De-crypto-ing Signals in Initial Coin Offerings: Evidence of Rational Token Retention

Tetiana Davydiuk, Deeksha Gupta, and Samuel Rosen*†

April 2021

Abstract

Using the market for initial coin offerings (ICOs) as a laboratory, we provide evidence that entrepreneurs use retention to alleviate information asymmetry. The underlying technology and the absence of regulation make the ICO market well suited to study this question empirically. Using a hand-collected dataset, we show that ICO issuers that retain a larger fraction of their tokens are more successful in their funding efforts and are more likely to develop a working product. Moreover, we find that retention is a stronger signal when markets are crowded and investors do not have as much time to conduct due diligence.

Keywords: asymmetric information, signaling, entrepreneurial financing, ICOs

JEL Classifications: D82, G32, L26

*Tetiana Davydiuk (tetianad@andrew.cmu.edu) and Deeksha Gupta (deekshagupta@cmu.edu) are at Carnegie Mellon University. Samuel Rosen (sam.rosen23@gmail.com) is at the Fox School of Business at Temple University.

†We would like to thank Gurdip Bakshi, Xiaohui Gao Bakshi, Hugo Benedetti, Jonathan Berk, Lin William Cong, Dean Corbae, Marvin Goodfriend, Todd Gormley, Hanna Halaburda, Zhiguo He, Burton Hollifield, Sabrina Howell, Kose John, Stephen Karolyi, Theresa Kuchler, Uday Rajan, Bryan Routledge, Fahad Saleh, Harald Uhlig, Ariel Zetlin-Jones; attendees at the 2019 Toronto Fintech Conference, the 2019 NBER meeting on Blockchain, Distributed Ledgers, and Financial Contracting, ESSFM Asset Pricing 2019 (Gerzensee), AFA 2020, and seminar participants at Temple University, the Office of Comptroller of the Currency, Carnegie Mellon University, the Securities and Exchange Commission, Boston University, and UCLA for their helpful comments and discussions.

1 Introduction

Asymmetric information can make it more difficult for early-stage ventures to obtain financing. The theoretical literature has long proposed that entrepreneurs can overcome such informational frictions by retaining a stake in their enterprise. Costly retention can allow entrepreneurs to signal their type (Leland and Pyle, 1977) and can generate incentives for them to exert effort, increasing the expected value of their venture (Holmstrom and Tirole, 1997).[1] Studying this question empirically, however, poses three key challenges. First, the degree of information asymmetry in the chosen setting must be severe enough to induce entrepreneurs to signal their quality through retention. This challenge is germane to any market with regulators, legal protections, and other institutional features that can reduce the problem of asymmetric information. Second, there is limited data available on early-stage companies such as those seeking venture capital funds. Moreover, the data that is available suffers from severe selection bias as researchers typically do not observe companies that were not successful in obtaining venture financing. Third, a researcher must be able to control for all relevant information that could jointly affect investors' perception of quality and retention. In this paper, we address these empirical challenges by studying the role of retention in a novel setting – the market for initial coin offerings (ICOs).

In simple terms, an ICO is the sale of digital assets, "tokens," by an entrepreneur or company to fund the creation of an online platform or ecosystem that uses blockchain technology.[2] The token being sold can be used on the platform that the issuer is creating. The value of the token is not derived from profit-sharing, as most tokens do not have any cash-flow rights. Rather, once the platform is completed, the token can be exchanged for the product or service that is being sold on the platform.[3] Investors can often sell their tokens in liquid secondary markets (crypto asset exchanges). ICO issuers raised $33.4 billion in aggregate through 2018, but activity has since declined significantly. While our empirical findings are based on the historical period in which

[1]In this paper, when we use the phrase "signaling through retention," we are not simply referring to information asymmetries due to adverse selection. This can additionally refer to retention as a signal that the entrepreneur will commit to exerting effort to overcome moral hazard problems.

[2]We refer to the digital assets sold in an ICO as "tokens" and reserve the word "coins" only to refer to cryptocurrencies. We realize this may seem confusing, given that the word "coin" is in the term ICO, but it is a growing convention to refer to these digital assets in this way. In fact, some parties in the crypto industry are pushing to replace the term ICO with either "token sale" or "token generation event" (Token Alliance, 2018).

[3]There are two broad types of tokens depending on whether the exchange rate between the token and the future good or service is pre-determined (product tokens) or is endogenously determined based on supply and demand (platform tokens). More details on token classifications can be found in Cong and Xiao (2021).

this market was quite active, the insights they deliver about how retention can help overcome information asymmetries faced by entrepreneurs do not require ICOs to be an important source of financing moving forward.

The ICO market addresses the first key empirical challenge of studying retention by entrepreneurs – a sufficiently high degree of information asymmetry – as it is effectively unregulated and inherently opaque. The lack of regulation is in part because ICOs take place globally over the internet, which makes legal enforcement difficult. The market is opaque because there are neither intermediaries nor face-to-face interactions between ICO investors and entrepreneurs. This combination of features is unique relative to other forms of financing (e.g., initial public offerings or venture capital). As a result, ICO investors face a uniquely high degree of information asymmetry with prospective issuers.

We address the second empirical challenge by building a comprehensive database of 5,644 ICOs which includes both ICOs that were successful in obtaining early-stage financing and those that failed to do so. We collect data on numerous ex-ante characteristics and ex-post outcomes of interest through a combination of web scraping, PDF scraping, collection by hand, and manual review. We focus on "utility" tokens whose value is derived from the ability to exchange the token for an economic good or service. Our primary sources are websites that provide information about ICOs to potential investors, but we also rely on documents provided by ICO issuers and manual validation. We additionally hand-collect data from the Apple Store and company websites to determine the performance of entrepreneurs following the ICO.

The technology underlying ICOs allow us to effectively control for information that jointly affects investors' perception of quality and retention – the third key empirical challenge. ICOs are a unique way of financing because their terms are enforced through software ("smart contract") instead of the legal system and because of the public nature of the distributed ledger. These features of ICOs create a clear distinction between promises that are enforceable and those that are not. Essentially, a promise is credible if it can be written into a smart contract which creates explicit commitment (e.g., the total number of tokens to be created) and/or monitored on the blockchain (e.g., transactions into and out of ICO teams' crypto wallets) which allows for implicit commitment.[4] However, any

[4]Token retention, for example, can be committed to even if it is not explicitly written into the smart contract due to the public nature of the blockchain. If the team tries to sell tokens it was supposed to retain, the movement out

piece of information that cannot be translated into computer code or verified on the blockchain (e.g., résumés of entrepreneurs) is generally not credible from an investor's point of view because ICO issuers face no ex-post repercussions for misrepresenting such information. The opacity and lack of regulation do not imply that all, or even most, available soft information about ICOs is misrepresented — legitimate ICOs with a competent entrepreneurial team and a promising idea will indeed convey this information to the public. However, the lack of legal repercussions creates an incentive for low-quality ICOs to mimic such behavior at relatively little, if any, cost, thereby limiting how much rational investors can condition their investment decisions on such information. Indeed, numerous examples exist of ICOs misrepresenting information about their entrepreneurial team and their project without facing any consequences.[5] The technology underlying ICOs and the lack of legal accountability therefore give us a limited set of observable promises that an entrepreneur can commit to, greatly reducing the confounding information that we need to account for.

We formalize the intuition for our empirical analysis and results using a stylized model of the ICO market in which heterogenous entrepreneurs with private information about their type can conduct an ICO to monetize their venture. The asymmetric information problems in the model are similar to standard models of retention, such as that of Leland and Pyle (1977). While its mechanisms are not novel, the model helps us establish theoretical bases for our empirical results. The applicability of the seminal theories on retention and asymmetric information to the ICO market is not clear a priori. Unlike equity value, which is based on profit-sharing, the value of the token derives from the expected value of the goods or services that it can be traded for if the platform is completed successfully. The model delivers two main empirical predictions. First, there is a positive relationship between the fraction of tokens retained by entrepreneurs and the quality of the ICO. Second, this relationship becomes stronger when the quality of public information about entrepreneurs decreases.[6]

of the team's crypto wallets is publicly observable and the price will fall in reaction to this. The threat of a price decrease creates implicit commitment. In support of this argument, we find that ex-post retention measures within six months after an ICO generally remain at or above the ex-ante commitment to retention.

[5]For example, the CEO of ICO Confido lied that he had worked at PepsiCo and Zalando (https://www.cnbc.com/2017/11/23/confido-ceo-who-allegedly-pulled-375k-scam-lied-about-employment.html). As a particularly egregious example, the Miroskii ICO's advertised team was composed of stock images including one of the famous actor Ryan Gosling (https://yourstory.com/2018/09/fake-ico-verification).

[6]For generality, in Appendix A, we extend our benchmark model to include a moral hazard friction so that entrepreneurs have private information about their type and have to exert unobservable effort following the ICO to successfully create a platform. The empirical predictions of the model are unchanged. While moral hazard alone can deliver the first empirical prediction of a positive relationship between the fraction of tokens retained by entrepreneurs and the quality of the ICO, adverse selection is required for the second empirical prediction that this relationship

We perform numerous tests to assess our model's predictions. First, we examine whether ICOs that retain a greater fraction of tokens are more likely to fundraise successfully. The fraction of tokens retained is public knowledge, is announced before the ICO takes place, and can be coded into a smart contract. Therefore it is a credible ex-ante commitment to retention, which is key to our empirical analysis. To provide further support for the credibility of the commitment to retention, we observe that ex-post measures of retention generally remain at or above the ex-ante commitment to retention for at least six months following the ICO. To get a continuous measure of fundraising success for each ICO, we divide the amount of funds raised by its "hard cap." The hard cap is the maximum amount of funds an ICO wants to raise and is willing to accept. In line with retention as a means of overcoming financing frictions, we show that a 1 percentage point increase in the fraction of tokens retained by entrepreneurs is associated with an increase in fundraising success by 0.2–0.4 percentage points. Our results are robust to a number of alternative measures of ICO fundraising success. We additionally control for a number of confounding factors that are public promises made by the ICO team that can be either coded into a smart contract or verified on the blockchain. We further control for variables related to reputation — whether the ICO has obtained venture funding in the past, if it has reputable experts as advisers, and if the ICO team members have previous experience in ICO fundraising — as this may arguably be an alternative way to credibly signal quality.

Second, we test whether the relationship between token retention and fundraising success is stronger when the quality of public information decreases. In a market with more ICOs or more content available per ICO, investors cannot undertake as much due diligence per ICO – they have less time to study the various ICOs, examine their prototypes, read analyst reports, etc. As a result, investors will rely more on signals of quality such as retention. To test this model prediction, we develop a measure of market crowdedness for each ICO using the number of concurrent ICOs and the amount of information provided by token issuers in ICO prospectuses (commonly referred to as "whitepapers"). In accordance with our hypothesis, we find that for each percentage point increase in our measure of market crowdedness, a 1 percentage point increase in token retention leads to an additional 0.2–0.4 percentage point increase in fundraising success.

Third, we also show that higher retention is related to better ex-post performance of ICOs

becomes stronger when the quality of public information about entrepreneurs decreases.

beyond the fundraising stage. We measure ex-post performance in a variety of ways. In line with the literature, we first consider market-based measures — whether the token begins to trade on a cryptocurrency exchange following the ICO and conditional on exchange listing, the average trading volume of tokens and the average market capitalization of the ICO. In accordance with our hypothesis, we find that ICOs that retain more tokens are more likely to be listed on exchanges. We additionally find that, conditionally on trading on an exchange, higher token retention predicts higher average trading volume and market capitalization. Market-based measures, however, may not be good indicators of long-term performance because most tokens begin trading soon after the ICO end date, while the product is still in the early stages of development. Therefore we additionally measure ICO success by whether the company continues to have a working (i.e., active) website after the fundraising campaign, has developed a live platform, and has an application available to download on the Apple Store. Howell, Niessner, and Yermack (2019) also consider a working website as an ex-post measure of interest. To the best of our knowledge, ours is the first study to evaluate ICOs based on performance metrics relating to actual product development. In line with previous results, we document that ICOs that retain more tokens are more likely to achieve these positive outcomes.

Finally, in order to validate the economic mechanism underlying our results, we explore the costs that entrepreneurs face when retaining tokens. When entrepreneurs are risk-averse, retention is costly because they are forced to hold a large portion of their wealth in a single risky asset. This is particularly costly if such risk cannot be diversified away through financial markets. Consequently, entrepreneurs with limited risk-sharing opportunities will sell a larger share of their project to mitigate their risk exposure. We capture variation in risk-sharing opportunities across ICO teams with the Financial Development (FD) Index of the country in which they are located. In line with our theoretical argument, we find that ICO teams located in countries with a higher level of financial development tend to retain more tokens. In our regressions, we control for the GDP per capita of each country. Therefore, this result does not appear to be driven by differences in the wealth across countries.

Related Literature. Our paper contributes to the empirical literature on the use of retention as a way to overcome information asymmetry problems. We are the first to test this mechanism in the ICO market, which allows us to study the role of retention for early-stage financing. Previous literature has mostly focused on markets for relatively late-stage funding such as the market for initial public offerings (IPOs) or for more developed products such as commercial real estate. The empirical findings in these studies are mixed, which may partially be attributed to the fact that information asymmetry problems are less severe in these markets due to legal protections and greater availability of credible information.[7] Data limitations and sample selection biases pose challenges to study signaling through retention in venture capital markets — the most common setting used to examine early-stage financing. Moreover, venture capital markets are commonly thought to have a double moral hazard issue (Repullo and Suarez, 2004; Schmidt, 2003; Casamatta, 2003; Dessí, 2005; Inderst and Müller, 2004; Kaplan and Strömberg, 2001, 2004; Hellmann and Puri, 2002) which requires creating incentives for venture capitalists to exert monitoring efforts and consequently changes the nature of the basic skin-in-the-game problem of entrepreneurs. By focusing on the ICO market, we can study retention in a setting with severe asymmetric information, driven by lack of regulation and limited verifiable information, which is quite similar to settings modeled in the theoretical literature. We not only find positive evidence for signaling through retention, but we also show how the strength of this signal depends on the market environment.

Our paper also contributes to the growing empirical literature that analyzes the ICO market. Most of these papers investigate the drivers of ICO success (Adhami, Giudici, and Martinazzi, 2018; Amsden and Schweizer, 2018; Boreiko and Sahdev, 2018; Bourveau, De George, Ellahie, and Macciocchi, 2018; Deng, Lee, and Zhong, 2018; Fisch, 2019; Florysiak and Schandlbauer, 2019; Howell, Niessner, and Yermack, 2019; Roosenboom, van der Kolk, and de Jong, 2020; Lee, Li, and Shin, 2019; Lyandres, Palazzo, and Rabetti, 2020) while others study post-ICO returns in the spirit of the IPO literature (Benedetti and Kostovetsky, 2021; Momtaz, 2020, 2021; Lyandres et al., 2020).[8] Similar to our study, the papers that study ICO success use the achievement of the stated

[7] Purnanandam and Begley (2016) find that higher levels of equity tranches in private-label residential mortgage-backed security deals are associated with lower delinquency rates and higher prices. Garmaise and Moskowitz (2003) find no evidence that informed sellers of commercial real estate signal their information through retention. Downes and Heinkel (1982) and Clarkson, Dontoh, Richardson, and Sefcik (1991) find positive support for the share of equity retained by insiders in an IPO as a signal, while Krinsky and Rotenberg (1989), Ritter (1984), and Schultz and Zaman (2001) do not. In the IPO setting, retention is defined as a fraction of ownership, which is similar to our measure of token retention in ICOs.

[8] Benedetti and Kostovetsky (2021) observe an average first-day return of 179%. They find that cryptocurrency prices and social media activity are significant determinants of ICO underpricing. While they believe their results could indicate an irrational bubble, they conclude that the observed ICO underpricing is also consistent with high compensation for risk. Momtaz (2020, 2021) explores both short-term and medium-term returns after an ICO. He argues that his results are consistent with a "market liquidity hypothesis" which proposes that ICO issuers want to underprice to generate ex-post market liquidity for the token. Lyandres et al. (2020) conclude that post-ICO

funding goal, amount of funds raised, or listing on an exchange as measures of success. In our study, we consider novel ex-post measures of performance, such as whether an ICO issuer has a product in the Apple Store. Moreover, we uniquely investigate how important the market environment is for the relationship between retention and success.

Broadly speaking, the other empirical ICO papers rely on the same data sources as we do. However, they focus on a different goal: uncovering the set of key factors associated with ICO success. Accordingly, they tend to consider dozens of explanatory variables. Similar to us, many of these studies consider token retention as an explanatory variable in their multivariate regressions. Early studies of the ICO market (e.g., Roosenboom et al., 2020; Fisch, 2019) do not find that retention is significant, which we believe is due to their smaller samples covering a shorter time period. For example, Roosenboom et al. (2020) look at 630 ICOs and their sample ends in December 2017. Our findings are in line with Lyandres et al. (2020) who have a comparable-sized sample to ours. In contrast to other papers, the goal at the heart of this paper is to empirically test the longstanding theoretical concept that retention helps entrepreneurs overcome information asymmetries. As such, we delve deeper into the drivers of retention in line with theoretical predictions. We show that retention is a stronger signal when markets are more crowded. Additionally, we explore the costs that entrepreneurs face when retaining tokens in order to validate the economic mechanisms in our model.

Our theoretical analysis complements a growing literature that analyzes the properties of ICOs. In general, these papers highlight channels through which the novel aspects of an ICO can improve or worsen outcomes relative to traditional forms of financing. Catalini and Gans (2018) highlight that the ICO mechanism can elicit the value that consumers place on the underlying project. Both Li and Mann (2018) and Sockin and Xiong (2018) find that an ICO can solve the coordination failure problem in creating a new platform. In a dynamic setting, Cong, Li, and Wang (2021) highlight endogenous user adoption in ICOs. They point out the nonlinear relationships between equilibrium values for token price, platform productivity, and network size. Sanchez (2017) proposes an optimal ICO mechanism to improve upon the current set of models. Bakos and Halaburda (2019) study how token tradability and broader crowdsourcing of due diligence affect the decision to use an ICO. Goldstein, Gupta, and Sverchkov (2019) develop a model in which they show that utility tokens

empirical patterns resemble those observed in the IPO market.

can allow a monopolist to commit to long-term competitive pricing. As part of our analysis, we develop a simple model to formally evaluate the intuition behind our key empirical findings. We regard our contribution as primarily empirical, as our theoretical model closely follows the analysis already done by seminal studies in the literature. Through our model we are able to formalize the intuition for our empirical findings. For a deeper theoretical examination of token retention in ICO markets, we recommend the models of Chod and Lyandres (2021), Garratt and van Oordt (2019), and Malinova and Park (2018), who analyze the decision of an entrepreneur on whether to use other forms of financing or to do an ICO.

The rest of this paper is arranged as follows. Section 2 gives an overview of the ICO market and provides key institutional details. Section 3 describes the model and its empirical predictions. Section 4 describes our data and provides summary statistics about the ICO market. Section 5 describes our empirical findings, and section 6 concludes.

2 Background on Initial Coin Offerings

In this section, we provide key institutional details and a brief overview of the regulatory treatment of ICOs. We start by describing the basics of blockchain which is the key technology underlying ICOs. A blockchain is a digital, decentralized, distributed ledger for a specified digital asset. Specifically, it is a file that contains the sequential list of transactions for that digital asset. Blockchain technology refers to the combination of this file with a network of computers (nodes) and software that dictates how the blockchain is monitored and updated. Using cryptography and an incentive scheme to reward participation on the network, this technology enables the direct and secure transfer of the underlying digital asset without a central authority to verify the transfer of ownership (Narayanan, Bonneau, Felten, Miller, and Goldfeder, 2016).

An ICO involves the sale of digital assets, or “tokens,” on the blockchain. The tokens we study, “utility” tokens, have a functional purpose in the product or platform that the issuer is creating. This utility aspect is an important characteristic in the current regulatory debate (see the discussion below). To provide a concrete example, let us consider the STORJ token issued by a company named Storj (https://storj.io/). The underlying product is a blockchain-based decentralized cloud storage solution. One can either rent space on the devices of others using the STORJ token or lease out one’s own space in exchange for STORJ tokens. The token is therefore an integral part of the blockchain-based platform. This example also highlights the early-stage nature of platforms that choose to do an ICO. The Storj project issued STORJ tokens in May 2017. As of December 2019, the Storj network was in a beta version.

The development of the Ethereum blockchain was a key milestone in the ability of entrepreneurs to conduct an ICO. The Ethereum blockchain, which is denominated in the cryptocurrency Ether, was itself funded through a crowdfunding effort in 2014. Ethereum was the "world's first programmable blockchain" that allowed users to design smart contracts. As a result, users could easily create proprietary tokens that existed on a platform on top of the Ethereum blockchain. Before this advancement, an ICO issuer would have to build and maintain their own blockchain.

The stylized process for an ICO is as follows:

1. **Announcement:** The team announces project details and the planned sale period on cryptocurrency forums/websites.

2. **Payment:** In the ICO period, prospective buyers submit orders by sending cryptocurrency payments to the crowdsale smart contract.

3. **Distribution:** At the sale's conclusion, the crowdsale smart contract automatically creates and sends the tokens to the digital wallets of buyers conditional upon a successful sale as defined within the contract itself.

Despite recent regulatory changes for digital assets in the United States, ICO markets for utility tokens remain largely unregulated. Since 2014, cryptocurrencies have been formally designated as commodities and, accordingly, fall under the regulatory jurisdiction of the Commodity Futures Trading Commission (CFTC).[9] The key issue for ICOs is whether the token being sold is a "security." If a token passes the Howey Test, it is regulated under existing securities laws with oversight from the Securities and Exchange Commission. In response, security token offerings (STOs) began to gain popularity. Tokens sold in an STO resemble traditional financial securities except that their ownership is enforced through blockchain technology. In contrast, utility tokens sold in an ICO do not offer profit-sharing or voting rights, and as such are not regulated as securities.

[9] Testimony of CFTC Chairman Timothy Massad before the U.S. Senate Committee on Agriculture, Nutrition and Forestry (Dec. 10, 2014), http://www.cftc.gov/PressRoom/SpeechesTestimony/opamassad-6.

ICO market participants have responded to the events of the past few years with their own forms of self-regulation. ICOs commonly consult with advisers and law firms to avoid potential legal issues from regulators. Also, many ICOs voluntarily comply with Know Your Customer (KYC) and Anti–Money Laundering (AML) regulations by gathering personal information from prospective buyers. In 2019, many entrepreneurs selling utility tokens opted to have an initial exchange offering (IEO) instead of conducting their own fundraising campaign via an ICO. In IEOs, entrepreneurs delegate token sales to a crypto exchange that performs due diligence before listing the project's tokens and implements identity verification procedures. These types of positive changes aim to reduce uncertainty about future regulatory actions.

3 Theory and Testable Implications

In this section, we develop a simple model to illustrate the relationship between retention and ICO quality. In the model, entrepreneurs are subject to an adverse selection problem. The asymmetric information problem we model is similar to that of standard models (e.g. Leland and Pyle, 1977), with the key difference being that, unlike equity, tokens do not entitle the holders to profits from the platform in the form of dividends. Instead tokens can either be exchanged for the services being sold on the platform or sold to someone on an exchange who wishes to purchase a service on the platform.

Since tokens are different from equity shares, it can be hard to understand what determines token value and cost of retention in ICO markets as compared to seminal models of informational frictions. We therefore explain these features of the model based on an example of an ICO used in an industry report (Deloitte, 2017) — a hypothetical platform called SpeedX.[10] Suppose that an entrepreneur is developing a platform for ride sharing called SpeedX with an associated token SPX. Customers can use SPX tokens to buy rides. Drivers can sell the SPX tokens they receive as a form of payment to new riders in exchange for other cryptocurrencies. The price of SPX tokens will therefore be determined by the future relative supply and demand for rides. Once the platform is developed, it can run without the entrepreneur and, as such, he will get no revenues from the platform.

To build the platform and monetize his idea, the entrepreneur conducts an ICO in which he

[10] This is an example of a platform token. Similar arguments would apply to a product token.

creates 50 SPX tokens and sells a fraction of them. Tokens may be purchased during the ICO either by potential future riders (who will later exchange these tokens for rides) or by investors who expect to profit by selling tokens to future SpeedX customers. Suppose that each token will be worth \$10 if the platform is developed successfully. For example, if the equilibrium number of rides is 100 and the equilibrium price per ride is \$5 then each ride would be worth 0.5 tokens. A token could buy you two rides and therefore would be worth \$10. Further assume that there is a 50% probability that the entrepreneur will successfully develop the SpeedX platform following the ICO. If investors are deep pocketed, risk-neutral and there is no time-discounting, they will be willing to pay \$5 per token during the ICO. Even if they do not expect to take rides themselves, investors can monetize each token by selling it to future riders for \$10 per token if the platform is successfully completed. In the model, we capture this token value with a payoff R received in the case of successful platform development.

