From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2019

Essays in Financial Innovation and FinTech

Shin, Donghwa

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01f7623g45d

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS IN FINANCIAL INNOVATION AND FINTECH

Donghwa Shin

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Economics
Adviser: Professor Wei Xiong

June 2019

© Copyright by Donghwa Shin, 2019.

All rights reserved.

Abstract

In my thesis, *Essays in Financial Innovation and FinTech*, I investigate two growing innovative asset classes, structured retail products (SRPs) and cryptocurrencies.

Chapter 1 examines the market for SRPs. This market has grown rapidly in sales volume and complexity in last two decades. Using a comprehensive dataset on SRPs, I examine investors extrapolation as an explanation of this phenomenon. I find that products with higher past returns have enjoyed higher sales growth, even though past returns do not predict future performance. This effect is stronger for more complex products, leading to the observation of greater popularity of more complex products that happen to deliver better past performance. While there is some evidence of financial intermediaries exploiting investors in early part of the sample, my results further suggest that the rapid market growth induced by extrapolation has led to more competition among intermediaries, which in turn disciplines exploitation.

In Chapter 2, my coauthors and I study Initial Coin Offerings (ICOs). Certification by a crowd of online analysts and early investors can generate excitement among potential token investors, leading to successful ICOs. We test "wisdom of crowds" using novel data on over 1,500 ICOs, including sequential investor subscriptions during token sales. We find that favorable analyst opinions on the underlying project are associated with aggressive initial token subscriptions, which predicts subsequent token sales. Overall, our results suggest that the wisdom of crowds could mitigate information asymmetry in the ICO market.

In Chapter 3, my coauthors and I study pervasive market manipulation activities Pump-and-dump schemes (P&Ds) in the cryptocurrency market. We study these events using trade-by-trade data and a sample of P&Ds with precisely identified starting time. We find that P&Ds lead to short-term cryptocurrency bubbles featuring dramatic increases in prices, volume and volatility. Prices reach the peak in one minute and a quick

reversal follows. Evidence, including a significant price run-up before the start of P&Ds, implies significant wealth transfers from insiders to outsiders. Bittrex, a cryptocurrency exchange, banned P&Ds on November 24, 2017. Using a difference-in-differences, we provide causal evidence that P&Ds are detrimental to the liquidity and price of cryptocurrencies.

Acknowledgements

First and foremost, I would like to thank God for giving me the strength, courage, and wisdom to undertake this research and and complete it satisfactorily. This achievement would not have been possible without his blessing and help.

In my long, painful but also enjoyable journey towards my Ph.D. degree, I have found a teacher, an inspiration, a great role model, Professor Wei Xiong. He has provided his support and guidance all the time and has given me invaluable suggestions, which have contributed greatly to improve the thesis.

This thesis has also benefited from the comments and suggestions by Professor Motohiro Yogo and Professor Jakub Kastl. My job market paper could not have started without Professor Kastl's strong encouragement. In the early stage of my job market paper, he has shown a great interest in structured retail products and found a great potential in it. Professor Yogo always has provided very sharp comments on my paper. My paper experienced a continuous improvement after weekly research meetings with him.

I also thank my coauthors for my two cryptocurrency papers, Tao Li, Jongsub Lee, and Baolian Wang. When I started working on my cryptocurrency projects, this area of research was still very new and I did not have a clear understanding in FinTech. However, through the continuous discussions with them and the constant effort in learning this market, I finally was able to complete two great papers, which greatly helped in the pursuit of my successful career as an academic researcher in finance. I would like to express my sincere appreciation to Tao Li again for his help not only for my two cryptocurrency papers but also for his invaluable advice on my job search. Tao was always willing to spend his time and resource for the success of my job search. I cannot express in words how grateful I am for what he has done for me. Tao and I have been a great team and we always will be.

I thank all who helped for the success of my papers. Fernandes Tiago, the managing director at StructuredRetailProducts, Ltd. has provided wonderful dataset and an invaluable insight for understanding the market for structured retail products. I am very lucky to have my friends at the Department of Operations Research and Financial Engineering and the Department of Economics: Nana Kobby Aboagye, Woon Sang Cho, Paul Ho, Michael Dobrew, Liyu Dou, Andrei Zeleneev, and many others. Especailly, Paul and Michael have been always happy to discuss about my research idea. Nana and Woon Sang are my best friends at Princeton who are always willing to help me with everything. They were very helpful in proof-reading my job market paper.

Last but not least, I would like thank my family, especially my wife, Minyoung, who has been always very supportive for the pursuit of my career. Although I know how stressed she has been as a wife of a Ph.D. student, she has never complained but rather she always has tried to show how happy she was to encourage me, which in the end leads to the great achievement in this thesis. She also gave me the greatest gift in my life last year, my dear son, Hajoon, who has been a great motivation for my research.

My dissertation is dedicated to my family, especially to my late father. My father passed away of cirrhosis in South Korea in February 2017. Although I am very sad that he could not see my achievement at Princeton while he was alive, I am sure the he is very proud of me in heaven as he always was. My efforts are all for him.

To my family, especially to my late father

Contents

Abstract . . . iii

Acknowledgements . . . v

List of Tables . . . xii

List of Figures . . . xiv

1 Extrapolation and Complexity **1**

1.1 Introduction . . . 1

1.2 Institutional Background . . . 7

1.2.1 What are structured retail products? . . . 7

1.2.2 Examples of SRPs . . . 9

1.2.3 Investor characteristics . . . 12

1.3 Data and Empirical Design . . . 13

1.3.1 Data . . . 13

1.3.2 Empirical Design . . . 15

1.4 Methodologies . . . 18

1.4.1 Text mining for payoff structures . . . 18

1.4.2 Estimation of main variables . . . 20

1.4.3 Validity of text mining . . . 25

1.4.4 Variables construction . . . 26

1.5 Empirical Results . . . 28

1.5.1 Ex-post performance of SRPs . . . 28
1.5.2 Extrapolation . . . 30
1.5.3 Extrapolation (Intensive and extensive margins) . . . 32
1.5.4 Extrapolation vs rational information extraction . . . 33
1.5.5 Extrapolation: Evidence from surveys . . . 34
1.6 What Explains the Rapid Growth in Sales Volume and Complexity? . . . 37
1.6.1 Explanation by extrapolation . . . 37
1.6.2 Other potential explanations . . . 39
1.7 Evidence of Increasing Competition and Improved Transparency . . . 41
1.7.1 Decreasing markup and increasing competition . . . 41
1.7.2 Improved transparency - availability of accurate markup . . . 43
1.7.3 Effect of extrapolation on markup and sales commission . . . 45
1.8 Conclusion . . . 49
Appendix A . . . 70
Appendix B . . . 72
Appendix C . . . 75

2 The Wisdom of Crowds in FinTech: Evidence from Initial Coin Offerings 75

2.1 Introduction . . . 75
2.2 Institutional Background . . . 83
2.2.1 What is an ICO? . . . 83
2.2.2 The ICO process and post-ICO events . . . 85
2.3 Data Sources and Sample Overview . . . 91
2.3.1 The ICO sample . . . 91
2.3.2 Analyst ratings . . . 93
2.3.3 Primary market subscription data . . . 94

2.3.4 Secondary market prices, volumes, and transactions 95
2.3.5 Sample overview 96
2.4 Fundraising Success and Campaign Duration 99
2.4.1 Successful fundraisers versus failed ICOs 99
2.4.2 Determinants of ICO success 102
2.4.3 Gross proceeds and ICO success 104
2.4.4 ICO duration 105
2.5 Investor Subscriptions and the Path to ICO Success 106
2.5.1 Initial and subsequent subscriptions 106
2.5.2 Initial investors' investment returns 107
2.5.3 Predicting primary market subscriptions 109
2.5.4 The effect of early subscriptions on fundraising success 111
2.6 Post-ICO Performance 112
2.6.1 ICO Underpricing 112
2.6.2 Long-run Performance 114
2.7 Conclusion 116

3 Cryptocurrency Pump-and-Dump Schemes 144

3.1 Introduction 144
3.2 Institutional Background 152
3.2.1 Cryptocurrencies 152
3.2.2 Pump-and-Dump Schemes 153
3.2.3 The current regulatory environment 155
3.3 Data 156
3.3.1 P&D events 156
3.3.2 Exchange trading data 158
3.3.3 CoinMarketCap data 159

3.3.4 Other data . . . 159
3.3.5 Summary statistics . . . 160
3.3.6 Potential problems with the data . . . 162
3.4 Empirical Results Pertaining to P&Ds . . . 163
3.4.1 Can P&Ds move token prices? . . . 163
3.4.2 Spillovers to Ethereum pairs and other exchanges . . . 166
3.4.3 Performance by purchase and sale time . . . 168
3.4.4 Other potential sources of unfairness . . . 170
3.4.5 Maximum returns, target returns, and viewership . . . 172
3.4.6 Potential mechanisms . . . 173
3.5 The Bittrex Ban . . . 176
3.5.1 The background . . . 176
3.5.2 Consequences of the ban . . . 177
3.6 Conclusion . . . 180

List of Tables

1.1 An example of data structure 59
1.2 Summary statistics 60
1.3 Ex-post performance of SRPs 61
1.4 Extrapolation and sales 62
1.5 Extrapolation and sales (Intensive and extensive margin) 64
1.6 Effect of rational information extraction 66
1.7 Evolution of ex-post performance 67
1.8 Effect of extrapolation on markup and sales commission 69

2.1 ICOs by Country and Industry 124
2.2 ICO Characteristics 125
2.3 Determinants of Fundraising Success 127
2.4 ICO Duration 129
2.5 Investor Subscriptions 130
2.6 Initial Investors' Investment Returns 131
2.7 Investor Subscriptions 132
2.8 Investor Subscriptions and ICO Success 133
2.9 Token Returns and Turnover 135
2.10 Short-Run and Long-Run Token Returns 136

3.1 Summary Statistics 195

3.2 Trade Performance by Purchase Time 197
3.3 Persistent Advantages for Certain Telegram Channels 201
3.4 Maximum Returns and Target Returns 203
3.5 Viewership 204
3.6 The Bittrex Ban 205

List of Figures

1.1 Sales dynamics 50
1.2 Examples of SRPs 51
1.3 Classification 53
1.4 Validity of text mining 54
1.5 Distribution of ex-post performance of SRPs 55
1.6 Autocalls vs Reverse convertibles 56
1.7 Estimated vs Reported 57
1.8 Evolution of markup 58

2.1 Timeline of ICO 117
2.2 The Aragon Token Sale 118
2.3 ICO Starts and Success 120
2.4 Primary Market Subscriptions in ICOs 121

3.1 An Example of Pump-and-Dump Schemes on Telegram 182
3.2 Distribution of Pump and Dumps and Trading Volume by Hour 184
3.3 Distribution of Maximum Returns and Time to Maximum Returns . . . 185
3.4 Return, Volume and Volatility Around Pump and Dumps 186
3.5 Return, Volume and Volatility Sorted by Liquidity 188
3.6 Spillovers to Ethereum Trading Pairs and Other Exchanges 190
3.7 Price Impact 192

3.8 A Sharp Decrease in P&Ds on Bittrex after the Ban 193
3.9 The Effects of the Bittrex Ban on Price and Volume 194

Chapter 1

Extrapolation and Complexity

1.1 Introduction

Understanding the market for complex financial products has been a long-standing challenge in financial economics. This paper studies the rapid growth in sales volume and complexity observed in the market for structured retail products (SRPs), which is a good laboratory to study the market for complex financial products. The existing literature on complex financial products focuses on the exploitation view that financial intermediaries deliberately mislead and missell complex financial products to investors (Carlin, 2009; Célérier and Vallée, 2017; Egan, 2018; Henderson and Pearson, 2011). While there is strong evidence of exploitation in the market for SRPs, it is difficult for this view to fully explain the persistent growth of this market for over two decades. For example, the market for global equity-linked SRPs has experienced sustained growth from a global sales volume of only $50 billion in 2001 to nearly $250 billion in 2015. Exploitation

This chapter is based on my job market paper, "Extrapolation and Complexity". This work was presented at Financial Innovation and Risk Management Conference in Memory of Peter Christoffersen (2019).

may explain the initial motive for financial intermediaries to offer such products at the beginning, but it remains a puzzle why the investor demand continued to increase after so many years, as one would expect investors to eventually suffer investment losses and realize the exploitation.

In this paper, I examine an alternative mechanism to address this puzzle. An extensive literature in behavioral finance has suggested that in predicting future asset returns, retail investors tend to extrapolate past returns (Amromin and Sharpe, 2009; Greenwood and Shleifer, 2014; Vissing-Jorgensen, 2003). This extrapolative behavior offers a potential explanation for investors' demand for SRPs. Specifically, I hypothesize that investors with extrapolative expectations estimate expected future returns of SRPs based on their past returns, rather than their characteristics that determine their expected future returns. The presence of such extrapolative investors in turn motivates financial intermediaries to offer a large variety of SRPs so that some of them would offer high returns and attract more investor demand over time. To the extent that complex SRPs are more difficult for investors to comprehend, they may rely even more on extrapolating past returns. Consequently, investor demand for complex SRPs would be more exposed to their past returns.

Financial institutions, such as investment banks, sell SRPs that package exotic options, linked to underlying assets such as equity, with bonds. The most notable feature of the SRP market is rapid growth in sales volume. In fact, the global equity-linked SRP sales volume nearly quadrupled from about $50 billion in 2001 to approximately $250 billion in 2015. Another interesting feature of this market is increasing complexity. Autocallable products, also known as autocalls, are the most complex SRPs. Notwithstanding their complexity, autocalls have experienced remarkable growth rates, higher than other SRP products, which resulted in a market share of nearly 50% in the global SRP market in 2015. Similar trends are also observed in the largest SRP market, the

United States. Given the risks and complexities associated with SRPs and the market's fast growth, which have been highlighted by the U.S. Securities and Exchange Commission (SEC), understanding what affects SRP investors' purchase decisions should prove highly valuable (Starr, 2015).[1]

Using a comprehensive data set comprising 71,300 individual SRPs issued in the United States between 2001 and 2016, I show that SRP investors possessed extrapolative expectations and they exhibited stronger extrapolation when purchasing more complex products by examining the relationship between sales and past returns. Extrapolation, combined with extensive studies on ex-post performance of SRPs, provides important clues for explaining the increasing complexity and the growing market size. In addition, I provide evidence that this market has become more competitive and transparent, and therefore exploiting investors has become more difficult.

In a formal test for extrapolation, I find that the past returns on SRPs have a statistically and economically significant effect on future sales of SRPs. Furthermore, sales of more complex SRPs are more sensitive to their past performance, compared with those of less complex products. This result is consistent with investors' extrapolation and it is suggestive that investors are subject to more severe extrapolation when purchasing more complex products because understanding products is harder for more complex products. This result is not explained by investors' rational expectations. Contrary to what investors with extrapolative expectations believe, I show that neither past returns nor sales volume predicts the performance of newly issued SRPs. This suggests that the positive relationship between SRPs' past returns and sales is not driven by rational information extraction from past returns on SRPs.

Survey results involving SRP investors further support my empirical findings. Surveys conducted in the United Kingdom (Hunt et al., 2015), Hong Kong (Securities and

[1]Amy M. Starr, Structured Products - Complexity and Disclosure - Do Retail Investors Really Understand What They Are Buying and What the Risks Are? - `https://www.sec.gov/news/speech/speech-amy-starr-structured-products-.html`

Futures Commission, 2006), and South Korea (Financial Supervisory Service, 2010; Cha and Park, 2015) offer evidence that SRP investors generally have only limited background knowledge about SRPs. A significant portion of investors do not have a good understanding of the products they purchase. This corroborates the hypothesis that investors do not fully comprehend the risk-and-return profile of SRPs. Two other surveys conducted in South Korea are indicative that investors possess extrapolative expectations. In particular, Cha and Park (2015) show that investors who buy more complex products exhibit higher subjective expected returns although more complex SRPs did not generate higher past returns in this survey.

Extrapolative expectation provides important clues regarding the increasing complexity, represented by the growing popularity of autocalls. I find that in my sample, more complex products perform better than other products by chance. Since extrapolation is more pronounced when purchasing more complex products, combined with the higher ex-post performance of the more complex products, extrapolation provides a clear explanation of the increasing complexity. In addition, the rapidly increasing market share of more complex products contributes to the rapid growth of the overall SRP market.

A natural follow up question at this juncture is how financial intermediaries react in the presence of investors' extrapolation bias. On the one hand, it is possible that exploitation by financial intermediaries persists and the market remains "dark," as consistently highlighted in the previous literature. In that case, markups charged by financial intermediaries should remain high over a long period and investors should keep buying the products. On the other hand, it is also plausible that the market has become more competitive, and therefore the effect of exploitation could have become weaker.

To answer this question, I first show that product markups have fallen over time based on S&P 500-linked SRPs sample. Early studies (Henderson and Pearson, 2011;

Stoimenov and Wilkens, 2005) provide evidence that markups for some SRPs are too high to be justified. In my sample, the median annualized markup was about 1.9% in 2004, which means that products with a one-year (three-year) term of maturity charged a markup of 1.9% (5.7%) in 2004. Such a high markup is consistent with findings from the early studies. However, the median annualized markup decreased to nearly 0.5% in 2016. During this time period, the number of issuers has consistently increased. This result is consistent with recent studies highlighting low markups of SRPs issued by members of Deutscher Derivate Verband in 2013 and 2016 (Deutscher Derivate Verband, 2015; Muller et al., 2017). Overall, issuers have lowered markups and the effect of exploitation has become weaker, which is consistent with the pattern that the market has become more competitive.

I provide evidence that the market has become more transparent as well. Since 2013, issuers in the United States have been required to provide the fair value of any SRP on the first page of its prospectus. The fair value is an essential element in calculating the annualized markup, and therefore one of the most important information in decision-making. I compare the annualized markups calculated based on reported fair values with annualized markups estimated based on a stochastic volatility model (Heston, 1993) and find that issuers have reported quite accurate fair values, which resulted in improved transparency of the SRP market.

If exploitation motives by financial intermediaries are strong, banks could charge high markups or brokers could charge high sales commissions when they expect that demand for particular SRPs will be high due to solid past performance. I find higher past returns predict slightly lower future markups and sales commissions. Taken together, my results suggest that the effect of exploitation is becoming weaker. Market based mechanisms, such as increased competition, could have prevented issuers and brokers from easily exploiting unsophisticated investors.

My research has real policy implications. Understanding how investors make a purchase decision of complex financial products is crucial in designing and implementing new regulations. In addition to implementing rules regulating the disclosure of SRPs' fair values, regulators have focused on how to make prospectuses more transparent to convey information more effectively (Starr, 2015). However, improved transprancy through the enhanced disclosure alone may not be sufficient to prevent investor extrapolation as my analysis shows that even after the SEC implemented a disclosure rule pertaining to SRPs, the effects of extrapolation persist. This result is consistent with a prediction by Brunnermeier and Oehmke (2009).

My research contributes to several strands of the literature. First, this study enriches the literature on complex financial products (Carlin, 2009; Carlin and Manso, 2010; Coval et al., 2009). The vast majority of existing papers focus on highlighting exploitation by financial intermediaries (Stoimenov and Wilkens, 2005; Gabaix and Laibson, 2006; Henderson and Pearson, 2011; Bordalo et al., 2015; Egan et al., 2016; Célérier and Vallée, 2017; Ghent et al., 2017; Egan, 2018) while recent studies find mixed results (Döhrer et al., 2013; Deutscher Derivate Verband, 2015; Calvet et al., 2017; Muller et al., 2017). My results are generally consistent with this literature in the sense that the markups were high in early 2000 in my sample. I add to this literature by showing the salient patterns of decreasing markups and increasing competition over time. More importantly, the accurate information on markups, which was hidden in early periods, has been publicly available since 2013, and therefore exploitation through charging high markups has become more difficult.

Second, my study introduces an important mechanism that drives financial innovation. Allen and Gale (1994) and Duffie and Rahi (1995) highlight the benefit of risk sharing in financial innovation through market completion. Simsek (2013), however, provides an alternative view. Simsek (2013) proposes that financial innovation

can encourage more speculation due to the belief disagreements between investors, and therefore makes investors' portfolios riskier. Gennaioli et al. (2012) and Arnold et al. (2018) emphasize the role neglected risk by investors plays in understanding financial innovation. My paper shows that another mechanism, extrapolation, might be essential to explaining financial innovation. More specifically, extrapolation explains what types of products are adopted and abandoned in financial innovation. My results show that products with high ex-post performance are adopted fast and products which provided very poor ex-post performance experience a sharp decline and are eventually abandoned.

Finally, my study provides empirical evidence on how investors deal with financial complexity. Brunnermeier and Oehmke (2009) suggest several approaches that bounded rational investors can take to deal with financial complexity. For example, investors can divide the complex products into several pieces that are relatively easy to analyze. Also, they can employ a simple pricing model, such as the Black-Scholes Model (Black and Scholes, 1973). I add to this literature by showing that if investors cannot use these systematic approaches because products are very complex, investors tend to rely heavily on a simple approach, *extrapolation*.

1.2 Institutional Background

1.2.1 What are structured retail products?

Structured retail products are retail financial products whose performance is determined by complex payoff structures that vary with the performance of underlying assets. The payoff structures of SRPs are generally nonlinear functions of the performance of the underlying assets. Additionally, these payoff structures of SRPs are generally more complex than those of plain vanilla options and they take various forms across multiple types of products. SRPs are based primarily on equity indexes or individual stocks but there

are also SRPs that provide exposure to other underlying assets such as commodities, exchange rates, or alternative indexes.

According to Structured Retail Products, Ltd.,[2] the SRP market has experienced steady growth and, as of 2016, assets under management on this market total is about $1.92 trillion globally and about $8 trillon worth of products have been issued since the market first opened. The largest market in terms of assets under management is the United States. As of 2016, the assets under management in the United States totaled $367 billion, followed by South Korea with $143 billion worth of assets under management.

When purchasing SRPs, investors consult with brokers. Throughout a meeting with investors, brokers comprehend the risk appetites of the investors and recommend suitable SRPs. Brokers are compensated by sales commissions, which represent fixed proportions of total amount invested by investors. A sales commission is decided through negotiation between an issuer and a brokerage firm before a product is marketed and it varies across products.

Since 2001, when the major markets began to become popular, SRPs markets have experienced significant change. The most notable change is the the rapid growth in sales volume and complexity.

[Insert Figure 1.1 here.]

Figure 1, Panel A displays global sales of equity-linked SRPs. While global equity-linked SRPs sales volume was approximately $50 billion in 2001, it increased to nearly $300 billion in 2007, followed by a sharp decrease in the midst of the financial crisis. SRP sales have recovered since then and the sales volume was approximately $200 billion in 2016. Autocalls, which are the most complex SRPs, have experienced higher growth rate than other products. Autocall sales accounted for only about 1% of all SRP in 2001 but

[2] `https://StructuredRetailProducts.com`

that share increased steadily and was about 45% as of 2016. The same trend is observed in the United States, the biggest equity-linked SRP market, as seen in Figure 1.1, Panel B. Autocalls have experienced steady growth in the United States as well. I also present proportions of sales taken by other kinds of products, which I explain in detail in the next subsection. Reverse convertibles, which were once popular, have experienced a sharp decline in sales and participation products show a slow growth pattern. Understanding these patterns in the sales dynamics of various types of products is one of the main goals of this research.

1.2.2 Examples of SRPs

In what follows, I provide descriptions and examples of four types of SRPs, issued in the United States, which cover a broad range of products. Understanding these examples is important for understanding the empirical design of my research, which I explain in detail in the next section.

A. Capital-protected participation products

The first type of SRP I introduce here is capital-protected participation product presented in Figure 1.2, Panel A. A participation product is a relatively simple product in the sense that the return on the product at maturity is determined by the return of the underlying asset at maturity and the continuous trajectory of the performance of the underlying asset does not contribute to the return on the product at maturity.

[Insert Figure 1.2 here.]

One example of a capital-protected participation product, called *Principal Protected Product*, was issued by the Canadian Imperial Bank of Commerce in the United States on June 13, 2005, maturing on June 15, 2010. The underlying asset of this product is the S&P 500 index. As the name of the product indicates, the main feature of this product

is that investors' principal (or capital) is protected irrespective of the performance of the S&P 500 index.

The return on this product at maturity is determined by the return on the S&P 500 index over the five-year investment period. For SRPs, returns on the underlying asset is measured based on the ex-dividend price of the underlying asset. More specifically, the underlying return is defined as $\frac{\text{Level of S\&P 500 on maturity date}}{\text{Level of S\&P 500 on issuance date}} - 1$. This product provides 70% of the underlying return at the maturity if the underlying return is greater than 0%. For example, if an investor invests \$100 on the issuance date and the underlying return is 10%, the investor receive \$107 at maturity. Otherwise, the investor receives the principal, \$100, irrespective of the size of any negative shock to the underlying asset. Therefore, the investor sacrifices 30% of the upside while his/her principal is fully protected in any case.

B. Non-capital-protected participation products

The second type of SRP is non-capital-protected participation product presented in Figure 1.2, Panel B. The name of this product is *Leveraged Buffered Notes* and was issued by Goldman Sachs on April 22, 2014. This product is linked to the Eurostoxx 50 index for a two-year term. The payoff structure of this product is similar to that of the previous example with the main difference being that the capital is not fully protected. Investors enjoy 159% of the underlying return if the underlying return is positive. Their capital is protected until the underlying return reaches -15% and investors begin to lose capital at a rate of 1.1765% for every 1% fall in excess of -15%. Although investors enjoy a high participation rate (159%), they could lose the capital they invest in a bear market.

C. Reverse convertibles

The third example of an SRP is reverse convertible, which was once popular in the United States. This is known as one of the most typical types of SRPs. In Figure 1.2, Panel C I present an example of a reverse convertible issued by JP Morgan Chase on May 21, 2015, which was linked to the Linkedin stock price. The term of this product is one-year and investors earn 8.5% annualized coupons paid every quarter during the tenure of the product, which means that a 2.15% of coupon is paid every quarter. This coupon rate is quite high considering that this product was issued in a low-interest environment. Such a high coupon rate is generated by the special structure of the reverse convertible: a significant downside risk. If the underlying return has ever fallen below the knock-in barrier, -30%, during the investment period, investors receive only the underlying return at maturity. Otherwise, investors receive their full invested money back at maturity. The payoff structure around the knock-in barrier is similar to the short position of a put option. Intuitively, when an investor buys a reverse convertible, he/she effectively sells a put option to the issuer, and therefore the issuer can generate the high coupon rate using the put option.

D. Autocalls

The last SRP type I introduce here is the autocall. One example of an autocall, called *Autocallable Optimization Securities with Contingent Protection*, was issued by HSBC on March 26, 2008 with a term of 1.5 years. The underlying asset of this product is the S&P 500 index. The underlying return is the defined similarly, $\frac{\text{Level of S\&P 500 at time t}}{\text{Level of S\&P 500 on issuance date}} - 1$, but I use the time subscription t for the definition of the underlying return. Unlike what occurs with the participation products, the underlying return here is monitored quarterly to determine the redemption amount. The quarterly dates are called call dates. On the first call date, if the underlying return is greater than the knockout barrier, which is 0% in this example, 6% of the coupon along with the principal is paid to the investor and

this product is terminated. I interpret this to mean that this product is automatically called by the issuer if a specified condition is met, which is why this product type is termed an *autocall*. If the product is not called because the underlying return is lower than the knockout barrier, then I check this condition on the second call date, which is six months after the issuance date. If the underlying return on the second call date is greater than the knockout barrier, 12% of the coupon along with the principal is paid and the product is terminated as on the first call date. The investor carries out this procedure again if the product is not automatically called on any call date because of low returns on call dates.

If an autocall fails to be automatically called by maturity, we need to determine whether the underlying return has ever fallen below the knock-in barrier, which is -50%, during the investment period. If it has ever fallen below the knock-in barrier, an investor receives only the underlying return at maturity. If the underlying return has never fallen below the knock-in barrier, the investor receives 36% over the term of the product (1.5 years). This translates to an annualized coupon rate of 24%. Such a high coupon rate can be provided because investors effectively sell put options to issuers on the issuance date as is the case with reverse convertible purchases.

1.2.3 Investor characteristics

Who buys SRPs? Unfortunately, to the best of my knowledge, there is no comprehensive study of SRP investors in major SRP markets. I utilize the results of a survey of 1,049 SRP investors conducted in August 2010 by the Financial Supervisory Service, a finance watchdog in South Korea, to provide a synopsis of investor characteristics. South Korea is the second biggest market next to the United States in terms of cumulative sales volume of equity-linked SRPs. Although the survey is confined to South Korea, I believe it nevertheless provides meaningful information pertaining to the characteristics

of investors in major SRP markets. I summarize the detailed results of the survey in Appendix C.

