columns (1) and (4), the reported coefficients are the marginal effects of each independent variable and the reported R-squared are pseudo R-squared. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

	(1) raised dummy	(2) log(amount raised+1)	(3) raised-to-hardcap	(4) listing dummy
% for sale	-0.075*	-1.500**	-0.129***	-0.196***
	(-1.891)	(-2.506)	(-4.527)	(-4.270)
presale	-0.018	-0.375	-0.038**	0.016
	(-0.861)	(-1.176)	(-2.523)	(0.616)
log (hardcap+1)	0.002	0.351***	-0.034***	0.011
	(0.230)	(3.046)	(-6.261)	(1.156)
whitelist	-0.010	0.138	0.052***	0.026
	(-0.499)	(0.441)	(3.471)	(0.988)
kyc	0.154***	2.503***	0.120***	0.207***
	(7.284)	(7.675)	(7.711)	(7.807)
white paper	0.048***	0.857***	0.056***	0.073***
	(2.669)	(3.105)	(4.278)	(3.339)
log (team members+1)	0.069***	1.138***	0.038***	0.064***
	(5.823)	(6.234)	(4.385)	(4.175)
log (commits at ICO start+1)	0.020***	0.305***	0.017***	0.023***
	(5.022)	(5.200)	(6.198)	(5.337)
log (Twitter at ICO start+1)	0.001	0.017	0.003	-0.010**
	(0.188)	(0.270)	(1.140)	(-1.999)
log (Reddit at ICO start+1)	-0.007**	-0.114**	-0.012***	-0.033***
	(-2.012)	(-2.069)	(-4.585)	(-7.269)
log (Bitcointalk at ICO start+1)	0.028***	0.435***	0.010***	0.016***
	(8.073)	(7.773)	(3.646)	(3.683)
log (Medium at ICO start+1)	0.009	0.207*	0.024***	0.022**
	(1.234)	(1.871)	(4.458)	(2.545)
bonus	0.098***	1.614***	0.033**	0.015
	(5.428)	(5.681)	(2.447)	(0.625)
bounty	-0.033	-0.670**	-0.082***	-0.126***
	(-1.608)	(-2.120)	(-5.427)	(-5.046)
Observations	2,331	2,349	2,349	1,328
R-squared	0.209	0.285	0.277	0.312
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Table 7. Alternative data sources. This table reports estimates of regressions in which the dependent variable is log (amount raised+1). In column (1), we report the baseline specification using the full data, see column (2) of Table 6. In Column (2), we only use data from CoinGecko. In column (3), we only use data from CryptoCompare. In column (4), we only use data from ICObench. In column (5) we only use data from ICOdrops. In column (6), we identify for each of the four variables–amount raised, hardcap, token supply, and tokens for sale–three sources with the highest estimated data quality for that variable, and use data only from top-three sources for that variable. In column (7), we only use ICOs in the top tercile of observation-level overall data quality. See Table 1 for the definitions of all independent variables. In cases in which the amount raised is missing, it is assumed to equal zero. The regressions are estimated by OLS. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	CoinGecko	CryptoCom.	ICObench	ICOdrops	top3 sources	top quality
% for sale	-1.500**	-0.471*	-0.551*	-0.184	-0.215	-1.658***	-0.568**
	(-2.506)	(-1.882)	(-1.757)	(-0.985)	(-1.127)	(-2.870)	(-2.558)
presale	-0.375	-0.059	-0.228	0.008	-0.055	0.092	-0.024
	(-1.176)	(-0.408)	(-1.165)	(0.076)	(-0.561)	(0.298)	(-0.198)
log (hardcap+1)	0.351***	0.706***	0.535***	0.655***	0.673***	0.249**	0.668***
	(3.046)	(12.683)	(6.793)	(16.916)	(12.000)	(2.240)	(13.767)
whitelist	0.138	0.395***	0.086	0.262**	0.139	0.424	0.403***
	(0.441)	(2.870)	(0.421)	(2.551)	(1.154)	(1.401)	(3.272)
kyc	2.503***	1.211***	0.675***	0.443***	0.095	3.303***	0.734***
	(7.675)	(7.053)	(3.465)	(4.013)	(0.654)	(10.499)	(5.360)
white paper	0.857***	0.181	0.353**	0.147*	0.132	1.119***	0.104
	(3.105)	(1.529)	(2.268)	(1.676)	(1.388)	(4.204)	(0.993)
log (team members+1)	1.138***	0.155*	0.287***	0.328***	-0.055	1.092***	0.216***
	(6.234)	(1.962)	(2.811)	(5.364)	(-0.972)	(6.201)	(3.163)
log (commits at ICO start+1)	0.305***	0.044*	0.042	0.028	0.026*	0.323***	0.053***
	(5.200)	(1.922)	(1.401)	(1.625)	(1.720)	(5.714)	(2.748)
log (Twitter at ICO start+1)	0.017	0.027	0.033	0.003	-0.009	0.022	0.010
	(0.270)	(1.011)	(0.957)	(0.155)	(-0.436)	(0.351)	(0.436)
log (Reddit at ICO start+1)	-0.114**	-0.030	-0.062*	-0.049***	-0.017	-0.141***	-0.056***
	(-2.069)	(-1.262)	(-1.864)	(-2.814)	(-0.869)	(-2.665)	(-2.668)
log (Bitcointalk at ICO start+1)	0.435***	-0.020	0.022	0.016	-0.013	0.401***	-0.006
	(7.773)	(-0.845)	(0.684)	(0.893)	(-0.746)	(7.436)	(-0.291)
log (Medium at ICO start+1)	0.207*	0.037	0.045	0.072**	0.041	0.223**	0.086**
	(1.871)	(0.867)	(0.720)	(2.166)	(1.398)	(2.088)	(2.266)
bonus	1.614***	0.089	0.008	-0.180*	-0.100	0.689**	0.041
	(5.681)	(0.696)	(0.040)	(-1.907)	(-0.971)	(2.516)	(0.369)
bounty	-0.670**	-0.186	-0.098	-0.152	-0.066	-0.745**	-0.125
	(-2.120)	(-1.409)	(-0.590)	(-1.557)	(-0.547)	(-2.443)	(-1.081)
Observations	2,349	803	352	1,009	345	2,349	898
R-squared	0.285	0.358	0.361	0.383	0.466	0.335	0.364
Time dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 8. ICO success, white paper available. This table reports estimates of regressions of determinants of ICO success conditional on having a white paper. We use four measures of ICO success: (1) dummy variable that takes the value of one if some funds were raised in an ICO; (2) logarithm of total amount raised plus one; (3) ratio of total amount raised and ICO hardcap; (4) dummy variable that takes the value of one if the token ends up being traded on at least one cryptographic exchange. See Table 1 for the definitions of all independent variables. In cases in which the amount raised is missing, it is assumed to equal zero. We estimate OLS regressions in columns (2) and (3) and logit regressions in columns (1) and (4). In column (4) the sample includes only observations with positive amount raised. In columns (1) and (4), the reported coefficients are the marginal effects of each independent variable and the reported R-squared are pseudo R-squared. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