When the entrepreneur is risk averse, there is a cost of retaining tokens. If the entrepreneur sells a token during the ICO, he will receive \$5 with certainty. If he retains this token, he will receive \$10 dollars with probability 50% (if the platform development succeeds) and \$0 otherwise (if the platform development fails). Jensen's inequality implies that the entrepreneur would prefer to sell the token at the time of the ICO. Risk-aversion therefore makes retention costly for the entrepreneur. In the model, we capture this cost with a function $C(\cdot)$, which is increasing in the fraction of tokens retained. Note that what matters is the fraction and not the number of tokens retained. For example, if the entrepreneur created 100 SPX tokens, then each token would be worth \$2.50 at the ICO.

Through our model, we formalize the key theoretical relationships that we later test empirically. In the model equilibrium, high-quality entrepreneurs retain a higher fraction of tokens than low-quality entrepreneurs. As the market becomes more crowded, the quality of ex-ante information available to investors deteriorates. In such an environment, a larger fraction of high-type entrepreneurs, facing uncertainty from investors about their type, need to retain a high fraction of tokens to signal their quality, causing the relationship between retention and fundraising to become stronger. One concern in ICO markets is that the lack of legal recourse makes the market vulnerable to frauds and scams which may affect our theoretical predictions. Importantly, we show that our model results survive even in the presence of "scammers," as long as there is a limited measure of them.

While its mechanisms are not novel, the model is helpful in establishing a theoretical framework for our empirical results. Importantly, the theoretical propositions we establish for the ICO market and subsequently test are not a priori clear from the seminal theories of asymmetric information. In particular, we demonstrate how the relationship between token retention and quality changes as the market environment changes. Additionally, we use the model to show that our propositions continue to hold even in the presence of scammers which is a concern in the ICO market. For generality and completeness, we extend the model to include both moral hazard and adverse selection (Appendix A). Each friction can individually generate the prediction that in equilibrium there will be a correlation between ICO quality and retention. However, adverse selection is necessary for this correlation to increase during periods of high market noise.

3.1 The Model

The model has three dates, $t = 0, 1, 2$ and two types of agents - penniless, entrepreneurs and deep-pocketed, risk-neutral investors. Entrepreneurs can undertake a project at $t = 1$ to build a platform using blockchain technology. There are two types of entrepreneurs - a proportion q are high-type entrepreneurs (H) and a proportion $1 - q$ are low-type entrepreneurs (L). A project undertaken by a high-type entrepreneur succeeds with probability π_H. A low-type entrepreneur succeeds with probability $\pi_L < \pi_H$. If the project is successful it delivers a payoff of R. Otherwise, the project returns 0.

At $t = 1$, the entrepreneur can undertake an ICO to raise money for investment. Entrepreneurs can monetize their projects by either selling tokens today or holding on to the tokens and getting the payoff R at $t = 2$ in case of success. Entrepreneurs choose the fraction of tokens they wish to retain, s, and sell the remaining share of tokens to investors. Both entrepreneurs and investors can invest in an outside option that gives them a gross risk-free return of R_f. For simplicity, we normalize R_f to 1.

Entrepreneurs maximize the following utility function:

$$u(W_T, s) = E[W_T] - \gamma C(s),$$

where W_T is their period-2 wealth, $\gamma > 0$ can be interpreted as a coefficient of risk aversion and $C(s)$ is their cost of holding tokens. The cost function C has the following standard properties: $C^{'}(s) > 0$, $C^{''}(s) > 0$, $C(0) = 0$. We think of the holding costs $C(\cdot)$ as a reduced form way of capturing the cost to entrepreneurs of holding a large portion of their wealth in a single, risky asset and not being able to diversify (see e.g. Leland and Pyle (1977)). This cost implies that entrepreneurs are effectively risk-averse. Alternatively, this cost can be interpreted as creating a liquidity constraint on the entrepreneur since the entrepreneur has to keep tokens rather than monetize them.

At $t = 0$, before entrepreneurs issue tokens, there is a public signal which perfectly reveals the entrepreneur's type with probability α. With probability $1 - \alpha$ the signal is uninformative. The informativeness of the signal is the key parameter we vary when we think about markets being noisier. A decrease in α can be interpreted as a decrease in the information quality in the market. For example, when a small number of ICOs are taking place, investors have more time to read up on an ICO; study its whitepaper, business plan and/or prototype; and talk to other investors. In the model, such a time is proxied for by a high α. When many ICOs are happening simultaneously, it is harder for investors to undertake such detailed due diligence; we interpret this as a low α.

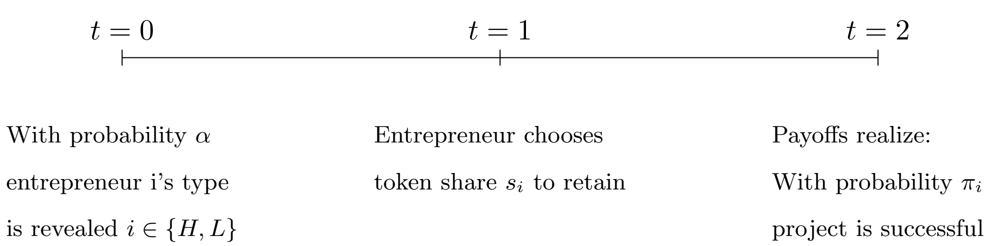


Figure 1: Model Timeline

Entrepreneur's Problem: At $t = 1$, there are 4 possible types of entrepreneurs - a high type whose type has been revealed, a low type whose type has been revealed, a high type whose type has not been revealed and a low type whose type has not been revealed. Let $i = \{H, L\}$ represent a high-type versus a low-type entrepreneur and let $j = \{R, U\}$ represent whether an entrepreneur's type is revealed or unknown. At $t = 1$, an entrepreneur of type $\{i, j\}$ chooses s_i^j by solving the following problem,

$$\max_{s\in[0,1]} \ (1-s)P(s) + s\pi_i R - \gamma C(s).$$

Equilibrium Definition: A perfect Bayesian equilibrium in pure strategies of the model is given by,

1. The share of tokens, s_H^{R*}, s_L^{R*}, s_H^{U*} and s_L^{U*}, retained by the each type of entrepreneur.
2. Investor beliefs, $\mu(i|s)$, that an entrepreneur is of type $\theta = \{H, L\}$, if he chooses to retain share s of tokens and his type has not been revealed.
3. The price investors pay per share of tokens sold to them such that their participation constraint is satisfied:

$$P^U(s) = \mu(H|s)\pi_H R + (1-\mu(H|s))\pi_L R,$$
$$P^R = \pi_i R.$$

3.2 Model Equilibrium and Empirical Predictions

In this section, we briefly describe the model equilibrium and our main empirical predictions. For greater detail on the model, please refer to Appendix A. In a benchmark equilibrium with full information, in which the entrepreneur type is observable, there is no correlation between retention and the entrepreneur quality. Moreover, entrepreneurs sell all tokens to minimize the cost of retention. Since in the full information benchmark all entrepreneurs' type is known, we can drop the superscript of whether entrepreneurs' type is revealed or unrevealed.

Proposition 1 *[Equilibrium with Full Information] Under full information, neither type of entrepreneur retains any tokens, i.e.,* $s_H = s_L = 0$*. There is therefore no correlation between ICO quality and retention.*

In the equilibrium with asymmetric information, with probability α an entrepreneur's type will be revealed. In this case, the entrepreneur's and investors' problems are the same as in the full information case. Both high- and low-type entrepreneurs whose type is revealed will select $s_H^{R*} = s_H^{R*} = 0$ and retain no tokens. Investors will buy all the tokens at a price of $\pi_i R$ where $i \in \{H, L\}$.

If an entrepreneur's type is not revealed, then there exists a separating equilibrium in which the high type chooses a retention fraction, $s_H^{U*} > 0$, to signal his type while the low type choose $s_L^{U*} = 0$. Since retention is costly, it is a credible signal of quality.[11] Under the Intuitive Criterion, we obtain a unique separating equilibrium.[12]

Proposition 2 *[Equilibrium with Asymmetric Information] Under the Intuitive Criterion, a unique separating equilibrium in which high-type entrepreneurs who have not had their type revealed at $t = 0$, signal their equality by retaining a fraction, $s_H^{U*} > 0$, tokens and low-type entrepreneurs who have not had their type revealed at $t = 0$, sell all tokens, $s_L^{U*} = 0$, exists and is unique.*

Thus, in the presence of asymmetric information, high-type entrepreneurs retain a higher fraction of tokens than low-type entrepreneurs. This creates a positive relationship between token retention and quality in the ICO market.

Proposition 3 *[Retention and Quality] In equilibrium, there is a positive covariance between the value of an ICO and the fraction of tokens retained, defined as $\rho \equiv Cov\,(\pi R, s^*)$.*

Retention when markets are crowded. The total fraction of tokens retained in equilibrium is given by

$$q(1-\alpha)s_H^{U*}.$$

A decrease in α can be interpreted as an increase in market crowdedness. If only a few ICOs are happening in a given month, investors have more time to study a company's whitepaper, business plan, examine its prototype, read analyst reports, talk to other investors, etc. These periods can be thought of as a time of high α, when a large fraction of entrepreneurs have their type revealed to investors. On the other hand, at a time when many ICOs are taking place it is harder for investors to undertake such detailed due diligence. We interpret these periods as a time of low α. As α decreases, a greater fraction of high-type entrepreneurs retain a large fraction of tokens, subsequently exerting more effort and producing higher quality products. We can establish the following proposition:

[11] Our model satisfies single crossing. A high-type entrepreneur whose type has not been revealed will face a higher *direct* cost of token retention than a low-type entrepreneur since $C(s_H^{U*}) > C(s_L^{U*})$ in a separating equilibrium. However his *net* cost of token retention, $s_H^{U*}\pi_H R - C(s_H^{U*})$, will be lower than that of a low-type entrepreneur who wants to copy the high-type entrepreneur by retaining a similar amount, $s_H^{U*}\pi_L R - C(s_H^{U*})$, since the high-type entrepreneur expects a greater payoff from the retained tokens.

[12] Since retention is a continuous variable, the Intuitive Criterion eliminates all pooling equilibria as well as all separating equilibria except for the least-cost separating equilibrium.

Proposition 4 *[Retention and Market Crowdedness] In the unique signaling equilibrium, ICO value covaries positively with the fraction of tokens retained, $\rho > 0$. As α decreases and the quality of public information deteriorates, the magnitude of this covariance increases,*

$$\frac{\partial \rho}{\partial \alpha} \leq 0.$$

Robustness to scammers. The ICO market is vulnerable to frauds and scams due to the lack of regulation and legal recourse. One might be concerned that our theoretical results do not hold in the presence of scammers. In our setting, we can define scammers as low-type entrepreneurs who are not developing a platform and face no costs of retention. In Appendix A, we show that our empirical predictions hold even in the presence of such entrepreneurs, as long as there is a limited measure of them.

Moral Hazard. For generality and completeness, we extend our model to additionally incorporate moral hazard in Appendix A. If the platform's success depends on the entrepreneur's unobservable effort, he will have to retain tokens to have enough skin-in-the-game to exert effort. Our key empirical predictions remain the same in this setting. Introducing moral hazard additionally causes all entrepreneurs to retain a strictly positive amount of tokens in equilibrium. While moral hazard alone can also generate the correlation between ICO quality and retention, adverse selection is necessary for this correlation to increase during periods of high market noise.

4 Data

We construct our list of ICOs from several ICO tracking websites that provide information to potential ICO investors and the general public. These websites include ICO Data (https://www.icodata.io/), Token Data (https://www.tokendata.io/), Cryptoslate (https://cryptoslate.com), ICO Bench (https://icobench.com), ICO Drops (https://icodrops.com), ICO Rating Agency (https://icorating.com), and ICO Checks (https://icocheck.io). We collect secondary market data from CoinMarketCap (https://coinmarketcap.com/). To merge ICO characteristics across different websites, we employ three identifiers: ticker symbol, token name, and website URL. We additionally perform extensive manual review to verify the list of unique ICOs. The combined database includes 5,644 distinct ICOs between January 1, 2016 and December 31, 2018, including 4,880 completed, 582 ongoing, and 182 planned ICOs. While our empirical analysis tends to focus on the subset of

1,501 ICOs with nonmissing data for key fields, we use the full set both to establish descriptive facts and also to construct our measure of market crowdedness.

We collect data on both ex-ante characteristics of ICOs that are known to investors at the time of the ICO (e.g., *Start Date*, *End Date*, *Hard Cap*, *Soft Cap*, *Token Retention*) and the performance of the project following the ICO (e.g., *Funds Raised*, *Listed Dummy*).[13] For a subset of ICOs, we augment this list of variables by hand-collecting data from ICO whitepapers, websites, and Apple's App Store (e.g., *Vesting Dummy*, *Product Live Dummy*).[14] Finally, we gather a list of alleged scams from the Dead Coins website (https://deadcoins.com/), which is a user-generated content forum that aims to catalog all failed cryptocurrencies and tokens. We merge this list with our ICO dataset to create an ex-post indicator of whether an ICO was reported as a scam (*Scam Dummy*). The full list of collected variables along with their definitions is provided in Table 1. To the best of our knowledge, ours is among the most comprehensive ICO databases to be used for academic research.[15] Additionally, our hand-collected ex-post measures of success are novel to the academic literature.

Panel (a) of Table 2 reports summary statistics for all ICOs in our dataset. Though we start with a sample of 5,644 ICOs, for most of our analysis we focus on ICOs which were completed by December 31, 2018, and which report funds raised and token retention. The summary statistics for these ICOs are listed in Panel (b) of Table 2. We use our full sample of ICOs when constructing our measure of the crowdedness of the ICO market. Also, in our robustness checks we include an additional 1,709 ICOs that did not report funds raised but did report token retention to address sample selection concerns (Appendix B, Table B.1).

[13] A number of ICOs in our sample appear on multiple ICO tracking websites. In some cases, the reported information on ICO characteristics differs across the websites. To resolve such conflicts, we proceed in the following manner. If there is more than one non-missing observation and two (or more) sources agree, we pick the matching value. If we have at least four non-missing observations such that we have two matching pairs and they do not agree between themselves, we pick the pair with the highest value. If we have at least five non-missing observations such that we have a matching pair and a matching set of three and they do not agree between themselves, we pick the value that the three websites agree on. If there is more than one non-missing observation and none of the sources agree, we narrow our choices to the pair with the smallest disagreement error (measured with the pair-wise percentage differences) and pick the highest value within that pair. Our results are robust to modifications in the above algorithm such as choosing the lowest value as a tie-breaker or taking the average.

[14] We hand-collect this data only for ICOs which have a non-missing start date, report token retention, and have raised a strictly positive amount of funds raised.

[15] At the largest end, Lyandres et al. (2020) gather data for 7,514 completed and ongoing ICOs between 2013 and 2019, Deng et al. (2018) use a sample of 3,573 completed ICOs between August 2015 and July 2018, Florysiak and Schandlbauer (2019) use a sample of 2,655 ICOs from ICO Bench, and Benedetti and Kostovetsky (2021) use a sample of 2,390 ICOs that were completed on or before April 30, 2018. The remaining empirical papers referenced in our literature discussion use samples that include anywhere from 64 to 1,549 ICOs.

The ICO market grew rapidly over our sample period (2016 through 2018). The number of ICOs increased from less than 100 in 2016 to over 1,500 in 2017 and then nearly to 4,000 in 2018. Through the end of 2018, the ICO market had raised over $33.4 billion in total. Since 2019, industry reports show that ICO activity has declined, although the number of STOs and IEOs has been growing (Token Alliance, 2019; Fitzpatrick, 2018; Aitken, 2019). For example, ICO Bench reports 1,023 ICOs occurred in 2019 and raised $3.4 billion in total.[16] In Figure 1, we see the steep increase in the aggregate amount of capital that was attracted by token issuers over our sample period. The ICO market reached its monthly peak of $6.8 billion in June 2018. Most of the funds raised that month are attributable to one of the largest ICOs in history – EOS, a developer of infrastructure for decentralized apps, – that completed its ICO on June 26, 2018, and raised $4.2 billion.[17] Another peak coincides with the completion date at which Telegram Open Network raised $1.7 billion. Excluding the 10 largest ICOs as measured by the amount of capital raised, ICO entrepreneurs raised around $2 billion per month throughout 2018. The average and median ICO in our full dataset raised $15.4 million and $4.9 million, respectively, though the distribution of funds raised is positively skewed (Table 2). The numbers are comparable across the full and analysis samples.

When announcing a project, token issuers typically produce a whitepaper that provides project details and the key terms of the ICO. We find that two-thirds of the ICOs in our full sample and 99% in our analysis sample produce such a whitepaper. A growing number of ICOs now require their contributors to pass KYC verification before purchasing tokens. This is a procedure of identity verification in which each buyer has to provide his credentials (e.g., a passport or driver's license number) in order to participate in the ICO. KYC ensures the transparency of transactions to assure regulatory compliance (e.g., that no illegally obtained money is being used to fund the platform). Such procedures are commonplace in traditional financial institutions. Given the recent adoption of KYC into the ICO market, only 38% (48%) of ICOs in our full (analysis) sample implement KYC. However its use rapidly grew from roughly 20% in early 2018 to 60% by early 2019 with the development of technology to facilitate its implementation.[18] An even smaller fraction of token

[16] These figures cover the period from January till November 2019. Note that the ICO activity might be underestimated as ICO Bench does not track all the ICOs occurring in the market.

[17] One unique feature of the EOS ICO is that it occurred over multiple rounds, raising $150 million in June 2017, $180 million in July 2017, $100 million in November 2017, $420 million in December 2017, $750 million in January 2018, and $500 million in February 2018, which is not reflected in the time-series of total funds raised. Importantly, since we conduct all empirical tests at the ICO level, our findings are unaffected even if we drop EOS from our sample.

[18] For example, a company called Civic provides decentralized KYC services for ICOs.

issuers (28% in the full sample and 38% in the analysis sample) request investors to preregister on a "whitelist" before the token sale launch. Only "whitelisted" customers are able to purchase tokens during the sale.

A fraction of ICOs (35% in the full sample and 51% in the analysis sample) also post their code to a GitHub repository, an online software development platform. Posting one's code to GitHub is a signal of transparency that prior studies have utilized (e.g., Amsden and Schweizer, 2018; Howell et al., 2019). Having open source code can indicate that entrepreneurs are further along in the development of their product or platform and, as such, it should be positively valued by investors.

A small subset of ICOs (3% in the full sample and 7% in the analysis sample) had prior venture capital funding.[19] These figures are similar to the share of venture backed ICOs reported in Howell et al. (2019). Securing funding from a venture capital firm can be viewed as a positive signal about the underlying project of the ICO. This signal is credible as long as the information can be verified on the website of the venture capital fund. Moreover, investors may anticipate that the ICO project will benefit directly from the assistance and expertise of venture capitalists.

We consider two measures of ICO team quality. First, we determine whether an ICO has reputable experts as advisers from its whitepaper or website. Similar to the use of venture capital funding, the presence of advisers can be viewed both as a positive signal about the project quality (i.e., an adviser would not want to be associated with a low-quality project or a scam) and an indication of the ICO's access to high-quality advice. Second, we determine whether any of the non-adviser ICO team members has previous experience in conducting an ICO. Investors may value such prior experience, especially in a new fundraising market. While one may be concerned that information about the ICO team members or advisers can be fake, preexisting reputations help create credibility. In particular, advertised team information about previous ICO experience can be verified through professional and social media platforms such as LinkedIn (Lyandres et al., 2020). For example, members and/or advisers with reputations typically already have a publicly known LinkedIn profile and social media accounts. They will therefore be expected to update these when they are working on new projects making it harder to falsely claim their involvement in an ICO.

[19] Data on prior funding are from ICO Rating Agency, which lists all funds that have been invested in a particular ICO. We define venture capital funds as those that explicitly list "venture" in their strategy.

Holding a pre-sale prior to the ICO itself has become common among token issuers (49% in the full sample and 58% in the analysis sample). A pre-sale is either a private sale of tokens to a small set of large investors or a smaller public sale that occurs prior to the ICO itself.[20] Although not part of the smart contract, these token sales are publicly observable. Holding a pre-sale can help entrepreneurs to assess demand for their product and showcase validation by the market to potential future investors.

In our analysis, we measure an issuer's signal of project quality and commitment to exert effort using the percentage of tokens that are retained by the ICO issuer and hence not available for sale during the fundraising campaign, which we denote as *Token Retention.* This measure is obtained from ICO whitepapers and tracking websites. It is important to note that token retention is declared before the ICO begins, can be written into the smart contract, and is verifiable on the blockchain. Moreover, 30% of ICOs in our analysis sample specify an explicit vesting schedule for retained tokens (Table 2). Panel (a) of Figure 2 shows that there is a lot of heterogeneity among token issuers in terms of the percentage of tokens they choose to sell to outside investors. Token retention ranges from 0% to close to 100%. An average ICO retains 44% of all issued tokens in both the full and analysis samples. Furthermore, there is no significant trend in the level of token retention over time, as shown in Panel (b) of Figure 2.

Ex-post measures of performance provide additional insights into the ICO market and the riskiness of the underlying projects. Through a manual review of the ICOs in our analysis sample (conducted in June 2019), we find that 76% of these projects continue to have a working website, 32% ultimately produce a working product, and 12% have a product in the Apple Store. Note that having a working product includes having an online platform or an application available to download. These figures support the notion that ventures funded by ICOs are very early stage and prone to failure.

Additionally, we find that 3% (6%) of the ICOs in our full (analysis) sample are reported as a

[20] In some pre-sales, investors were sold purchase agreements, which act as forward contracts for tokens. The general idea is that these contracts turn into tokens once the project has created a functional token. For example, Telegram raised $1.7 billion in two separate pre-sale rounds through selling such purchase agreements. Importantly, these types of agreements do not represent equity stakes in the venture, only a claim to future tokens. One may be concerned that these types of pre-sales would overstate our token retention measure. In Section 5.6, we investigate measures of ex-post retention and find that very few ICOs have ex-ante retention higher than our ex-post estimate.

scam on the Dead Coins website.[21] The relatively low rate of scams runs contrary to a common sentiment that most ICOs are scams. We reconcile this viewpoint with our data by noting the common conflation between token failure and a scam. However, we also acknowledge that the rate of scams may be underestimated because obvious scams would likely not have been listed on (or would have been quickly taken down from) the ICO aggregator websites that determine our sample.[22] For this reason and potential concerns about the quality of user-reported data, we do not consider *Scam Dummy* in our empirical analysis. Instead, we consider the rate of reported scams as a valuable summary statistic. For our theoretical mechanism to hold, we require that there be a limited measure of scammers who cannot be recognized as such ex ante by investors in ICO markets. The percentages of identified scams in our samples demonstrate that this assumption is well supported in the data.

5 Empirical Analysis

In this section, we test our main hypothesis that token issuers signal the quality of their project and a commitment to exert effort to potential investors by retaining a higher fraction of tokens during the ICO. We also test whether the strength of this signal depends on the market environment and investigate determinants of the cost of retention. Finally, we provide numerous robustness checks of our findings.

5.1 Empirical Challenges and Benefits of ICO Markets

As noted earlier in this paper, studying retention in the context of the ICO market allows us to overcome three key challenges germane to other empirical settings. First, the presence of regulators, intermediaries, and legal protections reduces the problem of asymmetric information in many markets. Therefore, the degree of information asymmetry may not be severe enough to induce entrepreneurs to signal through retention. ICOs remain largely unregulated, and there are essentially no legal repercussions for misrepresenting information. Moreover, the lack of intermediaries and face-to-face interactions between investors and entrepreneurs makes the market opaque. As a result, ICO investors face a uniquely high degree of information asymmetry.

[21] For more details about the underlying data scraped from Dead Coins, see Appendix C.

[22] Florysiak and Schandlbauer (2019) find a similar rate of potential scams in their sample (5.2%) and provide a similar explanation for the seemingly low rate.

Second, there is limited data on entrepreneurs seeking early-stage financing. The data that does exist, for example from venture capital markets, suffers from severe selection problems as we do not typically observe information on companies that were unsuccessful in obtaining financing. Our dataset, both the full and analysis sample, include comprehensive information on both ICOs that successfully obtained funding and the ones that failed to do so.

Third, a researcher must be able to control for all relevant information – both hard and soft – that could affect investors' perception of quality and retention. Confounding hard information includes, but is not limited to, the entrepreneur's schooling, résumé, and previous experiences with start-ups. In most markets, there are reputation costs and legal consequences for misrepresentation of such hard information, allowing investors to rationally condition their decisions on such data. If these variables are related to skill, then an investor may perceive an entrepreneur who, for example, has a better résumé to be of higher quality. Subsequently, the investor would ask for a smaller share of the project for a given amount of funds the entrepreneur is trying to raise. This would create a correlation between retention and project quality that would not be due to a signaling channel. Accounting and controlling for all such information empirically is not feasible. In ICO markets there are no legal repercussions for misrepresenting hard information. Hence, the only promises that ICO issuers commit to are ones that can be coded into smart contracts and/or are publicly observable on the distributed ledger. For example, in November 2017, the founder of ICO Confido lied about previously working for PepsiCo and Zalando but faced no legal consequences for this action. Because the underlying technology gives us a limited set of credible promises that an entrepreneur can commit to, the number of omitted variables we need to account for is greatly reduced.[23]

Token retention is one of the few verifiable actions an entrepreneur can take to convey information to investors. Retention can be coded explicitly into a smart contract along with a promised vesting schedule. Moreover, even without an explicit commitment, the nature of the public distributed ledger creates an implicit commitment to retention. If tokens that are meant to be retained by the ICO team are transferred out of the entrepreneur's wallet, investors can observe this action. They will rationally reduce the price they are willing to pay for these tokens, as this

[23]The ability to specify immutable governance terms ex ante through smart contracts is often cited as a benefit of ICOs as a financing mechanism (Cong and He, 2019; Howell et al., 2019).

means the entrepreneur is no longer committed to exerting effort. Entrepreneurs would prefer to hold on to these tokens, exert effort, and receive a higher price for them upon the successful completion of the platform. We verify in our data that ICO entrepreneurs typically do not violate the retention they announce before the ICO. Measures of ex-post retention tend to remain above ex-ante values until at least six months after the end of the ICO (see Section 5.6).