The average age of respondents is 42 years and 81.4% report having earned a college degree.[3] The average current investment in SRPs is around $35,000. Although the average asset value is hard to calculate based on this the survey data, the median asset value is in the range of $100,000 - $500,000. Considering that the median asset value of households in Korea in 2010 was $140,000 (according to Statistics Korea), it seems that SRP investors do not hold more value in assets than the average household in Korea. Regarding risk preference, 66.7% of respondents consider themselves to be "risk-seeking" or "extremely risk-seeking." Their occupations are not particularly skewed to finance-related jobs since only 29.2% of them have such jobs. One of the most common reasons why investors purchase SRPs is because expected returns of SRPs are high. Overall, the SRP investors comprising the sample have higher education attainment levels than the average Korean are generally risk-seeking investors, but their other characteristics are not particularly different from the South Korean population as a whole.

1.3 Data and Empirical Design

1.3.1 Data

A. SRP data

In this study, I analyze equity linked-SRPs issued in the United States between 2001 and 2016. Equity linked-SRPs have stock market indexes or individual stocks as underlying assets. I obtain the sample of SRPs from Structured Retail Products, a data provider that specializes in the structured retail products industry. Conducting academic research on SRPs involves several challenges. The first challenge is that there

[3]In 2010, the average college graduation rate in South Korea was about 40%.

is no publicly available database that includes all issued SRPs. The second challenge is that information pertaining to each SRP is available only in text format. Issuers of SRPs provide prospectuses that describe essential information about the SRPs to the public in text format.[4] However, because each issuer has its own forms for its prospectuses, it is impossible to extract essential information based on a simple technique. Due to these two bottlenecks, academic research on SRPs remains skimpy relative to research on other financial products.

The data provider, Structured Retail Products, addresses these challenges by collecting prospectuses from multiple sources, including issuers themselves, and analyzing each prospectus. Through this procedure, the data provider has constructed the most comprehensive database on SRPs that covers all SRPs issued in all countries. An example of the format of the data is provided in Table 1.1.

[Insert Table 1.1 here.]

Following Célérier and Vallée (2017), among the SRPs that the database covers, I choose publicly available retail tranche-type products that are considered non-standardized SRPs. Other types of products, such as flow products, are standardized products and are usually traded on exchanges. Such products share similar characteristics with plain vanilla options.[5] Altogether, my sample consists of 71,300 with and total sales volume of $332.87 billion.

B. Market data

I use Bloomberg and CRSP as a source of market data for prices and dividend yields of underlying assets. To measure SRP markups and performance of SRPs, having

[4]See Appendix A for an actual example of the prospectus of an *Autocallable Optimization Securities with Contingent Protection* product, offered by HSBC on March 26, 2008.

[5]I also exclude call overwriting and warrant products because these are also considered standardized products. The proportion of these products is very small, and therefore excluding them does not change the main results of my analysis.

data on exchange-traded options is essential. I obtain exchange-traded option data from Optionmetrics USA for options for individual stocks and indexes in the United States. I obtain credit default swap spread data from Markit to measure issuers' credit risk. Finally, to obtain market risk premiums for the S&P 500 index, I access the data calculated on an annual basis based on the free-cash-flow approach described in Damodaran (2018). The data are available on his website.[6]

1.3.2 Empirical Design

A. Complexity and Extrapolation Hypothesis

How do investors deal with financial complexity? Brunnermeier and Oehmke (2009) introduce several approaches that bounded rational investors can take to deal with financial complexity. First, investors divide complex products into several pieces that are relatively easy to analyze. Second, investors use "building blocks." For example, when making a decision to buy Goldman Sachs stocks, investors combine the stock price of Goldman Sachs, a building block, with his own private information rather than trying to conduct a full-fledged bottom-up valuation exercise. Finally, investors can use a simple model such as Black Scholes formula to understand complex financial products although such simple model doesn't fully capture complicated nature of reality.

It is worth noting that the three SRP examples that I introduced in the previous section except autocalls are easily analyzed using the above approaches. For example, the capital-protected participation product example in Section 1.2.2 is divided into a call option and a bond. Similarly, the non-capital-protected participation product example is decomposed into a call option, a short position in put option, and a bond. The reverse convertible example can be decomposed into a coupon bond and a short position in a knock-in put option, whose closed-form solution is available (Derman and Kani, 1997).

[6]Equity risk premium `http://pages.stern.nyu.edu/~adamodar/`

However, how can investors deal with complexity if none of the above approaches can be employed because the products are very complex? Autocalls are good examples of such products. It is known that autocalls cannot be decomposed into several pieces that are easy to analyze, and there is no building blocks to use. In addition, there does not exist a closed-form solution of the price for an autocall even under very simple pricing assumptions (Hansson, 2012).

I conjecture that investors might depend on extrapolative expectations in purchasing complex financial products, under which investors estimate expected returns of products based on past ex-post returns. Extrapolation has been studied extensively in behavioral finance (Amromin and Sharpe, 2009; Greenwood and Shleifer, 2014; Vissing-Jorgensen, 2003). Extrapolation has been successful in explaining various asset pricing puzzles and particularly successful in explaining behaviors of stock returns compared to other behavioral finance theories (Barberis, 2018). Since equity-linked SRPs have stock indexes or individual stocks as underlying assets, it is natural to conjecture that investors exhibit extrapolative extrapolations in the market for equity-linked SRPs. In addition, the tendency to rely on extrapolation might be stronger for more complex products because it is harder for investors to rely on other approaches to deal with complexities for those products. Taken together, I propose the following testable hypothesis, which I term *Extrapolation Hypothesis*.

Hypothesis (Extrapolation Hypothesis)

(i) Investors buy more SRPs that provided higher past returns.

(ii) Investors respond more aggressively to the past returns in purchasing more complex products.

B. Construction of group-level data

Testing the *Extrapolation Hypothesis* is challenging in the market for SRPs. SRPs are not standardized and a SRP is not issued repeatedly. Therefore, measuring the past return for a product, which is essential for testing the *Extrapolation Hypothesis* is impossible. To resolve this challenge, I categorize SRP products into several groups and construct "group-level data" at each time based on "product-level data."

First, I systematically categorize SRPs into five types using simple criteria, motivated by several benchmark categories including SEC.[7]

[Insert Figure 1.3 here.]

The most important criterion that I use to split the SRP product space into two groups is the *yield-enhancement property*. Products with the yield-enhancement property deliver high coupons relative to benchmark rates by exposing investors to the tail risk. Investors buying these products implicitly sell a type of put options to issuers and issuers offer high coupon rates in normal times through sales of the put options. The yield-enhancement products group is further subdivided into two groups based on *the fixed maturity criterion*. Products with fixed maturities are called *reverse convertibles* and products with non-fixed maturities are called *autocalls*.

In the non-yield-enhancement products group, there are diverse product types and it is challenging to categorize them systematically. Therefore, I choose to introduce one criterion that defines a big product group called *participation products* and define the rest as *'Others'*. Participation products are products whose returns at their maturities are determined by underlying returns at maturities and do not depend on the intermediate underlying returns over the investment period. Participation products are divided further into two groups: capital-protected, and non-capital-protected participation products, depending on whether the investors' capital is protected in any case

[7]SEC. `https://www.sec.gov/oiea/investor-alerts-bulletins/ib_structurednotes.html`

or not. The 'Others' product group is much smaller than other groups and, therefore, grouping them into one separate category does not significantly affect the main results.

I further divide each category into several groups based on underlying assets and construct the group-level variables at each time by taking the medians of the product-level variables. I present the detailed procedures in the next section.

1.4 Methodologies

1.4.1 Text mining for payoff structures

One of the bottlenecks in SRP research is that raw data are collected only in unstructured text format. This is particularly true in SRP markets, where the prospectuses that contain information on payoff descriptions of the products are provided by issuers in text formats. They are quite heterogeneous across issuers and across products. Fortunately, Structured Retail Products collects all prospectuses and summarizes the payoff descriptions in plain English, as seen in Table 1.1. The next challenge is to fully interpret payoff descriptions. Purely statistical learning-based approaches are prone to error, and more importantly, even if errors occur, it is hard to explain why such errors occur. In addition, analyzing each payoff description is unrealistic because of a large sample size, 71,300. Therefore, I take a simple and precise approach to fully interpret the payoff structures via the following steps.

First, I decompose the payoff descriptions into three parts: *coupon description*, *early termination description*, and *maturity payoff description*. Then, for each part, I group the products whose descriptions have the same structure. I explain the basic idea of my text mining approach using the following hypothetical example for the maturity payoff description.

Payoff description of product 1: *At maturity, the product offers a capital return of 100%, plus 300% of the rise in the index over the investment period, subject to a maximum overall return of 114%. If the final index level is lower than its initial level, the capital return is 100% minus 1% for every 1% fall.*

Payoff description of product 2: *At maturity, the product provides a capital return of 100%, plus 200% of the rise in the underlying asset over the period of the investment, subject to the maximum cap of 124%. If the final underlying asset is lower than its initial value, the capital return is 100% minus 1% for every 1% fall.*

The two paragraphs above describe the maturity payoff descriptions of two products. However, these seemingly different descriptions actually have the same structure. The main differences are in the expressions that each description employs. For example, the word *offers* in the first paragraph has the same meaning as the word *provides* in the second paragraph. Likewise, the following pairs of words have the same meanings: *(offers, provides), (index, underlying asset), (maximum overall return, maximum cap), (a, the), (initial value, initial level).* By analyzing random samples of product descriptions, I construct a comprehensive dictionary that groups expressions that have similar meanings or communicate the same idea. Finally, I replace expressions that have similar meanings with a single expression. For example, I map *offers* and *provides* into *offers*. Through this approach, I form a manageable number of groups of product descriptions. By analyzing these groups of product descriptions, I fully interpret almost the entire sample of products. Of the 71,300 entire sample of products, I fully interpret 69,877 product descriptions. The remaining 1,423 product descriptions were incomplete and could not be analyzed.

1.4.2 Estimation of main variables

In this subsection, I explain how to estimate main variables of interest: annualized realized return, markup, past return, and volatility. Estimating markup, past return, and volatility, requires making sophisticated modeling assumptions. More assumptions are needed for SRPS that have multiple underlying assets, and these estimations are heavily driven by the assumptions. Therefore, in this work, I exclude the SRPs that have multiple underlying assets and focus on SRPs that have a single underlying asset. I use full sample to estimate annualized realized returns because it requires no modeling assumptions.

Annualized realized returns

I calculate annualized realized return by combining the text mining described in the previous section and market data on underlying assets. A typical SRP provides a stream of coupons on top of a redemption amount at maturity. In the spirit of the bond yields, I define the annualized realized return, r^{ARR}, for an SRP based on the following formula,

$$0 = -P + \frac{c_{T_1}}{(1+r^{ARR})^{T_1}} + \frac{c_{T_2}}{(1+r^{ARR})^{T_2}} + ... + \frac{c_{T_N} + X_{T_N}}{(1+r^{ARR})^{T_N}}, \tag{1.1}$$

where c_{T_i} represents intermediate coupon payments at time T_i, X_{T_N} is the redemption amount at maturity (T_N), and P is the price of the product. T_N is a fixed time for non-autocalls and it is a stopping time, a random variable that is a function of the performance of underlying assets, for autocalls. The solution of the above nonlinear equation is unique because c_{T_i} and X_{T_N} are nonnegative and the sign of the payoff stream changes only once when the first payment occurs. Therefore, r^{ARR} is a well-defined measure. After applying the text mining techniques to determine the stream of coupons (c_{T_i}) and the redemption amount (X_{T_N}), I solve the equation to obtain r^{ARR}.

Markups

Estimating markup requires modeling assumptions regarding the dynamics of underlying assets. I employ the stochastic volatility model proposed by Heston (1993).[8] The stochastic volatility model is commonly used in practice in derivative pricing. It is proposed to explain anomalies generated from simple assumptions regarding the process of underlying assets and it is a good alternative for resolving the issues arising from simple pricing assumptions such as Geometric Brownian Motion (GBM). It assumes the following dynamics of the price (S_t) and variance (V_t) of an underlying asset under a risk-neutral measure, Q:

$$\begin{aligned} \frac{dS_t}{S_t} &= (r_f - q)dt + \sqrt{V_t}dW_t^1 \\ dV_t &= a(\bar{V} - V_t)dt + \eta\sqrt{V_t}dW_t^2 \\ dW_t^1 dW_t^2 &= \rho dt \end{aligned} \tag{1.2}$$

where r_f is the risk-free rate, q is the dividend yield, $\bar{V}$ is long-run variance, η is the volatility of V_t, and ρ is instantaneous correlation coefficient between variance and the underlying asset. I assume that the risk-free rate (r_f) and dividend yield (q) are constant over time. I use the yield of treasury bonds with the same term of the corresponding SRP to obtain the risk-free rate. For dividend yield (q), I use the implied dividend yield provided by Optionmetrics USA if it is available. If the implied dividend yield is unavailable and options are European style, I use put-call parity to find the implied dividend yield of options proposed by Van Binsbergen et al. (2012), as follows:

$$q = -\frac{1}{T}\log\left(\frac{C(K,T) - P(K,T) + Ke^{-r_f T}}{S_0}\right) \tag{1.3}$$

[8]The seminal papers by Black and Scholes (1973) and Merton (1973) derive the price of call options under the assumption that prices follow Geometric Brownian Motion (GBM). The limitation of this approach is highlighted in many subsequent studies including Aït-Sahalia and Lo (1998). In particular, Geometric Brownian Motion (GBM) fails to explain the pattern of an implied volatility surface across different strike prices and maturities.

where $C(K,T)$ and $P(K,T)$ are the prices of the call and put options with the strike price and time to maturity, K and T, respectively. Most individual stocks options are American style and the implied dividend yield from Equation (1.3) is not valid since put-call parity does not hold for American options. In such a case, I calculate the dividend yield by dividing the 12-month trailing dividends of individual stocks by individual stock prices.

I first calibrate the model using available exchange-traded option data. Heston (1993) derives the closed-form solutions of European call options, $C^{Heston}(K,T;a,V_0,\bar{V},\eta,\rho)$, and they are presented in Appendix B. I calibrate the stochastic volatility model for an underlying asset on each issuance date of a corresponding SRP by minimizing the following least square of the difference between market-implied volatility and implied volatility derived from the Heston model, denoted as σ^{Heston}:

$$(a^*, V_0^*, \bar{V}^*, \eta^*, \rho^*) = \underset{a,V_0,\bar{V},\eta,\rho}{\arg\min} \frac{1}{N} \sum_i \left(\sigma^{Heston}\left(K_i, T_i; a, V_0, \bar{V}, \eta, \rho\right) - \sigma(K_i, T_i)\right)^2$$

where $\sigma(K_i,T_i)$ is the implied volatility of call options traded in secondary markets with strike price and time to maturity K_i and T_i, $\sigma^{Heston}\left(K_i,T_i;a,V_0,\bar{V},\eta,\rho\right)$ is the implied volatility of the call option price derived from the Heston model with a given set of parameters, $(a,V_0,\bar{V},\eta,\rho)$, and N is the number of options available in the market.

After I find the optimal set of parameters, $(a^*,V_0^*,\bar{V}^*,\eta^*,\rho^*)$, I generate 10,000 paths for the underlying asset price, S_t, from the stochastic volatility model at the optimal set of parameters. I employ an Euler Scheme to generate the prices by iterating the following two equations:

$$\log(S_{t+\Delta}) = \log(S_t) + \left(r_f - q - \frac{1}{2}V_t\right)\Delta + \sqrt{V_t}Z_X\sqrt{\Delta}$$

$$V_{t+\Delta} = V_t + a(\bar{V} - V_t)\Delta + \eta\sqrt{V_t}Z_V\sqrt{\Delta}$$

where Z_X and Z_V are standardized Gaussian variables with correlation coefficient ρ and Δ is the time interval. I use $\Delta = \frac{1}{252}$ assuming that there are 252 trading days a year. I define markup and annualized markup as follows:

$$\begin{aligned} Markup &= \frac{P - FV_0}{P} \\ Annualized\ Markup &= \frac{1}{T}\left(\frac{P - FV_0}{P}\right) \end{aligned} \tag{1.4}$$

where P is the price of an SRP and FV_t is the fair value of an SRP at time t, with the following form:

$$FV_t = E_t^Q\left(\int_t^T e^{-r_f(s-t)} f(S_s) ds\right) \tag{1.5}$$

where f is the payoff function of an SRP. Here, subscript 0 in FV_0 indicates the issuance date.

Past returns

To test extrapolative expectations, measuring past returns on SRPs is essential. Because there are no secondary markets for SRPs in my sample, it is impossible to measure returns using historical prices. In addition, the varying lengths of terms across products make the direct use of annualized realized returns unreliable as a measure for past returns. To resolve this issue, I measure the fair value of a product, FV_t, at the end of every quarter. At the end of each quarter, t, I estimate the past return (r_t^{Past}) over the last one-year period, by solving the following equation for r_t^{Past}:

$$0 = -FV_{t-4\ quarters} + \frac{c_{T_1}^{Intermediate}}{(1 + r_t^{Past})^{T_1}} + \frac{c_{T_2}^{Intermediate}}{(1 + r_t^{Past})^{T_2}} + ... + \frac{c_{T_M}^{Intermediate} + FV_t}{(1 + r_t^{Past})^{T_N}}. \tag{1.6}$$

where $c_{T_i}^{Intermediate}$ $(i = 1, 2, ..., M)$ is the coupon paid between (t-4 quarters, t] and FV_t and $FV_{t-4\ quarters}$ are calculated based on Equation (1.5). Intuitively, r_t^{Past} measures

the holding period return between t-4 quarters and t, assuming that the market price at time, t, is the same as FV_t.

If the product matures in (t-1 quarter, t], I assume that the product matures at time, t. If the term of an SRP is not longer than one year, I use P, the price of an SRP at issuance, instead of $FV_{t-4\ quarters}$ in Equation (1.6). Then, Equation (1.6) reduces to Equation (1.1) and in that case, r_t^{Past} is equal to annualized realized return. r_t^{Past} is an important input to estimate *Past return* of a group of SRPs, which I explain in detail in Section 1.4.4.

To evaluate E_t^Q, I generally follow the approach for estimating markups described in the previous subsection. I use the stochastic volatility model calibrated at issuance and update the risk-free rate (r_f) and dividend yield (q) when I evaluate E_t^Q in each quarter t. If exchange-traded option data are not available for some underlying assets, I estimate historical volatility based on past one-year monthly returns on underlying assets and use GBM to generate sample paths of prices. I use a relatively short period for the data sample, one year, to estimate volatility to capture a short-term trend in volatility.

Volatility

I calculate the volatility of an SRP to measure its risk. Since there are no secondary markets for SRPs in our sample, it is impossible to estimate the volatility using historical returns. I take advantage of estimates for the volatility and the beta of underlying assets to resolve this issue. I estimate the volatility of SRPs linked to the S&P 500 index or individual stocks of the companies in the United States.

I estimate volatility for an underlying asset of an SRP on its issuance date by taking a standard deviation of the past three years of monthly returns. To estimate the beta for the underlying asset, I regress, excess returns on the underlying asset, defined

as returns on the underlying asset minus the three-month treasury bill rate, on the contemporaneous excess market using the past three years monthly data.

I denote estimates for volatility and beta of underlying asset as $\hat{\sigma}$ and $\hat{\beta}$. To estimate the volatility and beta of an SRP, I generate 10,000 sample paths of the price of each underlying asset based on the following equation:

$$\frac{dS_t}{S_t} = (\hat{\mu} - q)dt + \hat{\sigma}dW_t. \tag{1.7}$$

where $\hat{\mu} = r_f + \hat{\beta}\ E(r_m - r_f)$ and q is dividend yield. For dividend yield q, I divide the 12-month trailing dividend of the underlying asset by the underlying asset price at issuance. I use the estimate for the market risk premium, $E(r_m - r_f)$, which is available on Damodaran's website.[9] I use the treasury bond yield with the term that closely matches the term of the SRP to obtain risk-free rate. For each sample path, I calculate annualized realized return using Equation (1). Finally, to estimate the volatility of an SRP, $\hat{\sigma}^{SRP}$, I take the standard deviation of the annualized realized returns calculated from each sample path.

1.4.3 Validity of text mining

I provide evidence that my text mining technique reliably extracts the full information on payoff structures by comparing the annualized realized returns on SRPs provided by the data provider, which does not cover entire sample, and my estimates obtained from the text mining technique.

[Insert Figure 1.4 here.]

[9] `http://pages.stern.nyu.edu/~adamodar/`

Figure 1.4 is based on 7,411 products, a subset of the whole sample for which annualized realized returns are provided by the data provider.[10] The correlation coefficient is 0.97 and this high correlation is strong evidence that my text mining technique reliably extracts the full information on the products.

1.4.4 Variables construction

I first define five types of SRPs for my main analysis according to Section 1.3.2: capital-protected participation products, non-capital-protected participation products, reverse convertibles, autocalls, and others. I have 1,821 unique underlying assets in my sample period and I group them according to the following criteria. I group individual stock prices, based on the Bloomberg Industry Classification Standard (BICS). I group stock market indexes, based on countries. For example, the Nikkei 225 index and the TOPIX index are grouped into the *Japanese stock market index* group and KOSPI 200 and KOSDAQ are grouped into the *Korean stock market index* group.[11] At each time (t), I measure the variables of interest at underlying asset group (i) by product type level (j).

The main outcome variable is *Normalized sales*$_{i,j,t}$. *Normalized sales*$_{i,j,t}$ is defined as $\log\left(1+\frac{Sales_{i,j,t}}{Outstanding_{i,j,t-1}}\right)$, where Sales$_{i,j,t}$ is the aggregated sales volume for underlying group (i) and product type (j) at time t, and *Outstanding*$_{i,j,t-1}$ is the aggregated volume of SRPs with underlying group (i) and product type (j) that have not yet matured at

[10]Here, I choose products that do not provide coupons during the investment periods because I find that the current data set from the data provider sometimes ignore coupon payments when calculating annualized realized returns.

[11]I make one exception here by separating the S&P, Russell, and NASDAQ indexes. These are all based on US stock prices but both are sufficiently large underlying assets. S&P indexes such as the S&P 500 and the S&P 100 are grouped into *S&P index* group and Russell indexes such as the Russell 2000 and the Russell 1000 are grouped into *Russell index* group. NASDAQ has only one index in my sample, the NASDAQ 100.

time $t-1$.[12] I take log in order to guarantee that my result is not heavily driven by a few of outliers. This measure is analogous to the fund-flow measure in Chevalier and Ellison (1997).

Another important variable is *Past return*$_{i,j,t-1}$. To calculate this measure, I take the median of past returns (r_{t-1}^{Past}) on products with underlying group (i) and product type (j) at time $t-1$ by following the approach described in Section 1.4.2.

I also measure other variables for controls in a similar manner: *Annualized realized return*$_{i,j,t}$, *Annualized markup*$_{i,j,t}$, *Volatility*$_{i,j,t}$, *Sales commission*$_{i,j,t}$, *Term*$_{i,j,t}$, *One-year CDS spread*$_{i,j,t}$, *Complexity (Autocall)*$_{i,j,t}$, and *Complexity (# characters)*$_{i,j,t}$. These variables are defined similarly. I first measure *Annualized markup (%), Sales commission (%), Term (years), One year CDS spread (%), Complexity (# characters)* and *Volatility (%)* at the product-level. In particular, I follow Célérier and Vallée (2017) to define *Complexity (# characters)* as the number of characters that are used in payoff descriptions and the unit is 1,000 characters.[13] After measuring these variables at the product-level, I take the median of the variables that have the same underlying group (i) and product type (j) and are issued at the same time t. I choose samples whose *Past return* information is available because this is the main explanatory variable. Finally, to prevent my results from being driven by very small SRPs, I further choose samples whose *Outstanding*$_{i,j,t-1}$ is larger than \$35 million. Basic product-level and group-level summary statistics for my sample are presented in Table 1.2. The group-level data for my main analyses is measured at quarterly frequency.

[Insert Table 1.2 here.]

[12] If an SRP has more than one underlying asset and they are in different underlying groups, I assume that this product contributes $\frac{1}{\text{number of underlying assets}}$ of sales volume to *Sales* and *Outstanding* of each underlying group.

[13] For autocall products, I also add the number of characters used in the *Early redemption information* part as in Table 1.1.

1.5 Empirical Results

1.5.1 Ex-post performance of SRPs

As emphasized in Amy Starr's speech at the SEC (Starr, 2015), understanding the ex-post performance of SRPs is the first step toward understanding SRP market. Using the text mining technique described in Section 1.4.1, I construct annualized realized returns on 51,606 products. This is the largest sample ever studied and, most importantly, this sample should not be susceptible to any serious selection problem. The basic results are presented in Table 1.3.

[Insert Table 1.3 here.]

Since the beginning of the market, SRPs issued in the United States have provided on average 2.07% of returns with a standard deviation of 22.73%. The sales-volume-weighted average return is 3.28% with a sales-volume-weighted standard deviation of 18.99%. I use the sales volume-weighted performance results extensively hereafter to prevent my results from being potentially driven by small SRPs.

During the same sample periods, the one-year Bank (AAA) Bond has provided 3.7% of average returns with a standard deviation of 1.43% and the S&P 500 index has provided 6.39% of average returns with a standard deviation of 14.6%. Therefore, SRPs on average have provided worse performance than other simple benchmarks such as bond and stock market index.

It is worth noting that each type of SRPs has performed differently. Autocalls have provided 7.59% of average returns with a standard deviation of 13.44%. Non-capital-protected participation products have produced 4.00% of average returns with a standard deviation of 16.32% and capital-protected participation products have provided 3.3% of average returns with a standard deviation of 5.19%. Therefore, some product types such as autocalls have performed well, outperforming the the S&P 500 index in terms

of average returns and standard deviations. Although it is difficult to compare the standard deviations of annualized realized returns of SRPs with the standard deviation of benchmark returns directly, I believe that such a comparison provides meaningful information.

Interestingly, the ex-post performance of reverse convertibles is under 0%. Reverse convertibles on average have provided -2.84% of returns with a standard deviation of 30.08% although such poor performance is driven mainly by products issued around the financial crisis. The poor performance of reverse convertibles is reflected in several litigation cases against issuers of reverse convertibles and special attention from the media.[14] Reverse convertibles have also been thoroughly studied in many academic papers and have served as typical examples of the "dark side of financial innovation" (Szymanowska et al., 2009; Wallmeier and Diethelm, 2008; Egan, 2018). In addition, the performance of 'other' type is also poor.

[Insert Figure 1.5 here.]

Figure 1.5 displays distribution of annualized realized returns for each type of SRPs. A notable feature is that reverse convertibles have a thick left tail compared with the other types. A significant portion of autocalls experienced gains and the left tail is thin. The distribution of non-capital-protected participation products looks relatively smooth compared with the other types. Beginning in the next subsection, I discuss extrapolative expectations in detail by investigating how past returns on SRPs explains higher sales volume.[15]

[14]FINRA Fines RBC $1.4 Million Over Unsuitable Reverse Convertibles. `https://www.securitieslawyersblog.com/finra-fines-rbc-1-4-million-over-unsuitable-reverse-convertibles/`

[15]Sources of the information on ex-post performance of SRPs vary across countries. Although investors can learn about the ex-post performance based on their own investments or by consulting with brokers, there are also publicly available sources of information. For example, in the United Kingdom, the UK Structured Products Association (UKSPA) keep track of all SRPs and provides monthly performance reports based on the ex-post performance of SRPs. Also, CompareStructuredProducts provides summary statistics of ex-post performance of SRPs.

1.5.2 Extrapolation

In this subsection, I test the *Extrapolation Hypothesis* that I propose in Section 1.3.2.

Hypothesis (Extrapolation Hypothesis)

(i) Investors buy more SRPs that provided higher past returns.

(ii) Investors respond more aggressively to the past returns in purchasing more complex products.

[Insert Table 1.4 here.]

The main dependent variable is *Normalized sales* whose unit is %. As I show in the first column of Table 1.4, Panel A, there is a strong univariate relationship between *Past return* on SRPs and *Normalized sales*. Based on the coefficient, 0.264, seen in the first column, one-standard-deviation increase in *Past return* translates into a 4.14% increase of *Normalized sales*, where the standard deviation of *Normalized sales* is 19.4%. The positive relationship stays strong even after controlling for the underlying group and quarter fixed effects.

To demonstrate the differential effects of extrapolation on sales of products with varying degree of complexity, I interact the two complexity measures, *Complexity (Autocall)* and *Complexity (# characters)*, with *Past return* in columns (3)-(6). The magnitude of the coefficient of the interaction term, 0.443, as seen in column (3), is sizable. This means that, for autocalls, which are considered the most complex SRPs, a one-standard-deviation increase in *Past return* translates into a 9.56% increase in *Normalized sales*. This effect stays strong even after controlling for underlying group and quarter fixed effects, as seen in column (4). I perform a similar exercise with another complexity measure, *Complexity (# characters)*. As explained in Section 1.4.4, this measure counts the number of characters in a unit of 1,000 characters that describes the payoff structures of SRPs. The coefficient of the interaction term, 0.564, that I report in column (5) implies that having 1,000 more characters in the payoff descriptions translates into an 8.85%

increase in *Normalized sales* for every one-standard-deviation increase in *Past return*. The results that I report in column (6) indicate that controlling for underlying group and quarter fixed effects generally does not change the results.