	(1) raised dummy	(2) log (amount raised+1)	(3) raised-to-hardcap	(4) listing dummy
log (# nlp words)	0.031	0.778*	0.087***	0.062*
	(1.048)	(1.714)	(3.817)	(1.789)
tech ratio	0.649***	9.701***	0.745***	0.427
	(2.655)	(2.648)	(4.032)	(1.467)
% for sale	-0.078	-1.704**	-0.136***	-0.122**
	(-1.422)	(-2.012)	(-3.177)	(-2.004)
presale	-0.024	-0.498	-0.042*	-0.019
	(-0.826)	(-1.099)	(-1.856)	(-0.576)
log (hardcap+1)	0.003	0.369**	-0.049***	0.022*
	(0.227)	(2.149)	(-5.653)	(1.717)
whitelist	0.009	0.555	0.086***	0.058*
	(0.301)	(1.241)	(3.812)	(1.699)
kyc	0.139***	2.370***	0.141***	0.210***
	(4.709)	(5.030)	(5.929)	(6.271)
log (team members+1)	0.036**	0.673***	0.016	0.024
	(2.166)	(2.596)	(1.191)	(1.208)
log (commits at ICO start+1)	0.021***	0.300***	0.016***	0.026***
	(3.711)	(3.810)	(3.949)	(4.483)
log (Twitter at ICO start+1)	-0.000	0.003	0.005	0.004
	(-0.021)	(0.037)	(1.153)	(0.657)
log (Reddit at ICO start+1)	-0.006	-0.100	-0.013***	-0.026***
	(-1.275)	(-1.322)	(-3.315)	(-4.446)
log (Bitcointalk at ICO start+1)	0.027***	0.408***	0.010**	0.017***
	(5.528)	(5.188)	(2.424)	(2.935)
log (Medium at ICO start+1)	0.011	0.257*	0.027***	0.040***
	(1.094)	(1.744)	(3.611)	(3.473)
bonus	0.116***	1.898***	0.041**	0.006
	(4.681)	(4.695)	(2.024)	(0.208)
bounty	-0.019	-0.473	-0.113***	-0.088***
	(-0.621)	(-1.018)	(-4.838)	(-2.668)
Observations	1,103	1,114	1,114	684
R-squared	0.250	0.334	0.368	0.394
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Table 9. Commits growth. This table reports the relation between the growth in cumulative Github commits relative to cumulative commits at the ICO end date and the percentage of tokens for sale. The dependent variable is the log ratio of cumulative commits 30 days (in column 1), 90 days (in column 2), 180 days (in column 3), and 365 days (in column 4) after the ICO end date on one hand and cumulative GitHub source commits at ICO end date on the other hand. Lagged commits growth is the log ratio of cumulative commits at ICO end date and cumulative commits 90 days before the ICO end date. Lagged commits level is log cumulative commits 90 days before ICO end date. Raised dummy takes the value of one for ICOs that raised a positive amount. The regressions are estimated with OLS. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