Besides hard information exchange, in other markets there can be significant soft-information exchange through in-person interactions between buyers and sellers. For example, consider the interaction between a venture capitalist and a potential entrepreneur. The venture capitalist may be influenced by a variety of "soft" variables such as a person's presentation skills, perceived intelligence, demeanor, or manner of speaking, which are nearly impossible to control for in an empirical design. Not accounting for this information can create an omitted variable bias similar to that in the hard-information case. Since ICO fundraising takes place online with little to no interaction between the team developing the platform and purchasers of the tokens, we have less confounding soft-information exchange to control for.

An alternative way to overcome the information asymmetry problem is to build reputation. As such, we control for variables related to reputation in this market such as whether an ICO previously obtained venture funding and whether an ICO team has members with previous experience in ICO fundraising or well-known ICO advisers. While information about the latter can be fake, it is verifiable to some extent through professional platforms as reputable team members and advisers would already have existing social media profiles that they can update with information about a new project. In general, due to the relative nascency of the market it can be difficult to build a reputation in the ICO market. For instance, the median ICO team has no members that have worked on any previous ICOs.

Another benefit of the ICO setting is that the current contracts between entrepreneurs and investors are relatively simple, making the interpretation of our results straightforward. In contrast, traditional financing contracts which rely on legal enforcement can be quite complicated, with multiple clauses and terms. Kaplan and Strömberg (2003) find that venture capital contracts often have anti-dilution provisions such as ratchet protections. Such clauses render tenuous any inference between the share retained by the entrepreneur and the price investors are willing to pay for shares.

The high degree of information asymmetry, limited set of credible promises, and simplicity of ICO contracts yield a unique setting in which to study signaling via retention, one that closely resembles a theoretical setting.

5.2 Signaling with Token Retention

We use ordinary least squares for the majority of our specifications, which we believe is well identified for the reasons outlined above. We consider various measures of ICO success in securing financing. Our results in this section provide information about whether investors are rationally reacting to retention by ICOs when they make investment decisions.

As a first step, we test whether ICOs that retain a larger fraction of tokens tend to raise a larger amount of funds. The specification is as follows, for ICO j:

$$ln\,(1 + Funds\ Raised_j) = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \tau_t + \epsilon_j, \tag{1}$$

where the dependent variable is the natural logarithm of the total funds raised during the ICO.[24] The coefficient estimate on *Token Retention* captures the relationship between retention and ICO success as measured by the total amount of funds raised. We control for a number of other ICO characteristics X_j that might affect ICO success. In particular, we control for ICO features that are credible – can be coded into the smart contract, verified on the public blockchain, or otherwise validated online (for example, through reputable online social media profiles) – since investors may rationally respond to such information. These controls include whether an ICO has a vesting schedule for its tokens, whether an ICO requires its investors to pass KYC verification and register on its whitelist, whether an ICO has obtained venture capital funding, whether an ICO posts its code to an online GitHub repository, whether an ICO lists advisers in its whitepaper or website, whether an ICO team member has worked on previous ICOs, whether an ICO has conducted a pre-sale, and whether an ICO has a hard cap and a soft cap. We do not control for whether an ICO has a whitepaper in our regressions as 99% of ICOs in our analysis sample have one. There is therefore not enough variation for the presence of a whitepaper to be a meaningful control. Finally,

[24] The total funds raised are denominated in US dollars. Some websites report funds raised in alternative currencies, e.g., Ethereum or Bitcoin. We convert these amounts to USD using the exchange rate prevailing on the last day of the ICO. This may introduce bias into our measure of the funds raised, as most successful ICOs manage to achieve their funding goal well before the announced end date of the ICO. This bias might be also attenuated due to the fact that cryptocurrency prices are highly volatile. For robustness, we also run all the regressions restricting our sample only to ICOs that report the amount of funds raised in USD. All results continue to hold.

we include time fixed effects based on the month in which an ICO begins.

Our measure of token retention is ex-ante and may differ from ex-post retention. When entrepreneurs have unsold tokens after the ICO is concluded, the typical practices are to burn these tokens or transfer them into the company's wallet. In either case, the ex-post token retention will be strictly higher than the ex-ante value announced by the issuers. In our analysis, we focus on retention announced prior to the start of the ICO because it is ex-ante and provides a lower bound on the ICO issuers' stake. Higher effective retention could encourage an entrepreneur to exert more effort and further improve ICO quality. From this perspective, our results provide a lower bound on the relationship between retention and ICO quality.

In Table 3, we report the estimates of the regression specified in (1). In Column (1), to analyze a larger sample, we set the funds raised for an ICO that does not report any funds raised to zero. In Column (2), we set the token retention of any ICO that does not report the amount/share of tokens retained to zero. In Columns (3)–(5), we focus on our analysis sample which includes ICOs that report both funds raised and token retention. The coefficient on token retention is positive and statistically significant across all specifications. Focusing on our analysis sample, we find that a 1 percentage point increase in token retention by the issuer, or equivalently a 1 percentage point decrease in the investor supply, leads to an approximately 0.9%–2.4% increase in the total funds raised. This result is consistent with the mechanisms described in our model. Importantly, this relationship is both economically and statistically significant. As noted previously, other studies of the ICO market also consider retention. Early studies (e.g., Roosenboom et al., 2020; Fisch, 2019) do not find a significant relationship between retention and ICO funds raised, which we believe is due to their smaller samples covering a shorter time period. For example, Roosenboom et al. (2020) look at 630 ICOs and their sample ends in December 2017. Our findings are in line with Lyandres et al. (2020) who have a comparable-sized sample to ours.

We additionally find that ICOs featuring a vesting schedule in their whitepapers raise 0.3%–0.5% more funds on average, suggesting that a long-term commitment to retention is valued by investors.[25] Note that even without an explicit commitment to vesting, there is still an implicit

[25]For simplicity and ease of interpretation, we use a dummy variable to control for the existence of a vesting schedule. Our results are robust to using a continuous measure that accounts for the duration of the vesting period.

commitment to retention due to the nature of a public distributed ledger.[26] An explicit vesting schedule can help further increase this implicit commitment by guaranteeing retention both for a longer period and in the event of market uncertainty or unexpected shocks.

We also find that ICOs with *KYC* or *Whitelist* raise 0.4%–0.6% more funds on average. Both *KYC* and *Whitelist* can be interpreted as barriers for investors to participate in ICOs. When token issuers choose to require their contributors to pass identify verification and/or register for a token sale in advance, they limit their potential pool of investors and in so doing may miss an opportunity to achieve their funding goals. At the same time, by creating these additional constraints, entrepreneurs reduce the risk that in the future their project will be shut down due to a regulatory infraction, which increases the ex-ante value of the project and, subsequently, the value of its tokens. These procedures therefore allow entrepreneurs to attract more credible investors, and as a result they are more successful in their fundraising campaign. The positive and statistically significant coefficients on *KYC* and *Whitelist* suggest that this second channel dominates.

Additionally, we find that ICOs with prior venture funding are more successful in their fundraising efforts. Specifically, they raise 1.2%–1.4% more funds on average. This result supports the view that investors consider venture capital funding to be both a positive and credible signal about project quality. It is in line with Howell et al. (2019) who find that prior venture funding is negatively associated with the failure of an ICO. Measures of team quality are also positively associated with funds raised. In particular, if an ICO team has advisers (*Advisers Dummy*) or members with prior ICO experience (*Experts Dummy*), it raises on average 0.3% and 0.4% more funds, respectively. These findings are in line with studies documenting that team quality is one of the most important factors in attracting venture capital (Bernstein, Korteweg, and Laws, 2017; Gompers, Gornall, Kaplan, and Strebulaev, 2020; Howell, 2020).

While this positive relationship between funds raised and token retention supports our hypothesis, it might also suggest that entrepreneurs who sell a smaller fraction of their tokens during an ICO can better monetize their project both at the time of sale and upon the project's

[26]As described earlier, if retained tokens are sold too quickly from the issuer's wallet, investors will observe this action and rationally reduce the price. Therefore, even without an explicit vesting promise, entrepreneurs have an incentive to retain their tokens due to the nature of a public ledger. In support of this argument, in Section 5.6, we show that measures of ex-post retention within six months following the ICO generally tend to remain above ex-ante values.

completion. In the model, future payoffs for both types of entrepreneurs always increase with token retention. If retaining more tokens also increases an entrepreneur's payoff at the ICO stage, a separating equilibrium would not be sustainable under any reasonable interpretation of the cost of holding tokens. A low-type entrepreneur would prefer to retain a larger fraction of tokens and copy the high type, thereby receiving a larger payoff at the time of sale and in the future. It is therefore important to distinguish between funds raised in an ICO that are invested in platform development and funds raised that are monetized by token issuers during an ICO.

To address this concern, we measure the total funds that are monetized by the ICO team as the difference between the ICO *Hard Cap* and *Soft Cap*. The *Hard Cap* is the maximum amount of funds the issuer wants to raise and is willing to accept. The *Soft Cap* is the minimum funding amount required for the issuer to continue the development of the product. The soft cap therefore captures the investment that ICO issuers have to make in their platform. Both amounts are explicit figures and, as such, can be coded into a token sale smart contract. In Table 4, we report the estimates of the following regression:

$$ln\left(Hard\ Cap_j - Soft\ Cap_j\right) = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j. \tag{2}$$

We find a negative relationship between total funds monetized and token retention. Furthermore, we find a strong positive relationship between the ICO soft cap and token retention, as reported in Table 5. These results support the idea that entrepreneurs that sell a smaller fraction of tokens during an ICO have projects with larger investment needs, causing them to raise more funds. However, these entrepreneurs do not pocket more money during the ICO. This finding provides evidence that it is costly for risk-averse entrepreneurs to retain a large fraction of tokens.

Our initial results based on specification (1) could be driven by a few large ICOs (as measured by the funds raised) that have high token retention but, at the same time, fail to achieve their fundraising goal. To address this concern, we can focus on token sales that feature a hard cap. A better success metric would be the amount of funds raised in the ICO as a fraction of the ICO hard cap, which we call *Fundraising Success*. This metric is continuous, ranges between 0 and 1, and captures how close token issuers are toward achieving their fundraising goal regardless of the ICO size. In Table 6, we report the estimates of the following regression:

$$Fundraising\ Success_j = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j, \tag{3}$$

where the dependent variable is the total funds raised as a fraction of the hard cap. We again find a strong positive relationship between token retention and ICO fundraising success, implying that investors do indeed respond to token retention and invest more in projects of higher quality. Our results are robust if we consider a larger sample by setting funds raised to zero for ICOs that do not report funds raised and by setting token retention to zero for ICOs that do not report the amount/share of tokens retained. Specifically, in our analysis sample, we find that a 1 percentage point increase in token retention results in a 0.2–0.4 percentage point increase in fundraising success. Again, this estimate is both economically and statistically significant. We can alternatively address the concern that our findings are driven by a few large ICOs by simply excluding them from our analysis. As expected, we find that our results are not driven by a few large ICOs which might have failed to achieve their fundraising goal despite raising a large amount of funds (Appendix B, Table B.2).

Compared to specification (1), we find that the same set of ICO characteristics positively affect fundraising success except for whether an ICO team has members with prior ICO experience. While we believe that team information is credible, our results in Table 6 imply that the value of team quality signals for fundraising success are diminished after taking into account ICO size.

We also estimate regression (3) with a rolling six-month time window. Figure 3 plots the estimates of the coefficient on *Token Retention* over time. We find that the relationship between signal quality and ICO fundraising success strengthens in the later months of the sample time period. Importantly, we control for time fixed effects. In the subsequent sections, we investigate potential reasons for this time pattern.

5.3 Signal Quality during Periods of Crowded Markets

In this section, we test whether the relationship between token retention and the quality of the ICO grows stronger during busier markets, as predicted by the theory. We proxy for crowdedness of the market by an increased number of ICOs, as well as by an increased amount of information provided by token issuers in their whitepapers. The intuition for this measure is that investors have less time to perform detailed due diligence per ICO when many ICOs are taking place and/or whitepapers become lengthier. They therefore rely more strongly on signals such as token retention. We propose using the number of ongoing ICOs at a given time scaled by the average number of words in the whitepapers of ICOs at that time as a measure of the crowdedness of the market. Specifically, for each ICO j, we count the number of ICOs that are active within a 15-day range from the launch date of ICO j and multiply that by the average number of words in the whitepapers of ICOs active at that time. In Figure 4), we plot the evolution of this measure over time.

We test whether the signaling mechanism is stronger during periods when there are a large number of active ICOs and/or a large amount of information provided in their whitepapers, using the following specification:

$$\begin{aligned} Fundraising\ Success_j = {} & \alpha + \beta_1 \cdot Token\ Retention_j + \beta_2 \cdot ln\,(Scaled\ \#\ of\ ICOs_j) \\ & + \beta_3 \cdot Token\ Retention_j \times ln\,(Scaled\ \#\ of\ ICOs_j) + \gamma X_j + \epsilon_j, \end{aligned} \tag{4}$$

where all variables, besides dummies, are demeaned to capture the average effect of token retention on the total funds raised as a fraction of the ICO hard cap. We are interested in the coefficient estimate on the interaction term that captures the differential effect of token retention on ICO success when market crowdedness is above its mean level. Note that for this analysis we restrict our sample to ICOs with a beginning date on or after June 1, 2017. This allows us to focus on the months when a sufficient number of ICOs are happening at the same time.

In Table 7 we see that the positive relationship between token retention and ICO fundraising success becomes stronger when ICO markets become more crowded. This is in line with our model's predictions that signaling a project's quality is more important when the ICO market is very crowded, because when this is the case it is harder for investors to separate out good investments from bad in the absence of such signals. As before, our results are robust to using a larger sample of ICOs by assuming funds raised and token retention when not reported are zero. In particular, in our analysis sample, for each percentage increase in the scaled number of ICOs, a 1 percentage point increase in token retention leads to an additional 0.2–0.4 percentage point increase in fundraising success. For reference, when the scaled number of ICOs is at its mean level, a 1 percentage point increase in token retention leads to a 0.1–0.3 percentage point increase in fundraising success. For robustness, we also consider alternative measures of market crowdedness in Section 5.6. Our results remain quantitatively similar.

5.4 Ex-Post Measures of ICO Performance

Reaching the fundraising goal during the ICO does not necessarily result in the successful development of a product. Thus, the amount of funds raised is only a proxy for the project's quality, and it is important to explore ex-post measures of ICO performance.

To start, in line with other studies, we measure ICO performance using a number of market based measures. First, we measure success by whether a token lists on an exchange after an ICO. Crypto exchanges engage in due diligence prior to accepting a token for listing to protect their reputation both with customers and regulators. Furthermore, the fees issuers must pay to list their tokens on an exchange are nontrivial and can range from \$1 million to \$3 million.[27] Since tokens are listed on an exchange after the ICO start date, an indicator of whether an ICO's tokens are listed on a reputable exchange gives us an ex-post measure of the project's quality. Second, conditional on a token being listed on an exchange, we measure its quality with the token's average trading volume and average market capitalization between its first listing date and December 31, 2018. We estimate the following specification:

$$\begin{aligned} \textit{Ex-Post Success Measure}_j = \alpha &+ \beta_1 \cdot \textit{Token Retention}_j \\ &+ \beta_2 \cdot \textit{Successful Dummy}_j + \gamma X_j + \epsilon_j, \end{aligned} \tag{5}$$

where the dependent variable is our ex-post measure of success. The control variable *Successful Dummy* is an indicator that equals 1 if ICO *Funds Raised* $\geq$ *Soft Cap*. For ICOs without an explicit soft cap, any funds raised during the ICO are not returned to investors, effectively implying a soft cap of 0. For these ICOs, the indicator is therefore set to 1 if token issuers have raised a strictly positive amount of funds, *Funds Raised* > 0. We include this variable to capture the idea that an ICO that has been successful in its fundraising efforts is more likely to develop a working product and, hence, to list its token on an exchange.

In line with our previous findings, we find that ICO investors indeed rely on token retention as

[27] https://www.bloomberg.com/news/articles/2018-04-03/crypto-exchanges-charge-millions-to-listtokens-autonomous-says.

a signal of the project's quality. As reported in Table 8, a 1 percentage point increase in token retention leads to a 0.1 percentage point increase in the probability of tokens being listed on an exchange. We additionally document that, conditional on trading on an exchange, token retention predicts how successful the ICO will be based on two other measures of market based performance. In particular, a 1 percentage point increase in token retention leads to a 0.9%–1.6% increase in the average daily trading volume of the token. Similarly, a 1 percentage point increase in token retention leads to a 0.6%–1.4% increase in the average market capitalization.

Many tokens begin trading soon after the ICO end date, while the product is still in the early stages of development. Therefore, market-based measures may not be good indicators of long-term performance. Thus, we additionally use three alternative measures of ex-post ICO success: (1) whether the issuer continues to have an active and working website, (2) whether the issuer has produced a product that is live to users, and (3) whether the issuer has an application for download in Apple's App Store. To the best of our knowledge, we are the first study to evaluate ICOs based on performance metrics related to actual product development.

In line with our hypothesis, we find that a 1 percentage point increase in token retention leads to a 0.2 percentage point increase in the probability of the company still having a working website (Table 9). We additionally find that a 1 percentage point increase in token retention leads to a 0.1 percentage point increase in the probability of the company having a live product and a 0.1 percentage point higher probability of having an application available in Apple's App Store. Hence, high-quality ICO issuers who have been successful in signaling their quality through token retention will be more likely to have their tokens traded and to develop their product.

Data on our alternative ex-post success measures were all recorded in June 2019. Given that we measured outcomes for all ICOs in the same calendar month, one might be concerned that we are underestimating the successful development of products funded by those in the latter part of our sample, which can bias the coefficient measuring the effect of token retention on ex-post success. In Figure 5, we indeed observe that there has been a weak downward trend in the fraction of ICOs with a live product and product available in Apple's App Store, with the exception of December 2018. However, the fraction of ICOs with a working website does not exhibit a similar trend. Nonetheless, to address this concern we include time fixed effects in the regressions to effectively compare the performance of ICOs with differential retention launching within the same month. Furthermore, we estimate regression (5) with a rolling six-month time window. The estimates of the coefficient on *Token Retention* are plotted in Figure 6. With the exception of the early sample in which we had fewer ICOs, the relationship between signal quality and ex-post measures of success is rather stable over time and similar in magnitude to the full sample estimates.

5.5 Costs of Retention

In this section, we explore the costs entrepreneurs face when retaining tokens. As noted earlier, if entrepreneurs are risk-averse, retention is costly because they are forced to hold a large portion of their wealth in a single risky asset. Retaining tokens is particularly costly for entrepreneurs who cannot diversify this risk away through financial markets. Entrepreneurs with limited risk-sharing opportunities therefore will sell a larger share of their project to mitigate their risk exposure. In the model, this differential cost of retention across entrepreneurs is captured with a parameter (γ).

To capture variation in risk-sharing opportunities across ICO teams, we use the Financial Development (FD) Index of the country in which they are located. The FD index, which was developed by the International Monetary Fund, ranges on a scale from 0 to 1 and summarizes the degree of development of the financial institutions and financial markets of a given country in terms of their size, liquidity, access, and efficiency. See Svirydzenka (2016) for more details on the methodology behind the index's construction.

There is substantial heterogeneity in the location of ICO teams. Though token issuers are mostly concentrated in the United States, Singapore, the United Kingdom, the Russian Federation, Estonia, and Switzerland, we nevertheless observe ICO activity across a wide range of countries (Figure 7). The full list of locations is provided in Appendix B, Table B.3. Token issuers tend to concentrate in the same six countries listed above although their relative shares of ICO activity have changed over time.

To assess the effect of risk-sharing opportunities on token retention, we estimate the following regression:

$$Token\ Retention_j = \alpha + \beta \cdot Financial\ Development\ Index_j + \delta X_j + \varepsilon_j. \tag{6}$$

In line with our theoretical argument, we find that ICO teams located in countries with a higher level of financial development tend to retain more tokens. As shown in Table 10, we find that a 1 percentage point increase in the FD Index is associated with a 0.1 percentage point increase in token retention. Note that both token retention and the FD Index range between 0 and 1 which suggests if the FD Index increases by 1 the associated change in token retention would be 5%–11%. The coefficient estimates are statistically and economically significant.

An alternative explanation for our results is that entrepreneurs located in more financially developed countries are wealthier. Under decreasing absolute risk aversion, retaining tokens is less costly for richer entrepreneurs, as the wealth allocated to an ICO constitutes a smaller share of their overall wealth. To test this, we additionally control for a country's GDP per capita and we find that GDP does not affect retention beyond the FD Index (Table 10).

5.6 Robustness

In this section, we present a variety of robustness checks to corroborate and sharpen our main empirical results. In particular, we address selection bias and consider alternative measures of token retention, fundraising success, and market crowdedness. We also analyze whether ICO teams commit to the ex-ante retention levels announced in their whitepapers.

Addressing Selection Bias. As we mentioned earlier, not all of the ICOs in our full sample report funds raised. One possible concern is that these ICOs with missing values for funds raised are more likely to have failed, which would explain why they did not report values for the capital they raised. Our coefficient estimates therefore may be subject to a selection bias, as our regressions might include only ICOs that have been relatively more successful and as such chose to report the amount they raised. About one-quarter of the ICOs in our sample achieved their financing goal, and this rate has been relatively stable over time (Figure 8). Note that about 30% of the ICOs included in our sample raised close to $0, demonstrating that our analysis sample includes both successful and unsuccessful ICOs. We further address this selection bias by setting funds raised and token retention to zero when these variables are not reported by ICOs. Our results remain robust and are reported in the first few columns of our main tables.

Alternative Measure of Retention. Some tokens that are not sold to ICO investors may be kept aside for additional expenses related to project development. Therefore, one might be concerned that our primary retention measure does not truly capture entrepreneurial retention as defined in our model. To address this concern, we estimate regression (3) using an alternative measure of retention, which we call *Team Retention*. We hand-collect data on the allocation of retained tokens from ICOs' whitepapers. These data allow us to compute an alternative measure of retention defined as the fraction of tokens not for public sale (i.e., our primary measure) less shares designated for the following business purposes: private sale to early investors, incentive schemes (e.g., bounty programs), operation and/or growth of the venture (e.g., marketing), and advisers. The remaining tokens are designated specifically for insiders (e.g., founders) or to a generic company account. Our results are effectively the same when using this alternative retention measure (Appendix B, Table B.4), and therefore we conclude that our primary measure does indeed capture entrepreneurial retention.[28]

Ex-Post Retention. Our primary data sources for the token retention variable are the aforementioned ICO tracking websites and ICO whitepapers (Section 4). These websites source their information from both the whitepapers and the publicly available smart contract codes. However, one might be concerned that this due diligence is not so rigorous and that, as a result, ICO issuers actually retain a smaller fraction of tokens than the promised amount. Similarly, one may be concerned that insiders quickly sell their tokens soon after the ICO in the absence of an explicit commitment to a vesting schedule. To address these concerns, we check the percentage of tokens in circulation after the sale as reported on CoinMarketCap.[29] Specifically, we compute the percentage of tokens not in circulation as a proxy for ex-post token retention and compare it against our ex-ante value. For the ICOs with these data available, we find that the ex-post proxy is always greater than or close to the ex-ante value within six months of the ICO end date (Appendix B, Figure B.1). As discussed earlier, higher ex-post retention can occur in ICO markets for reasons

[28]The sample size in Appendix B, Table B.4, is smaller than Table 6 because a breakdown of the token allocation needed to compute *Team Retention* is only available for a subset of the ICOs in our analysis sample.

[29]CoinMarketCap reports secondary market data for thousands of cryptocurrencies and tokens and is generally perceived to be the highest-quality source for such data. Circulating supply is their "best approximation of the number of coins that are circulating in the market and in the general public's hands." It is a direct input to the market capitalization of crypto assets and, given the importance of this metric, CoinMarketCap is quite careful about measuring and reporting the data. In fact, it does not report circulating supply when it is not confident in the accuracy of the data. The percentage of tokens in circulation is the circulating supply divided by the total token supply.

such as unsold tokens that are burned or transferred to the team's wallet. These observations confirm that ICO issuers did not sell a larger fraction of tokens to the public than advertised. After six months, however, we do observe several ICOs with a higher percentage of tokens in circulation. These observations are consistent with insiders selling their tokens to the public after a vesting period. In sum, we find no evidence that ICO issuers violate their promised retention during the token sale.