The overall results remain generally unchanged even with more controls. As can be seen in Panel B of Table 1.4 I perform the same exercise with control variables: *Sales commission*, *Term*, and *One-year CDS spread*. My *Annualized markup* and *Volatility* data cover roughly 50% of the entire sample. Therefore, I provide results for regressions with control variables, including *Annualized markup* and *Volatility* separately in Panel C. The statistical and economic significance of these estimates for my main variables of interest, *Past return* and *Past return* $\times$ *Complexity*, which I report in Table 1.4, Panel B and Panel C, remain close to those reported in Table 1.4, Panel A.

Among the coefficients of control variables, the coefficients of *Term* are generally negative and statistically significant, which implies that investors tend to prefer short-term over long-term SRPs. The coefficients of *One-year CDS spread* are also generally negative and statistically significant. This suggests that investors prefer purchasing SRPs with lower credit risk of issuers. The coefficients of *Annualized markup* are generally negative and statistically significant, whereas the coefficients of *Sales commission* and *Volatility* are generally insignificant.

In order to validate that investors' extrapolations persist over a long period of time, I run the same regression analysis on two subgroups of the full sample, split before and after December 31, 2012. I find that the results are similar to the results from the full dataset. This observation suggests that investors' extrapolations persist for this particular time period.

Overall, the results that I report are consistent with the *Extrapolation Hypothesis*. A similar phenomenon is well documented in other markets, such as mutual funds (Chevalier and Ellison, 1997; Sirri and Tufano, 1998). My results differ from results reported in

previous literature on other markets insofar as investors depend on extrapolation when deciding whether to purchase *complex* financial products. More importantly, investors are more sensitive to past returns when they purchase more complex products. One possible interpretation of this phenomenon is that unsophisticated investors may chase high past returns on products and this effect is stronger when they buy more complex products because dealing with complexity of such products is harder and extrapolation becomes a more important influence on decision-making (Brunnermeier and Oehmke, 2009).

1.5.3 Extrapolation (Intensive and extensive margins)

In this subsection, I briefly describe the dynamics of issuance patterns by looking closely at the intensive and extensive margins of sales volume. My main outcome variable, *Normalized sales*$_{i,j,t}$, is defined as $\log\left(1+\frac{Sales_{i,j,t}}{Outstanding_{i,j,t-1}}\right)$. I decompose $Sales_{i,j,t}$ into two parts, $Sales_{i,j,t}^{Incumbent\ banks}$ and $Sales_{i,j,t}^{New\ banks}$. $Sales_{i,j,t}^{Incumbent\ banks}$ measures the sales volume of products with underlying asset group i and product type j at time t sold by banks that marketed the same group of products in the previous period. $Sales_{i,j,t}^{New\ banks}$ measures the sales volume of products with underlying asset group i and product type j at time t sold by banks that did not sell the same group of products in the previous period.

[Insert Table 1.5 here.]

In Table 1.5, Panel A, I present summary statistics on the *Proportion of sales volume of incumbent banks*, defined as $\frac{Sales_{i,j,t}^{Incumbent\ banks}}{Sales_{i,j,t}}$. The average of the *Proportion of sales volume of incumbent banks* is 72.4% and the median is 87.7%, which means that most of the variation in sales is driven by incumbent banks and understanding the role of big banks is therefore important in this market.

I run regressions that are similar to those whose results are reported in Table 4, Panel A and report the results in Table 1.5, Panel B and Panel C but with the alternative dependent variables, $Normalized\ sales_{i,j,t}^{Intensive}$ and $Normalized\ sales_{i,j,t}^{Extensive}$, respectively. $Normalized\ sales^{Intensive}$ is defined as $log\left(1+\frac{Sales_{i,j,t}^{Incumbent\ banks}}{Outstanding_{i,j,t-1}}\right)$ and $Normalized\ sales^{Extensive}$ is defined as $log\left(1+\frac{Sales_{i,j,t}^{New\ banks}}{Outstanding_{i,j,t-1}}\right)$. The panel shows that the sales increase due to high *Past return* is covered mostly by incumbent banks. The coefficients for *Past return* and *Past return*×*complexity* are generally statistically and economically significant for $Normalized\ sales^{Intensive}$ and the coefficients are comparable to those reported in Table 1.4, Panel A. The coefficients are not statistically significant, though, for $Normalized\ sales^{Extensive}$. I also conduct the same analysis with more control variables as seen in Table 1.4, Panel B and find that the results are generally unchanged.

1.5.4 Extrapolation vs rational information extraction

The strong relationship between *Normalized sales* and *Past return* discussed in previous subsections might be explained by rational information extraction by retail investors. For example, if *Past return* contains information on returns of newly issued products with the same underlying group and same type, investors could demand more products that experienced high performance in the previous period and this explanation would be consistent with empirical patterns provided in previous subsections.

[Insert Table 1.6 here.]

In Table 1.6 I first present evidence confirming the relationship between annualized realized returns of newly issued products and *Past return*. The coefficients for *Past return* reported in columns (1) and (2) show that *Past return* does not explain the performance of newly issued products that have same underlying group and are of the same product type. I also study the relationship between annualized realized returns for

newly issued products and *Normalized sales* and report the results in columns (3) and (4). The coefficients for *Normalized sales* are statistically insignificant, which implies that high *Normalized sales* is not correlated with better performance.

Based on the results in this subsection, I rule out the possibility that investor decisions based on *Past return* are driven by rational information extraction. Also, the finding that high demand for product groups does not explain the performance of those product groups further strengthens my argument that investor decisions are not driven by rational information extraction.

1.5.5 Extrapolation: Evidence from surveys

In this subsection, I provide further evidence of investor extrapolation by introducing survey findings associated with SRP investors.

A. Investors have limited understanding of SRPs

The first important finding that is common to the multiple surveys is that investors have only limited understanding of SRPs and their investment decisions are not based on careful evaluation of the risk-and-return profiles of these products. This result is consistent with my empirical finding that investors' decisions are not driven by rational information extraction from past returns on SRPs.

A 2006 survey conducted by the Securities and Futures Commission (SFC) in Hong Kong is based on 207 SRP investors. To measure their understanding of the SRPs in which they invested, investors were asked what would happen if the price of the underlying asset were to drop below the lower barrier beyond which investors lose part or all of their invested capital. To this question almost half of the investors gave a wrong answer as most them did not realize they could lose part or all of their investment capital in such an adverse scenario. More interestingly, 48.9% of investors recalled that

they had received a prospectus for an SRP when they bought one. Also, only 10.7% of the investors responded that they had received and read and fully understood the document. Therefore, 89.3% of the investors either did not recall that they received a product prospectus or, even if they recalled receiving one, they did not read carefully or did not understand the details about the products in which they had invested.

Another survey was conducted in 2015 by Cha and Park (2015) with a sample of 336 SRP investors in South Korea. Investors were asked whether they understood the basic concepts related to SRPs: underlying assets, initial levels of underlying assets, volatility, knock-in barriers, and so on. Even though these concepts are essential to understanding the risks and returns associated with the products, investors on average answered that they understood 5.37 concepts out of 12 concepts. Similarly, Hunt et al. (2015) conducted a survey on 337 SRP investors in the United Kingdom in 2015. Investors were asked to answer questions about the basic concepts underlying SRPs. The proportion of correct answers ranged between 65% and 80% for most questions, but for some difficult question, the proportion of correct answers was only about 24%.

In summary, multiple surveys conducted in multiple countries consistently reveal that SRP investors are unsophisticated in the sense that they do not fully understand the details associated with products they purchase, suggesting that an explanation based on behavioral finance theory is a plausible alternative.

B. Suggestive evidence of extrapolation

Two surveys show that investors' past returns on SRPs are comparable to their expected returns of newly purchased SRPs, which is suggestive of extrapolation. In the survey conducted by Financial Supervisory Service (2010), investors were asked to provide the annualized realized returns on SRPs in which they invested in during the past two years and also about subjective expected returns on the products they

invested in recently. The average of annualized realized returns over the past 2 years is 10% and average of subjective expected returns for newly issued products is around 13%, which is suggestive of extrapolative expectation. More evidence is provided by Cha and Park (2015). In this survey, the average of annualized realized returns of the products in which investors invested over a period of time preceding the survey is 13.4% and average of subjective expected returns for newly issued products is around 15.5%, which is suggestive of extrapolative expectation as well. Cha and Park (2015) divide investors into two groups based on the type of products they purchased: capital-protected-products and non-capital-protected products. Over 90% of the non-capital-protected products are autocalls in South Korea and capital-protected products are in general considered to be the simplest SRPs.

Investors who purchased the simpler products experienced 14.84% of annualized realized returns in the past on average and the average of their subjective expected returns on the simpler products is 13.83%. In contrast, investors who purchased the more complex products experienced 13.11% of annualized realized returns in the past on average and the average of their subjective expected returns on the more complex products is 28.8%. Although the difference between the past annualized realized returns is not statistically significant, the difference between their subjective expected returns is sizable and statistically significant. This is consistent with the stronger extrapolation observed in more complex products. In this survey, the two groups of investors were also asked about their risk preference and level of overconfidence. Investors that purchased more complex products are relatively more risk seeking compared to the investors that purchased less complex products but the difference between their levels of overconfidence is not statistically significant. Also, it turns out that investors who buy more complex products are relatively more sophisticated and more experienced in the sense that their test scores on financial literacy are higher and their investment period is longer.

Caution is needed in interpreting the survey results because it is hard to establish a causal relationship solely based on survey results. However, I believe that my empirical results combined with the survey findings provide a useful starting point for understanding the role of extrapolative expectation in the market for complex financial products.

1.6 What Explains the Rapid Growth in Sales Volume and Complexity?

1.6.1 Explanation by extrapolation

In this subsection, I discuss how investors' extrapolative expectations explain the rapid growth in sales volume and complexity of the SRP market. I show that high ex-post performance of autocalls generated by chance in my sample period, combined with investors' more pronounced extrapolation for more complex products, explains the increasing complexity. Such rapidly increasing market share of autocalls contributes to the growth in overall sales volume.

The presence of investor extrapolation motivates issuers to offer a large variety of SRPs so that some of them would offer high returns and attract more investors over time. The large variety of SRPs have performed very differently over time and I report the sales-volume-weighted average of annualized realized returns on each type of SRP matured in each year in Table 1.7, Panel A. Autocalls have consistently outperformed other products in general over the sample period except that non-capital-protected products provided slightly higher returns during 2013 - 2015. Before 2013, reverse convertibles had performed poorly compared with other types of products but this trend becomes modest after 2013. The returns on capital-protected participation products are modest over the entire sample period. 'Others' generally performed poorly.

As shown in Section 1.5.2, investors exhibited stronger extrapolation in more complex products. Investors, after experiencing high ex-post performance of autocalls generated by randomness of underlying assets, aggressively buy these products more believing that they will provide high returns in the future. Therefore, the strong extrapolation in more complex products, combined with the good ex-post performance of autocalls, has led to the increasing complexity of the SRP market. The rapidly increasing autocall sales contributes to the growth of overall market as well.

Extrapolation also explains sales patterns of the other products. The poor ex-post performance of reverse convertibles and 'others' explains the sharp declines of the market shares of those products, as seen in Figure 1.1, Panel B. The modest performance of participation products is also consistent with the modest growth of this product type.

To provide sharper evidence supporting this explanation, I compare sales of reverse convertibles and autocalls. These types of products are designed to generate high yields by exposing investors to the tail risk. Investors enjoy high returns in normal times and are exposed to downside risk if the underlying assets crash. These similar products experienced opposite outcomes in sales. Autocalls have experienced steady growth whereas reverse convertibles that were once popular have experienced a sharp decline.

[Insert Figure 1.6 here.]

The trends in reverse convertibles and autocalls are more clearly illustrated in Figure 1.6. The blue bars in Figure 1.6 plot the relative performance of autocalls compared with that of reverse convertibles. The relative performance of autocalls (β_t) in each year t is estimated by the following equation:

$$\text{Annualized realized return}_{k,t} = \beta_t \text{ Complexity (Autocall)}_{k,t} + \text{Underlying FE}_{k,t} + \epsilon_{k,t}, \tag{1.8}$$

where *Annualized realized return*$_{k,t}$ is the annualized realized return on an SRP with product k that matured in year t and *Complexity (Autocall)*$_{k,t}$ is an indicator variable equal to 1 if an SRP is an autocall and 0 otherwise. This regression is run for each year t for the sample consisting of reverse convertibles and autocalls that matured in year t and I provide the result of the regression in Table 1.7, Panel B. The red (red dashed) line plots the proportion of sales for autocall (reverse convertible) in each year t compared with overall sales of SRPs. The results reported in the figure indicate that the performance of autocalls was much higher than that of reverse convertibles until 2012, and during the same period, autocalls experienced sharp growth while reverse convertibles experienced sharp decline. This pattern becomes modest after 2012 when the differential performance between autocalls and reverse convertibles becomes more modest. Therefore, the contrasting sales patterns of the two products are quite consistent with the extrapolation hypothesis, which explains the increasing complexity in the SRP market.

1.6.2 Other potential explanations

Reaching for yield

The growth in autocalls I observe might have been driven by investors seeking high yields in a low-interest rate environment. This phenomenon, called 'reaching for yield,' has been studied in several markets including the bond market (Becker and Ivashina, 2015) and the corporate bond mutual fund market (Choi and Kronlund, 2017).

Autocalls are designed to provide high returns in normal times when the performance of underlying assets does not experience any large negative shock. Therefore, it is natural to think that the growth in autocalls could be explained simply by 'reaching for yield' phenomenon. However, reverse convertibles, which are also designed to deliver high returns in normal times have experienced exactly the opposite paths. As discussed in

the previous subsection, Figure 1.6 shows a clearly declining pattern in the proportion of reverse convertibles sales of overall sales of SRPs. This implies that even if reaching for yield might be important for explaining the growth in autocall sales, it has limited utility in this respect because it does not clearly explain why the products that have a same goal as autocalls experienced a drop. Instead, the patterns of growth and decline that mark the sales of these products are more consistent with the extrapolation hypothesis.

Insurance demand by banks

As explained in Section 1.2.2, investors effectively sell put options to banks when purchasing autocalls. Therefore, it is possible that the growth of the autocalls might have been driven by insurance demand by banks to hedge their portfolio risks. However, this alternative explanation does not clearly explain the collapse of reverse convertibles. Banks can also secure put options by selling reverse convertibles. In addition, reverse convertibles are simpler products than autocalls because they do not have early termination properties, which allows banks to secure put options more easily by selling reverse convertibles than autocalls. Therefore, the salient opposite patterns of sales of reverse convertibles and autocalls are hard to be explained by the insurance demand by banks.

Hedging demand by investors

Investor demand for hedging offers another possible explanation. Investors might have purchased autocalls to hedge their portfolio risks. However, the evidence from the surveys of SRP investors conducted by Financial Supervisory Service in South Korea (Financial Supervisory Service, 2010) is inconsistent with the hedging demand hypothesis. In response to question 10 of the survey, which asks about investors' risk appetite, 66.7% of investors responded that they are "risk-seeking" or "extremely risk-seeking." In addition, in question 14, respondents were asked about the main reason for which they buy SRPs. The proportion of "others" (1.7%), which includes hedging purpose is

quite small. We see similar results from the other surveys. In the survey conducted by the Securities and Futures Commission in Hong Kong (Securities and Futures Commission, 2006), the proportion of answers that are related to hedging demand is also low (4.4%). Overall, it is hard to explain that the investors buy complex products to hedge their portfolio positions.

1.7 Evidence of Increasing Competition and Improved Transparency

1.7.1 Decreasing markup and increasing competition

A natural follow-up question is how financial intermediaries react in the presence of investor extrapolation. It is possible that this market remains "dark" and investors have been consistently exploited over time as highlighted in many early studies or this market might have become more competitive and such exploitation is not as strong as before. In this subsection, I provide evidence that markups of SRPs have consistently decreased and the competition between issuers has increased. This evidence suggests that exploitation through charging high markups has been getting weaker.

[Insert Figure 1.8 here.]

In Figure 1.8, I show how the median of annualized markups of SRPs evolves over time. To construct this figure, I choose S&P 500-linked SRPs because the S&P 500 is the most popular underlying asset for SRPs in the United States, and therefore many SRPs have been issued for a long time, which enables me to construct a long time series of median of annualized markups. In addition, option market for S&P 500 is quite liquid compared to others, and therefore S&P 500 options provide reliable volatility surface information, which reduces the estimation errors. I also restrict my sample to SRPs

with terms of less than five years. The sample size for this figure is 7,117. For the markup estimation, I do not consider the credit risk. Therefore, the estimated markup measures issuers' profits per dollar invested in SRPs conditional on their having not defaulted.

The median annualized markup in 2004 was 1.9%, which means that products with one-year (three-year) term of maturity charged a markup of 1.9% (5.7%) in 2004. Such a high markup is consistent with findings from the early studies highlighting high markups in the SRP market (Henderson and Pearson, 2011; Stoimenov and Wilkens, 2005). In 2004, six issuers issued S&P 500-linked SRPs. The annualized markup and the number of issuers experienced significant change during 2004 - 2016. The annualized markup decreased after 2004 while number of issuers increased. As of 2016, the average annualized markup was 0.46% and number of issuers was 16.[16] Overall, the time series patterns of markups and number of issuers show that the market gets competitive and taking advantage of investors by charging high markups is getting more difficult. I see the similar patterns using different measures such as weighted average instead of median, and therefore this result is robust. This result is consistent with the recent papers on markups on SRPs issued by members of Deutscher Derivate Verband (Deutscher Derivate Verband, 2015; Muller et al., 2017), which find that markups of SRPs issued by members of Deutscher Derivate Verband in 2013 (2016) are just about 0.36% (0.7%) on average and explain that such low markups are consistent with improved transparency and competition in the market for SRPs.

It is important to note that markups estimated from financial engineering models are generally overestimate actual profits that banks earn in selling SRPs. In estimating markup, financial engineering models assume that issuers do not affect the prices of derivatives when hedging their positions. However, there is ample evidence that risk

[16]When measuring number of issuers, I ignored very tiny issuers that issue less than 0.5% of market share in each year.

exposure generated by SRPs is large. As explained in Section 1.5.3, a few of large issuers account for a significant portion of this market and therefore, they sometimes suffer losses when they should take very big positions in derivative markets to hedge their positions because prices move against them (Cameron, 2013; Credit Suisse Equity Derivative Strategy, 2013; Credit Suisse Equity Derivative Strategy, 2014). Therefore, a high markup estimated from a financial engineering model does not necessarily imply that issuers earn high profits because a significant portion of the markup can be used for hedging cost.

1.7.2 Improved transparency - availability of accurate markup

Markup is known as "hidden price" because it is not observable but should be estimated using complex financial models. In this subsection, I show that the market has become more transparent in the sense that issuers have provided accurate information on markup since 2013, and therefore investors can easily access to this valuable information when they make a purchase decision.

In April 2012, the SEC Division of Corporate Finance Office of Capital Market Trends disseminated a letter to major issuers of SRPs. This letter was related to issues regarding disclosure of information in SRP prospectuses. Although there are several issues regarding disclosure, this letter primarily discussed about providing fair values of SRPs in prospectuses. Markups and annualized markups are directly calculated using fair values based on Equation (1.4). However, issuers were free to use their own internal pricing models and assumptions regarding inputs because they are not required to provide detailed information on such inputs. Therefore, issuers could have provided inaccurate information on fair values.

To test if issuers provide reliable information on fair values, I compare annualized markups calculated based on the reported fair values with annualized markups estimated

using the stochastic volatility model of Heston (1993). Similar exercises have been utilized in other papers in other settings with different samples, including Bauer et al. (2016). My result differs from others in the sense that (i) I use a larger sample size, (ii) my estimation error is smaller, and also (iii) the correlation between two markups is very strong, providing correspondingly strong evidence that banks provided accurate information on fair values.

[Insert Figure 1.7 here.]

The scatter plot displayed in Figure 1.7 shows the relationship between reported annualized markups and annualized markups estimated using the Heston model. The plot is drawn based on S&P 500-linked products that have maturities shorter than the five years. I choose S&P 500-linked products specifically because, again, S&P 500 options market is quite liquid compared to others, and therefore S&P 500 options provide reliable volatility surface information, which reduces estimation errors. I also choose products whose terms are shorter than five years to reduce estimation errors. When pricing products with very long maturity, assumptions put in the Heston model can significantly affect fair values. The correlation between the two annualized markups is 0.72. The OLS equation is as follows (N = 3,563):

$$\text{Annualized markup from Heston (\%)} = -0.553 + 0.998 \times \text{Reported annualized markup (\%)}$$
$$(0.021) \quad (0.016)$$

where standard errors for the coefficients are presented in parentheses. The slope coefficient is not different from 1 at 5% significance level. The coefficient for the intercept is -0.556%, which means that the annualized markup is slightly underestimated compared with the reported annualized markup. This is because I did not control for issuers' credit risk. To control credit risk, I use CDS spreads of issuers, with term of two years, which

is close the median of the terms of S&P 500-linked SRPs. Rather than discounting the payoffs with risk-free rate, I discount payoffs with risk-free rate + two-years term CDS spreads of issuers, as in Hull (2011), under a risk neutral measure. After controlling for credit risk, I obtain the following OLS equation:

$$\text{Annualized markup from Heston (\%)} = -0.029 + 0.992 \times \text{Reported annualized markup (\%)}$$
$$(0.023) \quad (0.017)$$

where standard errors for the coefficients are presented in parentheses. The coefficient of the intercept is not different from 0 at the 5% level significance and the coefficient of the slope is not different from 1 at the 5% level significance. Multiple explanations can support this phenomenon. For example, the competition between banks might curb the incentive to provide inaccurate information. This explanation is consistent with literature studying the effect of competition on analyst forecasting bias (Hong and Kacperczyk, 2010) and financial misreporting by managers (Balakrishnan and Cohen, 2013). Given that the issuers of SRPs are well-known global financial institutions that have operated over a long period of time and the number of issuers has been steadily increasing, the reputation concern and competition are consistent with the strong correlation pattern. Overall, investors have obtained this valuable information at no cost and this is suggestive of improved transparency.

1.7.3 Effect of extrapolation on markup and sales commission

In the presence of the extrapolative expectation, issuers (brokers) could charge higher markups (sales commission) for the products that provided high past returns to take advantage of extrapolation bias. In this subsection, I show that issuers (brokers) actually charge slightly lower markups (sales commissions).

[Insert Table 1.8 here.]

In Table 1.8, I report results indicating how *Past return* affects changes in annualized markups and sales commissions. More specifically, Table 1.8, Panel A and Panel B show the results of the following regressions.[17]

$$\Delta Annualized\ markup_{k,m,i,j,t} = \beta Past\ return_{i,j,t-1} + \gamma X_{i,j,t-1} + \epsilon_{i,j,t} \tag{1.9}$$

$$\Delta Sales\ commission_{k,m,i,j,t} = \beta Past\ return_{i,j,t-1} + \gamma X_{i,j,t-1} + \epsilon_{i,j,t} \tag{1.10}$$

where k is underlying asset, m is the term of SRPs[18], i is the underlying asset group, j is the product type and t is time. $Past\ return_{i,j,t}$ is the past return of SRPs at time t and is defined by product type (j) at the underlying asset group (i) level, as explained in Section 1.4.4. To measure $\Delta Annualized\ markup_{k,m,i,j,t}$ and, I calculate the difference between medians of the annualized markups of products that have the same term (m) and same underlying asset (k) issued at $t-1$ and t. Underlying asset (k) is a member of underlying asset group (i). I measure $\Delta Sales\ commission_{k,m,i,j,t}$ in similar fashion. As one of the control variables, I use *Past outstanding*, which is defined as $Outstanding_{i,j,t-1}$ (unit : \$ billion). In both panels, I provide the results of unweighted and sales-volume-weighted regressions to show that my results are not driven by small SRPs.

In Panel A, I show in columns (1) and (3) that *Past return* is negatively related to Δ*Annualized markup*. The result reported in column (3), implies that a one-standard-deviation increase in *Past return* translates into a 0.29% decrease in Δ*Annualized markup*. In columns (2) and (4) I show the results with interaction terms. The coefficients for *Complexity (Autocall)* $\times$ *Past return* and *Past outstanding* $\times$ *Past return* are positive. This implies that for more complex products and products of higher *Past*

[17] I do not find a clear result based on data at quarterly frequency. This is perhaps because issuers and brokers do not change markups or sales commissions very quickly and, even if they change them, the magnitudes of the changes might be small, and therefore hard to detect.

[18] Terms of SRPs are continuous and I round them to make them discrete.

outstanding, the negative relation between Δ*Annualized markup* and *Past return* is weaker.

In Panel B, I show in columns (1) and (3) that *Past return* is weakly and negatively related to Δ*Sales commission*. As seen in columns (2) and (4), the coefficients for *Complexity (Autocall)* $\times$ *Past return* and *Past outstanding* $\times$ *Past return* are negative, which means that for more complex products and products with higher *Past outstanding*, the negative relationship between Δ*Annualized markup* and *Past return* is stronger. However, the magnitude of the coefficient is small and it is hard to tell whether the effect on Δ*Sales commission* is economically significant.

Overall, the annualized markups and sales commissions of products that enjoyed high returns in the past tend to decrease slightly although demand for such products increases due to investors' extrapolative expectations. I suggest several explanations of this phenomenon. First, the competition between issuers for products with high past returns might have intensified. Unfortunately, it is hard to measure competition very precisely with my current data at product-level. Brokers typically contact multiple issuers to request issuance of particular types of products. Multiple issuers then compete with each other to provide a given product at a lower price. Our data do not contain information about those issuers who take part in the competition and, therefore precise measurement of competition at product-level is challenging. Second, the elasticity of demand by marginal investors of products that provided high past returns might have increased and therefore, markup could have decreased.[19] Due to a lack of available data, I am unable to fully pin down the mechanism that would explain lowered annualized markups and sales commission for products that experienced high past returns. I leave this question for future research.

[19]In standard oligopolistic competition model, $\frac{Price}{Marginalcost} = \left(1 - \frac{1}{\epsilon}\right)^{-1}$ where ϵ is elasticity of demand.

The positive coefficients of the interaction terms reported in columns (2) and (4) are worth noting. This result can be explained by the hedging cost that issuers face when hedging complex financial products. It is documented that issuers face high hedging costs when hedging complex SRPs such as autocalls and hedging costs rise even higher when the outstanding volume of SRPs is large (Credit Suisse Equity Derivative Strategy, 2013; Credit Suisse Equity Derivative Strategy, 2014). Because of the complex nature of the products, autocalls need to be dynamically hedged and in this process issuers face high hedging costs if the market for volatility or dividend is not liquid enough. Due to high hedging costs, issuers sometimes experience significant losses when hedging (Cameron, 2013). The positive coefficient of *Complexity (Autocall)* $\times$ *Past return* is consistent with higher hedging cost for autocalls than for other products. Although issuers charge lower annualized markups on average if *Past return* is high, issuers cannot lower annualized markups for autocalls because hedging costs are higher for autocalls. The positive coefficient of *Past outstanding* $\times$ *Past return* further strengthens my argument because as *Past outstanding* increases, issuers face stronger hedging pressures. Interestingly, the coefficients of interaction terms are not positive for changes of sales commission regressions, in Panel B. This may be because brokers do not face hedging pressures because their role is confined to marketing SRPs engineered by issuers without having any exposure to equity and derivative markets.

Overall, the negative relationship of annualized markups and sales commissions with past returns shows that financial intermediaries do not sell products that experienced high past returns more expensively although this effect is weaker for very complex products.

1.8 Conclusion

This is the first study that uses a comprehensive data set associated with complex financial products and investigates the role of investors' extrapolative expectations in understanding the rapid growth in sales volume and complexity for the SRP market. Sales volumes of the SRPs that provided higher past returns increase faster due to extrapolation. Importantly, the effects of extrapolation are stronger for more complex products, which provides important clues for understanding the rapid growth in sales volume and complexity of the market.

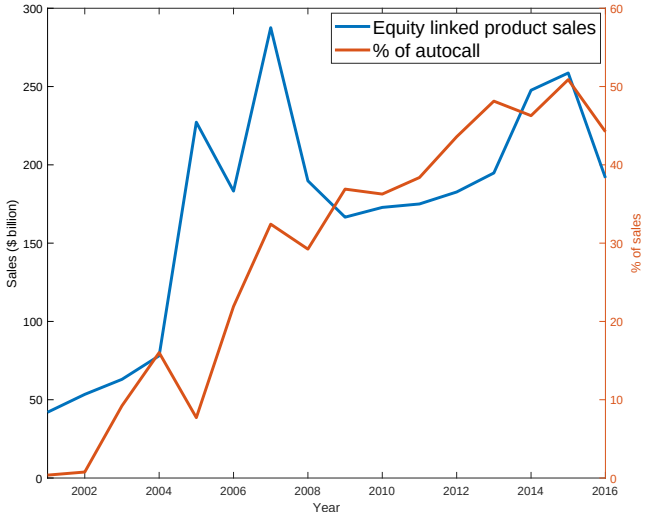
Understanding how investors form beliefs is crucial in designing and implementing new regulations that apply to complex finanical products. In addition to implementing rules for the disclosure of the fair values of SRPs, regulators have focused on how to make prospectuses more transparent to convey information more effectively. However, improved market transparency through the enhanced disclosure alone may not be sufficient to prevent investor extrapolation as my analysis shows that even after the implementation of the SEC disclosure rule in 2013, extrapolation persists.