	(1) 30-day	(2) 90-day	(3) 180-day	(4) 365-day
% for sale	-0.031**	-0.113***	-0.174***	-0.314***
	(-2.187)	(-3.236)	(-3.312)	(-4.036)
lagged commits growth	0.092***	0.197***	0.291***	0.376***
	(16.418)	(14.248)	(13.955)	(12.863)
lagged commits level	0.004***	0.006*	0.009*	0.019**
	(2.626)	(1.661)	(1.835)	(2.476)
raised dummy	0.002	0.051***	0.107***	0.158***
	(0.340)	(2.846)	(3.948)	(3.941)
Observations	1,241	1,241	1,241	1,096
R-squared	0.213	0.213	0.242	0.272
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Table 10. Social media growth. This table reports the relation between the growth in measures of cumulative social media activity relative to cumulative social media activity at ICO end date on one hand and the percentage of tokens for sale on the other hand. The dependent variable is the log ratio of a measure of cumulative social media activity 30 days (in column 1), 90 days (in column 2), 180 days (in column 3), and 365 days (in column 4) after the ICO end date and the measure of cumulative social media activity at ICO end date on the other hand. In Panel A, the measure of social media activity is the number of venture-initiated Twitter tweets; in Panel B, it is the number of Medium articles written by editors; in Panel C, it is the number of Reddit discussions; and in Panel D, it is the number of Bitcointalk posts. Lagged [Twitter, Medium, Reddit, Bitcointalk] growth is the log ratio of a measure of cumulative social media activity at ICO end date and that measure of social media activity 90 days before the ICO end date. Lagged [Twitter, Medium, Reddit, Bitcointalk] level is log measure of cumulative social media activity 90 days before the ICO end date. Raised dummy takes the value of one for ICOs that raised a positive amount. The regressions are estimated with OLS. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

Panel A: Twitter

	(1) 30-day	(2) 90-day	(3) 180-day	(4) 365-day
% for sale	-0.033**	-0.152***	-0.230***	-0.359***
	(-2.458)	(-4.479)	(-4.946)	(-6.029)
lagged Twitter growth	0.018***	-0.041***	-0.101***	-0.173***
	(4.346)	(-4.075)	(-7.323)	(-9.919)
lagged Twitter level	-0.016***	-0.083***	-0.137***	-0.193***
	(-6.964)	(-14.411)	(-17.482)	(-18.977)
raised dummy	0.031***	0.143***	0.239***	0.312***
	(4.609)	(8.371)	(10.225)	(10.446)
Observations	1,959	1,959	1,959	1,712
R-squared	0.132	0.181	0.240	0.312
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Panel B: Medium

	(1) 30-day	(2) 90-day	(3) 180-day	(4) 365-day
% for sale	-0.075***	-0.165***	-0.279***	-0.452***
	(-3.261)	(-3.629)	(-3.840)	(-4.079)
lagged Medium growth	0.014	0.013	-0.030	-0.115***
	(1.631)	(0.767)	(-1.126)	(-2.903)
lagged Medium level	-0.016***	-0.041***	-0.083***	-0.168***
	(-3.188)	(-4.223)	(-5.413)	(-6.957)
raised dummy	0.036***	0.101***	0.168***	0.336***
	(3.147)	(4.426)	(4.622)	(5.959)
Observations	621	621	621	522
R-squared	0.115	0.173	0.211	0.254
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Table 10. Social media growth. (Continued)