Alternative Regression Models. For binary outcomes, we can also estimate logistic and probit models in addition to OLS. For robustness, we verify whether our results continue to hold under these alternative regression models. We find that the coefficient estimates for token retention obtained from estimating logistic and probit models are both statistically and economically significant (Appendix B, Tables B.5–B.6).

Alternative Measures of Market Crowdedness. We also consider alternative measures of market crowdedness. Specifically, we alter the window over which we count the number of ICOs happening around the launch date of ICO j and additionally use alternative scaling measures (e.g., the average number of pages in the whitepapers). We also consider just the number of ICOs happening at the same time as ICO j without scaling this measure. The results are very similar both quantitatively and qualitatively to those in Table 7. Also, our findings continue to hold when we set fundraising success to zero for ICOs for which the funds raised is not reported to address the selection concerns discussed earlier (Appendix B, Table B.7).

Alternative Measures of Fundraising Success. For token sales that feature a soft cap, we use an alternative success metric – the amount of funds raised in the ICO as a fraction of the ICO soft cap. This measures whether entrepreneurs raised at least enough funds to continue the development of the product or platform. In Table B.8, we report the estimates of the following regression:

$$\text{``}Soft\text{''}\ Fundraising\ Success = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j, \tag{7}$$

where the dependent variable is the total funds raised as a fraction of the soft cap. Since not many token issuers set a minimum funding goal, the sample size shrinks to 639 ICOs. Nonetheless, we continue to find a positive, though not always statistically significant, relationship between token retention and ICO "soft" fundraising success. In columns (1) and (2), when we set the funds raised and token retention to zero for ICOs which do not report these values allowing us to increase the sample size, the relationship becomes statistically significant. Note that the coefficient estimate on token retention is 2–3 times larger than in specification (3) where the dependent variable is the total funds raised as a fraction of the hard cap. This is not surprising since, unlike "hard" fundraising success, which ranges between 0 and 1, "soft" fundraising success can exceed 1, and hence the effect of a 1 percentage point change in token retention should be stronger in specification (7).

6 Conclusions

In this paper, we provide evidence that entrepreneurs signal their quality and commitment to exert effort through retention in ICO markets. We first establish a strong positive relationship between token retention and fundraising success. Second, we demonstrate that market conditions affect the strength of the signal. Third, we find that ICO issuers that retain more tokens exhibit better ex-post performance across multiple measures. Finally, we show that risk-sharing opportunities affect the cost of retention. Collectively, these results confirm the hypotheses formalized in our stylized model of the ICO market.

Our study helps to advance two distinct areas of the literature. First, it contributes to the literature that empirically analyzes retention as a means of overcoming information asymmetry problems. The ICO market allows us to study the role of retention in early-stage financing. Our paper is the first to study the role of retention in the ICO market, which is unique in its lack of regulation and intermediation, absence of in-person interactions, and limited set of contractible promises. In this setting with a high degree of asymmetric information, we find positive evidence for retention as a signal. Furthermore, new to this literature, we show that the strength of this signal depends on the market environment. Second, our paper contributes to the general study of the ICO market. In particular, we show how investors and entrepreneurs rationally respond to the information asymmetries inherent in ICO financing.

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.

Aitken, R. (2019, March). After 'Crypto's Winter,' ICOs Growing Less But Maturing With Shift To STOs. *Forbes*.

Amsden, R. and D. Schweizer (2018). Are Blockchain Crowdsales the New "Gold Rush"? Success Determinants of Initial Coin Offerings. Working paper.

Bakos, Y. and H. Halaburda (2019). Funding new ventures with digital tokens: Due diligence and token tradability. Working paper.

Benedetti, H. and L. Kostovetsky (2021). Digital tulips? returns to investors in initial coin offerings. *Journal of Corporate Finance 66*, 101786.

Bernstein, S., A. Korteweg, and K. Laws (2017). Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment. *The Journal of Finance 72*(2), 509–538.

Boreiko, D. and N. Sahdev (2018). To ICO or not to ICO – Empirical analysis of Initial Coin Offerings and Token Sales . Working paper.

Bourveau, T., E. T. De George, A. Ellahie, and D. Macciocchi (2018). Initial Coin Offerings: Early Evidence on the Role of Disclosure in the Unregulated Crypto Market. Working paper.

Casamatta, C. (2003). Financing and Advising: Optimal Financial Contracts with Venture Capitalists. *The Journal of Finance 58*(5), 2059–2085.

Catalini, C. and J. S. Gans (2018). Initial Coin Offerings and the Value of Crypto Tokens. Working paper.

Chod, J. and E. Lyandres (2021). A theory of icos: Diversification, agency, and information asymmetry. *Management Science*.

Clarkson, P. M., A. Dontoh, G. Richardson, and S. E. Sefcik (1991). Retained ownership and the valuation of initial public offerings: Canadian evidence. *Contemporary Accounting Research 8*(1), 115–131.

Cong, L. and Z. He (2019). Blockchain Disruption and Smart Contracts. *The Review of Financial Studies 32*(5), 3412–3460.

Cong, L., Y. Li, and N. Wang (2021). Tokenomics: Dynamic Adoption and Valuation. *The Review of Financial Studies 34*(3), 1105–1155.

Cong, L. W. and Y. Xiao (2021). Categories and functions of crypto-tokens. Working paper.

Deloitte (2017). Initial Coin Offering: A new paradigm. Technical report, Deloitte.

Deng, X., Y.-T. Lee, and Z. Zhong (2018). Decrypting Coin Winners: Disclosure Quality, Governance Mechanism, and Team Networks. Working paper.

Dessí, R. (2005). Start-up Finance, Monitoring, and Collusion. *The RAND Journal of Economics 36*(2), 255–274.

Downes, D. H. and R. Heinkel (1982). Signaling and the Valuation of Unseasoned New Issues. *The Journal of Finance 37*(1), 1–10.

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

Fitzpatrick, L. (2018, October). Crypto Exchanges And ICOs Are Evolving Into New Entities. *Forbes*.

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

Garmaise, M. J. and T. J. Moskowitz (2003). Confronting Information Asymmetries: Evidence from Real Estate Markets. *The Review of Financial Studies 17*(2), 405–437.

Garratt, R. and M. R. C. van Oordt (2019). Entrepreneurial incentives and the role of initial coin offerings. Working paper.

Goldstein, I., D. Gupta, and R. Sverchkov (2019). Initial Coin Offerings as a Commitment to Competition. Working paper.

Gompers, P. A., W. Gornall, S. N. Kaplan, and I. A. Strebulaev (2020). How do venture capitalists make decisions? *Journal of Financial Economics 135*(1), 169 – 190.

Hellmann, T. and M. Puri (2002). Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence. *The Journal of Finance 57*(1), 169–197.

Holmstrom, B. and J. Tirole (1997). Financial Intermediation, Loanable Funds, and The Real Sector. *The Quarterly Journal of Economics 112*(3), 663–691.

Howell, S., M. Niessner, and D. Yermack (2019). Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales. *The Review of Financial Studies*.

Howell, S. T. (2020). Reducing information frictions in venture capital: The role of new venture competitions. *Journal of Financial Economics 136*(3), 676 – 694.

Inderst, R. and H. M. Müller (2004). The effect of capital market characteristics on the value of start-up firms. *Journal of Financial Economics 72*(2), 319 – 356.

Kaplan, S. N. and P. Strömberg (2001, May). Venture Capitals As Principals: Contracting, Screening, and Monitoring. *American Economic Review 91*(2), 426–430.

Kaplan, S. N. and P. Strömberg (2003). Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts. *The Review of Economic Studies 70*(2), 281–315.

Kaplan, S. N. and P. Strömberg (2004). Characteristics, Contracts, and Actions: Evidence from Venture Capitalist Analyses. *The Journal of Finance 59*(5), 2177–2210.

Krinsky, I. and W. Rotenberg (1989). Signalling and the Valuation of Unseasoned New Issues Revisited. *Journal of Financial and Quantitative Analysis 24*(2), 257–266.

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

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

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

Lyandres, E., B. Palazzo, and D. Rabetti (2020). ICO Success and Post-ICO Performance. Working paper.

Malinova, K. and A. Park (2018). Tokenomics: When tokens beat equity. Working paper.

Momtaz, P. P. (2020, 05). Initial coin offerings. *PLOS ONE 15*(5), 1–30.

Momtaz, P. P. (2021). The pricing and performance of cryptocurrency. *The European Journal of Finance 27*(4-5), 367–380.

Narayanan, A., J. Bonneau, E. Felten, A. Miller, and S. Goldfeder (2016). *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton University Press.

Purnanandam, A. and T. A. Begley (2016). Design of Financial Securities: Empirical Evidence from Private-Label RMBS Deals. *The Review of Financial Studies 30*(1), 120–161.

Repullo, R. and J. Suarez (2004). Venture Capital Finance: A Security Design Approach. *Review of Finance 8*(1), 75–108.

Ritter, J. R. (1984). Signaling and the Valuation of Unseasoned New Issues: A Comment. *The Journal of Finance 39*(4), 1231–1237.

Roosenboom, P., T. van der Kolk, and A. de Jong (2020). What determines success in initial coin offerings? *Venture Capital 22*(2), 161–183.

Sanchez, D. C. (2017). An Optimal ICO Mechanism. Working paper.

Schmidt, K. M. (2003). Convertible Securities and Venture Capital Finance. *The Journal of Finance 58*(3), 1139–1166.

Schultz, P. and M. Zaman (2001). Do the individuals closest to internet firms believe they are overvalued? *Journal of Financial Economics 59*(3), 347 – 381.

Sockin, M. and W. Xiong (2018). A Model of Cryptocurrencies. Working paper.

Svirydzenka, K. (2016, January). Introducing a New Broad-based Index of Financial Development. IMF Working Papers 16/5, International Monetary Fund.

The Economist (2018, August). From one cryptocurrency to thousands. *The Economist*.

Token Alliance (2018, July). Understanding Digital Tokens: Market Overviews and Proposed Guidelines for Policymakers and Practitioners. White paper, Chamber of Digital Commerce.

Token Alliance (2019, August). Market Overviews and Trends in Token Project Fundraising Events. White paper, Chamber of Digital Commerce.

Tables and Figures

Table 1: List of Variables

Variable Name	Description
Whitepaper Dummy	An indicator that equals 1 if an ICO has a whitepaper
Whitelist Dummy	An indicator that equals 1 if an ICO has a whitelist (i.e., requires its buyers to register in advance)
KYC Dummy	An indicator that equals 1 if an ICO requires its investors to pass Know Your Customer (KYC) verification
Venture Funding Dummy	An indicator that equals 1 if the ICO has prior venture capital funding
Advisers Dummy	An indicator that equals 1 if the ICO lists advisers in their whitepaper or website
Experts Dummy	An indicator that equals 1 if an ICO team member has previously worked on another ICO
Listed Dummy	An indicator that equals 1 if ICO tokens are traded on an exchange
Trading Volume	The sum of the value of all token transactions over a 24 hour period
Market Capitalization	The token price multiplied by the token's circulating supply (the number of coins circulating in the market and in the hands of the general public)
GitHub Dummy	An indicator that equals 1 if a token issuer posts its code to a GitHub repository
Pre-sale Dummy	An indicator that equals 1 if an ICO has a pre-sale
Funds Raised > 0 *Dummy*	An indicator that equals 1 if an ICO has raised a positive amount of funding
Funds Raised	The amount of capital raised during an ICO
Hard Cap	The maximum amount of capital that can be raised during an ICO
Soft Cap	The minimum amount of capital required by token issuers to develop a product
Fundraising Success	The ratio of ICO *Funds Raised* to *Hard Cap*
"Soft" Fundraising Success	The ratio of ICO *Funds Raised* to *Soft Cap*
Successful Dummy	An indicator that equals 1 if token issuers have raised the minimum amount of funds required to develop a product
Investor Supply	The fraction of all issued tokens that can be sold to outside investors during an ICO
Token Retention	The fraction of all issued tokens that cannot be sold to outside investors during an ICO
Vesting Dummy	An indicator that equals 1 if a whitepaper contains a vesting schedule for ICO tokens
Working Website Dummy	An indicator that equals 1 if an ICO's website is still active
Product Live Dummy	An indicator that equals 1 if an ICO issuer has a live product or platform
Product Apple Dummy	An indicator that equals 1 if an ICO issuer has an application available for download in the Apple Store
Scam Dummy	An indicator that equals 1 if an ICO is reported as a scam with explanation on the Dead Coins website (see Appendix C for details)

Table 2: Summary Statistics

Panel (a): Full Sample

	N	Mean	SD	1%	10%	50%	90%	99%
Whitepaper Dummy	5644	0.68	0.47	0.00	0.00	1.00	1.00	1.00
GitHub Dummy	5644	0.35	0.48	0.00	0.00	0.00	1.00	1.00
KYC Dummy	5644	0.38	0.48	0.00	0.00	0.00	1.00	1.00
Whitelist Dummy	5644	0.28	0.45	0.00	0.00	0.00	1.00	1.00
Venture Funding Dummy	5644	0.03	0.16	0.00	0.00	0.00	0.00	1.00
Advisers Dummy	5644	0.14	0.35	0.00	0.00	0.00	1.00	1.00
Experts Dummy	5644	0.19	0.40	0.00	0.00	0.00	1.00	1.00
Pre-sale Dummy	5644	0.49	0.50	0.00	0.00	0.00	1.00	1.00
(Funds Raised> 0) Dummy	5644	0.35	0.48	0.00	0.00	0.00	1.00	1.00
Funds Raised, $ mil	1989	15.39	105.40	0.00	0.15	4.93	30.00	117.02
Hard Cap, $ mil	3269	49.06	363.03	0.37	4.21	20.00	70.00	400.00
Soft Cap, $ mil	1568	5.46	10.85	0.06	0.50	2.90	11.60	50.00
Token Retention, %	3704	43.39	20.61	0.00	20.00	40.00	70.00	94.00
Fundraising Success, %	1567	46.11	39.36	0.01	1.68	33.17	100.00	100.00
Listed Dummy	5644	0.15	0.36	0.00	0.00	0.00	1.00	1.00
Scam Dummy	5644	0.03	0.18	0.00	0.00	0.00	0.00	1.00

Panel (b): Analysis Sample

	N	Mean	SD	1%	10%	50%	90%	99%
Whitepaper Dummy	1501	0.99	0.11	0.00	1.00	1.00	1.00	1.00
GitHub Dummy	1501	0.51	0.50	0.00	0.00	1.00	1.00	1.00
KYC Dummy	1501	0.48	0.50	0.00	0.00	0.00	1.00	1.00
Whitelist Dummy	1501	0.38	0.49	0.00	0.00	0.00	1.00	1.00
Venture Funding Dummy	1501	0.07	0.26	0.00	0.00	0.00	0.00	1.00
Advisers Dummy	1501	0.53	0.50	0.00	0.00	1.00	1.00	1.00
Experts Dummy	1501	0.31	0.46	0.00	0.00	0.00	1.00	1.00
Pre-sale Dummy	1501	0.58	0.49	0.00	0.00	1.00	1.00	1.00
(Funds Raised> 0) Dummy	1501	1.00	0.00	1.00	1.00	1.00	1.00	1.00
Funds Raised, $ mil	1501	17.10	120.58	0.00	0.28	5.50	30.00	100.00
Hard Cap, $ mil	1304	45.22	348.26	0.42	5.10	20.50	66.59	250.00
Soft Cap, $ mil	639	5.17	7.73	0.08	0.53	3.00	12.00	35.00
Token Retention, %	1501	43.58	20.46	0.00	20.00	40.00	70.00	91.43
Fundraising Success, %	1303	47.75	39.25	0.05	2.16	35.28	100.00	100.00
Listed Dummy	1501	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Scam Dummy	1501	0.06	0.24	0.00	0.00	0.00	0.00	1.00
Vesting Dummy	1501	0.30	0.46	0.00	0.00	0.00	1.00	1.00
Working Website Dummy	1501	0.76	0.42	0.00	0.00	1.00	1.00	1.00
Product Live Dummy	1501	0.32	0.47	0.00	0.00	0.00	1.00	1.00
Product on Apple Store Dummy	1501	0.12	0.32	0.00	0.00	0.00	1.00	1.00

This table reports summary statistics for the sample of ICOs from January 1, 2016, through December 31, 2018, including 4,880 completed, 582 ongoing, and 182 planned ICOs. In Panel (b), we restrict our sample only to ICOs which were completed before December 31, 2018, and which report funds raised and token retention. We obtain our ICO sample by combining the data from the following ICO tracking websites: ICO Data, Token Data, CoinMarketCap, Cryptoslate, ICO Bench, ICO Drops, ICO Rating Agency, and ICO Checks. For each variable, the table shows the number of nonmissing observations, along with the cross-sectional mean, standard deviation, 1^{st}, 10^{th}, 50^{th}, 90^{th}, and 99^{th} percentiles. The description of variables is provided in Table 1.

Table 3: Total Funds Raised and Token Retention

	(1)	(2)	(3)	(4)	(5)	(6)
Token Retention	2.414***	1.834***	1.562***	1.565***	0.966***	0.860***
	(3.79)	(7.99)	(5.42)	(5.46)	(3.53)	(3.19)
Vesting Dummy				0.459***	0.341***	0.348***
				(4.56)	(3.59)	(3.67)
KYC Dummy					0.393***	0.378***
					(2.91)	(2.84)
Whitelist Dummy					0.604***	0.536***
					(4.92)	(4.36)
Venture Funding Dummy					1.410***	1.249***
					(11.13)	(9.20)
Advisers Dummy					0.315***	0.321***
					(3.15)	(3.21)
Experts Dummy					0.424***	0.434***
					(4.63)	(4.77)
GitHub Dummy						0.063
						(0.62)
Pre-sale Dummy						-0.025
						(-0.22)
Hard Cap Dummy						0.734***
						(3.80)
Soft Cap Dummy						-0.614***
						(-5.67)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.112	0.099	0.080	0.090	0.191	0.213
N	3210	1946	1501	1501	1501	1501

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$ln\,(1 + Funds\ Raised_j) = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j.$$

The dependent variable in each regression is the natural logarithm of one plus the total funds raised in the ICO. The total funds raised are denominated in U.S. dollars. Token retention is expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. In column (1), the total funds raised are set to zero if the company does not report the amount of funds raised. In column (2), token retention is set to zero if the company does not report the amount/share of tokes retained. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 4: Total Monetized Funds and Token Retention

	(1)	(2)	(3)	(4)	(5)
Token Retention	-0.043	-0.122	-0.185	-0.224	-0.199
	(-0.28)	(-0.68)	(-0.74)	(-0.88)	(-0.78)
Vesting Dummy			-0.043	-0.101	-0.107
			(-0.44)	(-1.00)	(-1.05)
KYC Dummy				0.211*	0.207*
				(1.76)	(1.73)
Whitelist Dummy				-0.023	-0.024
				(-0.19)	(-0.20)
Venture Funding Dummy				-0.067	-0.076
				(-0.29)	(-0.32)
Advisers Dummy				0.105	0.102
				(0.94)	(0.92)
Experts Dummy				0.038	0.028
				(0.41)	(0.30)
GitHub Dummy					0.036
					(0.36)
Pre-sale Dummy					0.128
					(1.09)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.042	0.044	0.093	0.100	0.102
N	1393	1221	627	627	627

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$ln\,(Hard\ Cap_j - Soft\ Cap) = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j.$$

The dependent variable in each regression is the natural logarithm of the difference between the ICO hard cap and soft cap. Both the hard cap and soft cap are denominated in U.S. dollars. Token retention is expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. In column (1), token retention is set to zero if the company does not report the amount/share of tokes retained. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Soft Cap and Token Retention

	(1)	(2)	(3)	(4)	(5)
Token Retention	0.529***	0.533***	0.732***	0.465*	0.477*
	(3.12)	(2.66)	(2.75)	(1.74)	(1.78)
Vesting Dummy			0.079	-0.077	-0.076
			(0.78)	(-0.74)	(-0.73)
KYC Dummy				0.386***	0.386***
				(3.28)	(3.27)
Whitelist Dummy				0.224*	0.222*
				(1.92)	(1.91)
Venture Funding Dummy				0.599***	0.589***
				(3.29)	(3.18)
Advisers Dummy				0.208*	0.206*
				(1.95)	(1.92)
Experts Dummy				0.078	0.073
				(0.79)	(0.73)
GitHub Dummy					-0.025
					(-0.26)
Pre-sale Dummy					0.096
					(0.82)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.028	0.023	0.060	0.122	0.123
N	1420	1247	639	639	639

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$ln\,(Soft\ Cap) = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j.$$

The dependent variable in each regression is the natural logarithm of the ICO soft cap. The soft cap is denominated in U.S. dollars. Token retention is expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. In column (1), token retention is set to zero if the company does not report the amount/share of tokes retained. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Fundraising Success and Token Retention

	(1)	(2)	(3)	(4)	(5)	(6)
Token Retention	0.239***	0.310***	0.348***	0.351***	0.203***	0.157***
	(6.40)	(7.24)	(6.35)	(6.46)	(3.84)	(3.08)
Vesting Dummy				0.077***	0.072***	0.079***
				(3.36)	(3.23)	(3.62)
KYC Dummy					0.063**	0.063**
					(2.28)	(2.34)
Whitelist Dummy					0.125***	0.114***
					(4.78)	(4.47)
Venture Funding Dummy					0.386***	0.345***
					(13.35)	(11.90)
Advisers Dummy					0.021	0.034*
					(0.98)	(1.67)
Experts Dummy					0.015	0.029
					(0.70)	(1.40)
GitHub Dummy						-0.005
						(-0.26)
Pre-sale Dummy						-0.044**
						(-1.99)
Soft Cap Dummy						-0.163***
						(-7.85)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.117	0.084	0.079	0.087	0.215	0.258
N	2427	1533	1303	1303	1303	1303

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$Fundraising\ Success_j = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j.$$

The dependent variable in each regression is the ratio of ICO total funds raised to hard cap. Both fundraising success and token retention are expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. In column (1), fundraising success is set to zero if the company does not report the amount of funds raised. In column (2), token retention is set to zero if the company does not report the amount/share of tokes retained. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 7: Token Retention and Market Environment

	(1)	(2)	(3)	(4)	(5)	(6)
Token Retention	0.234***	0.317***	0.340***	0.342***	0.193***	0.147***
	(6.23)	(7.31)	(6.15)	(6.24)	(3.63)	(2.87)
ln(Scaled # of ICOs)	-0.079	-0.211*	-0.075	-0.069	-0.017	-0.042
	(-0.73)	(-1.68)	(-0.52)	(-0.50)	(-0.13)	(-0.33)
Token Retention × ln(Scaled # of ICOs)	0.324***	0.240***	0.398***	0.391***	0.316**	0.362***
	(2.91)	(2.67)	(2.97)	(3.04)	(2.57)	(3.09)
Vesting Dummy				0.074***	0.073***	0.080***
				(3.24)	(3.23)	(3.62)
KYC Dummy					0.064**	0.064**
					(2.32)	(2.39)
Whitelist Dummy					0.122***	0.110***
					(4.66)	(4.34)
Venture Funding Dummy					0.390***	0.349***
					(13.34)	(11.94)
Advisers Dummy					0.012	0.026
					(0.58)	(1.27)
Experts Dummy					0.016	0.031
					(0.77)	(1.50)
GitHub Dummy						-0.011
						(-0.56)
Pre-sale Dummy						-0.045**
						(-2.03)
Soft Cap Dummy						-0.161***
						(-7.79)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.097	0.070	0.068	0.076	0.206	0.249
N	2397	1476	1273	1273	1273	1273

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$\begin{aligned} Fundraising\ Success_j = \alpha + \beta_1 \cdot Token\ Retention_j + \beta_2 \cdot ln\,(Scaled\ \#\ of\ ICOs_j) \\ + \beta_3 \cdot Token\ Retention_j \times ln\,(Scaled\ \#\ of\ ICOs_j) + \gamma X_j + \epsilon_j \end{aligned}$$

The dependent variable in each regression is the ratio of ICO total funds raised to hard cap. Both fundraising success and token retention are expressed in percentage points. All the variables besides dummies are demeaned. Only ICOs that started on or after June 1, 2017, and were completed on or before December 31, 2018, are included. In column (1), fundraising success is set to zero if the company does not report the amount of funds raised. In column (2), token retention is set to zero if the company does not report the amount/share of tokes retained. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 8: Post-ICO Market-Based Outcomes and Token Retention