Although I find evidence that it is becoming more difficult for financial intermediaries to easily exploit investors' extrapolative expectations because the market has become more competitive and transparent, understanding how such extrapolation affects overall investor welfare is not clear yet. I leave this important question for future research.

Figure 1.1: Sales dynamics of SRPs

These figures display the sales dynamics of the SRP market. In Panel A, I plot the sales dynamics of the SRP market based on all equity-linked SRPs issued in all countries. The blue line plots the annual sales volume of equity-linked SRPs. The red line plots the proportion of the annual sales of equity-linked autocalls compared with the annual sales of equity-linked SRPs. Panel B plots the sales dynamics of the SRP market based on SRPs issued in the United States. The blue line plots the annual sales volume of equity-linked SRPs. The red line plots the proportion of the annual sales of equity-linked autocalls compared to the annual sales of equity-linked SRPs. The red dashed (dotted) line plots the proportion of the annual sales of equity-linked reverse convertibles (participation products) compared with the annual sales of equity-linked SRPs. The red dotted line with a marker plots the proportion of the annual sales of 'others' equity-linked products compared with that of equity-linked SRPs. The units along the x axis are years and the units along the left (right) y axis are $ billions (%).

Panel A. Global trends of sales of SRPs



Panel B. Trends of sales of SRPs in the United States

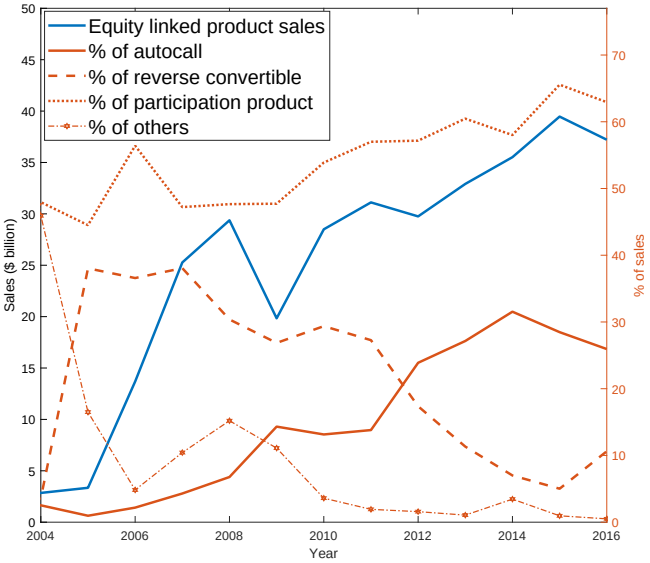
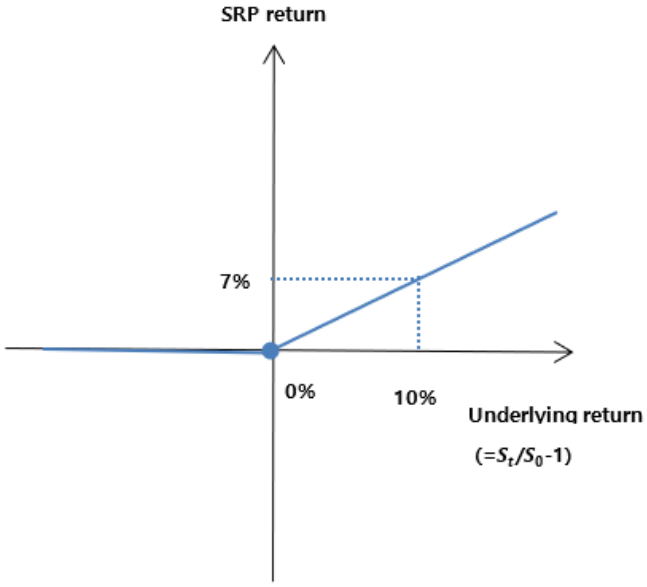
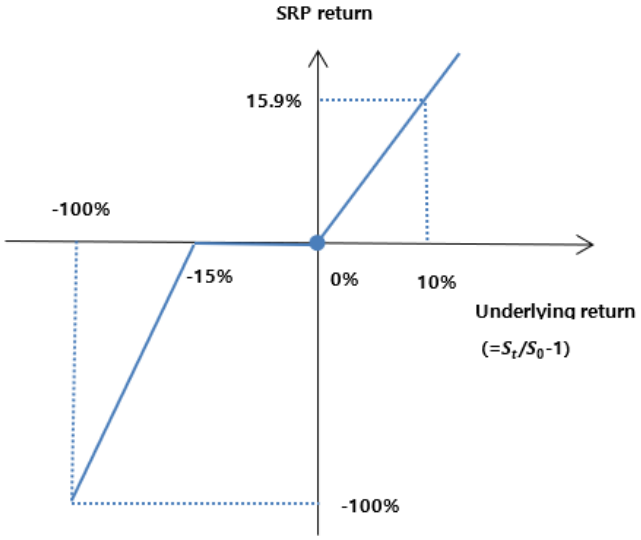


Figure 1.2: Examples of SRPs
The following figures display four real-world examples of SRPs. Panel A represents an example of a capital-protected-participation product linked to the S&P 500. This product was issued by the Canadian Imperial Bank of Commerce on June 13, 2005 with a five-year term. Panel B represents an example of a non-capital-protected participation product linked to the Eurostoxx 50. This product was issued by Goldman Sachs on April 22, 2014 with a two-year term. Panel C represents an example of a reverse convertible linked to the Linkedin stock. This product was issued by JP Morgan Chase on May 21, 2015 with one-year term. Panel D features and an example of an autocall linked to the S&P 500. This product was issued by HSBC on March 26, 2008 with a 1.5-year term. The characteristics of the product examples are described in detail in Section 1.2.2.

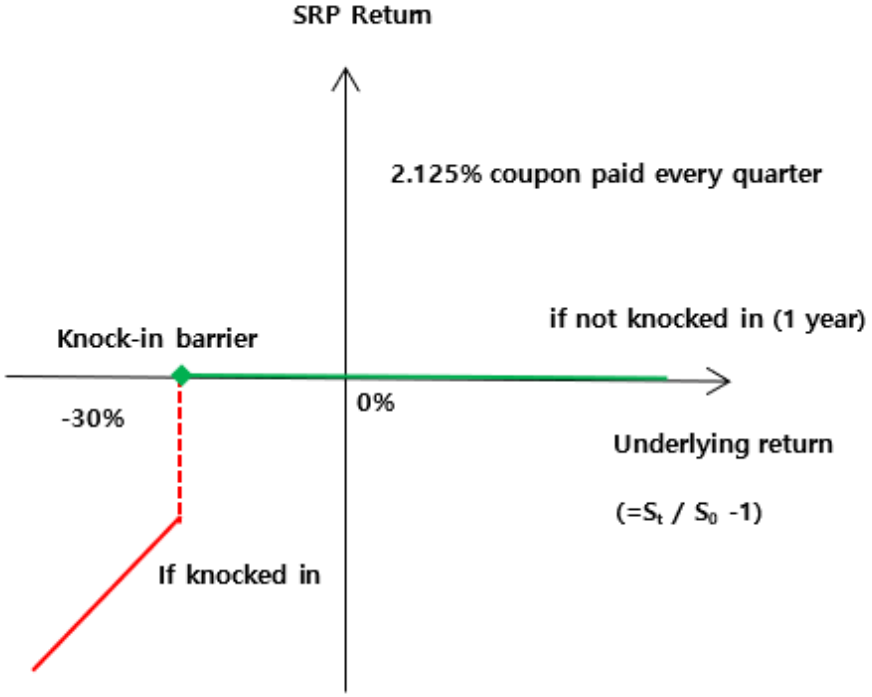
Panel A. Capital-protected participation product



Panel B. Non-capital-protected Participation product



Panel C. Reverse convertible



Panel D. Autocall

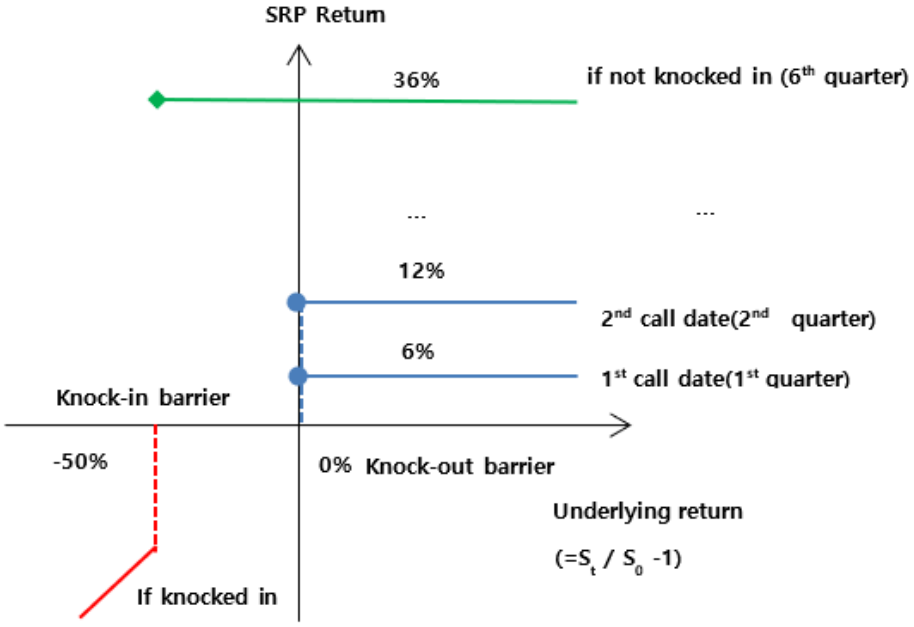


Figure 1.3: Categorization of SRPs
In this figure I present the categorization of SRPs according to the main features of their payoff structures, as described in Section 1.3.2.

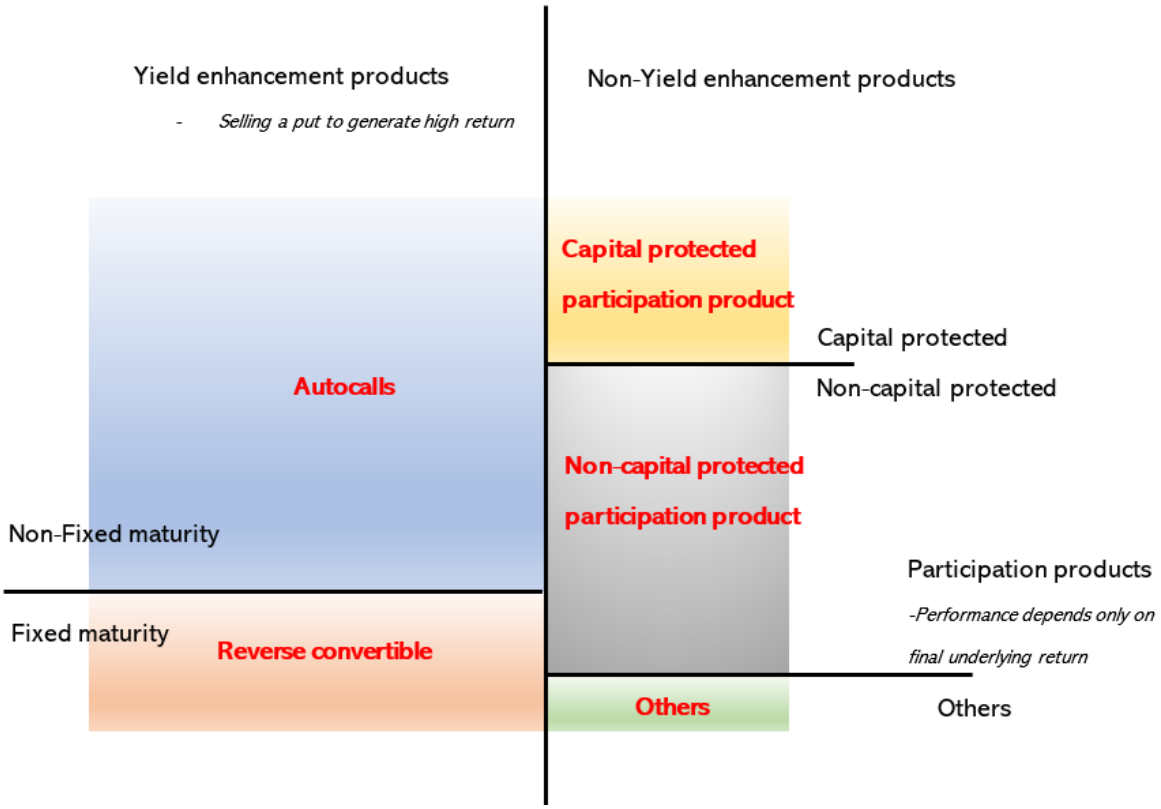


Figure 1.4: Validity of text mining techniques
In this figure I compare the annualized realized returns provided by the data provider and the annualized realized returns calculated based on the text mining technique described in Section 1.4.1. The sample size for this figure is 7,411 and the correlation coefficient between these two variables is 0.97.

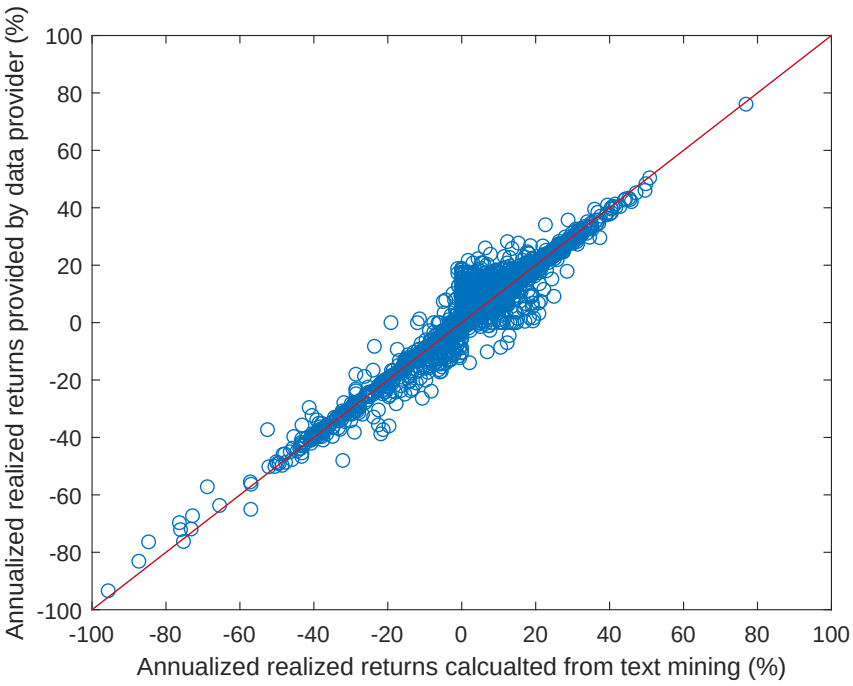


Figure 1.5: Distribution of ex-post performance of SRPs
This figure plots histograms of annualized realized returns across various types of SRPs: capita-protected participation products, non-capital-protected participation products, reverse convertibles, autocalls, and others. Annualized realized returns are calculated based on the text mining technique described in Section 1.4.1. The units along the x axis (y axis) are percentages (frequency).

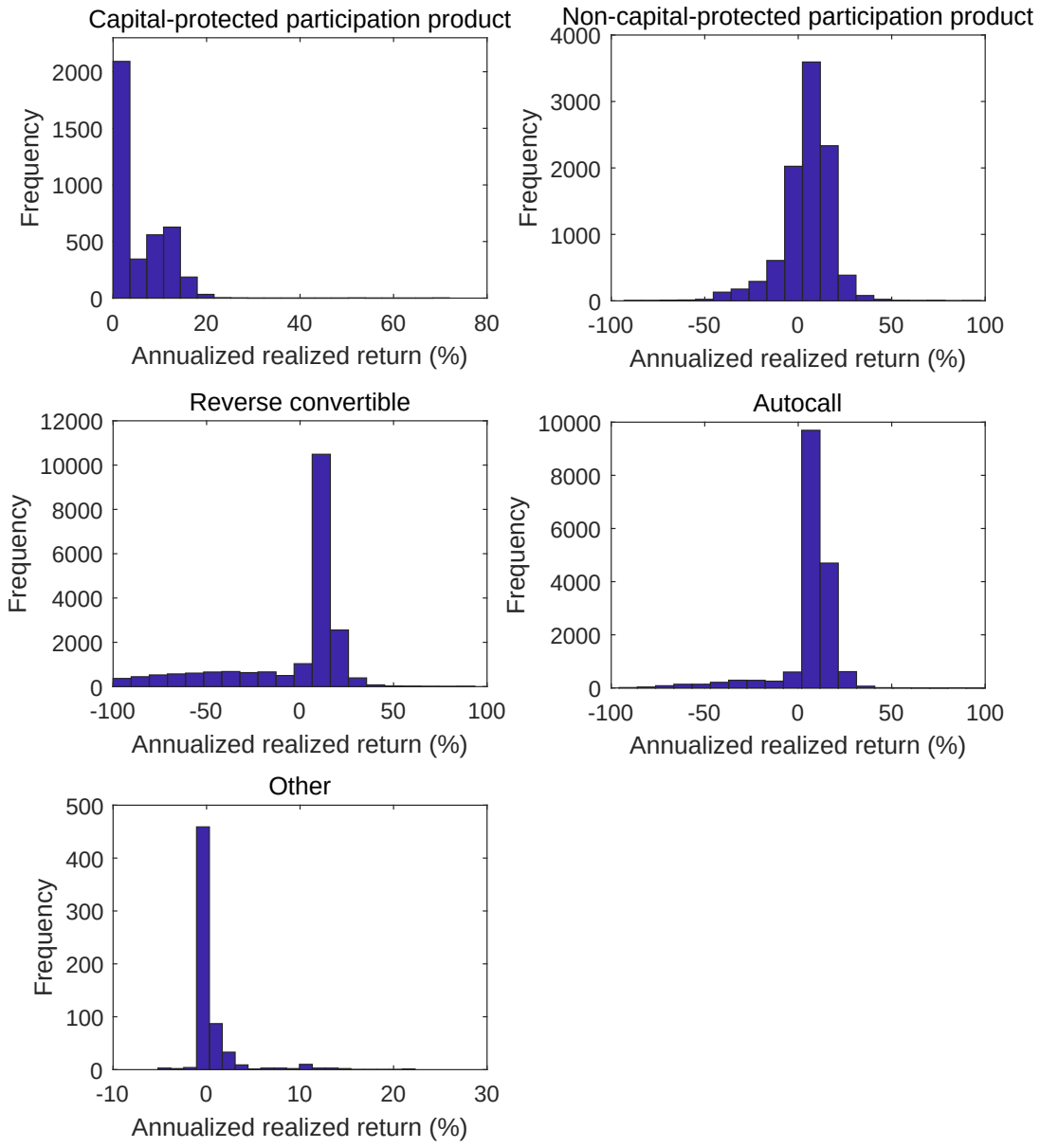


Figure 1.6: Autocalls vs Reverse convertibles
This figure displays the market shares of autocalls and reverse convertibles and the relative performance of autocalls compared with that of reverse convertibles. The red line (dashed line) plots the market shares for autocall (reverse convertible) in each year. The blue bar plots the relative performance of autocalls compared with that of reverse convertibles estimated with Equation (1.8).

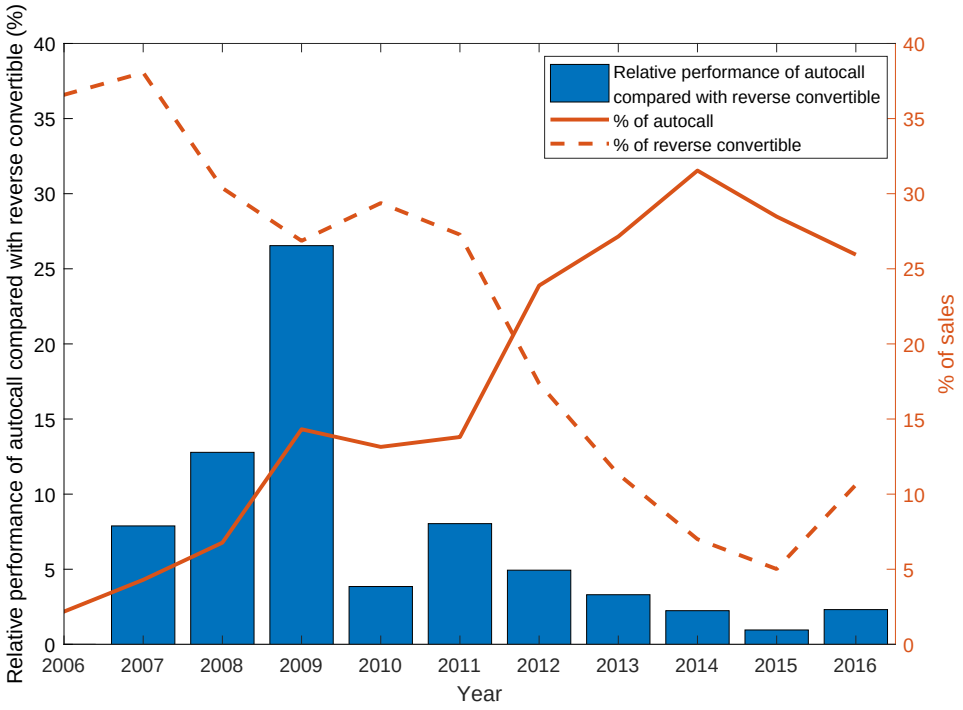


Figure 1.7: Comparison between estimated markup and reported markup
In this figure I compare the annualized markups estimated with the stochastic volatility model (Heston, 1993) and annualized markups based on the fair values of SRPs reported by issuers. The sample for this figure is based on short and medium term (term < 5 years) S&P 500-linked SRPs issued in the United States during 2013 - 2016. The red line is a fitted line estimated from the least square method. The size of the sample is 3,563 and the correlation coefficient between the two variables is 0.72.

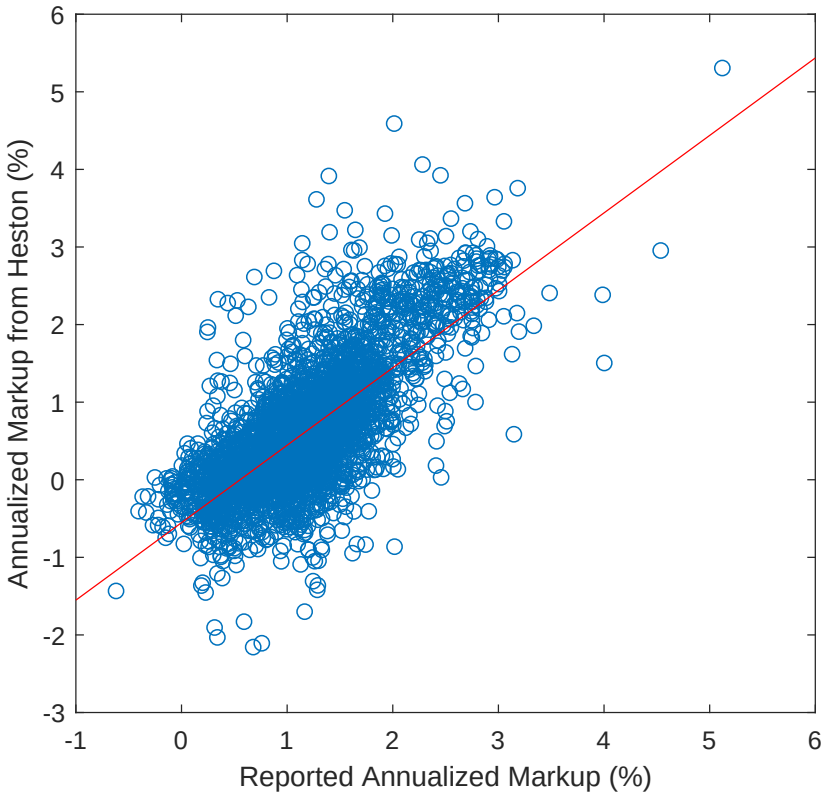


Figure 1.8: Evoluation of annualized markup of S&P 500-linked SRPs
In this figure I trace the evolution of annualized markup over the sample period of 2004 - 2016. This figure is constructed based on all SRPs that have the S&P 500 as underlying asset with terms of less than five years. The sample size for this figure is 7,117. The blue bars are the medians of annualized markups for all SRPs issued in each year. The red line plots the number of issuers that issue S&P 500-linked SRPs in each year.

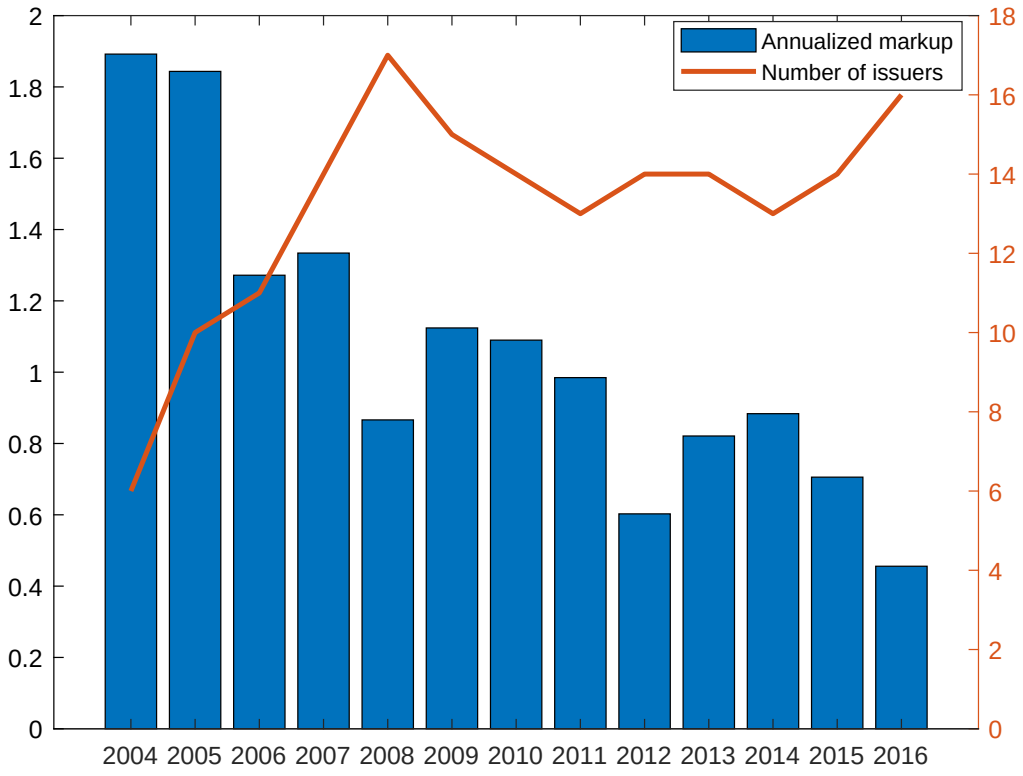


Table 1.1: An example of data structure
In this table I present the format of a product description in the SRP database. The example product, *Autocallable Optimization Securities with Contingent Protection*, is described in detail in Section 1.2.3.

Term	Detail
Product name	Autocallable Optimization Securities with Contingent Protection
ISIN code	US40428H5990
CUSIP	40428H599
Date of issuance	26 Mar 2008
Date of final valuation	25 Sep 2009
Date of maturity	30 Sep 2009
Underlying asset	S&P 500
Distributor	UBS
Issuer	HSBC Bank
Sales volume	$16.40 million
Sales commmission	1.5%
Early redemption information	*Knockout Date* / *Knockout Level* / *Payout* 25 Jun 2008 / 0% / 106.00% 25 Sep 2008 / 0% / 112.00% 25 Dec 2008 / 0% / 118.00% 26 Mar 2009 / 0% / 124.00% 25 Jun 2009 / 0% / 130.00% 25 Sep 2009 / 0% / 136.00%
Payoff description	This is a growth product linked to the S&P500. The product terminates early if the index level at the end of each quarter is equal to or greater than the initial index level. In this case, the capital return is calculated using an annualized return of 24%. Otherwise, the product is held till maturity. At maturity, if the index level never closed below 50% of the initial level, the product offers a capital return of 100%. If, however, the index closed below 50% of the initial index level on any trading day during the observation period, the capital return is 100% minus 1% for every 1% fall in the index over the investment period.

Table 1.2: Summary statistics

In this table I present product-level and group-level summary statistics on product characteristics of SRPs issued in the United States during the sample period of 2001 - 2016. In Panel A (Panel B) I describe product-level (group -level) summary statistics. The construction of group-level data using product-level data is summarized in Section 1.4.4. *Complexity (# characters)* is the number of characters that describe the payoff structure. *Complexity (Autocall)* is a dummy equal to 1 if a product has the autocallable feature and 0 otherwise. *Sales volume* is dollar amount of sales at issuance. *Annualized realized return* is ex-post performance measure of an SRP defined in Section 1.4.2. *Annualized markup* is defined as $\frac{1}{\text{Time to Maturity}}\left(\frac{\text{Price of a SRP - Fair value of a SRP}}{\text{Price of a SRP}}\right)$, where the fair value of an SRP is estimated from the stochastic volatility model of Heston (1993). *Sales commission* is a fixed proportion of sales volume paid to brokers. *Term* is years between the issuance date and the specified maturity date. *One-year CDS spread* is the one-year CDS spread of an issuer of an SRP. *Volatility* is the volatility of an SRP. *p25, p50* and *p75* are the 25%, 50%, and 75% percentile, respectively.