Panel C: Reddit

	(1) 30-day	(2) 90-day	(3) 180-day	(4) 365-day
% for sale	0.017	-0.075	-0.185*	-0.344**
	(0.575)	(-0.932)	(-1.876)	(-2.331)
lagged Reddit growth	0.038***	-0.012	-0.033	-0.100***
	(6.213)	(-0.707)	(-1.642)	(-3.428)
lagged Reddit level	0.002	-0.044***	-0.067***	-0.120***
	(0.633)	(-4.276)	(-5.294)	(-6.165)
raised dummy	0.042***	0.047	0.067	0.117*
	(3.081)	(1.285)	(1.489)	(1.786)
Observations	710	710	710	562
R-squared	0.168	0.181	0.217	0.265
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Panel D: Bitcointalk

	(1) 30-day	(2) 90-day	(3) 180-day	(4) 365-day
% for sale	-0.033**	-0.084***	-0.105***	-0.152***
	(-2.261)	(-3.019)	(-3.155)	(-3.863)
lagged Bitcointalk growth	0.014***	0.022***	0.025***	0.033***
	(4.016)	(3.233)	(3.025)	(3.451)
lagged Bitcointalk level	-0.009***	-0.023***	-0.028***	-0.028***
	(-3.937)	(-4.991)	(-5.173)	(-4.249)
raised dummy	0.033***	0.071***	0.089***	0.104***
	(4.545)	(5.002)	(5.220)	(5.099)
Observations	1,237	1,237	1,237	1,119
R-squared	0.171	0.163	0.188	0.218
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Table 11. Wallets growth. This table reports the relation between the growth in measures of product/platform adoption by users relative to product/platform adoption at ICO end date on one hand and the percentage of tokens for sale on the other hand. The dependent variable is the log ratio of a measure of product/platform adoption 90 days (in column 1), 180 days (in column 2), and 365 days (in column 3) after the ICO end date on one hand and the measure 30 days after ICO end date on the other hand. In Panel A, the measure of product/platform adoption is the number of wallets containing the token issued in the ICO; in Panel B, it is the cumulative number of transfers involving the token across wallets. Lagged # wallets growth is the log ratio of the number of wallets containing the token 30 days after ICO end date and the number of wallets at ICO end date. Lagged # wallets level is log number of wallets containing the token at ICO end date. Lagged # wallet transfers growth is the log ratio of the cumulative number of wallet transfers involving the token 30 days after ICO end date and cumulative number of transfers at ICO end date. Lagged # wallet transfers level is log cumulative number of wallet transfers at ICO end date. Raised dummy takes the value of one for ICOs that raised a positive amount. The regressions are estimated with OLS. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

Panel A: Wallets

	(1)	(2)	(3)
	90-day	180-day	365-day
% for sale	-0.231***	-0.335***	-0.377***
	(-3.340)	(-3.612)	(-3.131)
lagged # wallets growth	0.099***	0.128***	0.109***
	(4.853)	(4.666)	(3.114)
lagged # wallets level	-0.063***	-0.076***	-0.109***
	(-4.792)	(-4.314)	(-4.698)
raised dummy	0.008	-0.027	-0.176
	(0.120)	(-0.326)	(-1.525)
Observations	610	610	580
R-squared	0.218	0.234	0.225
Time dummy	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes

Panel B: Wallet transfers

	(1)	(2)	(3)
	90-day	180-day	365-day
% for sale	-0.164**	-0.296***	-0.333***
	(-2.529)	(-3.238)	(-2.768)
lagged # wallet transfers growth	0.179***	0.241***	0.249***
	(7.539)	(7.185)	(5.635)
lagged # wallet transfers level	-0.067***	-0.083***	-0.108***
	(-5.480)	(-4.796)	(-4.700)
raised dummy	0.004	-0.052	-0.207*
	(0.061)	(-0.633)	(-1.802)
Observations	610	610	580
R-squared	0.249	0.252	0.250
Time dummy	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes

Table 12. Post ICO returns. This table reports regressions of determinants of post-ICO cumulative returns over different horizons: 30 days, 90 days, 180 days, and 365 days. Δlog # wallets is the log ratio of the number of wallets containing the token 30 (90, 180, 365) days after the end of the first day of trading in the token on a crypto exchange and log # wallets is the log number of wallets at the end of the first trading day. Δlog # wallet transfers is the log ratio of the cumulative number of transfers involving the token across wallets 30 (90, 180, 365) days after the first trading day and log # wallet transfers is log cumulative number of wallet transfers involving the token at the end of first trading day. ICO end-to-open return is the log ratio of the token value at opening during the first trading day and the token value on the ICO end date. The latter quantity is calculated by dividing the amount raised by the circulating supply of tokens 7 days after the beginning of trading. ICO first-day return is the log ratio of the closing and opening prices of the first day of trading. log (amount raised+1) is the log amount raised at ICO plus one. Return btc is the contemporaneous return on Bitcoin. Return eth is the contemporaneous return on Ether. The regressions are estimated with OLS. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	30-day		90-day		180-day		365-day	
Δlog # wallets	0.083**		0.080*		0.149***		0.266***	
	(2.149)		(1.935)		(2.920)		(4.385)	
log # wallets	0.093***		0.123***		0.140***		0.188***	
	(3.561)		(3.597)		(3.103)		(3.238)	
Δlog # wallet transfers		0.178***		0.131***		0.204***		0.260***
		(3.739)		(2.764)		(3.801)		(4.169)
log # wallet transfers		0.079***		0.109***		0.121***		0.134**
		(3.094)		(3.209)		(2.764)		(2.397)
ICO end-to-open return	-0.143***	-0.141***	-0.203***	-0.205***	-0.237***	-0.236***	-0.296***	-0.296***
	(-6.465)	(-6.367)	(-7.212)	(-7.247)	(-6.641)	(-6.635)	(-6.484)	(-6.471)
ICO first-day ret.	-0.653***	-0.648***	-0.788***	-0.779***	-0.851***	-0.845***	-1.016***	-1.020***
	(-5.707)	(-5.666)	(-5.401)	(-5.331)	(-4.583)	(-4.559)	(-4.261)	(-4.266)
log (amount raised+1)	-0.201***	-0.205***	-0.237***	-0.243***	-0.294***	-0.295***	-0.328***	-0.312***
	(-6.760)	(-6.876)	(-6.260)	(-6.374)	(-6.113)	(-6.116)	(-5.427)	(-5.134)
return btc	0.110	0.091	0.083	0.083	0.490***	0.490***	1.217***	1.210***
	(0.942)	(0.787)	(1.171)	(1.176)	(7.005)	(7.086)	(10.611)	(10.529)
return eth	0.588***	0.575***	0.577***	0.565***	0.104***	0.101***	-0.218***	-0.218***
	(6.550)	(6.387)	(10.329)	(10.223)	(3.698)	(3.588)	(-7.203)	(-7.184)
Constant	2.129***	2.124***	2.159***	2.171***	2.152***	2.095***	1.623	1.645
	(4.287)	(4.290)	(3.416)	(3.429)	(2.671)	(2.606)	(1.601)	(1.615)
Observations	703	703	702	702	691	691	628	628
R-squared	0.241	0.245	0.391	0.390	0.401	0.404	0.376	0.374
Time dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo. region dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

A Additional results

Table A.1. Missing observations. This table reports estimates of regressions of determinants of ICO success under various assumptions about missing observations. In Panel A, the dependent variable is an indicator that takes the value of one if some funds were raised in an ICO. In Panel B, the dependent variable is the logarithm of total amount raised plus one. In Panel C, the dependent variable is the ratio of total amount raised and ICO hardcap. In Panel D, the dependent variable is an indicator that takes the value of one if the token ends up being traded on at least one cryptographic exchange. Column (2) of each panel presents results of the baseline specifications, in which missing values of amount raised are converted to zeroes (see the four columns of Table 6). Column (1) of each panel presents results of estimating regressions as in Table 6, while not replacing missing values of amount raised by zeroes. In column (3) of Panels B and C, in addition to assuming that amount raised is zero if missing, we assume that hardcap is zero if missing . In column (3) of Panels A and D and in column (4) of Panels B and C, we assume, in addition, that the percentage of tokens for sale and the number of team members are zero if missing. See Table 1 for the definitions of all independent variables. Missing % for sale, missing hardcap, and missing team members are indicators equalling one if % for sale, hardcap, and the number of team members, respectively, are missing. We estimate OLS regressions in Panels B and C and logit regressions Panels A and D. In Panel D, the sample includes only observations with positive amount raised. In Panels A and D, the reported coefficients are the marginal effects of each independent variable and the reported R-squared are the pseudo R-squared. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

Table A.1. Missing observations. (Continued)