	Listed Dummy		Ln(Trading Volume)		Ln(Market Capitalization)	
	(1)	(2)	(3)	(4)	(5)	(6)
Token Retention	0.145***	0.141**	1.599***	0.932*	1.429***	0.612*
	(4.53)	(2.34)	(2.96)	(1.87)	(3.61)	(1.68)
Successful Dummy		0.190***		0.725		0.469
		(5.97)		(1.65)		(1.47)
Vesting Dummy		0.032		-0.284		-0.243*
		(1.23)		(-1.45)		(-1.69)
KYC Dummy		0.129***		0.971***		0.791***
		(4.36)		(3.86)		(4.25)
Whitelist Dummy		0.092***		0.271		0.214
		(3.12)		(1.08)		(1.18)
Venture Funding Dummy		0.230***		2.012***		1.299***
		(5.47)		(8.88)		(8.33)
Advisers Dummy		0.074***		-0.262		-0.167
		(2.97)		(-1.33)		(-1.13)
Experts Dummy		0.059**		0.612***		0.144
		(2.31)		(3.28)		(1.09)
GitHub Dummy		0.111***		0.155		0.070
		(4.58)		(0.79)		(0.51)
Pre-sale Dummy		-0.030		-0.362		-0.457***
		(-1.12)		(-1.62)		(-2.98)
Soft Cap Dummy		-0.056**		-0.190		-0.249
		(-1.96)		(-0.88)		(-1.62)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.145	0.311	0.114	0.315	0.248	0.418
N	3210	1303	635	529	575	492

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$y_j = \alpha + \beta_1 \cdot Token\ Retention_j + \gamma X_j + \epsilon_j$$

In columns (1) and (2), the dependent variable is a dummy variable that equals 1 if the tokens of ICO j are listed on an exchange. In columns (3) and (4), the dependent variable is the natural logarithm of ICO j's average daily trading volume between its first listing date and December 31, 2018. In columns (5) and (6), the dependent variable is the natural logarithm of ICO j's average market capitalization between its first listing date and December 31, 2018. Token retention is expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 9: Post-ICO Real Outcomes and Token Retention

	Working Website		Product Live		Product Apple	
	(1)	(2)	(3)	(4)	(5)	(6)
Token Retention	0.244***	0.175***	0.117*	0.131*	0.106***	0.103**
	(4.53)	(3.18)	(1.90)	(1.89)	(2.67)	(2.26)
Successful Dummy		0.066*		0.141***		0.075***
		(1.84)		(3.79)		(3.47)
Vesting Dummy		0.077***		0.014		0.013
		(3.43)		(0.47)		(0.63)
KYC Dummy		0.028		-0.025		0.006
		(0.92)		(-0.75)		(0.29)
Whitelist Dummy		0.052*		0.021		0.030
		(1.77)		(0.64)		(1.34)
Venture Funding Dummy		0.096***		0.003		-0.010
		(3.38)		(0.07)		(-0.26)
Advisers Dummy		0.252***		0.083***		0.017
		(10.77)		(3.02)		(0.91)
Experts Dummy		0.045*		0.035		-0.026
		(1.94)		(1.23)		(-1.36)
GitHub Dummy		0.061***		0.024		-0.017
		(2.73)		(0.90)		(-0.90)
Pre-sale Dummy		-0.003		0.012		0.020
		(-0.14)		(0.42)		(0.95)
Soft Cap Dummy		0.007		0.057*		0.004
		(0.28)		(1.83)		(0.19)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.029	0.174	0.031	0.067	0.035	0.057
N	1501	1303	1501	1303	1501	1303

This table reports the estimated coefficients from cross-sectional regressions using OLS :

$$y_j = \alpha + \beta_1 \cdot Token\ Retention_j + \gamma X_j + \epsilon_j$$

The dependent variable is a dummy variable that equals 1 if the ICO issuer j as of June 2019 (i) has an active and working website in columns (1) and (2), (ii) has an associated live product or platform in columns (3) and (4), or (iii) has an associated application available for download on the Apple Store in columns (5) and (6). Token retention is expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 10: **Token Retention and Financial Development**

	(1)	(2)	(3)	(4)	(5)	(6)
FD Index	0.052***	0.102***	0.114***	0.088***	0.077***	0.098**
	(2.79)	(5.98)	(4.60)	(3.53)	(3.05)	(2.32)
Vesting Dummy			-0.000	-0.004	-0.004	-0.004
			(-0.00)	(-0.37)	(-0.35)	(-0.31)
KYC Dummy				0.006	0.006	0.006
				(0.36)	(0.42)	(0.41)
Whitelist Dummy				0.061***	0.060***	0.061***
				(4.24)	(4.19)	(4.20)
Venture Funding Dummy				0.071***	0.060***	0.060***
				(3.28)	(2.73)	(2.72)
Advisers Dummy				0.009	0.013	0.013
				(0.73)	(1.11)	(1.09)
Experts Dummy				-0.004	0.001	0.000
				(-0.32)	(0.04)	(0.02)
GitHub Dummy					0.015	0.015
					(1.36)	(1.35)
Pre-sale Dummy					-0.034***	-0.035***
					(-2.73)	(-2.77)
Soft Cap Dummy					-0.022*	-0.023*
					(-1.84)	(-1.89)
Hard Cap Dummy					-0.007	-0.007
					(-0.35)	(-0.36)
ln(GDP Per Capita)						-0.007
						(-0.70)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.043	0.037	0.056	0.088	0.097	0.097
N	3751	2849	1336	1336	1336	1334

This table reports the estimated coefficients from the OLS regressions:

$$Token\ Retention_j = \alpha + \beta \cdot Financial\ Development\ Index_j + \delta X_j + \varepsilon_j.$$

The dependent variable in each regression is token retention expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. In column (1), token retention is set to zero if the company does not report the amount/share of tokes retained. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

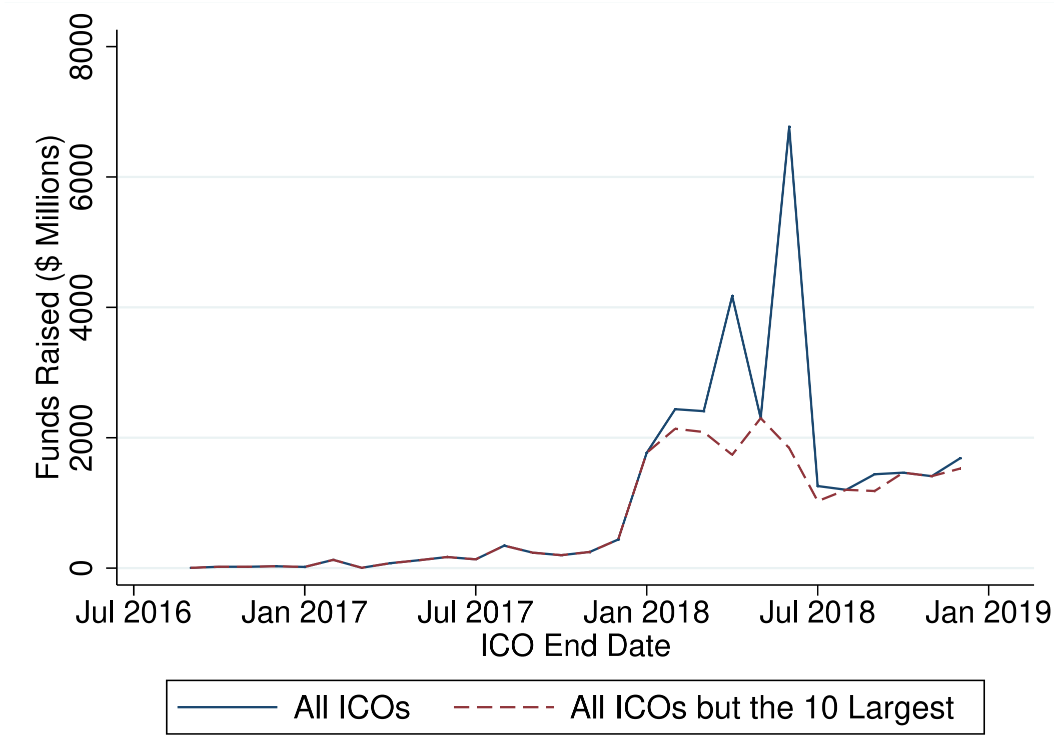
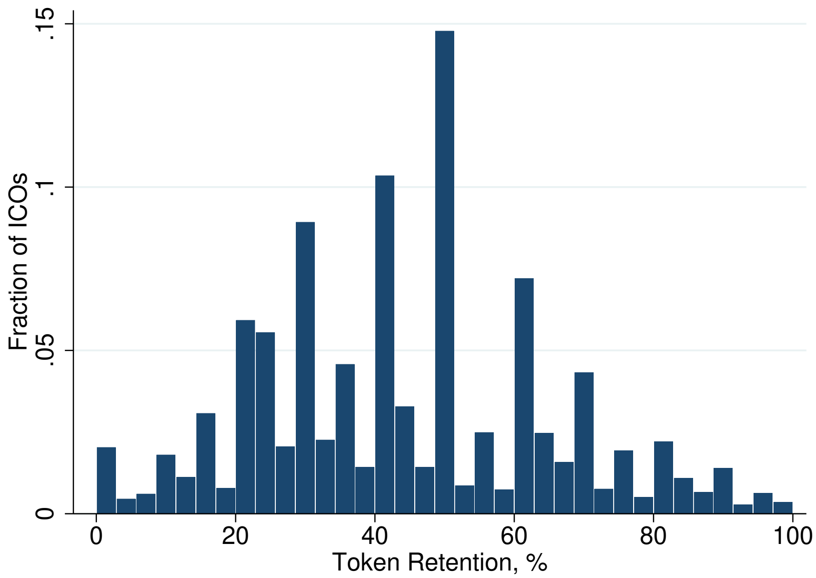
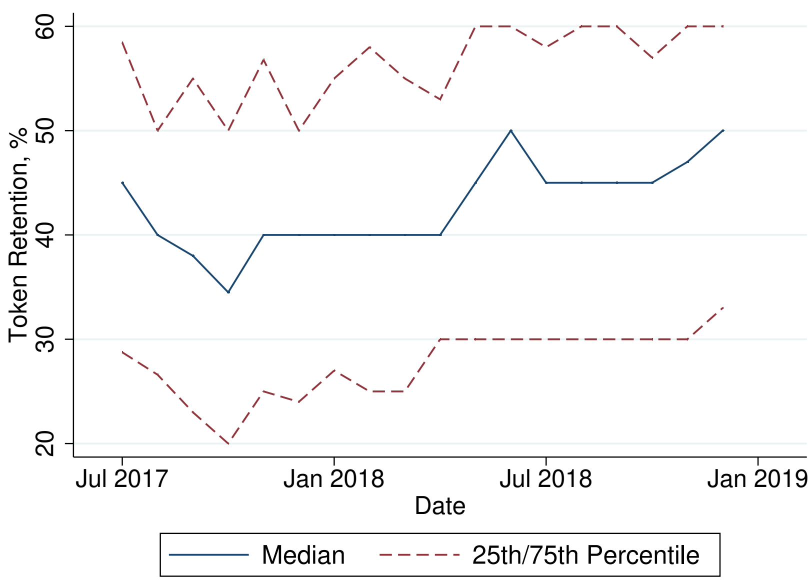


Figure 1: Total Funds Raised Over Time

This figure depicts the aggregate capital raised in the ICO market from July 2016 through December 2018. Raised funds are denominated in U.S. dollars.



(a) Distribution of ICO Token Retention



(b) ICO Token Retention Over Time

Figure 2: ICO Token Retention

Panel (a) depicts the distribution of ICO token retention. Panel (b) plots the cross-sectional median, as well as the 25^{th} and 75^{th} percentiles of ICO token retention. Only ICOs that were completed before or on December 31, 2018, are included.

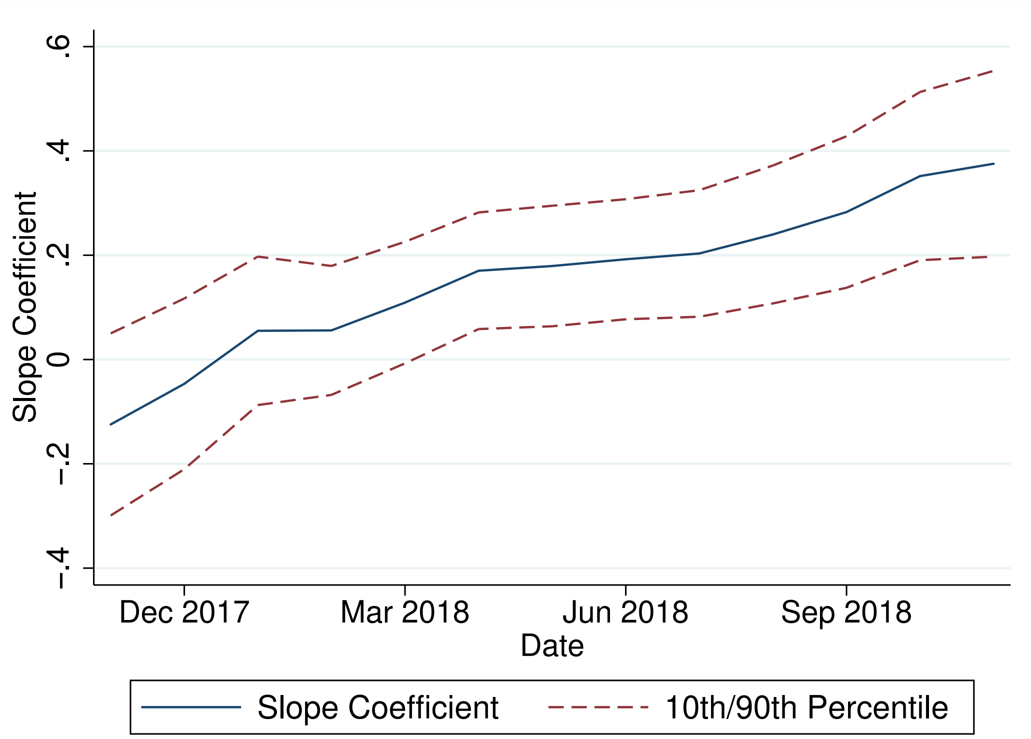


Figure 3: Fundraising Success and Token Retention Over Time

This figure plots the estimates of the slope coefficient, β, from the following 6-month rolling-window regression using OLS:

$$Fundraising\ Success_j = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j.$$

We include the full set of controls as in Table 3 as well as time fixed effects.

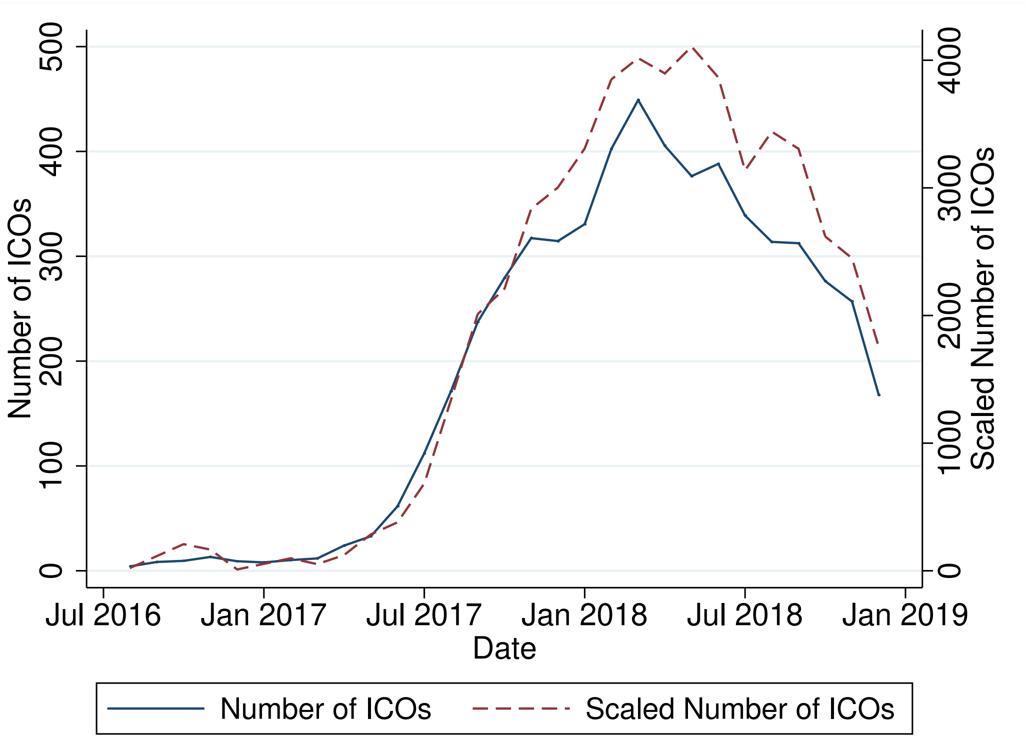


Figure 4: Market Crowdedness

This figure depicts the cross-sectional average of the number of ICOs and the number of ICOs scaled by the average number of words in the whitepaper over time. For each ICO j, we count the number of ICOs that are active within the 15 days from its launch date. The scaled number of ICOs is expressed in thousands.

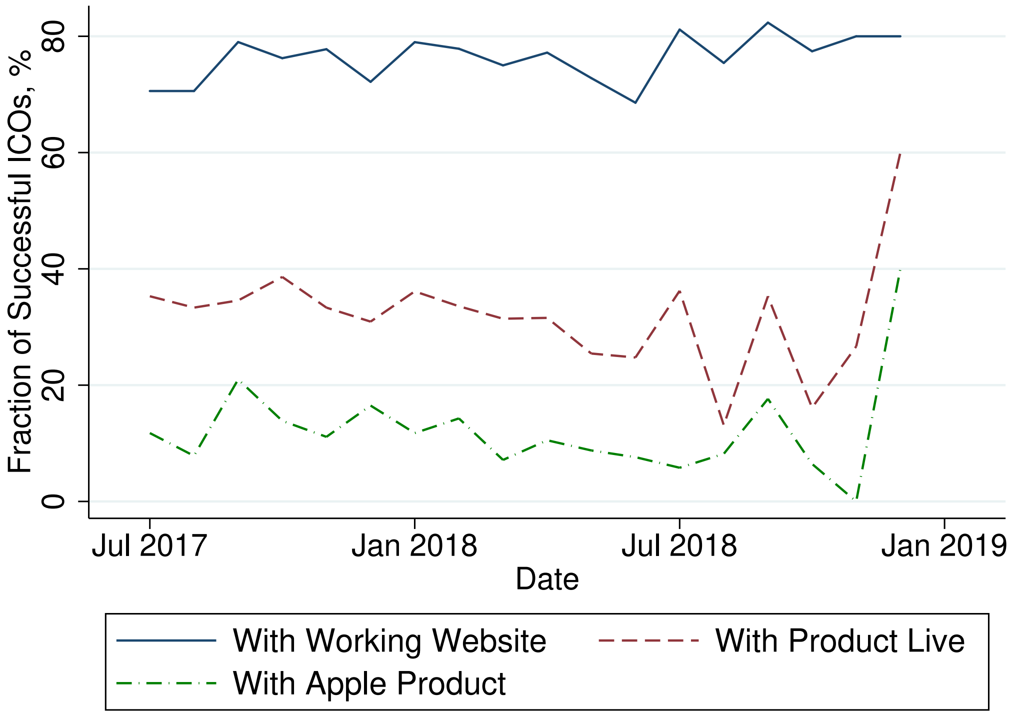
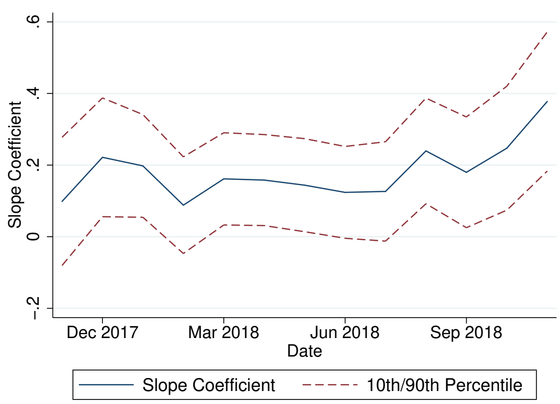
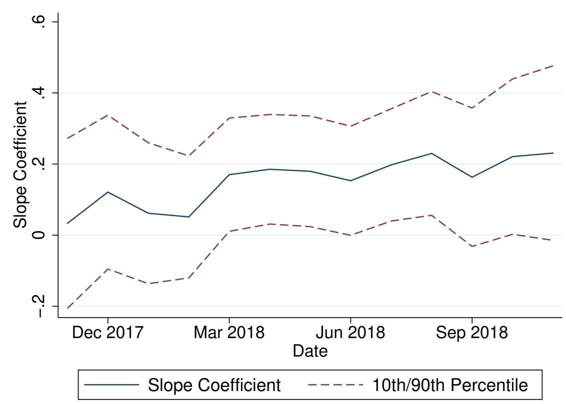


Figure 5: Fraction of Successful ICOs Over Time

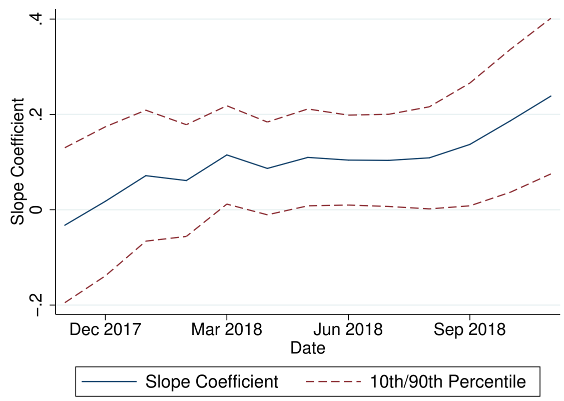
This figure depicts the fraction of ICOs with (i) an active and working website, (ii) an associated live product or platform, and (iii) an associated application available for download on the Apple Store as of June 2019. Only ICOs in our analysis sample are included, i.e., ICOs which were completed by December 31, 2018, and which report funds raised and token retention. The fractions are expressed in percentages.



(a) Working Website



(b) Product Live



(c) Product Apple

Figure 6: Post-ICO Real Outcomes and Token Retention Over Time

This figure plots the estimates of the slope coefficient, β, from the following 6-month rolling-window regression using OLS:

$$y_j = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j.$$

The dependent variable is a dummy variable that equals 1 if the ICO issuer j as of June 2019 (i) has an active and working website in Panel (a), (ii) has an associated live product or platform in Panel (b), or (iii) has an associated application available for download on the Apple Store in Panel (c). We include the full set of controls as in Table 3 as well as time fixed effects.