Panel A. Summary statistics (Product-level)

Variables	Average	Std. dev.	Minimum	p25	p50	p75	Maximum	*N*
Complexity measures								
- Complexity (# of characters)	587.5	225.7	50	413	576	707	3488	71,300
- Complexity (Autocall)	0.25	0.43	0	0	0	0	1	71,300
Sales volume ($ million)	4.69	12.57	0.001	0.34	1.50	4.61	547.56	71,300
Annualized realized return (%)	1.97	22.97	-100	0.77	9.31	12.82	99.56	51,606
Annualized markup (%)	2.53	4.14	-9.50	0.38	1.36	3.26	19.88	29,273
Sales commission (%)	1.91	1.26	0	1.25	1.65	2.50	9.90	62,164
Term (years)	2.37	2.41	0.09	1.01	1.27	3.02	30.26	71,300
One-year CDS spread (%)	0.63	0.72	0.01	0.27	0.40	0.77	20.21	56,667
Volatility (%)	20.05	10.51	1.01	12.16	18.01	25.13	113.71	39,032

Panel B. Summary statistics (Group-level)

Variables	Average	Std. dev.	Minimum	p25	p50	p75	Maximum	*N*
Complexity measures								
- Complexity (# of characters)	562.2	187.5	232	418.5	511	689	1074	1,575
- Complexity (Autocall)	0.21	0.41	0	0	0	0	1	1,575
Past return (%)	4.57	15.7	-56.5	0.18	8.50	12.1	47.2	1,575
Normalized sales (%)	20.4	19.4	0.03	0.04	15.4	27.7	213.3	1,575
Annualized realized return (%)	4.22	13.67	-58.58	0.52	8.56	10.91	29.57	1,365
Annualized markup (%)	2.20	2.71	-8.76	0.78	1.54	3.13	13.2	947
Sales commission (%)	1.89	0.81	0	1.51	1.76	2.04	6.98	1,473
Term (years)	2.18	1.78	0.26	1.02	1.52	3.01	7.52	1,575
One-year CDS spread (%)	0.64	0.60	0.02	0.29	0.41	0.82	6.88	1,554
Volatility (%)	19.6	8.13	2.15	13.96	18.15	23.20	66.29	769

Table 1.3: Ex-post performance of SRPs

In the following table, I summarize the annualized realized returns of equity-linked SRPs issued in the United States during the sample period of 2001 - 2016 that matured before January 1, 2017. In Panel A I report the unweighted and sales-volume-weighted average and standard deviation of annualized realized returns for capital-protected participation products, non-capital-protected participation products products, reverse convertibles, autocalls, and others. *Proportion of gain* is the number of products that provided positive returns as the percentage of total products. In Panel B I report the annualized realized returns for other benchmark asset classes during the same period.

Panel A. Annualized realized returns and proportion of gain of SRPs

	Annualized realized returns (%)				Proportion of gain (%)	N
	Unweighted		Sales-volume -weighted			
	Average	Std. dev	Average	Std. dev		
Capital-protected participation product	5.00	6.03	3.30	5.19	69.98	3,847
Non-capital-protected participation product	5.05	13.69	4.00	16.32	69.03	9,716
Reverse convertible	-4.20	31.27	-2.84	29.03	71.11	20,250
Autocall	6.86	15.53	7.59	13.44	88.32	17,168
Others	0.75	2.43	0.77	2.34	25.28	625
Total	1.97	22.97	3.20	19.16	75.81	51,606

Panel B. Annualized realized returns of other benchmark asset classes

	Average	Std. dev
Bank (AAA) Bond Yield(1 year)	3.7	1.43
S&P 500	6.39	14.6

Table 1.4: Extrapolation and sales

In the following table I report the effects of extrapolative expectations on SRP sales. In Panel A I report the coefficients of regressions in which dependent variables are *Normalized sales*. In Panel B I report the coefficients of regressions of *Normalized sales* on explanatory variables with control variables excluding *Annualized markup* and *Volatility*. In Panel C I report the coefficients of regressions of *Normalized sales* on explanatory variables with control variables including *Annualized markup* and *Volatility*. Explanatory variables and control variables in the tables are defined in Section 1.4.4. In each column of the regression tables, I report coefficients and their heteroscedasticity-robust standard errors in parentheses. Standard errors are clustered at the underlying asset group level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Normalized sales and past return

	Normalized sales					
	(1)	(2)	(3)	(4)	(5)	(6)
Past return	0.264***	0.204***	0.168***	0.102**	-0.0580	-0.130
	(0.0430)	(0.0424)	(0.0353)	(0.0390)	(0.127)	(0.100)
Complexity (Autocall)			0.132***	0.134***		
			(0.0149)	(0.0179)		
Complexity (Autocall) × Past return			0.443***	0.367***		
			(0.106)	(0.0805)		
Complexity (# Character)					0.299***	0.303***
					(0.0390)	(0.0430)
Complexity (# Character) × Past return					0.564**	0.549***
					(0.209)	(0.157)
Quarter FE		✓		✓		✓
Underlying FE		✓		✓		✓
N	1575	1570	1575	1570	1575	1570
adj. R^2	0.052	0.196	0.196	0.320	0.174	0.299

Panel B. Normalized sales and past return with controls (without annualized markup and volatility)

	Normalized sales					
	(1)	(2)	(3)	(4)	(5)	(6)
Past return	0.229***	0.211***	0.123***	0.0961***	-0.149	-0.173*
	(0.0399)	(0.0377)	(0.0281)	(0.0341)	(0.106)	(0.0904)
Complexity (Autocall)			0.0958***	0.116***		
			(0.0136)	(0.0165)		
Complexity (Autocall) × Past return			0.484***	0.392***		
			(0.107)	(0.0774)		
Complexity (# Character)					0.177***	0.227***
					(0.0488)	(0.0489)
Complexity (# Character) × Past return					0.639***	0.627***
					(0.182)	(0.150)
Sales commission	0.721	-1.704	1.392*	-0.778	1.022	-1.182
	(0.984)	(1.145)	(0.725)	(1.105)	(0.842)	(1.096)
Term	-4.064***	-2.629***	-3.676***	-2.479***	-3.227***	-1.855***
	(0.348)	(0.609)	(0.334)	(0.558)	(0.356)	(0.634)
One-year CDS spread	-4.346***	-1.403	-4.436***	-2.113	-4.526***	-1.548
	(1.146)	(1.756)	(1.137)	(1.857)	(1.185)	(2.070)
Quarter FE		✓		✓		✓
Underlying FE		✓		✓		✓
N	1468	1460	1468	1460	1468	1460
adj. R^2	0.188	0.255	0.289	0.361	0.235	0.317

Panel C. Normalized sales and past return with controls (with annualized markup and volatility)

	Normalized sales					
	(1)	(2)	(3)	(4)	(5)	(6)
Past return	0.214***	0.181***	0.0727*	0.0526	-0.393**	-0.413**
	(0.0550)	(0.0486)	(0.0387)	(0.0485)	(0.134)	(0.157)
Complexity (Autocall)			0.126***	0.148***		
			(0.0136)	(0.0153)		
Complexity (Autocall) × Past return			0.510***	0.388***		
			(0.132)	(0.101)		
Complexity (# Character)					0.316***	0.338***
					(0.0534)	(0.0399)
Complexity (# Character) × Past return					0.974***	0.921***
					(0.240)	(0.260)
Annualized markup	-1.023***	-1.077***	-0.582	-0.656*	-0.920**	-0.938**
	(0.304)	(0.331)	(0.331)	(0.356)	(0.298)	(0.326)
Sales commission	-1.247	-5.947*	3.302	-0.672	2.411	-1.025
	(3.029)	(2.656)	(2.296)	(2.364)	(2.649)	(2.541)
Term	-5.043***	-2.808*	-4.570***	-2.881***	-3.528***	-1.931**
	(0.911)	(1.299)	(0.687)	(0.870)	(0.636)	(0.835)
One-year CDS spread	-4.305**	-6.000*	-5.022***	-5.271*	-4.722**	-4.927*
	(1.531)	(2.923)	(1.486)	(2.780)	(1.618)	(2.685)
Volatility	0.201	-0.380	0.428**	0.136	0.310*	-0.120
	(0.153)	(0.223)	(0.138)	(0.205)	(0.140)	(0.156)
Quarter FE		✓		✓		✓
Underlying FE		✓		✓		✓
N	746	739	746	739	746	739
adj. R^2	0.159	0.247	0.302	0.386	0.267	0.359

Table 1.5: Extrapolation and sales (Intensive and extensive margin)
In the following table I report the effect of extrapolative expectations on SRP sales in the intensive and extensive margins. In Panel A I report the summary statistics on *Proportion of sales volume of incumbent banks*. *Proportion of sales volume of incumbent banks* is defined in Section 1.5.3. In Panel B (Panel C) I report the coefficients of regressions in which the dependent variable is $Normalized\ sales^{Intensive}$ ($Normalized\ sales^{Extensive}$). $Normalized\ sales^{\text{Intensive}}$ and $Normalized\ sales^{\text{Extensive}}$ are defined in Section 1.5.3. In each column of the regression tables, I report coefficients and their heteroscedasticity-robust standard errors in parentheses. Standard errors are clustered at the underlying asset group level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Proportion of sales volume of incumbent banks (%)

Average	Std. dev.	Minimum	p25	p50	p75	Maximum	N
72.4%	33.2%	0%	54.9%	87.7%	99.3%	100%	1,575

Panel B. Normalized sales and past return (Intensive margin)

	Normalized sales^{Intensive}					
	(1)	(2)	(3)	(4)	(5)	(6)
Past return	0.220***	0.193***	0.134***	0.102***	-0.0998	-0.109
	(0.0369)	(0.0378)	(0.0269)	(0.0323)	(0.102)	(0.0981)
Complexity (Autocall)			0.138***	0.133***		
			(0.0158)	(0.0192)		
Complexity (Autocall) × Past return			0.383***	0.307***		
			(0.0883)	(0.0761)		
Complexity (# Characters)					0.323***	0.317***
					(0.0420)	(0.0465)
Complexity (# Characters) × Past return					0.562***	0.492***
					(0.174)	(0.155)
Quarter FE		✓		✓		✓
Underlying FE		✓		✓		✓
N	1575	1570	1575	1570	1575	1570
adj. R^2	0.050	0.196	0.246	0.354	0.242	0.345

Panel C. Normalized sales and past return (Extensive margin)

	Normalized sales^{Extensive}					
	(1)	(2)	(3)	(4)	(5)	(6)
Past return	0.0540***	0.0211	0.0408**	0.00456	0.0354	-0.0245
	(0.0169)	(0.0184)	(0.0183)	(0.0183)	(0.0725)	(0.0610)
Complexity (Autocall)			0.00222	0.00932		
			(0.00511)	(0.00713)		
Complexity (Autocall) × Past return			0.0765	0.0732		
			(0.0482)	(0.0433)		
Complexity (# Characters)					-0.00614	0.00415
					(0.0134)	(0.0191)
Complexity (# Characters) × Past return					0.0321	0.0787
					(0.120)	(0.114)
Quarter FE		✓		✓		✓
Underlying FE		✓		✓		✓
N	1575	1570	1575	1570	1575	1570
adj. R^2	0.007	0.106	0.008	0.109	0.006	0.105

Table 1.6: Effect of rational information extraction
In this table I report the effects of rational information extraction on the sales dynamics of the SRP market. *Annualized realized return of newly issued products* measures annualized realized returns of newly issued products, which are realized later depending on the lengths of terms. Explanatory variables, *Past return* and *Normalized sales*, are defined in Section 1.4.4. I round *Term* and use it for fixed effect to control different terms of SRPs. In each column, I report coefficients and their heteroscedasticity-robust standard errors in parentheses. Standard errors are clustered at the underlying asset group level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Annualized realized return of newly issued products			
	(1)	(2)	(3)	(4)
Past return	-0.0311	-0.0307		
	(0.0239)	(0.0365)		
Normalized sales			0.0248	0.0363
			(0.0225)	(0.0221)
Quarter FE		✓		✓
Underlying FE		✓		✓
Term FE		✓		✓
N	1365	1365	1365	1365
adj. R^2	0.001	0.301	0.000	0.303

Table 1.7: Evolution of ex-post performance

In the following table I report the evolution of ex-post performance of SRPs. In Panel A I report the sales-volume-weighted average of annualized realized returns of products of five types that matured in each year. In Panel B I report the coefficients of regressions of *Annualized realized return* on *Complexity (Autocall)* dummy. I run the regressions with the sample of autocalls and reverse convertibles that matured in each year. The coefficients under the column *Complexity (Autocall)* represent relative performance of autocalls compared with reverse convertibles. In each regression in Panel B, I report coefficients and their heteroscedasticity-robust standard errors. Standard errors are clustered at the underlying asset level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Annualized realized returns of SRPs (sales-volume-weighted)

Year	Capital protected participation product	Non-capital protected participation product	Reverse convertible	Autocall	Others	Total
2006	N.A.	10.73	-5.55	14.45	N.A.	0.15
2007	8.49	8.54	8.30	10.99	N.A.	8.46
2008	0.80	-8.47	-19.77	1.41	1.65	-13.62
2009	0.18	-15.65	-16.92	16.46	0.31	-10.69
2010	0.94	10.60	8.08	10.91	0.19	8.55
2011	0.54	8.40	-1.54	10.09	0.69	4.07
2012	1.47	6.60	1.49	7.02	1.72	4.68
2013	4.55	12.70	8.46	8.70	2.50	9.95
2014	6.38	12.06	10.11	9.49	1.21	10.38
2015	5.77	8.59	1.32	5.49	1.72	6.76
2016	4.28	-0.85	1.09	3.28	1.48	0.69
Total (including products matured before 2006)	3.30	4.00	-2.84	7.59	0.77	3.20

Panel B. Relative performance of autocalls compared with reverse convertibles

Year	Complexity (Autocall)	Standard error	Underlying FE	N	adj. R^2
2007	7.88*	4.68	✓	1,236	0.389
2008	12.78	13.84	✓	3,654	0.243
2009	26.54***	4.28	✓	2,046	0.215
2010	3.85	3.51	✓	2,795	0.278
2011	8.04***	2.04	✓	2,983	0.212
2012	4.94***	0.91	✓	3,720	0.308
2013	3.30***	0.83	✓	4,664	0.322
2014	2.24**	0.89	✓	4,467	0.412
2015	0.96	1.13	✓	3,591	0.719
2016	2.31**	1.06	✓	3,414	0.556

Table 1.8: Effect of extrapolation on markup and sales commission
In this table I report the effects of investor extrapolation on changes in annualized markups and sales commissions based on annual frequency data. In Panel A (Panel B) I show the coefficients of regressions in which the dependent variable is Δ*Annualized markup* (Δ*Sales commission*). Δ*Annualized markup* is the changes in median annualized markups between current and previous period and Δ*Sales commission* is the changes in median sales commissions between current and previous periods described in Section 1.7.3. The explanatory variables are defined in Section 1.4.4. In each column, I report coefficients and their heteroscedasticity-robust standard errors in parentheses. Standard errors are clustered at the underlying asset level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Change of annualized markup and past return

	ΔAnnualized markup			
Sample	Unweighted		Sales-volume-weighted	
	(1)	(2)	(3)	(4)
Past return	-0.041*	-0.075***	-0.019***	-0.051***
	(0.023)	(0.021)	(0.0060)	(0.008)
Complexity (Autocall)		-0.761**		-0.547**
		(0.310)		(0.234)
Complexity (Autocall)× Past return		0.077***		0.060***
		(0.021)		(0.007)
Past outstanding		-0.026*		-0.010
		(0.015)		(0.008)
Past outstanding × Past return		0.004*		0.002***
		(0.001)		(0.0005)
Year FE×Underlying FE	✓	✓	✓	✓
Term FE	✓	✓	✓	✓
N	717	717	717	717
adj. R^2	0.254	0.257	0.215	0.217

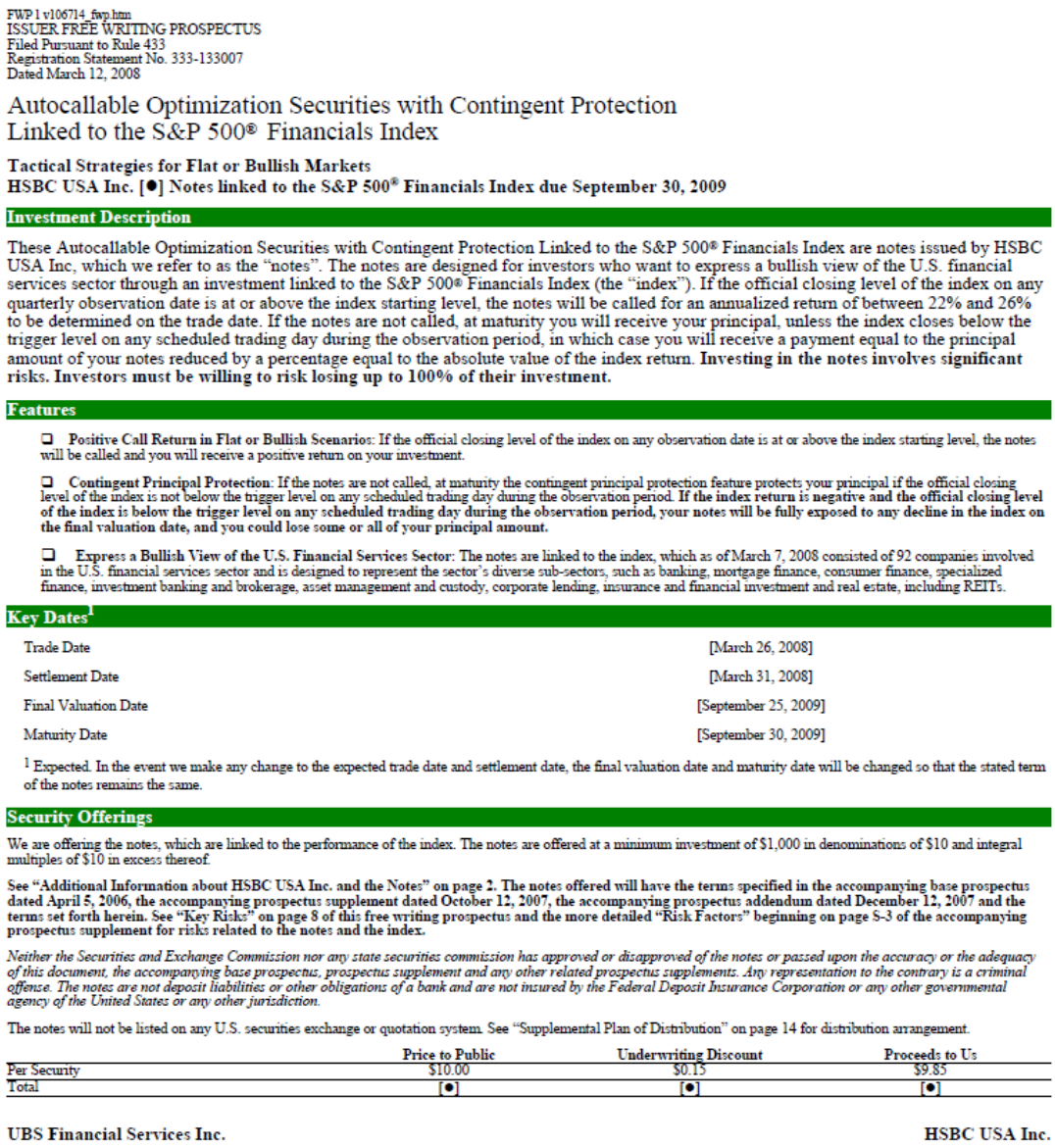
Panel B. Change of sales commission and past return

	ΔSales commission			
Sample	Unweighted		Sales-volume-weighted	
	(1)	(2)	(3)	(4)
Past return	-0.004	0.003	-0.001	0.008
	(0.007)	(0.010)	(0.007)	(0.012)
Complexity (Autocall)		0.067		0.068
		(0.051)		(0.045)
Complexity (Autocall)× Past return		-0.013**		-0.015**
		(0.006)		(0.006)
Past outstanding		0.015***		0.014***
		(0.003)		(0.003)
Past outstanding × Past return		-0.001		-0.001*
		(0.0006)		(0.001)
Year FE×Underlying FE	✓	✓	✓	✓
Term FE	✓	✓	✓	✓
N	1,564	1,564	1,564	1,564
adj. R^2	0.081	0.086	0.035	0.042

Appendix

Appendix A : An example of prospectus

This figure is the first two pages of prospectus for *Autocallable Optimization Securities with Contingent Protection*. The detailed description of this product is provided in Section 2.2.



FWP 1 v106714_fwp.htm
ISSUER FREE WRITING PROSPECTUS
Filed Pursuant to Rule 433
Registration Statement No. 333-133007
Dated March 12, 2008

Autocallable Optimization Securities with Contingent Protection
Linked to the S&P 500® Financials Index

Tactical Strategies for Flat or Bullish Markets
HSBC USA Inc. [●] Notes linked to the S&P 500® Financials Index due September 30, 2009

Investment Description

These Autocallable Optimization Securities with Contingent Protection Linked to the S&P 500® Financials Index are notes issued by HSBC USA Inc, which we refer to as the "notes". The notes are designed for investors who want to express a bullish view of the U.S. financial services sector through an investment linked to the S&P 500® Financials Index (the "index"). If the official closing level of the index on any quarterly observation date is at or above the index starting level, the notes will be called for an annualized return of between 22% and 26% to be determined on the trade date. If the notes are not called, at maturity you will receive your principal, unless the index closes below the trigger level on any scheduled trading day during the observation period, in which case you will receive a payment equal to the principal amount of your notes reduced by a percentage equal to the absolute value of the index return. **Investing in the notes involves significant risks. Investors must be willing to risk losing up to 100% of their investment.**

Features

- **Positive Call Return in Flat or Bullish Scenarios**: If the official closing level of the index on any observation date is at or above the index starting level, the notes will be called and you will receive a positive return on your investment.
- **Contingent Principal Protection**: If the notes are not called, at maturity the contingent principal protection feature protects your principal if the official closing level of the index is not below the trigger level on any scheduled trading day during the observation period. **If the index return is negative and the official closing level of the index is below the trigger level on any scheduled trading day during the observation period, your notes will be fully exposed to any decline in the index on the final valuation date, and you could lose some or all of your principal amount.**
- **Express a Bullish View of the U.S. Financial Services Sector**: The notes are linked to the index, which as of March 7, 2008 consisted of 92 companies involved in the U.S. financial services sector and is designed to represent the sector's diverse sub-sectors, such as banking, mortgage finance, consumer finance, specialized finance, investment banking and brokerage, asset management and custody, corporate lending, insurance and financial investment and real estate, including REITs.

Key Dates[1]

Trade Date	[March 26, 2008]
Settlement Date	[March 31, 2008]
Final Valuation Date	[September 25, 2009]
Maturity Date	[September 30, 2009]

[1] Expected. In the event we make any change to the expected trade date and settlement date, the final valuation date and maturity date will be changed so that the stated term of the notes remains the same.

Security Offerings

We are offering the notes, which are linked to the performance of the index. The notes are offered at a minimum investment of $1,000 in denominations of $10 and integral multiples of $10 in excess thereof.

See "Additional Information about HSBC USA Inc. and the Notes" on page 2. The notes offered will have the terms specified in the accompanying base prospectus dated April 5, 2006, the accompanying prospectus supplement dated October 12, 2007, the accompanying prospectus addendum dated December 12, 2007 and the terms set forth herein. See "Key Risks" on page 8 of this free writing prospectus and the more detailed "Risk Factors" beginning on page S-3 of the accompanying prospectus supplement for risks related to the notes and the index.

Neither the Securities and Exchange Commission nor any state securities commission has approved or disapproved of the notes or passed upon the accuracy or the adequacy of this document, the accompanying base prospectus, prospectus supplement and any other related prospectus supplements. Any representation to the contrary is a criminal offense. The notes are not deposit liabilities or other obligations of a bank and are not insured by the Federal Deposit Insurance Corporation or any other governmental agency of the United States or any other jurisdiction.

The notes will not be listed on any U.S. securities exchange or quotation system. See "Supplemental Plan of Distribution" on page 14 for distribution arrangement.

	Price to Public	Underwriting Discount	Proceeds to Us
Per Security	$10.00	$0.15	$9.85
Total	[●]	[●]	[●]

UBS Financial Services Inc. **HSBC USA Inc.**

Indicative Terms

Issuer	HSBC USA Inc. (Aa3/AA-)[1]
Principal Amount	$10 per note
Term	18 months, unless earlier called
Index	The S&P 500® Financials Index (Ticker: S5FINL) (the "index")
Call Feature	The notes will be called if the official closing level of the index on any observation date is at or above the index starting level
Observation Dates	On or about June 26, 2008, September 26, 2008, December 26, 2008, March 26, 2009, June 26, 2009 and September 25, 2009
Call Settlement Dates	Three business days following the applicable observation date
Return on Call Date	If the notes are called, on a call settlement date, investors will receive a cash payment per $10 principal amount note equal to the call price for the applicable observation date. The return on call date will be based upon an annualized return of between 22% and 26%. The table below assumes an annualized return of 24% (the midpoint of 22% and 26%). The actual annualized return upon which the return on call date is based will be determined on the trade date
Observation Date June 25, 2008 September 25, 2008 December 26, 2008 March 26, 2009 June 25, 2009 Final Valuation Date (September 25, 2009)	Return on Call Date Call Price (per $10.00) [6.00]% $[10.60] [12.00]% $[11.20] [18.00]% $[11.80] [24.00]% $[12.40] [30.00]% $[13.00] [36.00]% $[13.60]
Payment at Maturity (per $10 note)	**If the notes are not called and the official closing level of the index is not below the trigger level on any scheduled trading day during the observation period**, you will receive a cash payment on the maturity date equal to $10 per $10 principal amount note. **If the notes are not called and the official closing level of the index is below the trigger level on any scheduled trading day during the observation period**, you will receive a cash payment on the maturity date equal to: $10 × (1 + index return); **In this case, you may lose all or a substantial portion of your principal, depending on how much the index declines**
Index Return	$\frac{\text{index ending level - index starting level}}{\text{index starting level}}$
Trigger Level	[●], representing 50% of the index starting level
Observation Period	The period from, but excluding, the trade date to, and including, the final valuation date
Index Starting Level	The official closing level of the index on the trade date
Index Ending Level	The official closing level of the index on the final valuation date
Official Closing Level	The official closing level on any scheduled trading day during the observation period will be the closing level of the index as determined by the calculation agent based upon determinations with respect thereto made by the reference sponsor and displayed on Bloomberg Professional® service page "S5FINL<INDEX>".
CUSIP / ISIN	40428H 599 / US40428H5990

Determining Payment at Maturity

Appendix B : Stochastic volatility model (Heston, 1993)

Under the risk neutral measure, the price (S_t) and variance (V_t) have the following dynamics.

$$\begin{aligned}
\frac{dS_t}{S_t} &= (r_f - q)dt + \sqrt{V_t}dW_t^1 \\
dV_t &= a(\bar{V} - V_t)dt + \eta\sqrt{V_t}dW_t^2 \\
dW_t^1 dW_t^2 &= \rho dt
\end{aligned}$$

Under the above assumptions, Heston (1993) derives the closed-form solutions of the call option price, $C_0^{Heston}(K,T)$:

$$\begin{aligned}
C_0^{Heston}(K,T) &= S_0\Pi_1 - e^{-(r_f-q)T}K\Pi_2 \\
\text{where } \Pi_1 &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-iw\log(K)}\psi_{log(S_T)}(w-i)}{iw\psi_{\log(S_T)}(-i)}\right]dw \\
\Pi_2 &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-iw\log(K)}\psi_{log(S_T)}(w)}{iw}\right]dw
\end{aligned}$$

where K is the strike price of the call option, $\psi_{\log(S_t)}(w)$ is the characteristic function of $\log(S_t)$, and the function Re returns a real part of a complex number. A closed-from of the characteristics function, $\psi_{\log(S_t)}(w)$, is expressed as follows:

$$\begin{aligned}
\psi_{\log(S_t)}(w) &= e^{C(t,w)\bar{V} + D(t,w)V_0 + iw\log(S_0e^{(r-q)T})} \\
where\ C(t,w) &= a\left[r_- t - \frac{2}{\eta^2}\log\left(\frac{1-ge^{-ht}}{1-g}\right)\right] \\
D(t,w) &= r_-\frac{1-e^{-ht}}{1-ge^{-ht}} \\
r_\pm &= \frac{\beta \pm h}{\eta^2}; h = \sqrt{\beta^2 - 4\alpha\gamma}; g = \frac{r_-}{r_+} \\
\alpha &= -\frac{w^2}{2} - \frac{iw}{2}; \beta = a - \rho\eta wi; \gamma = \frac{\eta^2}{2}
\end{aligned}$$

Appendix C : Survey on SRP investors' characteristics

In this table I display results of a survey on the characteristics of SRP investors conducted by the Financial Supervisory Service (FSS), a finance watchdog in South Korea, in August 2010. The FSS selected random sample of investors who visited one of 17 major SRP brokerage firms in South Korea. The total number of respondents is 1,049.