Panel A: raised dummy

	(1)	(2)	(3)
% for sale	-0.075*	-0.071**	-0.088***
	(-1.891)	(-1.988)	(-3.066)
missing % for sale			0.007
			(0.311)
presale	-0.018	-0.014	-0.010
	(-0.861)	(-0.760)	(-0.795)
log (hardcap+1)	0.002	0.003	-0.001
	(0.230)	(0.384)	(-0.225)
missing hardcap		-0.137	-0.153*
		(-1.107)	(-1.777)
whitelist	-0.010	-0.021	-0.027*
	(-0.499)	(-1.101)	(-1.945)
kyc	0.154***	0.152***	0.131***
	(7.284)	(8.013)	(9.680)
white paper	0.048***	0.044***	0.053***
	(2.669)	(2.730)	(4.331)
log (team members+1)	0.069***	0.061***	0.078***
	(5.823)	(5.711)	(7.457)
missing team members			0.106***
			(4.033)
log (commits at ICO start+1)	0.020***	0.020***	0.016***
	(5.022)	(5.312)	(5.838)
log (Twitter at ICO start+1)	0.001	0.001	0.002
	(0.188)	(0.195)	(0.746)
log (Reddit at ICO start+1)	-0.007**	-0.005	-0.005**
	(-2.012)	(-1.380)	(-2.154)
log (Bitcointalk at ICO start+1)	0.028***	0.028***	0.023***
	(8.073)	(8.857)	(9.660)
log (Medium at ICO start+1)	0.009	0.011	0.011**
	(1.234)	(1.581)	(2.073)
bonus	0.098***	0.108***	0.109***
	(5.428)	(6.643)	(8.760)
bounty	-0.033	-0.030	-0.028*
	(-1.608)	(-1.597)	(-1.741)
Observations	2,331	2,775	5,086
Pseudo R-square	0.209	0.238	0.266
Time dummy	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes

Table A.1. Missing observations. (Continued)

Panel B: log (amount raised+1)

	(1)	(2)	(3)	(4)
% for sale	-0.563***	-1.500**	-1.469***	-1.614***
	(-2.963)	(-2.506)	(-2.748)	(-3.730)
missing % for sale				-0.234
				(-0.698)
presale	-0.157	-0.375	-0.315	-0.261
	(-1.479)	(-1.176)	(-1.109)	(-1.319)
log (hardcap+1)	0.678***	0.351***	0.382***	0.275***
	(17.210)	(3.046)	(3.381)	(3.482)
missing hardcap			3.772**	2.515*
			(1.973)	(1.904)
whitelist	0.260**	0.138	-0.044	-0.149
	(2.467)	(0.441)	(-0.155)	(-0.713)
kyc	0.673***	2.503***	2.494***	2.099***
	(6.059)	(7.675)	(8.514)	(10.257)
white paper	0.241***	0.857***	0.786***	0.915***
	(2.683)	(3.105)	(3.141)	(4.871)
log (team members+1)	0.232***	1.138***	1.014***	1.511***
	(3.886)	(6.234)	(6.257)	(9.405)
missing team members				2.124***
				(5.464)
log (commits at ICO start+1)	0.050***	0.305***	0.308***	0.254***
	(2.831)	(5.200)	(5.662)	(6.293)
log (Twitter at ICO start+1)	0.013	0.017	0.020	0.035
	(0.624)	(0.270)	(0.338)	(0.842)
log (Reddit at ICO start+1)	-0.040**	-0.114**	-0.075	-0.092**
	(-2.197)	(-2.069)	(-1.471)	(-2.415)
log (Bitcointalk at ICO start+1)	-0.015	0.435***	0.436***	0.376***
	(-0.840)	(7.773)	(8.509)	(9.809)
log (Medium at ICO start+1)	0.063*	0.207*	0.244**	0.249***
	(1.831)	(1.871)	(2.341)	(2.952)
bonus	0.090	1.614***	1.810***	1.798***
	(0.953)	(5.681)	(6.987)	(9.100)
bounty	-0.253**	-0.670**	-0.653**	-0.575**
	(-2.485)	(-2.120)	(-2.248)	(-2.350)
Observations	1,346	2,349	2,798	5,138
R-squared	0.335	0.285	0.310	0.334
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Table A.1. Missing observations. (Continued)