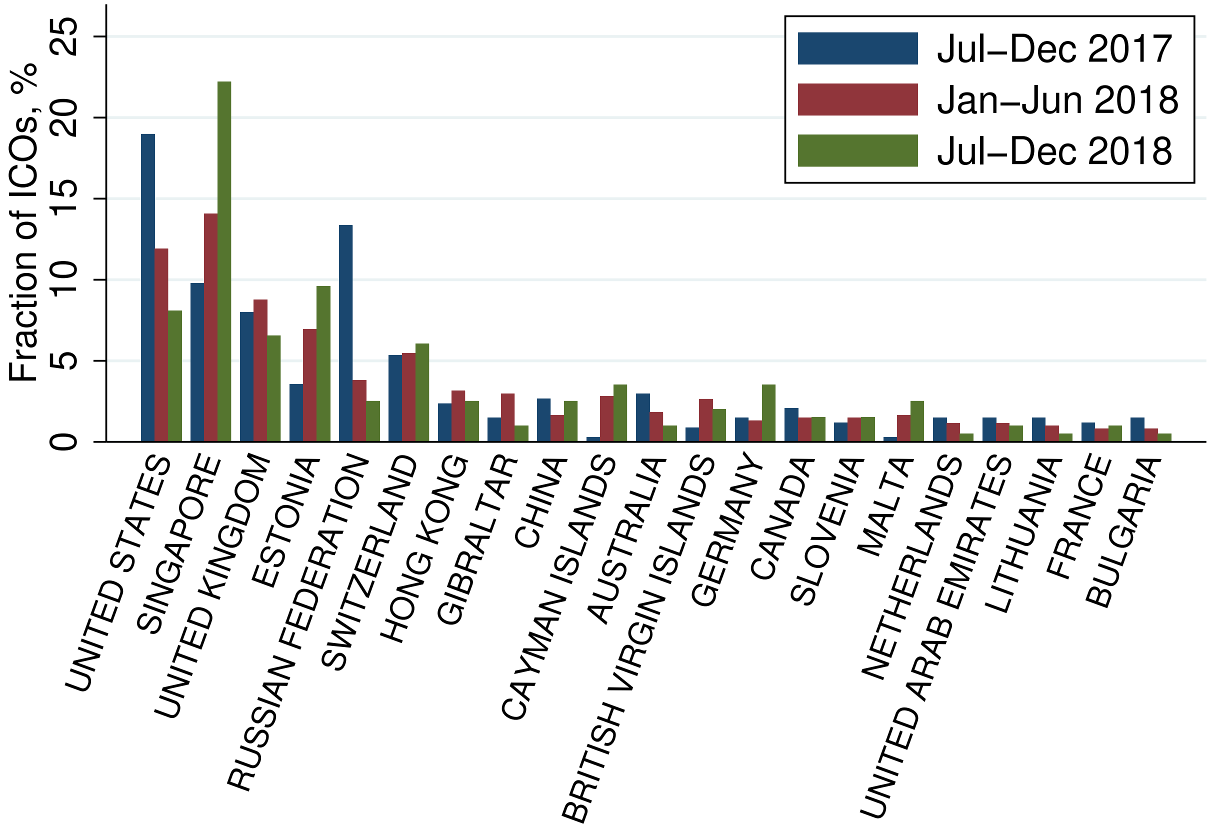
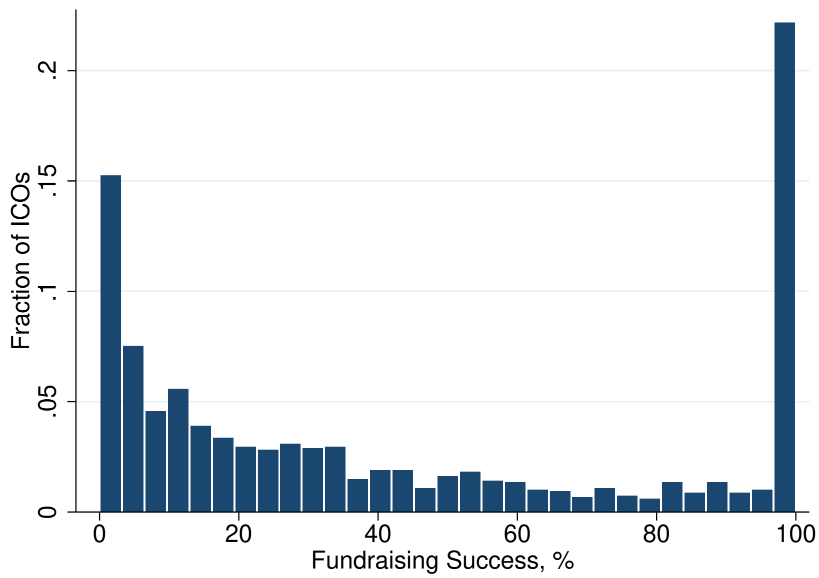
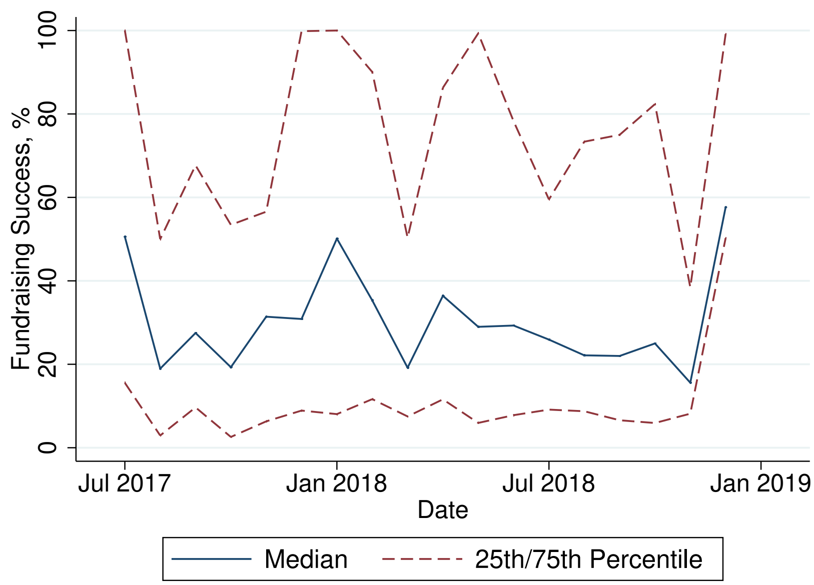


Figure 7: Number of ICOs by Country over Time

This figure depicts the distribution of the number of ICOs by ICO team location and ICO beginning date of the ICO. Only ICOs that were completed before or on December 31, 2018, are included.



(a) Distribution of ICO Fundraising Success



(b) ICO Fundraising Success Over Time

Figure 8: ICO Fundraising Success

Panel (a) depicts the distribution of ICO funds raised as a fraction of hard cap. Panel (b) plots the cross-sectional median, as well as the 25^{th} and 75^{th} percentiles of ICO funds raised as a fraction of hard cap. Only ICOs that were completed before or on December 31, 2018, are included. ICOs with nonreported funds raised are omitted.

Appendix

A Theoretical Proofs and Model Extensions

A.1 Proofs of Benchmark Model

Proof of Proposition 1. For a fraction $(1-s_i)$ of tokens sold by an entrepreneur of type $i \in \{L, H\}$, investors will pay $\pi_i R(1-s_i)$. An entrepreneur of type i choose $s \in [0,1]$ to solve,

$$\underset{s\in[0,1]}{max}\ (1-s)\pi_i R + s\pi_i R - \gamma C(s). \tag{A.1}$$

We can rewrite this as

$$\underset{s\in[0,1]}{max}\ \pi_i R - \gamma C(s).$$

This is always maximized when $s = 0$. With no asymmetric information, high and low-type entrepreneurs will choose $s_H = s_L = 1$.

□

Proof of Proposition 2. In the equilibrium with asymmetric information, if an entrepreneur's type is revealed, his maximization problem is the same as in the full information case given by (A.1). In this case, both high- and low-type entrepreneurs whose type is revealed will select $s_H^{R*} = s_L^{R*} = 0$. Investors will buy all the tokens at a price of $\pi_i R$ where $i \in \{H, L\}$.

We now show that a separating equilibrium in which a high-type entrepreneur who doe not have his type revealed chooses $s_H^{U*} > 0$ exists and is unique. Consider an equilibrium in which $s_H^{U*} > 0$ is given by the implicit solution to the following equation,

$$(1-s_H^{U*})R(\pi_H - \pi_L) = \gamma C(s_H^{U*}) \tag{A.2}$$

The RHS is increasing in s_H^{U*}. The LHS is decreasing in s_H^{U*}. Therefore there can be only one s_H^{U*} for which the two sides are equal. Such an equilibrium always exists as for $s_H^{U*} \in [0,1]$ the

LHS takes values between 0 and $R(\pi_H - \pi_L)$ while the right hand side takes values between 0 and infinity. Therefore they must intersect for some $s_H^{U*} \in [0,1]$.

The following investor beliefs can support such an equilibrium,

$$\mu(H|s) = \begin{cases} 1 & \text{if} \quad s \geq s_H^{U*} \\ \\ 0 & \text{if} \quad s < s_H^{U*} \end{cases}$$

First, we show that given these beliefs, a low-type entrepreneur whose type has not been revealed will choose $s_L^{U*} = 0$. The low-type entrepreneur chooses $s_L^U \in [0,1]$ to solve,

$$\underset{s \in [0,1]}{max}\, s\pi_H R \mathbb{1}_{s \geq s_H^{U*}} + s\pi_L R \mathbb{1}_{s < s_H^{U*}} + (1-s)\pi_L R - C(s)$$

If a low-type entrepreneur chooses an $s \in [0, s_H^{U*})$, it is a dominant strategy to choose $s = 0$. Similarly, if a low-type entrepreneur chooses an $s \in [s_H^{U*}, 1]$, it is a dominant strategy to choose $s = s_H^{U*}$. It follows that a low type-entrepreneur will choose $s_L^{U*} = 0$ iff,

$$(1 - s_H^{U*})\pi_H R + s_H^{U*}\pi_L R - \gamma C(s_H^{U*}) \leq \pi_L R$$

This just holds for s_H^{U*} from (A.2).

A high-type entrepreneur whose type has not been revealed chooses $s_H^U \in [0,1]$ to solve,

$$max_s \in [0,1]\, s\pi_H R \mathbb{1}_{s \geq s_H^{U*}} + s\pi_L R \mathbb{1}_{s < s_H^{U*}} + (1-s)\pi_H R - C(s)$$

Using the same reasoning as for the low-type entrepreneur, if a high-type entrepreneur chooses an $s \in [0, s_H^{U*}]$, it is a dominant strategy to choose $s = 0$. Similarly, if a high-type entrepreneur chooses an $s \in [s_H^{U*}, 1]$, it is a dominant strategy to choose $s = s_H^{U*}$. Then a high type-entrepreneur will choose $s = s_H^{U*}$ iff,

$$(1 - s_H^{U*})\pi_H R + s_H^{U*}\pi_H R - \gamma C(s_H^{U*}) \geq \pi_L R$$

Since $\pi_H > \pi_L$, this is always satisfied for s_H^{U*} from (A.2).

Therefore a high-type entrepreneur whose type has not been revealed will choose to sell s_H^{U*} fraction of tokens to investors. This is the least-cost separating equilibrium since the low-type's incentive compatibility constraint just binds. Under the Intuitive Criterion, this is the unique equilibrium. □

Proof of Proposition 3. The covariance between ICO value and retention is given by,

$$\begin{aligned}
\rho_v &= Cov(pR, s) \\
&= E[pR \cdot s] - E[pR]E[s] \\
&= (1-\alpha)q\pi_H R s_H^{U*} - (q\pi_H R + (1-q)\pi_L R)(1-\alpha)q s_H^{U*} \\
&= (1-\alpha)q s_H^{U*} R(\pi_H - \pi_L)(1-q) \qquad \text{(A.3)}
\end{aligned}$$

Since $\pi_H > \pi_L$, $q, \alpha < 1$, ρ_v is always positive.

□

Proof of Proposition 4. Using (A.3), it is straightforward to show that

$$\frac{\partial \rho_v}{\partial \alpha} < 0.$$

□

A.2 Robustness to the Presence of Scammers

We can additionally show that our main results will hold even with the presence of scammers. In fact, the presence of scammers in the model can strengthen the relationship between retention and ICO value.

We model scammers as follows. We assume a fraction ν of low-type agents are scammers and have $\pi_{sc} = 0$. We further assume that for these agents there is no cost of retention, as they are not holding their wealth in a single asset.

In our model, scammers will simply try to copy whichever agent gets the most revenue, i.e., $(1 - s_i^U)P^U$, at $t = 1$ and whose type is not revealed. Since retention is not costly for scammers, there is no way for the agent they copy to separate themselves.

Note that we are now considering equilibrium strategies in the presence of scammers, in which all agents rationally consider scammers' influence on equilibrium prices. These will therefore differ from the benchmark model without scammers. We focus on the case in which the low-type agents receive the largest payoff at $t = 1$. This is the case when

$$(1 - s_L^{U^*})\pi_L R > (1 - s_H^{U^*})\pi_H R.$$

Our results would also hold if high-type agents whose type is not revealed received the largest payoff at $t = 1$. However, under any reasonable interpretation of the cost of retention coming from risk aversion, the relevant case would be when low-type agents receive a relatively larger total payoff during the ICO than high-type agents, and high-type agents receive a larger payoff at $t = 2$ than low-type agents.[30] Therefore, we focus on showing robustness when scammers pool with the low-type. Our purpose is to show that our main results hold even if scammers are part of the model. This works even if scammers would rather pool with high-types. The key intutition is that as long as there is a limited measure of scammers, our results hold.

When scammers pool with low-types whose type is not revealed, the market now prices this in, and therefore the low-types IC constraint gets modified to

$$(1 - s_L^{U^*})(1 - \nu)\pi_L R + s_L^{U^*}\pi_L R - \gamma C(s_L^{U^*}) \geq (1 - s_H^{U^*})\pi_H R + s_H^{U^*}\pi_L R - \gamma C(s_H^{U^*}).$$

Since the low-type's payoff upon the end of the ICO is now reduced, they will have a greater incentive to copy the high type. Therefore $s_H^{U^*}$ will be higher with scammers than without them to deter the low-type from mimicking the high-type. We can follow the same steps as in the proof of Proposition 2 to show that a separating equilibrium exists in which the high-type will choose $s_H^{U^*}$ until the low-type is just indifferent between mimicking the high-type and revealing himself as the

[30]Retaining more always results in a larger payoff at $t = 2$ than at $t = 1$. If a high-type agent receives more at $t = 1$ than a low type agent, by copying him the low-type agent will receive more at $t = 1$ and at $t = 2$ than by retaining a lower amount. If the cost of retention is coming from risk aversion, the low-type agent would always want to copy the high-type agent in this case.

low-type, i.e.,

$$(1 - s_L^{U^*})(1-\nu)\pi_L R + s_L^{U^*}\pi_L R - \gamma C(s_L^{U^*}) = (1 - s_H^{U^*})\pi_H R + s_H^{U^*}\pi_L R - \gamma C(s_H^{U^*}).$$

The basic relationship between token retention and quality will stay the same as long as the measure of scammers is not too high, and we can follow steps similar to those in Proposition 3 to show that ρ is always positive. The key to preserving the relationship between retention and quality is that there is a limited measure of scammers. If there were an infinite measure of scammers, the entrepreneur's ICO price would be zero, causing him to forego an ICO and retain all tokens.

A.3 Extension to Incorporate Moral Hazard

We can extend the benchmark model to incorporate moral hazard. We assume that a project undertaken by an entrepreneur of type $i \in \{H, L\}$ succeeds with probability $\phi_i e$ (instead of π_i in the benchmark model), where $e \in [0,1]$ is the amount of effort chosen by the entrepreneur. We assume that, for every unit of effort put in, the high-type entrepreneur obtains a larger increase in the probability that the project will be successful ($0 < \phi_L < \phi_H < 1$). The entrepreneurs face a convex cost of exerting effort, $\frac{e^2}{2}$. For ease of exposition, we also assume that the proportion of high- and low-type entrepreneurs is the same.[31] The timeline of events is illustrated in the figure below.

$t = 0$	$t = 1$	$t = 2$
With probability α entrepreneur i's type is revealed, $i \in \{H, L\}$	1. Entrepreneur chooses token share s_i to retain. 2. Entrepreneur chooses effort e_i.	Payoffs are realized: With probability $\phi_i e_i$ project is successful

Entrepreneur's problem. After the ICO has taken place, entrepreneurs choose how much effort to exert, conditional on token retention. Specifically, an entrepreneur who has retained a fraction of tokens s_i^j chooses a level of effort to maximize,

[31] There may be a concern that the proportion of high-type entrepreneurs in this market is very low. All our results hold if the proportion of high-type entrepreneurs is less than half.

$$\max_{e\in[0,1]} s_i^j \phi_i e R - \frac{e^2}{2}.$$

Before the ICO takes place, entrepreneurs choose how many tokens to retain, taking into account how the token retention will affect their choice of how much effort to exert. In particular, an entrepreneur of type $\{i, j\}$ solves the following problem:

$$\max_{s\in[0,1]} (1-s)P^j(s) + s\phi_i e(s)R - \frac{e(s)^2}{2} - \gamma C(s).$$

We assume that $R\phi_H < 1$ to ensure an interior solution for the optimal effort level.

Equilibrium Definition: A perfect Bayesian equilibrium in pure strategies of the model is given by

1. The share of tokens, $s_H^{R^*}$, $s_L^{R^*}$, $s_H^{U^*}$, and $s_L^{U^*}$, retained by each type of entrepreneur.
2. The effort levels, $e_H^{R^*}$, $e_L^{R^*}$, $e_H^{U^*}$, and $e_L^{U^*}$, chosen by each type of entrepreneur.
3. Investor beliefs, $\mu(i|s)$, that an entrepreneur is of type $i = \{H, L\}$, if he chooses to retain share s of tokens and his type has not been revealed.
4. The price investors pay per share of tokens sold to them such that their participation constraint is satisfied:

$$P^U(s) = \mu(H|s)\phi_H e_H^{U^*} R + (1-\mu(H|s))\phi_L e_L^{U^*} R,$$
$$P^R(s) = \phi_i e_i^{R^*} R.$$

A.3.1 Model Equilibrium and Empirical Predictions

In this extended model, all the propositions in the benchmark model continue to hold. The key difference in predictions, relative to the benchmark, is that low-type entrepreneurs and high-type entrepreneurs whose type has been revealed will also retain tokens in equilibrium. Conditional on retention, s, the optimal choice of effort chosen by an entrepreneur of type i is $s\phi_i R$. Since a high-type entrepreneur gets a larger increase in the probability of success per unit of effort, he chooses a higher effort level than a low-type entrepreneur, conditional on retention. Therefore, high-type entrepreneurs whose type is revealed and who do not face an adverse selection problem retain more in equilibrium than low-type entrepreneurs. The adverse selection problem faced by high-type entrepreneurs whose type has not been revealed can further exacerbate the moral hazard problem. These entrepreneurs may have to retain even more to signal their quality to investors. Since retention is costly, it is a credible signal of quality. Under the Intuitive Criterion, we obtain a unique equilibrium.

Proposition 5 *[Equilibrium with Asymmetric Information] Under the Intuitive Criterion, a separating equilibrium in which the high-type entrepreneurs retain more tokens than low-type entrepreneurs,* $s_H^{U*} \geq s_H^{R*} > s_L^{U*} = s_L^{R*}$*, exists and is unique.*

Thus, in the presence of asymmetric information, high-type entrepreneurs retain a higher fraction of tokens than low-type entrepreneurs. This creates a positive relationship between token retention and quality in the ICO market. We can establish an analagous proposition to Proposition 3.

Proposition 6 *[Retention and Quality] In equilibrium, there is a positive covariance between the value of an ICO and the fraction of tokens retained, defined as* $\rho \equiv Cov\left(\phi e\left(s^*\right) R, s^*\right)$.

Retention when markets are crowded. In the model equilibrium, high-type entrepreneurs who have their type revealed by the public signal retain a smaller proportion of tokens than those who do not have their type revealed. Since equilibrium effort is increasing in retention, high-type entrepreneurs who do not have their type revealed are the most likely to produce a successful platform. The total fraction of tokens retained in equilibrium is given by

$$\frac{1}{2}(1-\alpha)s_H^{U^*} + \frac{1}{2}\alpha s_H^{R^*} + \frac{1}{2}s_L^*.$$

As α decreases, a greater fraction of high-type entrepreneurs retain a large number of tokens, subsequently exerting more effort and producing higher quality products. We can establish the following proposition analogous to Proposition 4.

Proposition 7 *[Retention and Market Crowdedness] In the unique signaling equilibrium, ICO value covaries positively with the fraction of tokens retained,* $\rho > 0$*. As* α *decreases and the quality of public information deteriorates, the magnitude of this covariance increases,*

$$\frac{\partial \rho}{\partial \alpha} \leq 0.$$

A.3.2 Proofs of the Extended Model

Proof of Proposition 5. We start by solving for the optimal choice of effort conditional on retention. An entrepreneur of type $i \in \{H, L\}$ who has retained s will choose effort level e to maximize,

$$max_{e \in [0,1]} \; s(\phi_i e)R - \frac{e^2}{2}$$

The first-order condition is

$$s\phi_i R - e = 0.$$

The optimal effort is given by

$$e^* = s\phi_i R.$$

Therefore, conditional on retention, a high type chooses greater effort given that $\phi_H > \phi_L$.

Next, we solve for the equilibrium retention of an entrepreneur whose type is revealed at $t = 0$. If entrepreneurs have had their type revealed, a type $i \in \{H, L\}$ entrepreneur will choose retention s_i^R to maximize,

$$max_{s_i^R \in [0,1]} \; (1 - s_i^R)s_i^R \phi_i^2 R^2 + s_i^R s_i^R \phi_i^2 R^2 - \frac{(s_i^R \phi_i R)^2}{2} - \gamma C(s_i^R).$$

The first-order condition is

$$\phi_i^2 R^2 - \phi_i^2 R^2 s_i^R - \gamma C'(s_i^R) = 0.$$

We can solve for optimal retention,

$$s_i^{R*} = 1 - \frac{\gamma}{\phi_i^2 R^2} C'(s_i^{R*}).$$

Because s_i^{R*} is increasing in ϕ_i, a high-type entrepreneur whose type has been revealed therefore will retain more than a low-type entrepreneur, $s_H^{R*} > s_L^{R*}$.

Now, consider the problem of an entrepreneur who has not had his type revealed. Adverse selection will exacerbate the moral hazard problem. We first show that a separating equilibrium exists in which the high-type retains a fraction of tokens $s_H^{U^*} \geq s_H^{R^*}$ to signal his type while the low type chooses $s_L^{U^*} = s_L^{R^*}$. The following investor beliefs can support such an equilibrium,

$$\mu(H|s) = \begin{cases} 1 & \text{if} \quad s \geq s_H^{U^*} \\ 0 & \text{if} \quad s < s_H^{U^*} \end{cases}.$$

There are two cases.

Case I:

$$(1 - s_H^{R^*}) s_H^{R^*} \phi_H^2 R^2 + \frac{(s_H^{R^*} \phi_L R)^2}{2} - \gamma C(s_H^{R^*}) \leq (1 - s_L^{R^*}) s_L^{R^*} \phi_L^2 R^2 + \frac{(s_L^{R^*} \phi_L R)^2}{2} - \gamma C(s_L^{R^*})$$

Note that in a separating equilibrium, the low-type's utility is maximized when he holds $s_L^{R^*}$. If the low-type agent does not have an incentive to copy the high-type when he holds $s_H^{R^*}$, then in the least-cost separating equilibrium $s_H^{U^*} = s_H^{R^*}$ and $s_L^{U^*} = s_L^{R^*}$. If the above inequality does not hold at $s_H^{U^*} = s_H^{R^*}$, then the adverse selection problem will affect equilibrium choices. We solve this case below.

Case 2:

$$(1 - s_H^{R^*}) s_H^{R^*} \phi_H^2 R^2 + \frac{(s_H^{R^*} \phi_L R)^2}{2} - \gamma C(s_H^{R^*}) > (1 - s_L^{R^*}) s_L^{R^*} \phi_L^2 R^2 + \frac{(s_L^{R^*} \phi_L R)^2}{2} - \gamma C(s_L^{R^*})$$

Since this is a fully separating equilibrium in which entrepreneurs are exerting effort, the prices investors are willing to pay per fraction of tokens sold to them conditional on retention revealing the entrepreneur type are the same as those when they know their types. We define U_i as the utility functions for type i when their types are revealed, i.e.,

$$U_i(s) = (1 - s) s \phi_i^2 R^2 + s^2 \phi_i^2 R^2 - \frac{(s\phi_i R)^2}{2} - \gamma C(s).$$

Finally, we define U_{LH} as the low-type's utility from mimicking the high-type,

$$U_{LH}(s) = (1 - s) s \phi_H^2 R^2 + s^2 \phi_L^2 R^2 - \frac{(s\phi_L R)^2}{2} - \gamma C(s).$$

Then, the low-type will have an incentive to copy the high-type if,

$$U_{LH}(s_H^U) - U_L(s_L^U) > 0. \tag{A.4}$$

In the least-cost separating equilibrium, the high-type will want to increase retention $U_{LH}(s_H^U) = U_L(s_L^U)$, and the low-type is just indifferent between revealing his type by retaining s_L^U and copying the high-type by retaining s_H^U. We show this is possible if the high-type increases s_H. This requires us to show that the left-hand side of (A.4) is decreasing in retention. The proof is below.