1. Gender

Male	Female
45.9%	54.1%

2. Age

20 - 29	30 - 39	40 - 49	50 - 59	Older than 60
18.1%	32.7%	23.2%	17.4%	8.6%

3. Occupation

Self Employed	Employee	Finance	Public sector	Others
11.6%	32.5%	29.2%	3%	23.8%

4. Education background

Elementary school	Middle school	High school	College	Graduate school
0.2%	1%	17.4%	75.2%	6.2%

5. Assets ($)

Less than 50k	50k - 100k	100k - 500k	500k - 1m	More than 1m
21.6%	12.9%	35.8%	16.5%	13.3%

6. Financial assets ($)

Less than 50k	50k - 100k	100k - 500k	500k - 1m	More than 1m
32.3%	26.2%	28.5%	8.0%	5.0%

7. Investment experience in SRPs

None	Less than 1 year	1-3 years	3-5 years	More than 5 years
20.0%	14.1%	44.2%	16.1%	5.6%

8. Amount of current investment in SRPs ($)

Less than 10k	10k - 50k	50k - 100k	100k - 500k	More than 500k
36.5%	35.8%	13.6%	11.7%	2.4%

9. Annualized realized returns in SPRs over the last 2 years

Less than -30%	-30 ∼ -15%	-15% ∼ -0%	0% ∼ 10%	More than 10%
7%	6.6%	8.9%	39.1%	38.4%

10. What is your risk preference?

Extremely risk averse	Risk averse	Risk neutral	Risk seeking	Extremely risk seeking
1.3%	14%	18.1%	37.6%	29.1%

11. Investment experience in the last 6 months

Stock	Bond	Option Futures	Deposit	SRP
76.5%	34.8%	5.7%	59.7%	65.6%

12. (Subjective) Expected return on SRPs

Less than 4%	4∼6%	6∼9%	9∼12%	More than 12%
0.9%	1.8%	13.7%	37.7%	45.9%

13. How did you start your first investment in SRPs?

Media	Brochures	Brokers	Friends	Internet
11.8%	14.5%	57.2%	9.3%	7.2%

14. Why do you invest in SRPs? (Multiple answers are possible.)

Reputation of issuers	Safe underlying	High expected returns	Principal is protected	Others
9.2%	43.7%	42.8%	25.9%	1.7%

15. Why do you think is the main risk factors of SRPs? (Multiple answers are possible.)

Downside risk	Market manipulation	Defaults of issuers	Others
76.2%	22.4%	8.3%	14.3%

16. Source of information for SRPs

Media	Brochures	Brokers	Internet	Term sheets
10.7%	26.4%	53.7%	5.7%	3.4%

17. How much are you satisfied with explanations by brokers?

Very satisfied	Satisfied	Neutral	Dissatisfied	Very dissatisfied
29.2%	57%	13%	0.6%	0.2%

Chapter 2

The Wisdom of Crowds in FinTech: Evidence from Initial Coin Offerings

2.1 Introduction

In recent years, initial coin offerings ("ICOs") have emerged as a popular way to fund blockchain-related startups, raising $13.7 billion in the first five months of 2018 (PwC, 2018). Through this new form of crowdfunding, an entrepreneur raises capital by creating and selling a virtual currency or "token," which provides a set of rights to its holders, including access to a platform, and can be resold in the secondary market.

In such initial offerings, information asymmetry between an entrepreneur and outside investors and the resulting adverse selection are two inherent problems that could hinder successful fundraising (e.g., Akerlof, 1970; Leland and Pyle, 1977; Rock, 1986).

This chapter is based on my paper, "The Wisdom of Crowds in FinTech: Evidence from Initial Coin Offerings," which is coauthored with Jongsub Lee (Seoul National University) and Tao Li (University of Florida). This work was presented at 18th FDIC/JFSR Banking Research Conference (2018), Florida Blockchain Think Tank (2018), 13th CAFM Conference (2018) and Fintech, Social Finance, and Financial Stability (2018).

To overcome asymmetric information problems, traditional primary market issuances, in particular initial public offerings ("IPOs"), are delegated to financial intermediaries that perform due diligence, determine the value and riskiness of the businesses, and price and sell the new securities (Diamond, 1984). For example, IPO underwriters conduct a bookbuilding process that helps gauge investor demand and price the new equity effectively. However, in blockchain-based crowdfunding, a fast-growing FinTech sector, such financial intermediation is absent by design thanks to decentralized bookkeeping enabled by blockchain networks. What makes ICOs so successful without financial intermediation? Rather than piggybacking on underwriter-driven intermediation, the ICO market relies heavily on the "wisdom of crowds" – the collective action of a large group of individuals rather than that of a single expert.

In this paper, we explore how the wisdom of crowds helps overcome the information asymmetry associated with an ICO. This mechanism works in two stages by (1) certifying the quality of the underlying venture before a token sale starts, and (2) harnessing the wisdom of crowds during the fundraising period. The diverse opinions from a number of online analysts can lead to an aggregate signal that closely reflects the true quality of the risky startup. Such market-based certification by individual analysts not only screens out "lemons" (e.g., Akerlof, 1970; Leland and Pyle, 1977), but is also associated with significant investor interest during initial periods of the sale, encouraging subsequent investors to invest regardless of their own information (Welch, 1992). Unlike in IPO bookbuilding, each token subscription in an ICO is broadcast to all potential investors through a blockchain network. With a critical mass of supporters, harnessed wisdom of crowds would quickly result in meeting the pre-specified funding target (Cong and Xiao, 2018; Li and Mann, 2018).

Our analysis, built on a novel and comprehensive sample of 1,549 completed ICOs from January 2016 to March 2018, aims to uncover how the crowds' wisdom affects

fundraising success, an event in which an ICO reaches its minimum fundraising target ("soft cap"). Equally importantly, we examine whether the "wisdom of crowds" phenomenon is related to tokens' secondary market performance. Our study is unique in its reliance upon primary market subscription information *during* the fundraising period, and only this allows for an empirical analysis of investor subscription dynamics in ICOs (Cong and Xiao, 2018; Li and Mann, 2018). To the best of our knowledge, we are the first to utilize these sequential transaction records to study primary market offerings.[1]

First, the probability of a successful fundraising campaign increases by 19.8 percentage points (relative to the unconditional success probability of 42.7%) with every one-standard-deviation increase in the average analyst rating, controlling for ICO characteristics. This result supports the positive intermediary role these experts play in a market without traditional underwriters. Due to reputational concerns, the majority of our analysts from ICObench, the most comprehensive database on ICOs, are likely to be unbiased in issuing their ratings.[2] The result remains qualitatively similar, however, when we exclude analysts who are founders of other ICOs, and thus are more likely to be biased. Similarly, the number of experts covering a token sale also positively predicts fundraising success. These results suggest that outside investors tend to follow a large crowd of informed analysts when making risky investment decisions.

The fraction of tokens for sale to outside investors would be inversely related to management's skin in the game. Without formal governance and incentive mechanisms post-sale, such as voting power to oust directors and stock incentive plans for managers, a significant stake retained by management could play a crucial governance role to align incentives of the management team and token holders (Leland and Pyle, 1977).

[1] In a concurrent paper, Howell et al. (2018) show summary statistics of investor subscriptions in a case study on Filecoin, which is included in our sample.

[2] Most of the experts in our sample are repeat analysts, who are likely concerned about their reputation in the long run. They use real names and disclose biographic information. Moreover, ICObench assigns each analyst an ICO Success Score, which is publicly accessible. See Section 2.3.2 for more details.

Consistent with this signaling theory, we find that a one-standard-deviation increase in the fraction of tokens for sale is associated with a 6.1 percentage point decrease in the marginal probability of ICO success.

In addition, various other ICO mechanisms could bolster investor participation by promoting consumer demand and governance. We find that arranging a (private) presale can boost ICO success likelihood by 15.2 percentage points, indicating that successful presales are interpreted by later investors as evidence that earlier investors held favorable information (Cong and Xiao, 2018). In the absence of underwriters, a presale is also a clever way for insiders to gauge demand from informed investors such that they can set a more informed price in the main sale. This in turn leads to a higher likelihood of success.

Although early-bird bonus tokens could attract investors early in the process, excessive bonuses or discounts can result in credit rationing by rational investors as many of these sales are believed to be scams (Stiglitz and Weiss, 1981; Sagar, 2017). We find that ICOs providing large bonuses or discounts are 10.9 percentage point *less* likely to conclude successfully. This echoes the U.S. Securities and Exchange Commission's (SEC) recent warnings against token sales with overly generous bonuses.

ICOs featuring multi-language whitepapers/websites or accepting multiple currencies are significantly more likely to succeed, reflecting the global nature of token sales and the ease of transactions with an expanded set of payment methods. However, anti-money laundering measures, such as a Know Your Customer policy, negatively predict fundraising success. This suggests that these investor governance proxies are associated with higher transaction costs, reducing the participation incentives of outside investors in token sales.

Consistent with our main analysis above, we find that favorable analyst ratings positively predict total gross proceeds, while the fraction of tokens for sale, high bonuses,

and a Know Your Customer policy are associated with a lower amount raised. We also use a Cox proportional hazards model to perform a duration analysis, which largely yields consistent results. Interestingly, token sales that feature multi-language websites or whitepapers or accept multiple currencies take a longer time to complete. This could be explained by the fact that these ICOs mainly target (small) retail investors, who typically participate in later stages of the sale by "following the crowd."

To investigate the channel through which quality certification by informed analysts leads to successful fundraising, we resort to the unique primary market subscription data collected from Etherscan, a leading "block explorer" that provides information on every transaction for Ethereum-based ICOs. For each token sale, we download information for all the transactions taking place between the ICO start and end dates. We obtain a transactions sample of over 650 ICOs, and for each ICO transactions are aggregated to the daily frequency.

This novel data set reveals that successful token sales attracted more than 2,000 backers on average, compared to the 39 supporters in failed ICOs. In successful ICOs, early investors also purchase tokens more aggressively, with the first-day per capita subscription at nearly 0.5% of token supply. This highlights the importance of the certification role large investors play in helping coordinate between investors (Sockin and Xiong, 2018). Furthermore, investor subscriptions during initial periods (even at the block-level) strongly predict subsequent token sales and eventual fundraising success, which is consistent with an "up-cascade" phenomenon described by Cong and Xiao (2018). Importantly, better analyst ratings ex ante not only predict stronger first-day token sales, but are also associated with faster sales in earlier periods. Decentralized blockchain-based fundraising is also in line with the notion of "crowd"-funding illustrated by Strausz (2017).

We also find that early token investors achieve better investment returns than their followers. This ability of earning superior returns corroborates their role in leading the crowd of subsequent investors, and justifies their participation decisions in early stages of an ICO.

Before turning our attention to whether and how our proposed certification mechanisms affect post-ICO token performance, we present key statistics about secondary market trading. Borrowing the terminology from the IPO literature, we first document that the first-day return or short-run underpricing (e.g., Ibbotson and Jaffe, 1975; Ritter, 1984; Beatty and Ritter, 1986) is significantly positive for listed tokens. The average underpricing is nearly 160%, substantially higher than the average IPO underpricing of 18% for the U.S. (Ritter, 2004). However, the median ICO underpricing is just 24%. Although newly issued tokens on average do not appear to underperform in the long run (up to one year), the median long-run excess return over a cryptocurrency index is -125%, statistically significant at the 1% level. This is consistent with the well-documented phenomenon of long-run IPO underperformance in small stocks (Ritter, 1991). The average (median) first-day turnover is 6.7% (1.3%), considerably lower than that for IPOs (Ritter, 2018).

Regarding cross-sectional patterns, we find that ICOs with generous bonuses have significantly lower first-day returns, all else being equal. This is consistent with Habib and Ljungqvist (2001) as bonuses can be viewed as marketing tools to attract early investors. Rational secondary market investors will request a lower price that is close to the intrinsic value of tokens sold in the primary market. Analyst ratings, however, do not appear to predict underpricing. This is because any extra demand induced by good ratings is likely to be fully absorbed during the sale, as most ICOs do not reach the maximum funding goal ("hard cap"). Importantly, the average analyst rating positively predicts 3-month returns, suggesting that market-based certification by large

crowds effectively signals the genuine quality of the startups in the long run (beyond the primary market phase).

Overall, we find strong empirical support for market-based certification mechanisms that are in play behind the emerging success of ICOs. In particular, decentralized analysts' opinions appear to mitigate information asymmetry and predict profitable investments in the long run. These results are consistent with the general notion of wisdom of crowds in the FinTech era (Strausz, 2017; Cong and Xiao, 2018; Li and Mann, 2018).

We make the following three contributions to the literature. First, we carefully examine the determinants of successful ICO fundraising campaigns. Following industry convention, we consider an ICO successful if it meets the minimum fundraising target. Existing empirical research on ICOs, represented by Amsden and Schweizer (2018) and Momtaz (2018), use exchange listing as their main criterion for ICO success. Because it can take a startup up to several months to list its token post-sale, this exchange trading-based definition of ICO success is likely to suffer from substantial measurement errors. In fact, among successful ICOs in our sample, only 60% saw their tokens listed as of May 31, 2018. Collecting information on soft caps and gross proceeds for all ICOs in our sample requires a tremendous amount of manual work, which enables our paper to make a significant contribution to the empirical literature on ICOs.[3]

Second, we analyze how our proposed two-tier mechanisms of wisdom of crowds affect both ICO fundraising and secondary market performance. We not only identify what explains ICO fundraising success, but also factors that affect tokens' short-run underpricing and long-run performance. The fact that analyst ratings predict long-term superior investment performance is a deep cross-sectional insight on token pricing.

[3]Amsden and Schweizer (2018) use only hard cap information for presales, which can not help identify eventual fundraising success. Momtaz (2018) defines ICO success based on positive returns measured from the first-day opening price (not the offer price) to first-day closing price. His return-based definition of success misclassified nearly 25% of successful ICOs in our sample, including several most successful ICO, such as Dragon Coins (the largest ICO ever), EOS, and Status, among others.

Last but not least, we are the first to analyze primary market subscription patterns *during* token sales. Using novel data on sequential investor subscriptions, we find evidence that is consistent with the existence of up-cascaded wisdom of crowds as predicted in Cong and Xiao (2018). That is, first-day token sales strongly predict subsequent token purchases and eventual success. Relatedly, positive analyst ratings ex ante not only predict stronger first-day subscriptions, but is also related to faster sales in earlier periods.

In addition to Cong and Xiao (2018), Li and Mann (2018), Sockin and Xiong (2018), our paper is related to several other studies in the context of ICOs. Catalini and Gans (2018a) show that by eliciting consumers' willingness to pay, ICOs may increase entrepreneurial returns beyond what can be achieved through traditional equity financing. Chod and Lyandres (2018) demonstrate that ICOs can facilitate risk-sharing without diluting control rights. Canidio (2018) derives an equilibrium with a positive probability that the entrepreneurs may not develop any products post-ICO.

Our analysis on post-ICO token performance relates to Cong et al. (2018), who develop a dynamic asset-pricing model of tokens and that features inter-temporal feedback effects. Howell et al. (2018) find that liquidity is higher for listed tokens when the issuers offer voluntary disclosure. Studies in this area also include Gandal and Halaburda (2014), Gans and Halaburda (2015), Athey et al. (2016), Ciaian et al. (2016), Fernández-Villaverde and Sanches (2016), and Pagnotta and Buraschi (2018).

Another related area is the burgeoning literature on the economics of the blockchain technology. Yermack (2017) considers how the blockchain technology can lead to changes in corporate governance. Harvey (2016) discusses the mechanics of cryptofinance and their applications including Bitcoin. Cong and He (2018) show that blockchain-based decentralization can mitigate information asymmetry and improve welfare. Malinova and Park (2017) show that the transparent trading environment enabled by blockchain

technology improves investor welfare. Biais et al. (2018) and Eyal and Sirer (2018) analyze cryptocurrency mining games, while Easley et al. (2019), Huberman et al. (2017), and Cong et al. (2019) analyze the compensation and organization of miners.

Lastly, our paper relates to an emerging literature on the wisdom of crowds, which include Surowiecki (2005), Kovbasyuk (2011), Da and Huang (2016), Kremer et al. (2014), and Dindo and Massari (2017).

2.2 Institutional Background

In this section, we briefly introduce the concept and process of an ICO and describe important events that take place after the ICO is completed. We proceed with a brief discussion of the current regulatory environment before concluding with two examples of ICOs. Our overarching goal is to highlight inherent information asymmetry and governance challenges associated with ICOs.

2.2.1 What is an ICO?

An ICO is a new fundraising method made possible by the development of blockchain technology and cryptographic tokens. Through an ICO, a technology startup creates and distributes its (decentralized) platform's digital tokens in exchange for cryptocurrencies, such as Ether ("ETH") or Bitcoin ("BTC"), or fiat currencies to raise public capital to fund their operations and product development. The token typically provides a specific set of rights to its holders, including access to a platform or network, rights to create or develop features for an ecosystem, the right to cast a vote on governance issues, among others.

This approach is radically different from the traditional corporate IPO. With an IPO, investors exchange money for equity shares and voting rights in a relatively established company. In the U.S. and abroad, the process is underwritten by an investment bank and

tightly regulated by securities regulators. To begin the IPO process, a firm is required to file a registration statement, which is a set of documents including a prospectus. Company management and the underwriter conduct a "road show" to meet potential investors and gauge demand for the stock. The underwriter then "builds a book" by accepting orders from investors, who indicate the number of shares they desire and the price they are willing to pay. After the offer price is determined, the management files a final prospectus with the regulatory authority and shares are allocated to investors. The issue is typically "closed" a few days later and shares begin their secondary market trading on a stock exchange. The underwriter also commits to making a liquid secondary market by assigning analysts to cover the stock, and when necessary, it will step in to support the price.

In the freewheeling world of ICOs, however, none of these features exist. There is no investment bank to underwrite the token, conduct a bookbuilding, or support secondary market trading. Token sales usually are open to investors around the world, regardless of where the startup is based. Unlike such arrangement in an IPO, insiders' tokens are often not subject to a lock-up period of 90 to 180 days after an ICO. The vast majority of tokens are deemed as utility tokens as opposed to securities or equity stakes. They typically lack voting rights hence control. As of today, regulatory oversight is minimal in that blockchain startups are not required to file any regulatory documents. In most cases, the startups have no corporate track records or even products, although more established technology firms are increasingly using ICOs to fund their operations. Table A1in the Appendix compares the fundraising steps between ICOs and IPOs.

One major difference between ICOs and traditional crowdfunding is that the latter investment is less liquid. Even in equity crowdfunding in which investors obtain a financial stake in the company they support, it is difficult to resell their securities due to a lack of liquidity. However, when tokens generated in an ICO are listed on exchanges,

they provide the buyers liquidity and a potentially positive rate of return when they are sold at a higher price.

2.2.2 The ICO process and post-ICO events

A typical ICO begins with the presentation of a whitepaper, which describes the business idea and model, the team, and the technical specifications of a project before the ICO. The entrepreneurs lay out a timeline for the project and describe how raised funds will be spent, such as on marketing, and research and development. They often specify a "soft cap" that is the minimum amount received at which the initial offering will be considered a success. Startups usually specify a "hard cap" as well, which is the maximum fundraising goal for a crowdsale. Most projects set very high hard caps that they are unlikely to achieve. In our sample, only 12.2% of ICOs hit their hard caps.

An ICO fixes the number of tokens on offer before the sale. The whitepaper and/or the project website features a discussion of how the tokens will be distributed, including how many tokens are for sale, and how many tokens the insiders will keep.[4] Retaining a reasonably high fraction of tokens with the firm can send a signal to the market that the entrepreneurs have more skin in the game, thus are more likely to expend serious efforts in developing the project (Leland and Pyle, 1977; Downes and Heinkel, 1982).

Investors who purchase tokens early may be given preferential terms, in the form of an "early bird" bonus or discount. One purpose of the bonus or discount is to compensate for the higher risks early buyers bear. Some ICOs include a presale period or also known as a pre-ICO, a token sale event that startup enterprises run before the official crowdsale campaign goes live. Presales generally target larger investors, many of whom are institutional investors. The fundraising targets for presales are usually lower than

[4]Some whitepapers also specify how many (reward) tokens are reserved for bounty programs. During a bounty program, the startup provides compensation for a number of tasks including marketing on social networks, bug reporting or even improving aspects of the cryptocurrency framework. The bounty tokens usually are a small percentage of the total supply of tokens.

those of the main sales and tokens are typically sold at a steep discount. Some companies run presales to collect funds to pay for the expenses incurred for launching their main ICOs. Investors and regulators may be wary of ICOs that provide extremely high bonuses, which sometimes exceed 100% (equivalent to a 50% discount). For example, the SEC has warned investors against token sales that offer high discounts to early buyers.[5] Offering too high a discount can lead to credit rationing from investors (Stiglitz and Weiss, 1981).

Due to the open-source nature of blockchain-based projects, before ICOs start, many startups choose to publish part or all of their initial codes that utilize smart contracts. Once the ICO period is set, the marketing campaign starts, which often begins with an announcement of the token sale on BitcoinTalk.org, a favorite social website of cryptocurrency enthusiasts. Other social networks, such as Medium, Steemit, Reddit, and Twitter, are often used as well.

By industry convention, an ICO is considered as a success if the amount it collects surpasses the soft cap. If a token sale does not reach its soft cap, funds are usually returned to investors. This is the "all-or-nothing" arrangement commonly used in ICOs. In rare cases, the team may decide to move forward regardless. If the hard cap is reached, additional subscriptions will be rejected and the funds will be returned.

After an ICO is successfully completed, the entrepreneurs typically begin to plan for an exchange listing. Most cryptocurrency exchanges require an application and a listing fee, and depending on each case, listing can take from several days to several months. Secondary market trading starts immediately after listing. If the project is implemented successfully and more capital is needed, the startup may return to the ICO market for a seasoned offering. Figure 2.1 presents the timeline for a typical ICO.

[5]In May 2018, the SEC set up a website, HoweyCoins.com, that imitates a fake ICO to educate retail investors about red flags of a potential scam. This ICO features a "double 25% discount" for early investors.

[Insert Figure 2.1 here.]

A changing regulatory environment

During the past few years, ICOs and cryptocurrency exchanges have operated in a legal and regulatory grey area. The first regulatory warning came from the SEC in July 2013, in the form of an investor alert about Ponzi schemes that involved Bitcoin and other virtual currencies. Since then, the SEC has issued a series of warnings suggesting that many token sales may have violated U.S. securities laws, including a July 2017 Report of Investigation that determined that the Ethereum-based DAO tokens were securities, and offers and sales of the DAO tokens were subject to the federal securities laws. In addition to issuing dozens of subpoenas and information requests in February 2018 to technology startups involved in ICOs, the SEC has recently halted several high-profile ICO frauds, such as Centra and AriseBank. In May 2018, more than 40 state and provincial jurisdictions in the U.S. and Canada announced one of the largest coordinated series of enforcement actions to crack down on fraudulent ICOs, resulting in almost 70 open investigations and 35 pending or completed enforcement actions.

Through these regulatory actions, the SEC has made clear that (1) ICO issuers must be able to demonstrate that their tokens are not securities or that they follow securities laws, (2) market participants must ensure that their cryptocurrency activities do not undermine their anti-money laundering and "Know Your Customer" obligations, the latter of which refers to a process of identifying and verifying the identity of potential clients.

Among major economies, China appears to be the most stringent cryptocurrency regulator, banning ICOs and shutting down exchanges in September 2017. The crack-down has recently broadened to Bitcoin miners, forcing some of the industry's biggest players to shift operations overseas. In neighboring South Korea, securities officials in January 2018 disallowed anonymous accounts from trading cryptocurrencies. European

Union ("EU") countries, together with Switzerland, Singapore, and Japan, have taken a relatively friendly stance toward cryptocurrency regulation. However, in April 2018 the EU approved a regulation proposed in December 2017 that requires cryptocurrency exchanges to register with authorities and apply due diligence procedures, including a Know Your Customer policy.

Due to such regulatory pressure and demand from cryptocurrency exchanges to combat money laundering, startups that launch ICOs increasingly ask their clients who participate in token sales to go through a Know Your Customer process. Many recent ICOs have also routinely prevented investors in the U.S., China and certain other countries from participating in their ICOs.

Two examples of ICOs

To give the reader a flavor of how an ICO actually works, we provide a description of two such cases. The first ICO illustrates features of successful ICOs, while the second highlights issues associated with a failed ICO.

The Aragon Token Sale

Founded by Luis Cuende and Jorge Izquierdo in Spain, the Aragon Network is a decentralized application built on the Ethereum Blockchain that allows users to create and manage decentralized companies. It enables users to implement basic features such as governance, fundraising, payroll and accounting, among other features. Aragon also includes a token (ticker ANT), which grants voting rights for making decisions about the direction of future development.

Aragon published a whitepaper in both English and Chinese on April 20, 2017, introducing its business model, functioning of the organization and features of the token.[6] Aragon is among the few ICOs that require a relatively long vesting period for founders,

[6]Aragon's whitepaper is available through https://github.com/aragon/whitepaper.

who will vest 25% of their tokens every six months after the sale (two-year vesting with six-month cliffs). Aragon is also a leading startup that publishes how it uses the funds raised, detailing each expenditure on its website, including the addresses of the company's accounts and the vendors'.[7]

On the same day, the token sale was officially announced in a blog post on Aragon's website. The sale was originally planned for four weeks, from May 17 to June 14, 2017. Aragon sought to sell 70% of tokens to investors, and accepted only ETH. In the first two weeks, one ANT token was priced at 0.01 ETH (equivalent to $0.90 on May 17, 2017), and the price would increase to 0.015 ETH per token in the remaining weeks. Aragon also implemented a hidden cap of 275,000 ETH (or roughly $25 million), which was not revealed at the time of the sale.

Due to overwhelming demand, the hard cap was reached in about 30 minutes and the sale ended. There were 6,593 transactions from 2,616 unique addresses, spanning 134 Ethereum blocks. Proposed transactions valued over $8 million did not go through before the sale ended. Figure 2.2, Panel A plots minute-by-minute investor contributions and the cumulative contributions, which indicates that within seven minutes Aragon raised over half of the hard cap. Panel B shows the value of tokens held by top investors. The top 10% of holders purchased about 80% of sold tokens. ANT began trading the next day, May 18, 2017, with an opening price of $1.49 per token and closing price of $1.52. The closing price on May 18, 2018, one year later, was $3.99.

[Insert Figure 2.2 here.]

Ebitz's ICO

[7]Each post-ICO expenditure Aragon incurs can be viewed through http://transparency.aragon.one/#/. Aragon stated that it would use the funds raised to further develop its software, implementing security audits, and hiring additional developers and operational staff.

In November 2016, a group of self-described "ethical hackers" announced the launch of Ebitz cryptocurrency, a clone version of ZCash, the 21st largest cryptocurrency by market value. Both platforms aim to protect privacy by publishing only each transaction ID on a public blockchain, but information on the sender, recipient, and amount of the transaction remains private. Unlike ZCash, however, Ebitz did not support large rewards to the founders or the standard consensus-based mining algorithm. The Ebitz ICO went live on November 28, 2016 and would end on December 26, 2016 or when the hard cap of 500 BTC was reached.

Ebitz planned to sell 95% of the 21 million emitted tokens to participants, while allocating the remaining 5% to developers and bounty programs. The platform offered an annual interest of 3% to its token holders. The ICO accepted both BTC and ETH as valid currencies for payment. Participants who invested during the first two days were promised a 25% early-bird bonus, while it was fixed at 20% for the remainder of the week. Bonuses for the second and third weeks were 15% and 10%, respectively.

Two days after the sale started, an investor revealed on BitcoinTalk that the email server for Ebitz actually belonged to the domain of Opair, a dubious platform that promoted a decentralized debit card system using its own token. The Opair platform was shut down in the summer of 2016 after users discovered that the LinkedIn profiles of some of the team members were fabricated.

Ebitz's website was quickly removed. However, the ICO still managed to raise about 200 BTC which were valued at $156,000 at the time. There was some speculation that these BTC mostly came from the developers themselves in an attempt to start a cascade and entice outside investors to purchase their tokens.

2.3 Data Sources and Sample Overview

2.3.1 The ICO sample

Our sample of ICOs announced between January 1, 2016 and March 31, 2018 is constructed using ICObench.com.[8] ICObench is one of the oldest rating platforms on ICOs and arguably maintains the most comprehensive database on ICOs. From the website, we collect the following information: startup name, token ticker, country of incorporation, ICO status (completed, ongoing, or upcoming), start and end dates of an ICO, soft and hard caps, gross proceeds, types of currencies accepted for an ICO, bonus/discount terms, token price, the number and percent of tokens for sale, whether an ICO includes a presale, whether an ICO has a Know Your Customer policy, and headline and individual ratings.[9] Excluding seasoned token offerings, we obtain an initial sample of 2,633 ICOs, which include all completed, ongoing, and upcoming sales. We then update the ICO status variable through May 31, 2018 by checking startups' websites for ongoing and upcoming offerings as of March 31, 2018. Our final sample includes 1,549 completed ICOs.