Panel C: raised-to-hardcap

	(1)	(2)	(3)	(4)
% for sale	-0.147***	-0.129***	-0.110***	-0.098***
	(-3.613)	(-4.527)	(-4.542)	(-5.271)
missing % for sale				-0.070***
				(-4.926)
presale	-0.040*	-0.038**	-0.036***	-0.041***
	(-1.745)	(-2.523)	(-2.790)	(-4.816)
log (hardcap+1)	-0.065***	-0.034***	-0.032***	-0.027***
	(-7.680)	(-6.261)	(-6.221)	(-8.114)
missing hardcap			-0.723***	-0.623***
			(-8.317)	(-11.028)
whitelist	0.083***	0.052***	0.044***	0.024***
	(3.663)	(3.471)	(3.363)	(2.630)
kyc	0.110***	0.120***	0.100***	0.080***
	(4.625)	(7.711)	(7.531)	(9.143)
white paper	0.062***	0.056***	0.046***	0.038***
	(3.232)	(4.278)	(3.999)	(4.737)
log (team members+1)	0.027**	0.038***	0.033***	0.054***
	(2.096)	(4.385)	(4.478)	(7.825)
missing team members				0.115***
				(6.927)
log (commits at ICO start+1)	0.011***	0.017***	0.017***	0.015***
	(2.856)	(6.198)	(6.671)	(8.461)
log (Twitter at ICO start+1)	0.006	0.003	0.004	0.000
	(1.366)	(1.140)	(1.466)	(0.072)
log (Reddit at ICO start+1)	-0.014***	-0.012***	-0.011***	-0.009***
	(-3.640)	(-4.585)	(-4.810)	(-5.487)
log (Bitcointalk at ICO start+1)	-0.002	0.010***	0.008***	0.008***
	(-0.460)	(3.646)	(3.383)	(4.744)
log (Medium at ICO start+1)	0.025***	0.024***	0.022***	0.022***
	(3.416)	(4.458)	(4.685)	(6.070)
bonus	-0.003	0.033**	0.032***	0.019**
	(-0.149)	(2.447)	(2.740)	(2.267)
bounty	-0.101***	-0.082***	-0.070***	-0.057***
	(-4.627)	(-5.427)	(-5.314)	(-5.472)
Observations	1,346	2,349	2,798	5,138
R-squared	0.245	0.277	0.289	0.280
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Table A.1. Missing observations. (Continued)

Panel D: listing dummy

	(1)	(2)	(3)
% for sale	-0.196***	-0.215***	-0.213***
	(-4.270)	(-4.843)	(-5.324)
missing % for sale			-0.144***
			(-4.518)
presale	0.016	0.027	-0.039**
	(0.616)	(1.051)	(-1.972)
hardcap	0.011	0.011	-0.001
	(1.156)	(1.137)	(-0.108)
missing hardcap		0.110	-0.096
		(0.645)	(-0.720)
whitelist	0.026	0.027	0.008
	(0.988)	(1.051)	(0.369)
kyc	0.207***	0.207***	0.217***
	(7.807)	(8.120)	(11.046)
white paper	0.073***	0.063***	0.086***
	(3.339)	(2.933)	(5.067)
log (team members+1)	0.064***	0.055***	0.061***
	(4.175)	(3.770)	(4.288)
missing team members			0.079**
			(2.022)
log (commits at ICO start+1)	0.023***	0.023***	0.020***
	(5.337)	(5.440)	(5.820)
log (Twitter at ICO start+1)	-0.010**	-0.012**	-0.020***
	(-1.999)	(-2.554)	(-4.987)
log (Reddit at ICO start+1)	-0.033***	-0.031***	-0.027***
	(-7.269)	(-7.054)	(-7.243)
log (Bitcointalk at ICO start+1)	0.016***	0.016***	0.019***
	(3.683)	(3.777)	(5.426)
log (Medium at ICO start+1)	0.022**	0.021**	0.025***
	(2.545)	(2.473)	(3.357)
bonus	0.015	0.018	0.001
	(0.625)	(0.807)	(0.069)
bounty	-0.126***	-0.121***	-0.093***
	(-5.046)	(-4.950)	(-4.282)
Observations	1,328	1,449	2,299
Pseudo R-square	0.312	0.292	0.277
Time dummy	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes

Table A.2. Raised net of presale. This table reports estimates of regressions in which the dependent variable is log (amount raised net of presale+1). In column (1), we report results of the baseline specification in which the dependent variable is log (amount raised+1), see column (1) of Table 6. In Column (2), we subtract from the total amount raised the amount raised in presale. In column (3), the dependent variable is log (amount raised+1), but (attempted) presale dummy is replaced by successful presale dummy, which equals one for ICOs that raised some funds during presale. In column (4), we subtract from the total amount raised the amount raised in presale, and use successful presale dummy instead of attempted presale dummy. See Table 1 for the definitions of all independent variables. In cases in which the amount raised is missing, it is assumed to equal zero. The regressions are estimated by OLS. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