Consider any $s_H > s_H^{R^*}$. U_H is maximized at $s_H^{R^*}$. Therefore, $U_H(s_H) < U_H(s_H^{R^*})$, i.e.,

$$(1-s_H)s_H\phi_H^2R^2 + \frac{s_H^2\phi_H^2R^2}{2} - \gamma C(s_H) < (1-s_H^{R^*})s_H^{R^*}\phi_H^2R^2 + \frac{{s_H^{R^*}}^2\phi_H^2R^2}{2} - \gamma C(s_H^{R*}).$$

Since $s_H > s_H^{R^*}$, $\frac{{s_H}^2\phi_H^2R^2}{2} > \frac{{s_H^{R^*}}^2\phi_H^2R^2}{2}$. Therefore, it must be the case that

$$(1-s_H^{R^*})s_H^{R^*}\phi_H^2R^2 - \gamma C(s_H^{R^*}) - \big((1-s_H)s_H\phi_H^2R^2 - \gamma C(s_H)\big) > \frac{{s_H}^2\phi_H^2R^2}{2} - \frac{{s_H^{R^*}}^2\phi_H^2R^2}{2} > 0.$$

Since $\phi_H > \phi_L$, this implies

$$(1-s_H^{R^*})s_H^{R^*}\phi_H^2R^2 - \gamma C(s_H^{R^*}) - \big((1-s_H)s_H\phi_H^2R^2 - \gamma C(s_H)\big) > \frac{{s_H}^2\phi_L^2R^2}{2} - \frac{{s_H^{R^*}}^2\phi_L^2R^2}{2} > 0.$$

Rearranging the above,

$$(1-s_H^{R^*})s_H^{R^*}\phi_H^2R^2 - \gamma C(s_H^{R^*}) + \frac{{s_H^{R^*}}^2\phi_L^2R^2}{2} > \frac{{s_H}^2\phi_L^2R^2}{2} + (1-s_H)s_H\phi_H^2R^2 - \gamma C(s_H).$$

Therefore,

$$U_{LH}(s_H^{R^*}) > U_{LH}(s_H).$$

Subtracting $U_L(s_L^U)$ from both sides,

$$U_{LH}(s_H^{R^*}) - U_L(s_L^U) > U_{LH}(s_H) - U_L(s_L^U).$$

Therefore, increasing s_H will decrease the low-type's incentive to copy the high-type. In a separating equilibrium the low-type whose type is not revealed will optimally choose to hold $s_L^{U^*} = s_L^{R^*}$. In the least-cost separating equilibrium, the high-type will increase retention until the low-type's incentive compatibility constraint just binds, i.e.,

$$U_{LH}(s_H^{U^*}) - U_L(s_L^{U^*}) = 0.$$

For this to be an equilibrium, we also require that the high-type's incentive constraint be satisfied, i.e.,

$$U_H(s_H^{U^*}) \geq (1 - s_L^{U^*})s_L^{U^*}\phi_L^2R^2 + \frac{s_L^{U^*2}\phi_H^2R^2}{2} - \gamma C(s_L^{U^*}).$$

Since $U_L(s_L^{U^*}) = U_{LH}(s_H^{U^*})$, this implies that

$$(1 - s_L^{U^*})s_L^{U^*}\phi_L^2R^2 + \frac{s_L^{U^*2}\phi_L^2R^2}{2} - \gamma C(s_L^{U^*}) = (1 - s_H^{U^*})s_H^{U^*}\phi_H^2R^2 + \frac{s_H^{U^*2}\phi_L^2R^2}{2} - \gamma C(s_H^{U^*}).$$

Rearranging the above, we get

$$\frac{s_H^{U^*2}\phi_L^2R^2}{2} - \frac{s_L^{U^*2}\phi_L^2R^2}{2} = (1 - s_L^{U^*})s_L^{U^*}\phi_L^2R^2 - \gamma C(s_L^{U^*}) - (1 - s_H^{U^*})s_H^{U^*}\phi_H^2R^2 + \gamma C(s_H^{U^*}).$$

Since $\phi_H > \phi_L$ and the right-hand side of the above equation is positive, this implies

$$\frac{s_H^{U^*2}\phi_H^2R^2}{2} - \frac{s_L^{U^*2}\phi_H^2R^2}{2} > (1 - s_L^{U^*})s_L^{U^*}\phi_L^2R^2 - \gamma C(s_L^{U^*}) - (1 - s_H^{U^*})s_H^{U^*}\phi_H^2R^2 + \gamma C(s_H^{U^*}).$$

Rearranging,

$$(1 - s_H^{U^*})s_H^{U^*}\phi_H^2R^2 + \frac{s_H^{U^*2}\phi_H^2R^2}{2} - \gamma C(s_H^{U^*}) > (1 - s_L^{U^*})s_L^{U^*}\phi_L^2R^2 + \frac{s_L^{U^*2}\phi_H^2R^2}{2} - \gamma C(s_L^{U^*}).$$

Therefore, the high-type's incentive compatibility constraint always holds in the separating equilibrium.

Under the Intuitive Criterion, the least-cost separating equilibrium described above is the unique equilibrium of the model. □

Proof of Proposition 6. We prove this proposition for both cases in the proof of Proposition 5.

Case 1: In Case 1, $s_H^{U^*} = s_H^{R^*}$ and $s_L^{U^*} = s_L^{R^*}$. Define these as s_H^* and s_L^*, respectively, dropping the identifiers for whether the type was revealed or not. Then,

$$\begin{aligned}\rho &\equiv Cov(\phi e(s^*)R, s^*) \\ &= \frac{1}{4}R(\phi_H e(s_H^*) - \phi_L e(s_L^*))(s_H^* - s_L^*).\end{aligned}$$

The value of this expression is always positive given that $\phi_H > \phi_L$ (assumption), $s_H^* > s_L^*$ (Proposition 5), and $e(s_H^*) > e(s_L^*)$ (see proof of Proposition 5).

Case 2: In Case 2, since $s_L^{U^*} = s_L^{R^*}$, we will drop the identifier of whether the type has been revealed. However, since the high type's retention differs depending on whether his type is revealed or not, we will continue to use $s_H^{U^*}$ and $s_H^{R^*}$ when describing equilibrium retention. Then,

$$
\begin{aligned}
\rho \equiv & Cov(\phi e(s^*)R, s^*) \\
= & \frac{1}{4}R\Big(\alpha(\phi_H e(s_H^{R^*}) - \phi_L e(s_L^*))(s_H^{R^*} - s_L^*) + (1-\alpha)(\phi_H e(s_H^{U^*}) - \phi_L e(s_L^*))(s_H^{U^*} - s_L^*) \\
& + \alpha(1-\alpha)\phi_H(e(s_H^{U^*}) - e(s_H^{R^*}))(s_H^{U^*} - s_H^{R^*})\Big).
\end{aligned}
$$

Since $s_H^{U^*} > s_H^{R^*} > s_L^*$ and effort is increasing in s, the value of this expression is always positive. $\square$

Proof of Proposition 7. In Case 1, retention does not change with α, since $s_H^{R^*} = s_H^{U^*}$. Therefore, $\frac{\partial\rho}{\partial\alpha} = 0$. We prove below that in Case 2, $\frac{\partial\rho}{\partial\alpha} < 0$.

$$
\begin{aligned}
\rho = & R\Bigg(\frac{1}{2}\alpha\phi_H e(s_H^{R^*})s_H^{R^*} + \frac{1}{2}(1-\alpha)\phi_H e(s_H^{U^*})s_H^{U^*} + \frac{1}{2}\phi_L e(s_L^*)s_L^* \\
& -(\frac{1}{2}\alpha\phi_H e(s_H^{R^*}) + \frac{1}{2}(1-\alpha)\phi_H e(s_H^{U^*}) + \frac{1}{2}\phi_L e(s_L))\left(\frac{1}{2}\alpha s_H^{R^*} + \frac{1}{2}(1-\alpha)s_H^{U^*} + \frac{1}{2}s_L\right)\Bigg) \\
= & R\Bigg(-\alpha\frac{1}{2}\phi_H(e(s_H^{U^*})s_H^{U^*} - e(s_H^{R^*})s_H^{R^*}) + \frac{1}{2}\phi_H e(s_H^{U^*})s_H^{U^*} + \frac{1}{2}\phi_L e(s_L^*)s_L^* \\
& -(-\alpha\frac{1}{2}(s_H^{U^*} - s_H^{R^*}) + \frac{1}{2}s_H^{U^*} + \frac{1}{2}s_L^*)(-\alpha\frac{1}{2}\phi_H(e(s_H^{U^*}) - e(s_H^{R^*})) + \frac{1}{2}\phi_H e(s_H^{U^*}) + \frac{1}{2}\phi_L e(s_L^*))\Bigg).
\end{aligned}
$$

Define $C = \frac{1}{2}\phi_H e(s_H^{U^*})s_H^{U^*} + \frac{1}{2}\phi_L e(s_L)s_L$, $A = \frac{1}{2}s_H^{U^*} + \frac{1}{2}s_L$, $B = \frac{1}{2}\phi_H e(s_H^{U^*}) + \frac{1}{2}\phi_L e(s_L^*)$. Then we can rewrite the above equation as

$$
\begin{aligned}
\rho = & R\Bigg(-\alpha\frac{1}{2}\phi_H(e(s_H^{U^*})s_H^{U^*} - e(s_H^{R^*})s_H^{R^*}) + C - \frac{\alpha^2}{4}\phi_H(s_H^{U^*} - s_H^{R^*})(e(s_H^{U^*}) - e(s_H^{R^*})) \\
& + \alpha\frac{1}{2}(s_H^{U^*} - s_H^{R^*})B + \frac{\alpha}{2}\phi_H(e(s_H^{U^*}) - e(s_H^{R^*}))A - AB\Bigg).
\end{aligned}
$$

Taking the derivative with respect to α,

$$
\begin{aligned}
\frac{1}{R}\frac{\partial\rho}{\partial\alpha} = & -\frac{1}{2}\phi_H(e(s_H^{U^*})s_H^{U^*} - e(s_H^{R^*})s_H^{R^*}) - \alpha\frac{1}{2}\phi_H(s_H^{U^*} - s_H^{R^*})(e(s_H^{U^*}) - e(s_H^{R^*})) + \frac{1}{2}(s_H^{U^*} - s_H^{R^*})B \\
& + \frac{1}{2}\phi_H(e(s_H^{U^*}) - e(s_H^{R^*}))A.
\end{aligned}
$$

Define $D = \alpha\frac{1}{2}\phi_H(s_H^{U^*} - s_H^{R^*})(e(s_H^{U^*}) - e(s_H^{R^*}))$. Then we can rewrite the above as

$$
\frac{1}{R}\frac{\partial\rho}{\partial\alpha} = -\frac{1}{2}\phi_H(e(s_H^{U^*})s_H^{U^*} - e(s_H^{R^*})s_H^{R^*}) - D + \frac{1}{2}(s_H^{U^*} - s_H^{R^*})B + \frac{1}{2}\phi_H(e(s_H^{U^*}) - e(s_H^{R^*}))A.
$$

We can write $A = s_H^{U^*} + \frac{1}{2}(s_L - s_H^{U^*})$ and $B = \phi_H e(s_H^{U^*}) + \frac{1}{2}(\phi_L e(s_L^*) - \phi_H e(s_H^{U^*}))$. Then,

$$\frac{1}{R}\frac{\partial \rho}{\partial \alpha} = (e(s_H^{U^*}) - e(s_H^{R^*}))(s_H^{U^*} - s_H^{R^*})(1-\alpha)\phi_H \frac{1}{2} + \frac{1}{2}\phi_H(e(s_H^{U^*}) - e(s_H^{R^*}))\frac{1}{2}(s_L^* - s_H^{U^*})$$
$$+ \frac{1}{2}(s_H^{U^*} - s_H^{R^*})\frac{1}{2}(\phi_L e(s_L^*) - \phi_H e(s_H^{U^*})).$$

Note that $\phi_H e(s_L^*) - \phi_H e(s_H^{U^*}) > \phi_L e(s_L^*) - \phi_H e(s_H^{U^*})$. Therefore with a bit of algebra we can show that,

$$\frac{2}{R\phi_H}\frac{\partial \rho}{\partial \alpha} < -\alpha(e(s_H^{U^*}) - e(s_H^{R^*}))(s_H^{U^*} - s_H^{R^*}) + \frac{1}{2}(e(s_H^{U^*}) - e(s_H^{R^*}))(s_L^* - s_H^{R^*})$$
$$+ (s_H^{U^*} - s_H^{R^*})\frac{1}{2}(e(s_L^*) - e(s_H^{R^*})).$$

Each term on the RHS of the above equation is negative. Therefore, $\frac{\partial \rho}{\partial \alpha} < 0$.

□

B Additional Tables and Figures

Table B.1: Summary Statistics

	N	Mean	SD	1%	10%	50%	90%	99%
Whitepaper Dummy	3210	0.82	0.38	0.00	0.00	1.00	1.00	1.00
GitHub Dummy	3210	0.41	0.49	0.00	0.00	0.00	1.00	1.00
KYC Dummy	3210	0.47	0.50	0.00	0.00	0.00	1.00	1.00
Whitelist Dummy	3210	0.36	0.48	0.00	0.00	0.00	1.00	1.00
Venture Funding Dummy	3210	0.04	0.19	0.00	0.00	0.00	0.00	1.00
Advisers Dummy	3210	0.25	0.43	0.00	0.00	0.00	1.00	1.00
Experts Dummy	3210	0.25	0.43	0.00	0.00	0.00	1.00	1.00
Pre-sale Dummy	3210	0.56	0.50	0.00	0.00	1.00	1.00	1.00
(Funds Raised>0) Dummy	3210	0.47	0.50	0.00	0.00	0.00	1.00	1.00
Funds Raised, $ mil	1507	17.03	120.34	0.00	0.26	5.50	30.00	100.00
Hard Cap, $ mil	2428	51.11	418.03	0.35	4.50	20.00	65.10	338.09
Soft Cap, $ mil	1247	5.35	10.78	0.06	0.50	2.85	11.05	50.00
Token Retention, %	3210	43.32	20.81	0.00	18.62	40.00	70.00	94.30
Fundraising Success, %	1307	47.60	39.27	0.04	2.07	35.17	100.00	100.00
Listed Dummy	3210	0.20	0.40	0.00	0.00	0.00	1.00	1.00
Scam Dummy	3210	0.04	0.19	0.00	0.00	0.00	0.00	1.00

This table reports summary statistics for the sample of ICOs which were completed before or on December 31, 2018, and which report token retention. We obtain the sample of ICOs by combining the data from the following ICO tracking websites: ICO Data, Token Data, CoinMarketCap, Cryptoslate, ICO Bench, ICO Drops, ICO Rating Agency, and ICO Checks. For each variable, the table shows the number of nonmissing observations, along with the cross-sectional mean, standard deviation, 1^{st}, 10^{th}, 50^{th}, 90^{th}, and 99^{th} percentiles. The description of variables is provided in Table 1.

Table B.2: Total Funds Raised and Token Retention

	All	< 99%	< 98%	< 95%	< 90%
Token Retention	0.860***	0.912***	0.895***	0.857***	0.965***
	(3.19)	(3.45)	(3.37)	(3.19)	(3.47)
Vesting Dummy	0.348***	0.339***	0.337***	0.351***	0.393***
	(3.67)	(3.69)	(3.68)	(3.84)	(4.16)
KYC Dummy	0.378***	0.381***	0.375***	0.362***	0.294**
	(2.84)	(2.90)	(2.87)	(2.74)	(2.17)
Whitelist Dummy	0.536***	0.539***	0.521***	0.518***	0.517***
	(4.36)	(4.43)	(4.29)	(4.25)	(4.18)
Venture Funding Dummy	1.249***	1.119***	1.148***	1.137***	1.107***
	(9.20)	(10.35)	(10.60)	(10.83)	(9.38)
Advisers Dummy	0.321***	0.360***	0.364***	0.374***	0.387***
	(3.21)	(3.66)	(3.72)	(3.80)	(3.83)
Experts Dummy	0.434***	0.441***	0.464***	0.468***	0.471***
	(4.77)	(4.90)	(5.17)	(5.17)	(5.07)
GitHub Dummy	0.063	0.074	0.088	0.085	0.099
	(0.62)	(0.74)	(0.87)	(0.84)	(0.96)
Pre-sale Dummy	-0.025	-0.034	-0.011	0.034	-0.002
	(-0.22)	(-0.30)	(-0.10)	(0.30)	(-0.01)
Hard Cap Dummy	0.734***	0.861***	0.850***	0.910***	0.730***
	(3.80)	(4.63)	(4.61)	(5.00)	(4.00)
Soft Cap Dummy	-0.614***	-0.594***	-0.588***	-0.546***	-0.482***
	(-5.67)	(-5.52)	(-5.46)	(-5.05)	(-4.33)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.213	0.219	0.220	0.224	0.206
N	1501	1489	1475	1427	1333

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$ln\,(Funds\ Raised_j) = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j.$$

The dependent variable in each regression is the natural logarithm of total funds raised in the ICO. In columns (2)–(5), we exclude ICOs with total funds raised above the 99^{th}, 98^{th}, 95^{th}, and 90^{th} percentiles, respectively. The total funds raised are denominated in U.S. dollars. Token retention is expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table B.3: ICO Counts by ICO Team Location

Rank	Country	Count	Rank	Country	Count
1	UNITED STATES	228	41	SERBIA	5
2	SINGAPORE	142	42	MEXICO	5
3	RUSSIAN FEDERATION	113	43	COSTA RICA	5
4	UNITED KINGDOM	110	44	LUXEMBOURG	5
5	ESTONIA	80	45	IRELAND	4
6	SWITZERLAND	74	46	ST. KITTS AND NEVIS	4
7	HONG KONG	43	47	ARGENTINA	4
8	AUSTRALIA	31	48	CAMBODIA	3
9	CANADA	26	49	CROATIA	3
10	GERMANY	26	50	NORWAY	3
11	CHINA	26	51	SWEDEN	3
12	SLOVENIA	23	52	INDONESIA	3
13	NETHERLANDS	21	53	NEW ZEALAND	3
14	MALTA	16	54	HUNGARY	3
15	LITHUANIA	16	55	MAURITIUS	3
16	INDIA	16	56	KAZAKHSTAN	2
17	FRANCE	15	57	GREECE	2
18	UNITED ARAB EMIRATES	14	58	MARSHALL ISLANDS	2
19	BELIZE	14	59	NIGERIA	2
20	SEYCHELLES	14	60	BARBADOS	2
21	UKRAINE	14	61	BRAZIL	2
22	BULGARIA	12	62	MALAYSIA	2
23	ISRAEL	12	63	TURKEY	2
24	POLAND	11	64	FINLAND	2
25	CZECH REPUBLIC	11	65	COLOMBIA	1
26	CYPRUS	11	66	ARMENIA	1
27	SOUTH AFRICA	10	67	VIETNAM	1
28	REPUBLIC OF KOREA	10	68	ST. VINCENT AND THE GRENADINES	1
29	JAPAN	10	69	GUINEA-BISSAU	1
30	PANAMA	9	70	SLOVAK REPUBLIC	1
31	AUSTRIA	8	71	PERU	1
32	LATVIA	8	72	KYRGYZ REPUBLIC	1
33	ROMANIA	7	73	BAHAMAS	1
34	ITALY	6	74	PAKISTAN	1
35	GEORGIA	6	75	SAMOA	1
36	SPAIN	6	76	ECUADOR	1
37	PHILIPPINES	6	77	DOMINICAN REPUBLIC	1
38	BELARUS	6	78	DENMARK	1
39	BELGIUM	5	79	EGYPT	1
40	THAILAND	5			

This table reports the list of countries where ICO teams are located. Only ICOs that were completed before or on December 31, 2018, are included in the corresponding counts.

Table B.4: Fundraising Success and Team Retention

	(1)	(2)	(3)	(4)	(5)
Team Retention	0.227**	0.227**	0.186*	0.142	0.144*
	(2.35)	(2.35)	(1.95)	(1.57)	(1.65)
Vesting Dummy			0.098***	0.088***	0.091***
			(3.64)	(3.44)	(3.63)
KYC Dummy				0.041	0.038
				(1.19)	(1.13)
Whitelist Dummy				0.115***	0.110***
				(3.57)	(3.50)
Venture Funding Dummy				0.433***	0.405***
				(11.01)	(10.50)
Advisers Dummy				0.011	0.019
				(0.40)	(0.69)
Experts Dummy				0.010	0.024
				(0.38)	(0.93)
GitHub Dummy					-0.027
					(-1.06)
Pre-sale Dummy					-0.039
					(-1.35)
Soft Cap Dummy					-0.146***
					(-5.64)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.063	0.063	0.078	0.195	0.233
N	826	826	826	826	826

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$Fundraising\ Success_j = \alpha + \beta \cdot Team\ Retention_j + \gamma X_j + \epsilon_j.$$

The dependent variable in each regression is the ratio of ICO total funds raised to hard cap. The team retention measure is computed from reported token allocation breakdowns, which we collected as available for all ICOs in our analysis sample. It is computed as the fraction of tokens not for public sale less shares designated for the following purposes: private sale to early investors, incentive schemes (e.g., bounty programs), the operation or growth of the venture (e.g., marketing), and advisors. The remaining tokens are designated specifically for insiders (e.g., founders) or to a generic company account. Both fundraising success and token retention are expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table B.5: ICO Outcomes and Token Retention Using Logit Model

	Listed	Working Website	Product Live	Product Apple
Token Retention	0.829**	1.044***	0.655**	1.098**
	(2.23)	(2.85)	(1.98)	(2.41)
Successful Dummy	1.372***	0.394*	0.729***	1.128***
	(5.33)	(1.81)	(3.67)	(3.14)
Vesting Dummy	0.130	0.579***	0.061	0.099
	(0.86)	(3.19)	(0.45)	(0.51)
KYC Dummy	0.837***	0.198	-0.126	0.051
	(4.28)	(0.99)	(-0.76)	(0.22)
Whitelist Dummy	0.582***	0.308	0.113	0.343
	(3.04)	(1.53)	(0.69)	(1.43)
Venture Funding Dummy	1.409***	1.227***	0.011	-0.096
	(4.39)	(2.79)	(0.05)	(-0.31)
Advisers Dummy	0.475***	1.575***	0.411***	0.206
	(3.20)	(10.13)	(3.09)	(1.06)
Experts Dummy	0.360**	0.307*	0.171	-0.263
	(2.40)	(1.76)	(1.27)	(-1.33)
GitHub Dummy	0.663***	0.372**	0.117	-0.178
	(4.63)	(2.45)	(0.92)	(-0.96)
Pre-sale Dummy	-0.180	-0.024	0.061	0.220
	(-1.13)	(-0.15)	(0.43)	(1.03)
Soft Cap Dummy	-0.289*	0.067	0.267*	0.053
	(-1.88)	(0.38)	(1.89)	(0.27)
Time Fixed Effects	Yes	Yes	Yes	Yes
Pseudo R^2	0.261	0.167	0.048	0.062
N	1251	1280	1295	1269

This table reports the estimated coefficients from cross-sectional regressions:

$$\text{Logit}\,(y_j) = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j.$$

The dependent variable in each regression is an indicator that equals 1 if the given outcome occurred. The variable definitions are provided in Table 1. Token retention is expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table B.6: ICO Outcomes and Token Retention Using Probit Model

	Listed	Working Website	Product Live	Product Apple
Token Retention	0.456**	0.613***	0.398**	0.609**
	(2.12)	(2.90)	(2.02)	(2.52)
Successful Dummy	0.767***	0.236*	0.436***	0.567***
	(5.39)	(1.91)	(3.73)	(3.35)
Vesting Dummy	0.091	0.317***	0.041	0.061
	(1.02)	(3.20)	(0.50)	(0.60)
KYC Dummy	0.503***	0.110	-0.077	0.022
	(4.43)	(0.96)	(-0.77)	(0.18)
Whitelist Dummy	0.321***	0.176	0.065	0.178
	(2.89)	(1.56)	(0.67)	(1.43)
Venture Funding Dummy	0.787***	0.649***	0.011	-0.052
	(4.43)	(2.94)	(0.08)	(-0.31)
Advisers Dummy	0.280***	0.894***	0.249***	0.102
	(3.23)	(10.31)	(3.14)	(1.01)
Experts Dummy	0.207**	0.166*	0.103	-0.145
	(2.34)	(1.73)	(1.27)	(-1.40)
GitHub Dummy	0.381***	0.199**	0.073	-0.094
	(4.56)	(2.32)	(0.95)	(-0.97)
Pre-sale Dummy	-0.103	-0.030	0.035	0.121
	(-1.10)	(-0.32)	(0.41)	(1.10)
Soft Cap Dummy	-0.191**	0.039	0.164*	0.022
	(-2.10)	(0.39)	(1.91)	(0.21)
Time Fixed Effects	Yes	Yes	Yes	Yes
Pseudo R^2	0.259	0.165	0.048	0.063
N	1251	1280	1295	1269

This table reports the estimated coefficients from cross-sectional regressions:

$$P\left(y_j = 1 | X_j\right) = \Phi\left(\alpha + \beta \cdot Token\ Retention_j + \gamma X_j\right).$$

The dependent variable in each regression is an indicator that equals 1 if the given outcome occurred. The variable definitions are provided in Table 1. Token retention is expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table B.7: Token Retention and Market Environment

	(1)	(2)	(3)	(4)	(5)
Token Retention	0.144***	0.142***	0.145***	0.147***	0.147***
	(2.82)	(2.76)	(2.83)	(2.85)	(2.84)
ln(Scaled # of ICOs)	-0.008	-0.569***	-0.044	-0.041	-0.038
	(-0.16)	(-2.74)	(-0.34)	(-0.33)	(-0.30)
Token Retention × ln(Scaled # of ICOs)	0.392***	0.407***	0.395***	0.361***	0.365**
	(3.70)	(3.38)	(3.36)	(3.13)	(2.45)
Vesting Dummy	0.079***	0.080***	0.079***	0.079***	0.079***
	(3.63)	(3.66)	(3.62)	(3.62)	(3.60)
KYC Dummy	0.062**	0.063**	0.065**	0.065**	0.063**
	(2.35)	(2.38)	(2.44)	(2.42)	(2.38)
Whitelist Dummy	0.111***	0.111***	0.110***	0.110***	0.111***
	(4.38)	(4.36)	(4.34)	(4.34)	(4.37)
Venture Funding Dummy	0.353***	0.345***	0.348***	0.350***	0.350***
	(12.15)	(11.86)	(11.93)	(11.95)	(11.94)
Advisers Dummy	0.026	0.030	0.026	0.026	0.027
	(1.27)	(1.47)	(1.28)	(1.27)	(1.32)
Experts Dummy	0.030	0.035*	0.031	0.031	0.031
	(1.43)	(1.68)	(1.52)	(1.51)	(1.47)
GitHub Dummy	-0.012	-0.009	-0.011	-0.011	-0.011
	(-0.62)	(-0.44)	(-0.57)	(-0.54)	(-0.53)
Pre-sale Dummy	-0.045**	-0.047**	-0.046**	-0.046**	-0.046**
	(-2.03)	(-2.10)	(-2.06)	(-2.05)	(-2.05)
Soft Cap Dummy	-0.161***	-0.163***	-0.162***	-0.161***	-0.161***
	(-7.81)	(-7.90)	(-7.83)	(-7.78)	(-7.76)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.251	0.254	0.250	0.249	0.247
N	1273	1273	1273	1273	1273