For certain variables, most notably soft and hard caps, gross proceeds, token price, bonus terms, and the number and percent of tokens for sale, ICObench misses information for a large number of ICOs.[10] Therefore, studies solely based on ICObench's data are subject to severe sampling bias. To ensure sample completeness, we manually collect such information from the startups' official websites and their ICO whitepa-

[8]ICObench's revenue mostly comes from listing fees for new ICOs. An entrepreneurs can choose to publish her ICO on ICObench using regular listing or priority listing. The former is free of charge, while the latter costs 0.1 BTC (equivalent to $643.1 as of June 30, 2018). As the website has "a very long queue of pending ICOs to be listed on ICObench it might take several days" for an ICO to be listed, entrepreneurs are encouraged to apply for priority listing.

[9]Gross proceeds, soft and hard caps, token price are often quoted in ETH, BTC, another fiat currency or a combination of these currencies, we convert the figures to dollars as of the ICO end dates. Converting the figures to dollars using the ICO start dates yield similar qualitative results.

[10]For example, ICObench collected information on the number and percent of tokens for sale for just 829 ICOs, while we collected such information for another 684 ICOs. Figures on gross proceeds were available for 484 ICOs, and we added such data for another 302 ICOs.

pers. For ICOs whose websites or whitepapers are not available, we rely on news searches on popular blogging and social networking websites such as Steemit, Medium and BitcoinTalk.org, and other credible data providers such as ICORating.com and TokenData.io.

Furthermore, we use the sources above to verify each data point that ICObench collects for these variables. In case of inconsistencies between data entries on ICObench and those from official websites and whitepapers, we use the latter.

As many ICOs seek to raise funds from investors around the globe, the startups often make their websites and whitepapers available in different languages. We visit each ICO's website and record the number of languages available. If a website or its associated whitepaper is no longer available, we take a conservative approach and assume only the English language is available for that ICO.

Lastly, as ICObench assigns multiple industry codes to each ICO, we rely on ICORating's industry classification. Matching on startup name and ticker yields 982 ICOs with industry classification. For the remaining 667 ICOs, we follow ICORating's methodology and manually collect industry data from their official websites, whitepapers, and news searches.

The bonus schedule

To precisely measure individual investors' investment returns (Section 2.5.2), we manually collect from whitepapers bonus schedules for the 298 listed tokens that offered bonuses or discounts. For each listed token, we record the start and end dates for each bonus period and their associated bonus/discount. In 102 ICOs, bonuses/discounts are conditional on sales milestones. We therefore calculate the end dates for each bonus ourselves. For robustness analysis, we also obtain bonus schedules during the presale periods. Figure A1 in the Appendix shows one example for each type of bonus structure.

2.3.2 Analyst ratings

One unique feature of ICObench's ratings is that in addition to its assessment algorithm that uses more than 20 different criteria (nicknamed the "Benchy" rating), a large number of experts provide their own ratings on each ICO. The assessment algorithm divides evaluation into four different groups: team, ICO information, product representation, and marketing and social media. All ICOs are rated with a scale from 1 to 5. Regarding the team, the algorithm records the number of team members, photos, full names, and social media links. A team is considered more trustworthy if any member has participated in multiple ICOs, either as an advisor or team member. ICO information refers to basic variables such as token ticker, ICO start and end dates, soft and hard caps, among others. For product presentation, the algorithm checks the availability of such information through whitepapers, milestones, and video presentations. The algorithm also monitors activity on various social networks to determine whether an ICO team reaches potential investors.

Independently, each expert assigns a rating from 1 to 5 to an ICO for team, vision, and product. Analysts disclose their real names and bigraphic information such as job titles and employers. Like the assessment algorithm, analysts consider a team strong and trustworthy if it keeps the community updated with project progress and/or has participated in other cryptocurrency-related projects. Vision mostly concerns what a platform aims to achieve in the mid-term and long-term future. When evaluating products, analysts consider the following aspects: (1) product maturity level – working products are easier to evaluate than concepts, (2) technology, both blockchain- and non-blockchain-related, (3) specific problems with their products/services, (4) product roadmap that shows short-term and long-term strategies and growth, and (5) projects' commitment to understanding the market environment.

The headline rating is a weighted average of all participants' ratings based on their expertise, years of experience in the field, and available publications.[11]

ICObench assigns an ICO Success Score to each analyst, which is the average score for all the ICOs they have rated. An analyst can receive up to 15 points for each ICO. Crucially, analysts receive up to 10 points if a token gets listed, which is an important milestone for ICO success. ICObench explains analysts can improve their success score by identifying successful ICOs.[12]

2.3.3 Primary market subscription data

One unique feature of any cryptocurrency is that each transaction needs to be broadcast to all participants on a blockchain network before it is validated. For example, each transaction involving ETH will be sent to every Ethereum node (a computer with an installed Ethereum program) on the Internet. Miners, a group of competitive bookkeepers, will validate the transaction by completing a cryptographic "proof of work," which involves a cryptographic hash function that takes an input and delivers an output. The purpose of the "proof of work" is to make sure that the transfer is genuine and there is no double spending or counterfeiting. If the majority of participants validate the transaction, it will be added to a "block" on the Ethereum public ledger, which is a decentralized database containing the entire history of every ETH transaction. All transactions on any Ethereum block are viewable by the public.

The same procedure applies to any ICO, where each token transfer/subscription during the primary market sale is recorded on the public ledger associated with the underlying blockchain platform. As 78% of our sample projects use the Ethereum platform to run their ICOs, we collect primary market subscription data on all Ethereum-based

[11]The reader is referred to https://icobench.com/ratings for a more detailed description of ICObench's rating methodology.

[12]Details about the success score can be accessed via https://icobench.com/faq#q-6-4.

ICOs from Etherscan, a leading "block explorer" that allows users to search for information about blocks and transactions on the Ethereum Blockchain. We manually search each Ethereum-based ICO's name and if there is a match, we record the ICO's contract hash address, a 42-character string. For each contract address, we download information for all transactions taking place between the ICO start and end dates. These include the transaction address, sender address, receiver address (all of which are 66-character strings), transaction time (e.g, May-20-2018 04:16:49 PM +UTC), quantity of tokens transferred, total dollar value of the transfer. We initially identify 952 ICOs that have primary market transactions available.

Our next step is to identify the addresses for ICO insiders and primary market investors. In most cases, it is straightforward to identify the insiders as all transfers are originated from one single address. The rest are primary market subscribers. However, when it is difficult to cleanly identify the insiders because multiple addresses are used to transfer tokens to investors, we take a conservative approach to exclude these ICOs in our main analysis. To facilitate the analysis on sequential investor subscriptions, we also drop the ICOs that distribute tokens *after* the token sales end. These criteria yield a transactions sample of 654 ICOs.

Our paper is unique in its reliance upon primary market subscription data, and only this allows for an empirical analysis of investor subscription dynamics during ICOs.

2.3.4 Secondary market prices, volumes, and transactions

Following a successful ICO, a token is expected to be listed on an exchange or several exchanges simultaneously. We collect its daily closing prices from CoinMarketCap.com, a website that is a top source for pricing data on hundreds of cryptocurrencies. For each token, CoinMarketCap aggregates pricing information from all major exchanges and produces one standard price quote in real time. It also publishes the 24-hour trad-

ing volume among major exchanges. We manually search each successful ICO's name on CoinMarketCap, and download its dollar price series, daily trading volumes, and circulating supply of tokens if available.[13] Our sample consists of 433 tokens that were listed following an ICO, and that had offer prices available.

To calculate individual token investors' investment returns, we collect from Etherscan secondary market sales data for each investor who purchased tokens in the primary market. Etherscan records both primary and secondary market transactions.

2.3.5 Sample overview

Figure 2.3 plots the frequency of ICO starts and the rate of fundraising success over our sample period. As in Mironov and Campbell (2018), we consider an ICO successful if its soft cap was reached or the project raised more than $0.5 million in the absence of a soft cap.[14] Just 19 ICOs opened in 2016. However, the market took off in 2017, enjoying a quarterly increase of 171% in 2017. The number of ICO starts continued to rise through the first quarter of 2018. The fundraising success rate was over 90% in the first half of 2017, and then dropped sharply in the second half of the year. The deteriorating success rate potentially reflects the "lemon's problem" that lower quality ICOs were created when the market became red hot (e.g., Akerlof (1970)). The decline in success rates also coincided with increasing regulatory scrutiny worldwide.

[Insert Figure 2.3 here.]

[13]To verify whether information from CoinMarketCap is accurate, we also download pricing and volume data from popular alternative pricing sites, such as Onchainfx.com, CryptoCompare.com, and CoinGecko.com. We find that for the vast majority of cryptocurrencies, the prices and volumes from CoinMarketCap are highly correlated with those from the alternative sites (the correlation is typically above 0.9).

[14]Instead of using this industry convention to define ICO success, Amsden and Schweizer (2018) and Momtaz (2018) use exchange listing as the criterion for ICO success. After a successful fundraiser, it can take a startup several months to list their token on an exchange. Some entrepreneurs may choose not to list their tokens. Since ICO is a recent phenomenon, both Amsden and Schweizer (2018) and Momtaz (2018) likely miss a substantial number of successful ICOs. In our sample, only about 60% of successful ICOs were listed as of May 31, 2018.

The top five largest ICOs to date are Dragon Coin, Huobi, HADC, Filecoin, and Tezos, all of which closed between July 2017 and March 2018.[15] Combined, these token sales raised nearly $1.4 billion, accounting for 15% of all ICO proceeds raised globally during the same period, according to EY Research. Appendix Table A2 shows the top 10 largest ICOs as of May 2018, with information on their fundraising periods and gross proceeds.

Table 2.1 reports the frequency of sales and fundraising success rate for ICOs from each of the top 10 countries and industries. As shown in Panel A, the U.S. is the most popular country for blockchain startups, followed by Russia. Interestingly, ICOs from Switzerland enjoy the highest fundraising success rate, assisted by "crypto-friendly" guidelines recently issued by Swiss regulators Atkins, 2018). However, Russia-based ICOs are least likely to succeed, followed by ICOs from Canada.

[Insert Table 2.1 here.]

Financial Services is the most popular industry for ICOs, attracting over 10% of all completed token sales. This is perhaps not surprising given that the original Bitcoin Blockchain was developed to replace the traditional centralized financial system. The average fundraising success rate for the top 10 industries is about 56%, substantially greater than the sample average of 45%. ICOs in the Blockchain Infrastructure industry achieve the highest success rate.

Analyst ratings

Our sample comprises 9,493 distinct ratings by 316 unique online analysts. Their average (median) ratings for team, vision, and product are 3.9 (4.0), 3.8 (4), and 3.6 (4.0), respectively. Of these analysts, 165 are founders or senior managers from blockchain-related companies, 96 are advisors to these firms, and 73 are blockchain analysts or

[15]We do not count Telegram's record-breaking $1.7 billion token sale in the first quarter of 2018 because it was structured as a private sale.

followers. Some analysts are also investors, with 58 being cryptocurrency/blockchain investors and 16 being venture capitalists or angel investors. Notably, there are also 52 engineers and technicians, who are potentially able to provide valuable insights into the technical aspects of ICOs. Other analysts are from fields such as finance/business (39), marketing (30), consulting (21), among others. Note that these categories are not mutually exclusive as a given analyst can play multiple roles.

Patterns in primary market subscriptions

In a bookbuilding process commonly used in IPOs, the underwriter solicits investors' bids, which are used to construct a demand curve and allocate shares to the investors. However, in ICOs, no underwriter is building a book for the sale. Rather, the price and offering period are set ex ante by entrepreneurs, and the gross proceeds equals investors' cumulative subscriptions by the end of the offering period. To understand the path to fundraising success, we resort to our unique second-by-second subscription data, which are aggregated at the daily frequency.

Figure 2.4, Panel A plots the time series patterns of daily token sales for both successful and failed ICOs. The blue bars (line) represent daily (cumulative) token sales as a percentage of total token supply for successful ICOs, while the orange bars and line represent the corresponding figures for failed sales. In successful ICOs, investors purchase nearly 15% of token supply on the first day, while the 30-day cumulative demand is about 30% of the total supply. In contrast, in failed sales, investors on the first day buy fewer than 1% of all tokens for sale, and the cumulative sales are just about 2.5% of token supply.

The plot in Panel B further shows hourly token sales on the first day of a sale for successful and failed campaigns. In the first hour alone, successful ICOs sell about 7% of tokens offered, while failed campaigns sell few coins during this period. Volumes in successful sales go down as time passes. Panel C repeats the analysis for each block

during the first hour of token sales. A block contains transactions within a 14-second interval. The time-series pattern is highly consistent with those in the daily and hourly plots. These patterns highlight the importance of "winning the battle" during the initial periods of an ICO, which often determines the outcome of the sale. Such a concave pattern of cumulative token sales is consistent with an information cascade mechanism proposed by Cong and Xiao (2018), Li and Mann (2018), and Welch (1992).

[Insert Figure 2.4 here.]

2.4 Fundraising Success and Campaign Duration

2.4.1 Successful fundraisers versus failed ICOs

Our first analysis examines the characteristics of successful ICO fundraising campaigns. Column (1) in Table 2.2 reports the attributes of ICOs that successfully raised funds, while column (2) compares the characteristics between successful and failed ICOs.

[Insert Table 2.2 here.]

Regarding ex ante ICO characteristics, most importantly, successful ICOs on average had a rating of 3.3 (out of 5) by online experts, 0.7 points higher than that for failed token sales. The difference is significant (at the 1% level), suggesting that analyst certification before an ICO goes live is an important predictor of fundraising success. In the absence of traditional underwriters who play a critical intermediary role in the IPO market, analysts fill the void and potentially help reduce information asymmetry in ICOs. These experts are likely unbiased as biased ratings may be uncovered in the long run, resulting in reputational damages. This could be a novel market solution for token sales, all of which feature decentralized fundraising platforms through blockchain technology. Successful token sales also attracted more analysts to initiate coverage,

with the number doubling that for failed ones. This is indicative of the "wisdom of crowds" phenomenon observed in the ICO market. The pattern based on the medians is qualitatively similar.

The average soft cap or minimum funding goal for successful ICOs was $6.8 million, nearly identical to that set by unsuccessful fundraising campaigns (the difference is insignificant at the 10% level). The hard cap or maximum goal for a successful ICO on average was $88 million, more than double the amount for failed ones (the difference is not significant at the 10% level). The median amounts tell a more consistent story, as both groups of ICOs have similar median soft and hard caps.

As an important governance indicator, the percent of tokens to be sold to investors measures management's skin in the firm. Successful ICOs sought to sell 57% of generated tokens to outsiders, compared to the target of over 61% in failed ones, with the difference being significant at the 1% level. This is a strong indication that investors embrace token sales in which management retains more stakes. Note that without formal governance and incentive mechanisms post-ICO, such as voting power to oust directors and performance compensation package commonly used after an IPO, management's stakes could play a vital governance role to align insiders' and outsiders' interests.[16] This is also consistent with Leland and Pyle's (1977) signaling hypothesis that firm value is positively related to the fraction of equity retained by the original stockholders.

Nearly 40% of successful ICOs included a presale before the main token sale, 18.2% higher than failed ICOs. Presales typically are open to only institutional or high-net wealth investors, and the proceeds raised are often used to cover the costs of launching the main ICOs. Successful initial sales are interpreted by subsequent investors as evidence that earlier investors held favorable information, encouraging later investors to

[16]Distinct from most IPOs in which management loses majority control of the firms (except some high-tech IPOs such as Facebook and Snap where management controls voting rights via dual class shares), in token sales management does not give away control as voting rights typically are not attached to tokens.

invest regardless of their own information. We posit that in the absence of underwriters, a presale is a clever way for insiders to gauge demand from informed investors such that they can set a more informed price for the main ICO. Given the market power these early investors enjoy, it is not surprising that presales often provide a steeper discount than the main sale. This is analogous to the analysis of informed IPO investors by Biais et al. (2002) and Benveniste and Spindt (1989).

Interestingly and perhaps counterintuitively, high bonus offers, defined as 20% or more, are more prevalent in failed ICOs (the difference is significant at the 10% level).[17] Although generous bonuses can attract investor subscriptions in early periods of an ICO, many of these token sales provide extremely high bonuses that sometimes exceed 100%. Wary investors may conclude that such ICOs are potential lemons or scams, and are reluctant to extend credit to the entrepreneurs (Stiglitz and Weiss, 1981).

ICOs that adopted a Know Your Customer policy were less likely to conclude successfully. Only 12% of successful token sales asked for customer identification, compared to nearly 17% for failed ICOs (the difference is significant at the 1% level). Similarly, ICOs that required advance registration or restricted sales in certain countries (*Participation restriction*) were also less likely to succeed. This suggests that Know Your Customer and related restrictive measures adopted by many ICOs since Q3 2017 tend to dampen investor demand, the bulk of which likely comes from countries like China, South Korea, and the U.S.

Lastly, token sales featuring multi-language websites or whitepapers tended to be more successful, reflecting the fact that potential token purchasers are not based in a single country and language barriers exist. ICOs that accepted multiple (digital) currencies were more likely to succeed (the difference is significant at the 10% level), compared to ICOs that took just one currency. Given that major digital currencies

[17]Sagar (2017) considers ICO bonuses on offer exceed 20% as a red flag. Using an alternative threshold of 30% yields consistent results in our main analysis.

such as BTC and ETH are drastically volatile, expanding the options of currencies, thus increased liquidity, can facilitate transactions.

Panel B compares key ex post ICO outcomes between successful and failed ICOs. On average, successful sales raised $18.7 million, far more than the $2.7 million failed ones raised, with the difference being significant at the 1%. In contrast, according to the Crowdfunding Center, successful crowdfunding campaigns on average raised just $29,900 in 2016, a tiny fraction of the amount raised in ICOs. Specialized crowdfunding platforms, such as Kickstarter, support even smaller fundraising campaigns (Xu, 2016). On the other hand, tech IPOs in 2017 grossed over $250 million on average (Ritter, 2018). Considering that most blockchain-related projects raise funds before actual launches and the companies are much smaller than IPO firms, ICOs have become an increasingly important source of alternative fundraising.

Perhaps more tellingly, successful ICOs on average achieved 59% of the hard cap, while unsuccessful ones obtained just 15.4% of the hard cap. Successful ICOs had more than 2,100 supporters on average, compared to the 39 subscribers in failed ICOs. Such a stark difference highlights the role of information cascades that can potentially lead to lopsided fundraising results.

Interestingly, successful ICOs on average took 30.0 days to complete, shorter than the 37.8 days failed fundraisers took (the difference is significant at the 1%). This is because although most ICOs are scheduled for about one month, successful ICOs often finish early (see the Aragon token sale in Section 2.2), usually at the time when they hit the hard cap.

2.4.2 Determinants of ICO success

Table 2.3, Panel A reports the results of predictive regressions where the dependent variable is ICO fundraising success, which equals one if an ICO reaches its soft

cap or the project raises more than $0.5 million in the absence of a soft cap (Mironov and Campbell, 2018).[18] The set of regressors are the same as those presented in Table 2.2 with the critical difference that all variables in the regressions are measured at the time of an ICO announcement. The sample includes all 1,461 ICOs for which we have the required information. Column (1) displays the probit coefficients and their associated marginal probabilities. Column (2) reports coefficients from a linear probability model with country and quarter fixed effects. Since the results in the two columns are qualitatively similar, we rely on the probit model for our discussion.

[Insert Table 2.3 here.]

Consistent with results shown in Table 2.2, the average analyst rating, all else being equal, has a significantly positive effect (at the 1% level) on the likelihood of a successful fundraising campaign. A one-standard-deviation increase in the average rating is associated with an increase in the marginal probability of 19.8 percentage points. Relative to the unconditional probability of ICO success of 42.7%, the incremental probability is remarkable. This finding is consistent with the positive intermediary role these experts play in a market where traditional underwriters are absent. In robustness analysis, we exclude experts who are founders of other ICOs, and therefore are potentially biased, and the results are qualitatively similar. The number of analysts covering an ICO also positively predicts fundraising success, suggesting a "wisdom of crowds" phenomenon in which investors tend to follow a crowd of analysts when making investment decisions.

The coefficients associated with three more ICO characteristics support their governance or signaling roles. For a one-standard-deviation increase in the fraction of tokens for sale, there is a 6.1 percentage point decrease in the marginal probability of ICO success (significant at the 1% level). This suggests that investors favor ICOs in which

[18]For ICOs that do not specify a soft cap, changing the target to $0.25 million or $0.75 million yields similar results.

management's and investors' incentives are more aligned through higher inside stakes (Leland and Pyle, 1977; Downes and Heinkel, 1982). On the other hand, token sales providing large bonuses or discounts are 10.9 percentage point less likely to successfully conclude the fundraising effort, reflecting credit rationing to ICOs with overly generous bonuses, many of which are believed to be potential scams (Sagar, 2017). Including a presale can boost the success likelihood by 15.2 percentage points (significant at the 1% level), suggesting that successful initial sales can promote the subsequent token sales by harnessing the widsom of crowds (Cong and Xiao, 2018; Li and Mann, 2018).

The probability of ICO success decreases by 5.7% percentage points (significant at the 1% level) when a token sale requires customer identification, a process that can deter some potential customers. Replacing this variable with *Whitelist*, an indicator equal to one if customers are required to register in advance for a sale, we obtain similar results (see Table A3 in the Appendix). Consistent with findings in Table 2.2, token sales that feature multi-language websites/whitepapers or accept multiple currencies are more likely to succeed, reflecting the global nature of token sales and the ease of transactions with an expanded set of currencies.

2.4.3 Gross proceeds and ICO success

Our analysis on fundraising success in subsection 2.4.1 ignores the degree of success, which can be measured by gross proceeds or gross proceeds as a percentage of the specified hard cap. In Table 2.3, Panel B, we repeat the same analysis by using these two alternative dependent variables. Our sample becomes smaller as this analysis requires that information on gross proceeds and/or hard cap is available.

As shown in column (1), gross proceeds increases by $4.7 million when the average analyst rating increases by one point. This is substantial given that the sample average gross proceeds is about $15.2 million. Consistent with results in subsection 2.4.1, the

fraction of tokens for sale and the availability of generous bonuses negatively predict the total amount raised in an ICO. The other covariates are not statistically significant at the 10% level, partly due to the smaller sample employed in this study.

Results reported in column (2) are largely consistent with those in column (1), with the coefficients on the number of analysts and Know Your Customer being statistically more significant.

2.4.4 ICO duration

Campaign duration is an alternative measure of ICO success in that highly successful ICOs often conclude at the time when the hard cap is reached. A longer fundraising campaign distracts management who need to focus on product development after an ICO. To assess the extent of such a cost, Table 2.4 reports results connecting ICO duration to key metrics identical to those shown in Table 2.3. In column (1), the dependent variable is the logarithm of the number of days between an ICO's start date and the completion of the sale.

[Insert Table 2.4 here.]

As expected, a favorable analyst rating is associated with a quicker sale. Interestingly, ICOs that feature multi-language websites or whitepapers or accept multiple currencies take a longer time to consummate. This may be explained by the fact that such ICOs mainly target retail investors. In Li and Mann (2018), a fundraising process is formalized as a multi-stage game where heterogeneous investors with private signals of a project's quality decide whether and when to participate in the ICO. In equilibrium, investors with stronger signals participate early and those with weaker signals "follow the crowd." Assuming that retail investors generally possess weaker signals, they are more likely to participate in later stages of the sale. Therefore, if the entrepreneurs aim to achieve decentralized ownership, they can target a retail investor base by making information on

the token sale available in multiple languages and accepting multiple currencies. These measures could potentially prolong the fundraising process. The same logic applies to ICOs that require a Know Your Customer policy.

The Cox (1972) proportional hazards model,[19] reported in column (2), yields qualitatively similar results. The estimated hazard ratio (equal to the exponentiated coefficient) associated with the dummy variable *Presale* implies that, conditional on an ICO being in process, the probability of a sale closure on a given day is 25.8% higher if the ICO has a presale. The coefficient estimates on other covariates are largely consistent with those in the OLS model.

2.5 Investor Subscriptions and the Path to ICO Success

2.5.1 Initial and subsequent subscriptions

Cong and Xiao (2018) show that the all-or-nothing mechanism used by crowdsales leads to uni-directional information cascades in which investors rationally ignore private signals and imitate preceding agents only if enough preceding investors decide to support the ICO. Li and Mann (2018) also feature such an information cascade in a simpler setting. We test whether this mechanism may exist in ICOs using a regression framework. Consistent with this prediction, in Table 2.5, we find that investor subscriptions on the first day strongly predict token sales on the second day, during the next four days or the next 14 days. Notably, other covariates only weakly predict subsequent token sales.

[Insert Table 2.5 here.]

[19] In the Cox model, the hazard function at a given time t (from initiation), conditional on the failure to complete an ICO, is characterized as $h_i(t) = h_0(t)e^{X_i\beta}$ where $h_0(t)$ is an unspecified (or nonparametric) function.

As theories on information cascades emphasize an all-or-nothing threshold, we replace first-day subscription with an indicator equal to one if total subscriptions on the first day are greater than 5% of total supply. The results are qualitatively similar. Using the hourly or block-level frequency also yields consistent results. Overall, the evidence presented in this subsection is consistent with the "wisdom of crowds" phenomenon, a unique feature of fundraising in the era of FinTech.

2.5.2 Initial investors' investment returns

As early investors appear to harness the wisdom of crowds during the primary sale, it is natural to examine whether these investors achieve a higher return than later participants. Moreover, any investment success among this group of investors can justify their incentives in investing with the risky startups.

Following Frazzini (2006) and Brav et al. (2018), we assume that investors use the purchase price as the base for calculating gains and losses when a token is purchased multiple times during a primary sale, and partially sold in the secondary market on different dates. Investors use a cost-based mental accounting method (first in, first out) to compute the quantity of tokens for the corresponding reference price. For any investor on any date t, the aggregate cost basis equals

$$Basis_t = \lambda^{-1} \sum_{n=0}^{t} Tokens_{t,t-n} Price_{t-n},$$

in which $Tokens_{t,t-n}$ is the number of tokens an investor acquires on date $t-n$ that she still holds on date t, λ is a normalizing constant such that $\lambda = \sum_{n=0}^{t} Tokens_{t,t-n}$. $Price_{t-n}$ is the token price on $t-n$, taking into account discounts/bonuses offered at that time. The beginning date is the first day of primary sale.

Our main return measure, an investor's basis-adjusted return, is then computed as the percentage deviation of her aggregate cost basis from the current secondary market

price. We calculate basis-adjusted returns at the end of the first day, first month, three months, and six months, respectively.

To measure the timing of investor subscriptions, we use the difference between the date of purchase and ICO open time. When an investor purchases a token on several different dates, we value weight the purchase time using the number of tokens bought. For example, an investor buys 100 tokens three days after an ICO opens, and purchases another 200 tokens six days after the open time. Her effective purchase time is five days after the ICO opens ($3 \times \frac{100}{300} + 6 \times \frac{200}{300} = 5$).

Table 2.6, Panel A reports results connecting investors' basis-adjusted returns to effective time of token purchase. ICO-level fixed effects control for any time-invariant unobservable factors involved with each ICO. As shown in column (1), buying tokens one day earlier is associated with a 0.3 percentage-point increase in basis-adjusted returns at the end of first day after tokens are listed, all else being equal. This relationship might be mechanical as earlier investors, who are supposed to sell out on the first day, enjoy more bonuses (the adjusted R-squared is abnormally high at 0.93). However, we obtain similar results when examining basis-adjusted returns by the end of the first month, three months, and six months. Investors have much more flexibility in timing sales when the holding horizon extends to six months. Thus, our results suggest that early investors possess at least some superior investment skills.[20] The model fit also becomes more reasonable.

[Insert Table 2.6 here.]

Basis-adjusted returns are not defined if investors have completely sold out their holdings by time t (Frazzini, 2006). In our sample, 4.0% of investors sell out their positions on the first day in the secondary market, while the fraction increases to 22.7%

[20] In unreported results, we find that when the discount an investor receives increases by one percentage point, basis-adjusted returns increase by 1.4 to 4.0 percentage points, depending on the investment horizon. This confirms that early investors' superior returns are not entirely due to higher bonuses.

in six months. For these investors, we create another variable, *Realized return*, which is defined below.

$$Realized\ return = \frac{Total\ proceeds\ in\ the\ secondary\ market}{Total\ primary\ market\ investment} - 1,$$

In Panel B, we repeat the analysis in Panel A by replacing the dependent variable as *Realized return*. The results in column (2) show that subscribing tokens one day earlier is associated with a 0.4 percentage-point increase in realized returns during the first month of secondary market trading (significant at the 5% level). We obtain similar results when extending the window to three months or six months.

Overall, our findings suggest that early token investors achieve better investment returns than their followers. This ability of earning superior returns corroborates their role in leading the crowd of subsequent investors, and justifies their participation decisions in early stages of an ICO.