	(1) log (amount raised+1)	(2) log (amount raised post-presale +1)	(3) log (amount raised+1)	(4) log (amount raised post-presale +1)
% for sale	-1.500**	-1.318**	-1.199**	-1.262**
	(-2.506)	(-2.181)	(-2.015)	(-2.083)
presale (attempted)	-0.375	-0.418		
	(-1.176)	(-1.298)		
presale (successful)			3.930***	1.045*
			(6.628)	(1.730)
log (hardcap+1)	0.351***	0.333***	0.333***	0.326***
	(3.046)	(2.865)	(2.921)	(2.803)
whitelist	0.138	-0.093	-0.148	-0.176
	(0.441)	(-0.294)	(-0.471)	(-0.552)
kyc	2.503***	2.364***	2.292***	2.308***
	(7.675)	(7.178)	(7.056)	(6.977)
white paper	0.857***	0.777***	0.734***	0.752***
	(3.105)	(2.790)	(2.677)	(2.693)
log (team members+1)	1.138***	1.172***	1.137***	1.160***
	(6.234)	(6.359)	(6.296)	(6.307)
log (commits at ICO start +1)	0.305***	0.271***	0.260***	0.261***
	(5.200)	(4.584)	(4.457)	(4.386)
log (Twitter at ICO start+1)	0.017	-0.001	-0.012	-0.013
	(0.270)	(-0.011)	(-0.192)	(-0.195)
log (Reddit at ICO start+1)	-0.114**	-0.127**	-0.102*	-0.125**
	(-2.069)	(-2.279)	(-1.874)	(-2.247)
log (Bitcointalk at ICO start+1))	0.435***	0.428***	0.403***	0.414***
	(7.773)	(7.576)	(7.279)	(7.349)
log (Medium at ICO start+1)	0.207*	0.125	0.157	0.114
	(1.871)	(1.120)	(1.424)	(1.019)
bonus	1.614***	1.506***	1.413***	1.421***
	(5.681)	(5.251)	(5.034)	(4.968)
bounty	-0.670**	-0.495	-0.519*	-0.454
	(-2.120)	(-1.551)	(-1.652)	(-1.419)
Observations	2,349	2,349	2,349	2,349
R-squared	0.285	0.264	0.298	0.264
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Geographical region dummy	Yes	Yes	Yes	Yes

Table A.3. Raised dummy: different thresholds. This table reports estimates of regressions in which the dependent variable is an indicator equalling one if the amount raised in the ICO exceeds a certain threshold. In column (1), we report the baseline specification, in which the threshold is zero, see column (1) of Table 6. In column (2), the threshold is $U.S. 10,000. In column (3), the threshold is $U.S. 100,000. In column (4), the threshold is $U.S. 1,000,000. See Table 1 for the definitions of all independent variables. In cases in which the amount raised is missing, it is assumed to equal zero. The regressions are estimated by logit. The reported coefficients are the marginal effects of each independent variable and the reported R-squared are the pseudo R-squared. We report t-statistics in parentheses. * Significant at 10 percent; ** Significant at 5 percent; *** Significant at 1 percent.

	(1)	(2)	(3)	(4)
	baseline	>10K	>100K	> 1M
% for sale	-0.075*	-0.066*	-0.087**	-0.133***
	(-1.891)	(-1.653)	(-2.208)	(-3.420)
presale	-0.018	-0.015	-0.015	-0.039*
	(-0.861)	(-0.709)	(-0.713)	(-1.890)
log (hardcap+1)	0.002	0.004	0.018**	0.057***
	(0.230)	(0.578)	(2.424)	(7.255)
whitelist	-0.010	-0.012	0.001	0.022
	(-0.499)	(-0.566)	(0.072)	(1.091)
kyc	0.154***	0.160***	0.176***	0.176***
	(7.284)	(7.576)	(8.388)	(8.333)
white paper	0.048***	0.049***	0.054***	0.054***
	(2.669)	(2.743)	(3.011)	(3.045)
log (team members+1)	0.069***	0.071***	0.075***	0.066***
	(5.823)	(5.979)	(6.302)	(5.549)
log (commits at ICO start+1)	0.020***	0.021***	0.022***	0.021***
	(5.022)	(5.181)	(5.558)	(5.474)
log (Twitter at ICO start+1)	0.001	0.001	0.001	0.001
	(0.188)	(0.329)	(0.277)	(0.173)
log (Reddit at ICO start+1)	-0.007**	-0.006*	-0.006*	-0.007*
	(-2.012)	(-1.742)	(-1.733)	(-1.932)
log (Bitcointalk at ICO start+1)	0.028***	0.027***	0.024***	0.023***
	(8.073)	(7.625)	(7.006)	(6.489)
log (Medium at ICO start+1)	0.009	0.009	0.012	0.014**
	(1.234)	(1.264)	(1.569)	(2.008)
bonus	0.098***	0.106***	0.112***	0.100***
	(5.428)	(5.884)	(6.289)	(5.582)
bounty	-0.033	-0.032	-0.037*	-0.056***
	(-1.608)	(-1.534)	(-1.783)	(-2.746)
Observations	2,331	2,331	2,336	2,343
Pseudo R-square	0.209	0.211	0.224	0.227
Time dummy	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes
Region dummy	Yes	Yes	Yes	Yes