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$Fundraising\ Success_j = \alpha + \beta_1 \cdot Token\ Retention_j + \beta_2 \cdot ln\,(Scaled\ \#\ of\ ICOs_j) \\ + \beta_3 \cdot Token\ Retention_j \times ln\,(Scaled\ \#\ of\ ICOs_j) + \gamma X_j + \epsilon_j$$

The dependent variable in each regression is the ratio of ICO total funds raised to hard cap. Both fundraising success and token retention are expressed in percentage points. The crowdedness in the market is measured in terms of (i) the number of ICOs that are active within 5 and 30 days from the launch date of ICO j scaled by the average number of words in the whitepapers reported in columns (1)–(2), respectively; (ii) the number of ICOs that are active within the 15 days from the launch date of ICO j scaled by the median number of words and average number of pages in the whitepapers reported in columns (3)–(4), respectively; and (iii) the raw number of ICOs that are active within 15 days from the launch date of ICO j, reported in column (5). All the variables besides dummies are demeaned. Only ICOs that started on or after June 1, 2017, and were completed before or on December 31, 2018, are included. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table B.8: "Soft" Fundraising Success and Token Retention

	(1)	(2)	(3)	(4)	(5)	(6)
Token Retention	0.750***	0.736*	0.789*	0.775*	0.506	0.453
	(2.63)	(1.70)	(1.67)	(1.68)	(1.10)	(0.98)
Vesting Dummy				0.644***	0.474**	0.502**
				(3.27)	(2.33)	(2.48)
KYC Dummy					0.169	0.186
					(0.70)	(0.77)
Whitelist Dummy					0.269	0.265
					(1.21)	(1.19)
Venture Funding Dummy					0.762	0.745
					(1.62)	(1.55)
Advisers Dummy					0.263	0.265
					(1.30)	(1.31)
Experts Dummy					0.358*	0.373*
					(1.81)	(1.87)
GitHub Dummy						-0.241
						(-1.32)
Pre-sale Dummy						-0.151
						(-0.69)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.084	0.060	0.079	0.096	0.117	0.121
N	1247	705	639	639	639	639

This table reports the estimated coefficients from cross-sectional regressions using OLS:

$$Soft\ Fundraising\ Success_j = \alpha + \beta \cdot Token\ Retention_j + \gamma X_j + \epsilon_j.$$

The dependent variable in each regression is the ratio of ICO total funds raised to soft cap. Both "soft" fundraising success and token retention are expressed in percentage points. Only ICOs that were completed before or on December 31, 2018, are included. In column (1), soft fundraising success is set to zero if the company does not report the amount of funds raised. In column (2), token retention is set to zero if the company does not report the amount/share of tokes retained. Time fixed effects are monthly. T-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

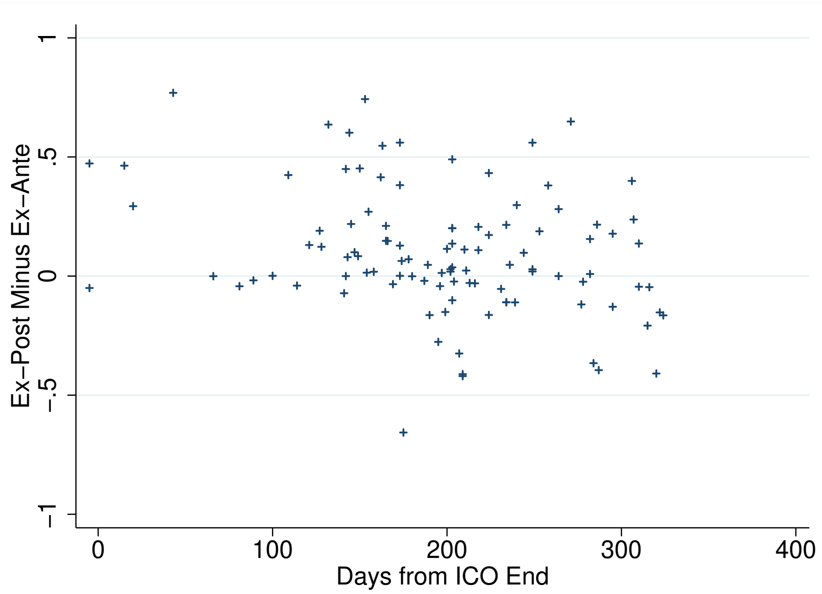


Figure B.1: Ex-Post versus Ex-Ante Retention

This figure plots the difference between our proxy for ex-post token retention and the ex-ante measure we use in our analysis. Each point represents an ICO, and the x-axis is the number of days from the end of the token sale event at which we measured ex-post retention. Our proxy for ex-post retention is the percentage of tokens not in circulation as computed from data on CoinMarketCap (https://coinmarketcap.com/. This measure is computed as 100% minus the circulating supply (CoinMarketCap's approximation of the number of coins that are circulating in the market and in the general public's hands) divided by total supply. Data for circulating supply are not available historically through the website, so we only have data for the two days that we scraped (June 30, 2018, and December 20, 2018). Hence we choose the observation closest to the end of the token sale for each ICO. We are able to compute our ex-post proxy for only 105 ICOs in our sample.

C Dead Coins Data Details

We collect data on alleged scams from the Dead Coins website (https://deadcoins.com/). Dead Coins is a user-generated content forum that aims to catalog all failed cryptocurrencies and tokens. This website is a trusted source in the crypto industry, and statistics from it are consistently referenced in news and reports (see, e.g., The Economist, 2018). Each forum post contains the report date, coin/token name, category, and summary description. Although our sample of ICOs ends in December 2018, we scraped this data most recently in September 2019 to allow for delays in reporting.

There are four categories of Dead Coins posts: *Deceased*, *Scam*, *Parody*, and *Hack*. In Table C.1, we see that Deceased and Scam are the most common report categories.

Table C.1: Frequency of Dead Coins Reports by Category

Category	N	%
Deceased	994	56
Scam	664	37
Parody	93	5
Hack	28	2
Total	1,779	100

There are no formal definitions for these categories and, because it is a user-generated forum, the categorization of a report is up to the user. We therefore reviewed the report descriptions manually to describe these categories. We interpret *Deceased* as capturing coins/tokens that are no longer supported by their creators and have market values close to zero. We interpret *Scam* as capturing coins/tokens for which investors were misled by entrepreneurs who never actually worked on the project as described. In Figure C.1, we see that the majority of *Deceased* reports occurred in September 2017, which is the first month that Dead Coins began its activity. It therefore appears that these reports represent a backlog of activity for coins/tokens before September 2017.

For the analysis in this paper, we focus on *Scam* reports, as these represent ex-post outcomes of interest. We want to make sure that these reports accurately represent scams as we understand them and therefore we investigate the descriptions more carefully. There are many very short or missing descriptions that do not provide any evidence or support that an ICO is a scam (see

Table C.3). For example, many reports simply write the word "scam." Because of insufficient explanation, we drop these reports from consideration when creating *Scam Dummy*. Specifically, we exclude reports with 25 or less characters in their description, which eliminates 15% of the scam report observations (Table C.4).

Before merging the Dead Coins data with our ICO database, we aggregate the reports to the coin/token level by using three identifiers: ticker symbol, token name, and website URL. We then merge these observations with our list of ICOs using the same identifying information. Over 50% of tokens designated as scams are matched, and this is the best-matched category (Table C.2). We do not expect to match all of the Dead Coins coins/tokens, given that they include many cryptocurrencies in addition to tokens. In Figure C.2, we see that the fraction of tokens reported as scams is consistent across our sample period.

Table C.2: Frequency of Tokens Matched to Our ICO Data by Dead Coin Category

Type	Dead Coins	Matched	%
Deceased	947	297	31
Scam	582	315	54
Parody	92	25	27
Hack	26	10	38
Any	1574	600	38

Our *Scam Dummy* indicator is set equal to 1 if an ICO is matched to a token in the Dead Coins data for which there was a scam report. As shown in Figure C.3, the percentage of tokens that are ultimately reported as a scam is consistent over most of the sample and equal to about 5% on average. The fraction of scams declines over time, which makes sense given that it may take several months for an ICO to be reported as a scam, and so the ICOs at the end of our sample are less likely to have been reported as scams as of September 2019, all else held equal.

Figures C.4 and C.5 show the distributions of token retention and fundraising success, respectively, conditional on whether a token is reported as a scam in Dead Coins. ICOs which are reported as scams appear to have lower token retention on average and disproportionately higher fundraising success. The latter finding is not surprising, given that an ICO must be successful in order to reach the point at which an investor would report it as a scam.

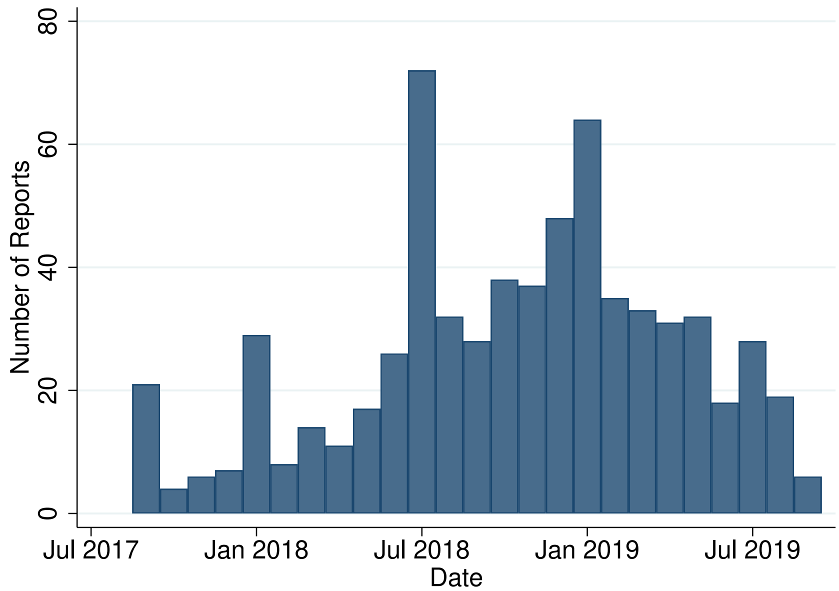
Table C.3: Dead Coins Scam Descriptions Less than 25 Characters

Token Name	Description	Token Name	Description
ROUND	Developers have run away	Confido	Exit scammed
JasonEdwardsCoin	Garbage Australian coin.	Arbitraging	Ponzi scheme
Gems	No monthly/weekly report	PRiVCY	Ponzi Scheme
Capra	Abandonment development	LOLIGO	Scam project
BITCOINNODE	Community Takeover Coin	Lotus	It's a scam
Corex	They scam it and exited	Eos	Total scam
monyx	Took the money and ran.	PChain	Total scam
BitConnect	A exposed ponzi scheme	asdasd	adsadsasd
ParkByte	company does not exist	asdasd	adsadsasd
travelflex	scam ceo delet account	INTERSTELLER HOLD	EXIT SCAM
chrysos	Took the money and ran	Travelblock	Exit scam
Assetereum	Asset backed currency	ETHERCONNECT	SCAM COIN
Menlo One	Dead ass a door nail.	Gainer Coin	Scam coin
ShipChain	https://shipchain.io/	Involve	ScAM cOIN
Pagarex	Lending Platform Scam	Refereum	Scam coin
FIBRECOIN	Probabily was a scam.	DigitexFutures	Scam ICO
ImmVRse	Scam. No development.	Daps	Scammers
GOLDPOWERCOIN	no effort and no job	Sficoin	Lending
Dentacoin	pump and dump scheme	Hextracoin	Ponzi
Playgame	The big Scam project	goldKash	Dead
Bitconnect	Bitcoonneeeeeeeect!	8coin	Scam
GT	This is a scam coin	AiOtoken	scam
RealtyCoin	THIS WAS A BIG SCAM	Befund	scam
SiaCashCoin	Fake Siacoin scam.	Bitcoin SV	Scam
feston	pretty surely scam	DimonCoin	Scam
gonetwork	This ico is a scam	Electroneum	SCAM
Hshare	A clone of decred	FDC	Scam
Ideacoin	Lending platform.	FirstCoin	scam
PlusCoin	Russian Scam Coin	GIZA Device	Scam
mktcoin	scam ponzi estafa	Gimli	Scam
bchconnect	lending scamcoin	INTcoin	Scam
OZTGToken (OZTG)	OZTGToken (OZTG)	Knoxcoin	SCAM
Bchsv	Scam for sure ..	Lionsea	SCAM
ethereumcashpro	ethereumcashpro	Netavo	Scam
bitbase coin	this is SCAMMMM	PIP	Scam
SMART Trade COIN	Arbitrage Coin	Pareto Network	Scam
Dropil	Bitconnect 2.0	STARDUST	SCAM
Dropil	Bitconnect 2.0	Storiqa	Scam
byteball	Dead mean dead	electroneum	scam
PORNX	Scam Scam Scam	eltrond	scam
Alphatech	AFX is A SCAM	fitrova	scam
play2live	fud. scamers!	xgs	SCAM
bitpaction	scam exchange		
Faceblock	Scam or hack.		
Icon	Shitcoin scam		

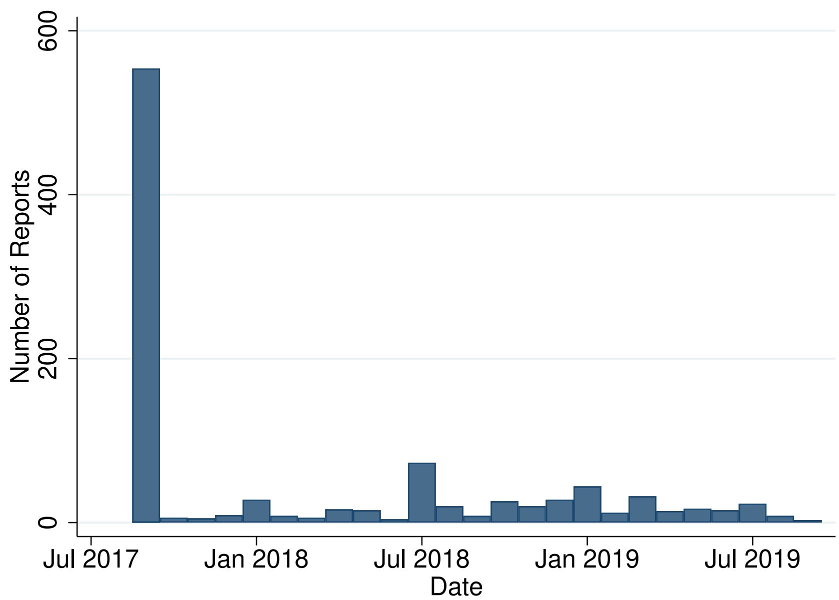
Table C.4: Dead Coins Reports by Description Length

Category	Stat.	>= 0	0	> 0	> 25	> 50	> 100	> 200	> 300
Deceased	Count	994	280	714	592	502	254	134	84
	Pct.	100	28	72	60	51	26	13	8
Hack	Count	28	0	28	26	23	15	11	8
	Pct.	100	0	100	93	82	54	39	29
Parody	Count	93	20	73	64	61	38	25	17
	Pct.	100	22	78	69	66	41	27	18
Scam	Count	664	12	652	563	492	374	213	138
	Pct.	100	2	98	85	74	56	32	21

This table shows the count and relative percentage of Dead Coins reports within each category according to the length of their description in terms of number of characters. The columns refer to ranges of description length. For example, "> 25" means that the description is longer than 25 characters.



(a) Scam



(b) Deceased

Figure C.1: Number of Dead Coins Reports by Category

This figure depicts the count of Dead Coins reports by category for each month during the period September 2017 through September 2019.

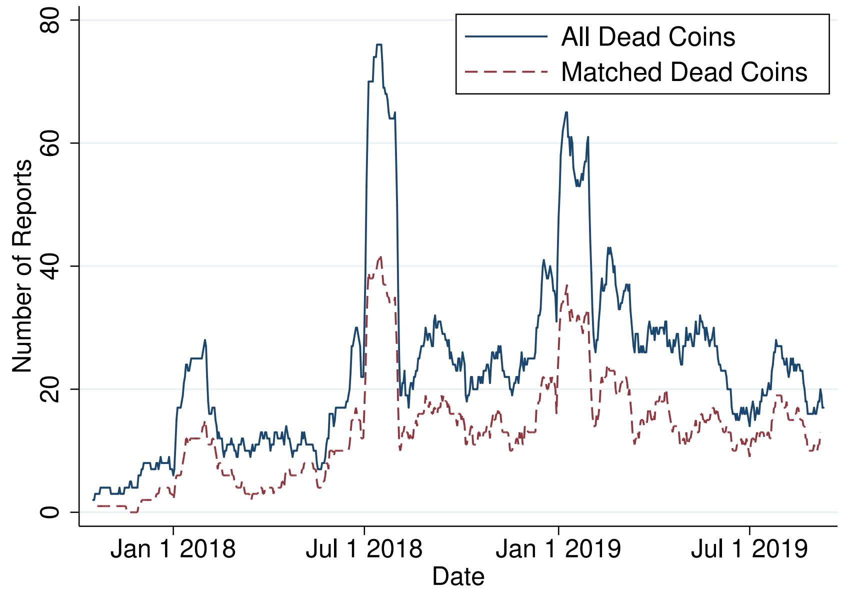


Figure C.2: Trailing 30-Day Scam Report Counts

This figure depicts the trailing 30-day count of Dead Coins scam reports for all tokens in the Dead Coins data (All Dead Coins) and for tokens matched with our full ICO database (Merged Dead Coins).

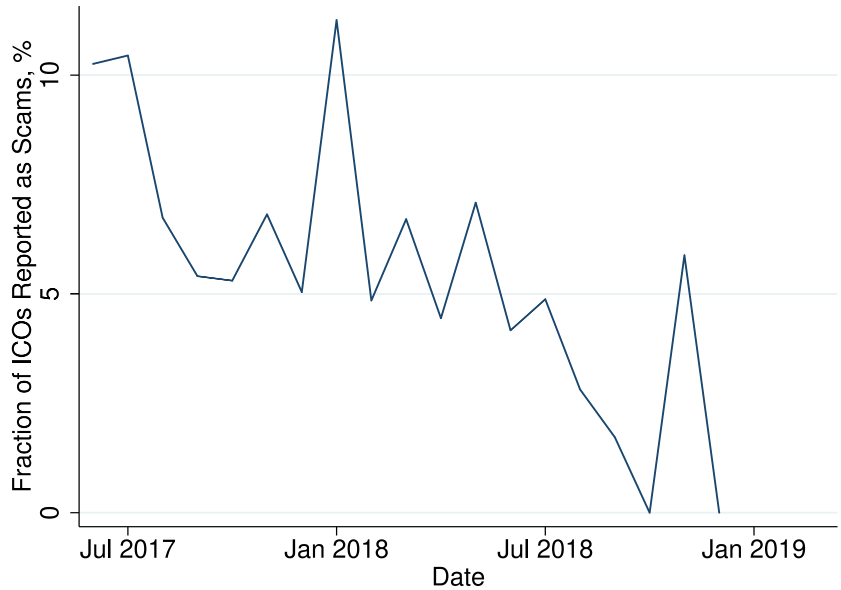


Figure C.3: Fraction of Scams in ICOs with Positive Funds Raised

This figure depicts the percent of ICOs with positive funds raised that are reported as a scam in the Dead Coins data by ICO launch month.

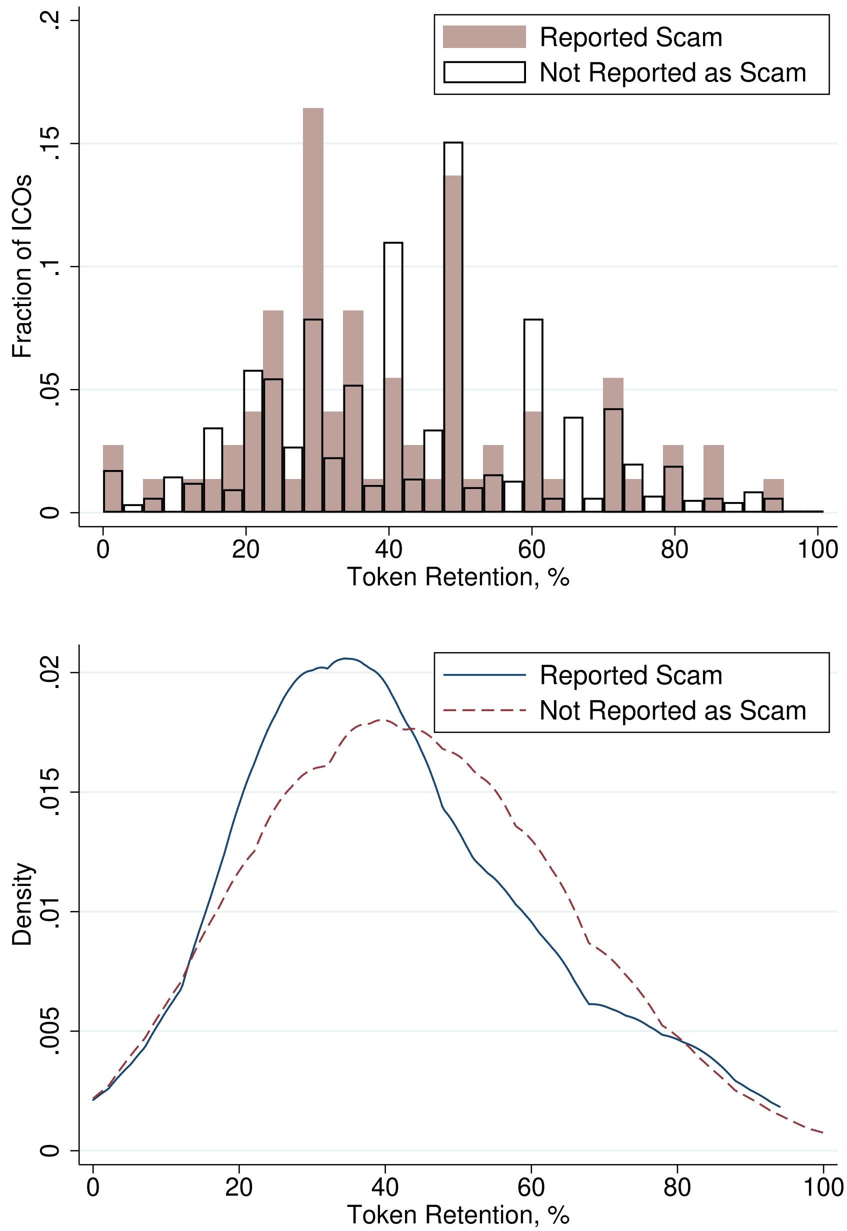


Figure C.4: Distribution of ICO Token Retention

This figure depicts the distribution of token retention for ICOs that were reported as scams ($Scam\ Dummy = 1$) and ICOs that were not reported as scams ($Scam\ Dummy = 0$). The top panel shows the histograms and the bottom panel shows the kernel densities. Only ICOs which were completed before or on December 31, 2018, and which report funds raised are included.

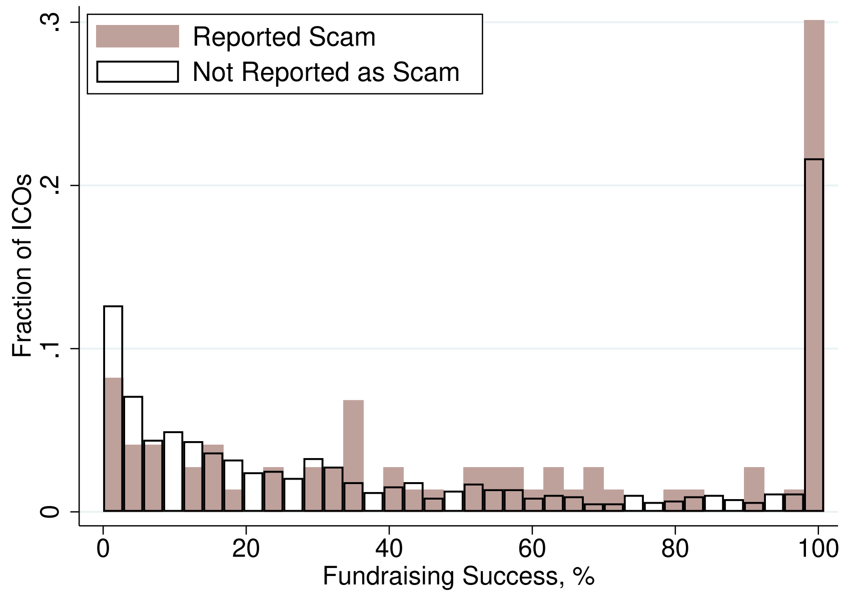


Figure C.5: Distribution of ICO Fundraising Success

This figure depicts the distributions of ICO raised funds as a fraction of hard cap for ICOs that were reported as scams ($Scam\ Dummy = 1$) and ICOs that were not reported as scams ($Scam\ Dummy = 0$). Only ICOs which were completed before or on December 31, 2018, and which report funds raised are included.