2.5.3 Predicting primary market subscriptions

To examine what factors may affect primary market subscriptions, we rely on the following empirical specification:

$$Subscription_{ijt} = \alpha + \beta Rating_{ijt} + \delta X_{ijt} + \lambda_t + \mu_j + \epsilon_{ijt} \tag{2.1}$$

in which $Subscription_{ijt}$ represents measures that gauge primary market subscriptions for token sale i that starts in quarter t in country j. $Rating_{ijt}$ is the average analyst rating. X_{ijt} is a vector representing ICO-level covariates. λ_t represent quarter fixed effects and μ_j are country fixed effects. ϵ_{ijt} is the error term. Standard errors are adjusted for heteroscedasticity and they are clustered along the quarter dimension.

In column (1) of Table 2.7, we examine the factors that predict first-day investor subscriptions as a fraction of token supply. The average rating from experts is a strong predictor of first-day subscriptions. A one-standard-deviation increase in the average rating is associated with an increase in first-day tokens sold of 2.3 percentage points (significant at the 1% level). Compared to the sample average first-day tokens sold of 3.9%, the incremental change is remarkable. This suggests that positive analyst ratings help harness demand in the absence of reputable underwriters in this decentralized market. On the other hand, generous bonuses are negatively correlated with first-day sales volumes, suggesting that the investor crowd regard such ICOs as potential frauds and thus are unwilling to extend credit to them.

Figure 2.4 suggests that due to higher demand in early periods of an ICO, the curve for cumulative daily token sales is concave. That is, tokens are sold faster in earlier periods. In column (2) of Table 2.7, we replace the dependent variable with a measure that captures the concavity of the cumulative demand curve. Our intuitive measure equals the cumulative token sales on the 15th day minus (1/2)×the cumulative token sales on the 30th day, with the latter date being the last day of an average ICO. In the benchmark of steady daily fundraising volumes, this concavity measure equals zero. However, it is strictly positive (negative) for a hot (cold) ICO. See Figure A2 in the Appendix for more details on how we construct this measure. Consistent with the results in column (1), a better analyst rating predicts faster token sales in earlier periods, all else being equal. The estimate is significant at the 1% level. ICOs with presales also see faster token sales early on, suggesting that successful initial sales can harness information cascades in a main sale. On the other hand, a Know Your Customer policy is found to slow down token sales, potentially causing delays for subscriptions by some customers.

An alternative measure for investor subscriptions during the ICO period is the number of days until 10% of token supply is sold, which approximates the soft cap for many token sales. As a number of ICOs never sold 10% of all the tokens to investors, there exists right-censoring. To overcome this issue, we apply a Tobit model with an upper limit of 30 days, which is approximately the average duration for ICOs. Using a 60-day upper limit yields similar results. As shown in column (3), the average rating and the number of analysts negatively predict the time to sell 10% of token supply. Consistent with the results in column (2), ICOs with presales take a shorter time to sell 10% of their token supply.

2.5.4 The effect of early subscriptions on fundraising success

To examine how early investor subscriptions affect ICO success, we first use a univariate probit model that relates the first-day token subscription with ICO success. The results in column (1) of Table 2.8, Panel A indicate that a half-standard-deviation increase in first-day subscription is associated with an increase in the marginal probability of 57 percentage points. The order of magnitude is large as the unconditional probability of ICO success is 42.7%. To control for other covariates, we use equation (1) to obtain the residual for *first-day subscription*, which by definition is orthogonal to all the covariates. Then we include this residual in our main regression equation for ICO success. Results in column (2) confirms that higher first-day subscriptions strongly predict a higher likelihood of success.

[Insert Table 2.8 here.]

Replacing the success indicator with gross proceeds, we obtain qualitatively similar results, as shown in Panel B. Overall, results in this section shed light on an important question: how do ICO fundraisers become successful? Using novel data on each investor's subscription, we are the first to show that early contributes (i.e, the first-day

subscriptions) are crucial to eventual successes of ICOs, possibly attributed to information cascade (Cong and Xiao, 2018; Li and Mann, 2018).

2.6 Post-ICO Performance

Given that ICO fundraising is just the first step for a successful blockchain-based project, it is crucial to analyze how major variables related to governance and certification mechanisms affect the short-run and longer-run token performance.

2.6.1 ICO Underpricing

IPO underpricing, a phenomenon that the share price jumps on the first day of trading relative to the offer price, has interested financial economists for decades (See Ritter and Welch (2002) and Ljungqvist (2007) for a review of theories and empirical findings on IPO underpricing). As explained in Section 2.2, although the market structure of ICOs is quite different from that of IPOs, it is useful to examine whether underpricing also exists in the ICO market.

[Insert Table 2.9 here.]

First-day return or underpricing equals $P_1/P_{ICO} - 1$, in which P_1 and P_{ICO} are first-day closing price and offer price, respectively. As shown in Table 2.9, the average ICO underpricing during our sample period is 158%, substantially higher than the average IPO underpricing of 18% for the U.S. (Ritter, 2018). However, the median ICO underpricing is only 24%, indicating that the distribution of underpricing is highly skewed to the right by very large first-day pops. Consequently, money left on the table is also highly right-skewed – the average figure is $42 million while the median is just about $1.3 million.

Since the average time from ICO completion to listing is 18.5 days, we adjust first-day returns by the returns on the value-weighted benchmark index of ETH and BTC during this "wait period." The average and median adjusted underpricing figures are slightly below the raw numbers. It is worth noting that to calculate first-day token returns, we rely on manual collection of primary market *offer prices* as the first-day return is measured from the token offer price to the first trading day closing price. This differentiates our study from Amsden and Schweizer (2018) and Momtaz (2018), the latter of which features a first-day return measured from the first trading day opening price to closing price.

First-day turnover, defined as the first-day trading volume divided by the number of tokens issued, has an average (median) value of 6.7% (1.3%). It is significantly lower than the average first-day turnover of 66.2% for IPOs between 2001 and 2016 (Ritter, 2018). This highlights the illiquid nature of many listed tokens.

Cross-sectional patterns in underpricing

To understand which factors may explain ICO underpricing, we start by regressing the first-day return on the wait-period benchmark return and the same covariates as in Table 2.3. As shown in column (1) of Table 2.10, a one-percent increase in the wait-period benchmark return is associated with a 2.1 percentage point increase in first-day returns, all else being equal (the coefficient is significant at the 1% level). This suggests that a hot overall cryptocurrency market substantially boosts investors' sentiment in these newly listed "alt-coins."

[Insert Table 2.10 here.]

On the other hand, ICOs featuring generous bonuses have a significantly lower first-day return than that for token sales without large bonuses. This is consistent with Habib and Ljungqvist (2001), who find that promotional/marketing costs of an IPO

can substitute for IPO underpricing. ICO bonuses can be viewed as marketing tools to attract early investors in the primary market. Knowing that early investors bought tokens at large discounts, rational secondary market investors would demand a lower price, which is closer to the intrinsic value these early investors' tokens represent.

Interestingly, analyst ratings do not appear to be correlated with underpricing. Any additional demand spurred by a good rating is likely to be fully absorbed in the primary market subscription process. This is reasonable as the majority of ICOs do not hit the hard cap. Another plausible reason is that unlike investment banks, who likely have incentives to underprice IPO stocks (Ritter and Welch, 2002; Ljungqvist, 2007), online analysts do not have financial stakes in startups undergoing an ICO and do not control the pricing process.

2.6.2 Long-run Performance

The median token return during the six month period after the first trading day significantly underperforms the benchmark of ETH and BTC by 37 percentage points, as shown in Table 2.9 (it is different from zero at the 1% level). However, the average six-month excess return is highly positive at 66%, reflecting the skewed nature of the underlying return distribution. To reduce risk mismeasurement, we use the top 10 digital currencies as an alternative benchmark, and obtain similar results.

The median one-year excess return is -125% (significant at the 1% level), while the average one-year excess return is not statistically significant from zero. This is consistent with Ritter (2018), who finds that in general, the longer is the horizon, the lower is the median return, due to the right-skewness of buy-and-hold returns.

It is worth noting that since ICOs are a recent phenomenon, many listed tokens still do not have a history of one year or more. Therefore, we have fewer observations for

one-year returns than the shorter-term returns. We plan to analyze longer-term returns in the future.

Cross-sectional patterns in long-run performance

To investigate possible explanations for the long-run performance of ICOs, this subsection documents various cross-sectional patterns. Columns (2)-(5) in Table 2.10 report the results of a multiple regression using the raw 1-month, 3-month, 6-month, and 1-year total returns on ICOs as the dependent variable. The explanatory variables are the market-adjusted first-day return, the corresponding total return on the cryptocurrency market, the relevant covariates, and country fixed effects. This model is similar to Ritter (1991), who uses it to study long-run underperformance of IPOs. Because the dependent variables, 1-month, 3-month, 6-month, and 1-year total returns, are so skewed, the residuals are also highly non-normal. Consequently, bootstrapped t-statistics are reported.

For all the horizons, long-run returns are inversely related to initial returns, exhibiting potential mean reversions. The coefficient on the market return of 0.91-1.16 (except the 2.02 for the six month window) is somewhat surprising, as many would expect that the average alt-coin betas would be substantially above 1.

Most importantly, we find that a good analyst rating positively predicts both 3-month and 1-year token returns (the estimates are significant the 10% and 5% levels, respectively), but it is not associated with shorter-term returns. Analyst ratings focus on team, vision and product, all of which are long-term indicators of the startup quality. This implies that analyst ratings are informative, even beyond the primary market stage. It suggests that analyst ratings are likely unbiased as biased ratings may be uncovered in the long run, resulting in a reputational cost. This result is also consistent with a key finding in Jia et al. (2018), who use a sample of China's IPOs, and find that

analyst coverage and earnings forecast optimism for an IPO before it starts is positively associated with IPO long-run returns.

Lastly, Know Your Customer policies have a dampening effect on 1-month and 3-month returns. This is not unexpected, as such policies potentially curb demand from wary investors who do not want to reveal their identity. [21]

2.7 Conclusion

This is the first study that uses token subscription data to examine how the wisdom of crowds can help overcome information asymmetry in ICOs. We find that favorable ratings on the underlying project from a crowd of online experts are associated with aggressive initial token subscriptions by investors, which in turn predicts subsequent token sales. Using blockchain-based subscription data, we find that successful token sales attract more than 2,000 investors on average, compared to the 39 investors in failed sales. Such "up-cascades" of wisdom of crowds is the key to successful ICO fundraising (Cong and Xiao, 2018). Analyst opinions also predict long-run token performance in the secondary market. This suggests that the wisdom of the two crowds – analysts and early token investors – can potentially substitute traditional underwriters' intermediary roles in financing blockchain-related ventures.

To prevent the "insanity of crowds," ICO analysts are required to be independent and possess a diversity of opinions (Sehra et al., 2017). Potential backdoor networks need be further investigated to ensure unbiasedness. How to set an optimal all-or-nothing threshold to ensure fundraising success without suffering from short-run underpricing is another important mechanism design problem. We aim to address these questions in future research.

[21] We believe that not only illegal demand is curbed but true believers in decentralized blockchain may find this policy counterproductive or intrusive.

Figure 2.1: An Illustration of the ICO Timeline
This timeline illustrates the timing of events for a typical ICO. A pre-announcement usually is a summary featuring the idea and team for a startup to the cryptocurrency community to gather interest and feedback. The documentation stage typically involves posting a whitepaper on the startup's website that describes the business model and technical specifications of the project. Many startups also publish initial codes for their ICOs. The ensuing marketing campaign often uses cryptocurrency forums and social network sites such as Medium, Steemit, Reddit, and Twitter. Before the official ICO goes live, there may be an optional presale of tokens. After the ICO, tokens are listed on exchanges.

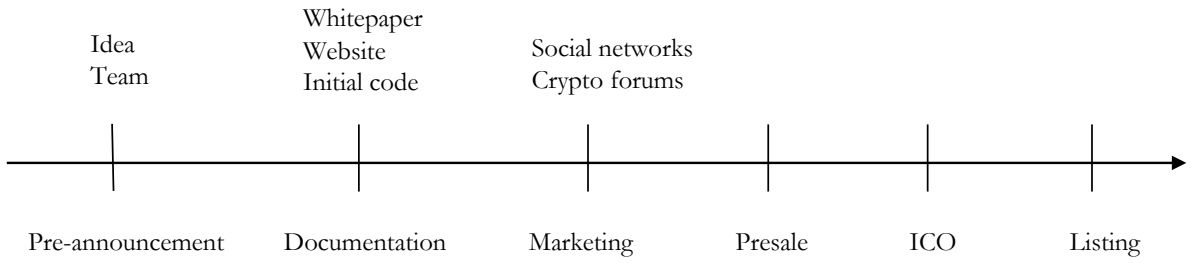
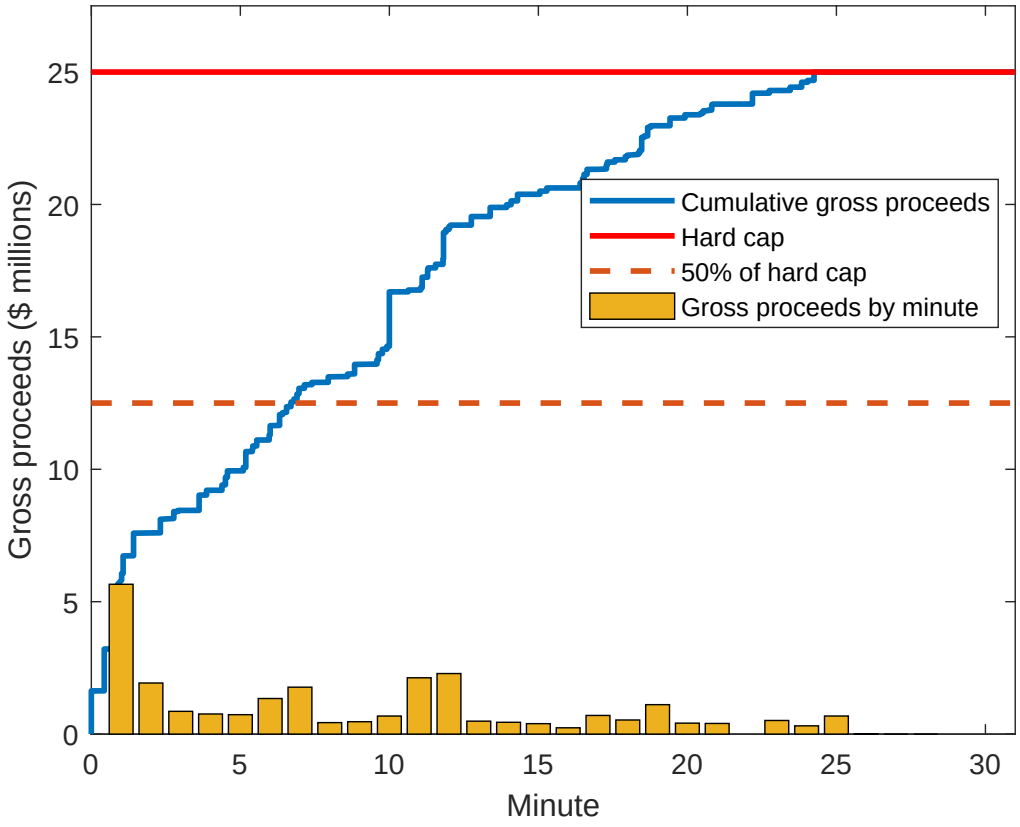


Figure 2.2: The Aragon Token Sale
Panel A shows time-series patterns of token subscriptions for Aragon Network, an ICO that concluded under 31 minutes. The orange bars plot the gross proceeds ($ million) by minute during the sale and the blue line plots the cumulative gross proceeds ($ million) by minute. The red line and red dotted line plot the hard cap and 50% of the hard cap, respectively. In Panel B, the blue curve plots the cumulative tokens held by percent of largest holders.

Panel A. Token subscription in Aragon Network



Panel B. Aragon token distribution by investor

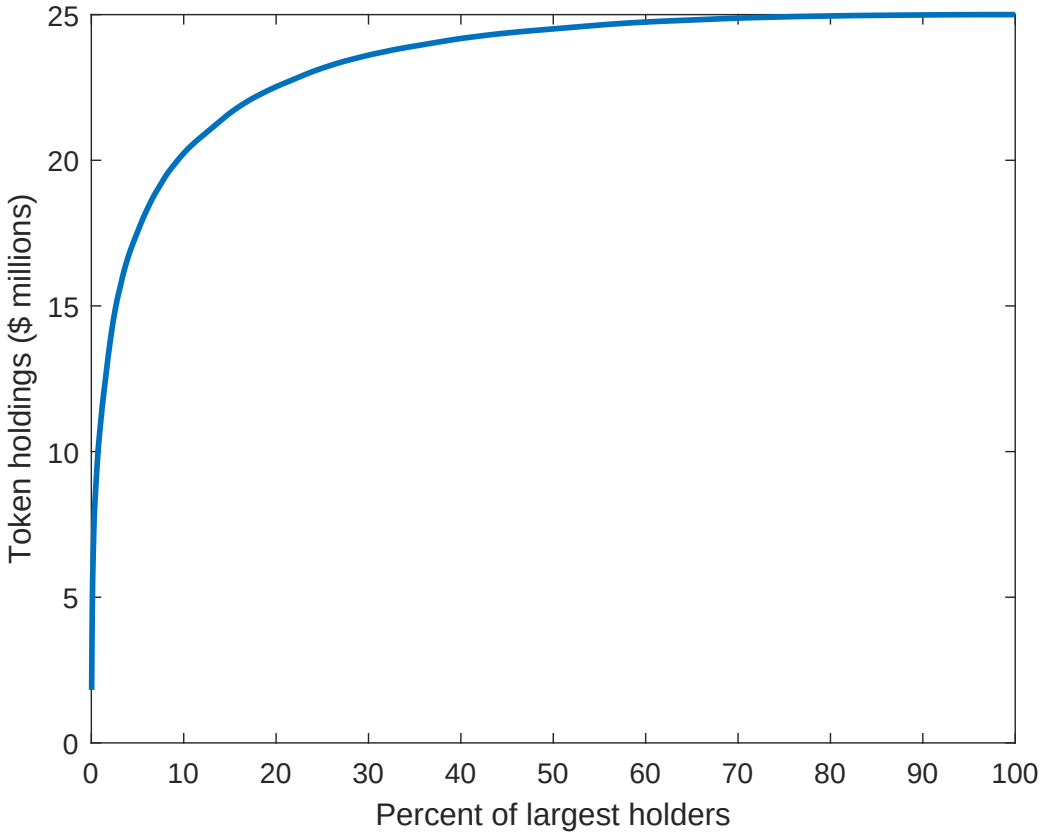


Figure 2.3: ICO Starts and Success
This figure features all ICOs that opened between Q1 2016 and Q1 2018. The blue bars (left axis) plot the number of opened ICOs in each quarter. The red line (right axis) plots the percentage of successful ICOs by quarter. We exclude ongoing ICOs as of May 31, 2018 when calculating the success rates. An ICO is considered successful if its soft cap was reached or the project raised more than $0.5 million in the absence of a soft cap.

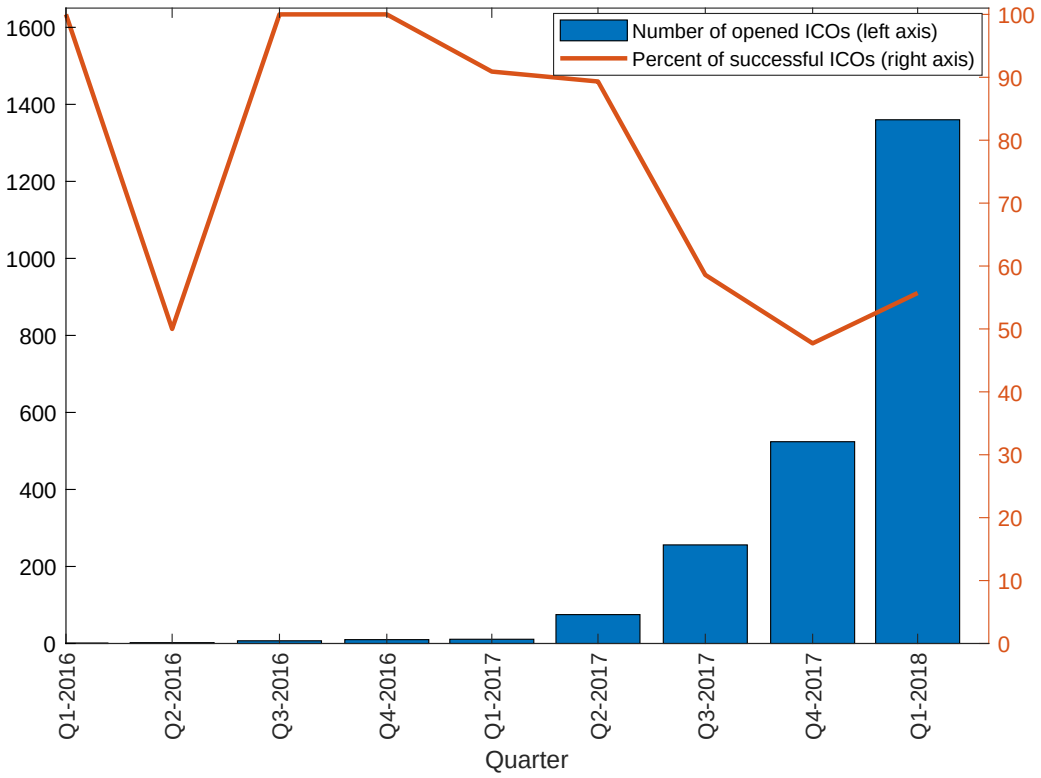
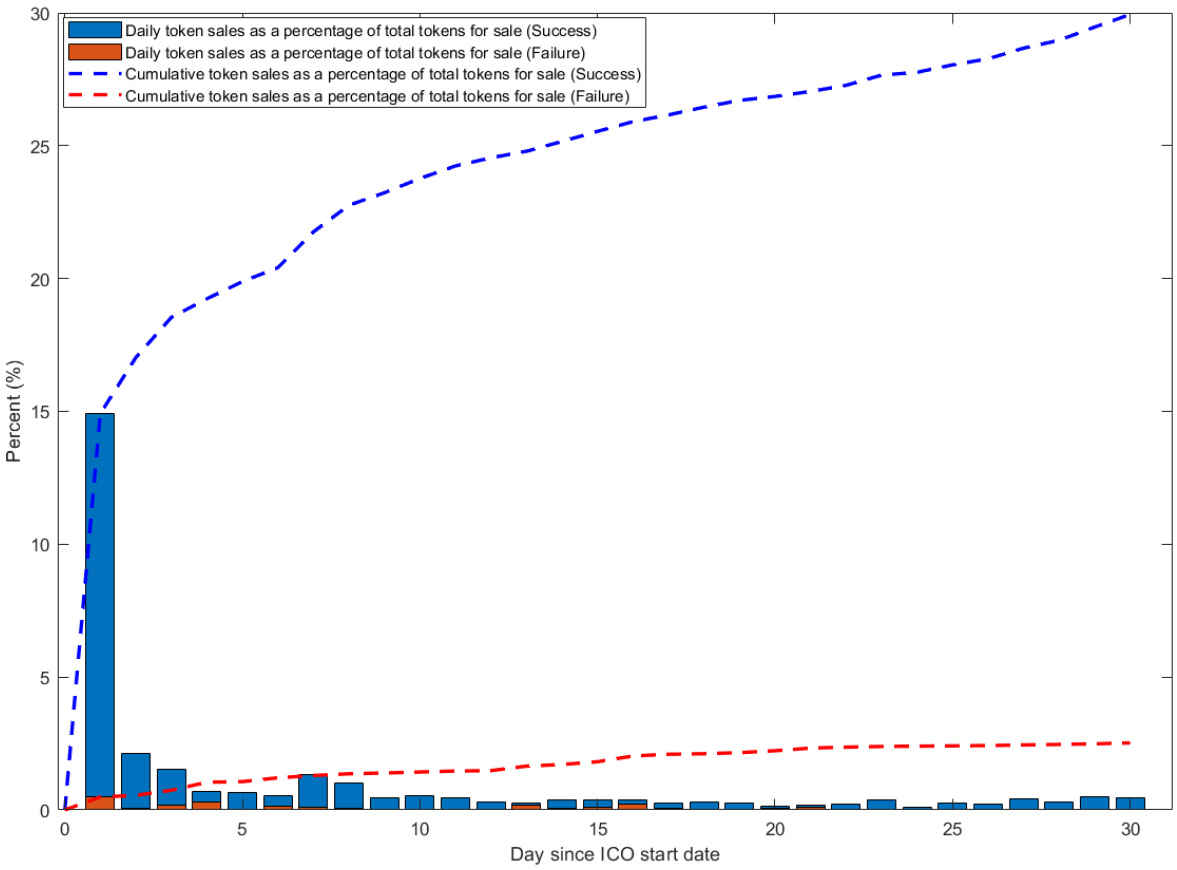
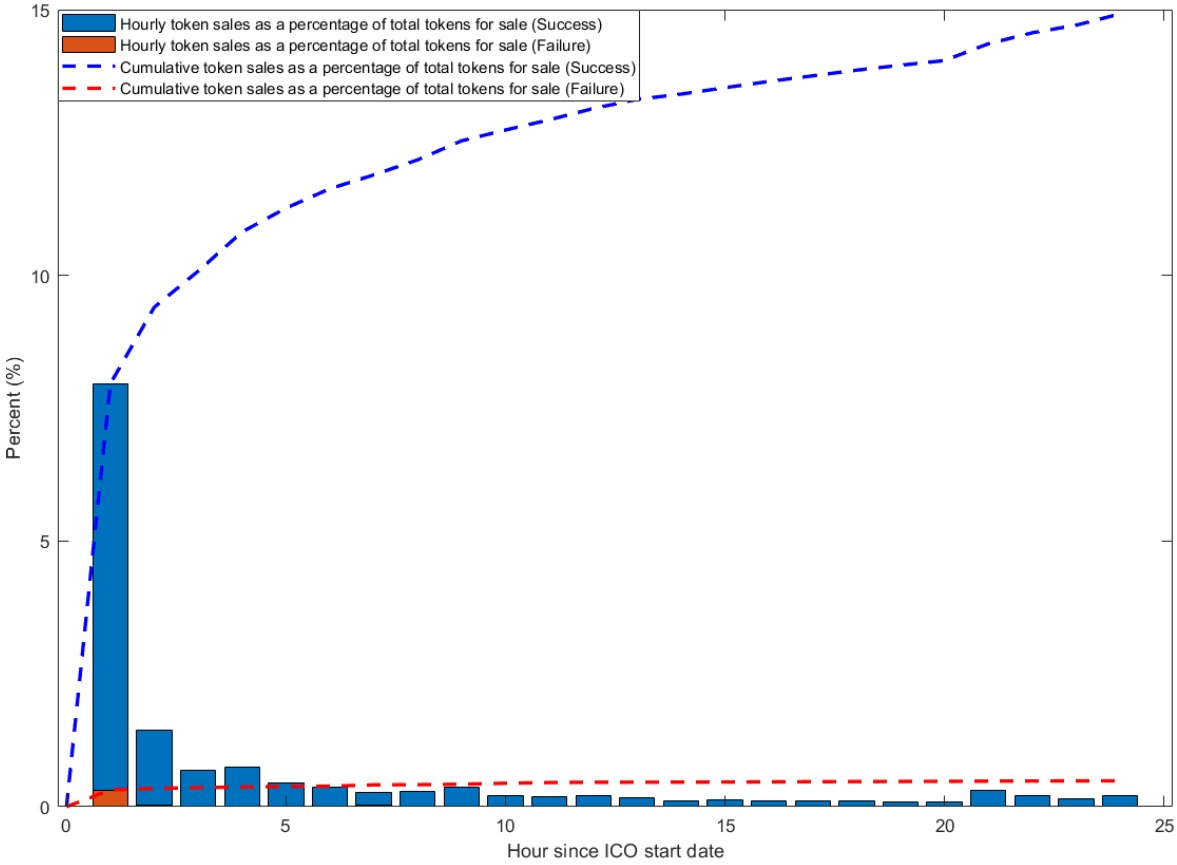


Figure 2.4: Primary Market Subscriptions in ICOs
This figure shows time-series patterns of token subscriptions during ICOs. Our sample includes all Ethereum-based ICOs that sold a positive number of tokens. In Panel A, the blue (red) bars plot the average daily token sales as a percentage of total tokens for sale in successful (failed) ICOs. The blue (red) dotted line plots the cumulative daily token sales as a percentage of total tokens for sale in successful (failed) ICOs. The plot in Panel B shows the average hourly token sales on the first day of ICO for successful and failed ICOs, while Panel C repeats the analysis for each block during the first hour of token sales. A block contains transactions within a 14-second interval. An ICO is considered successful if its soft cap was reached or the project raised more than $0.5 million in the absence of a soft cap.

Panel A. Daily token sales



Panel B. Hourly token sales on the first day of an ICO



Panel C. Block-by-block sales during the first hour of an ICO

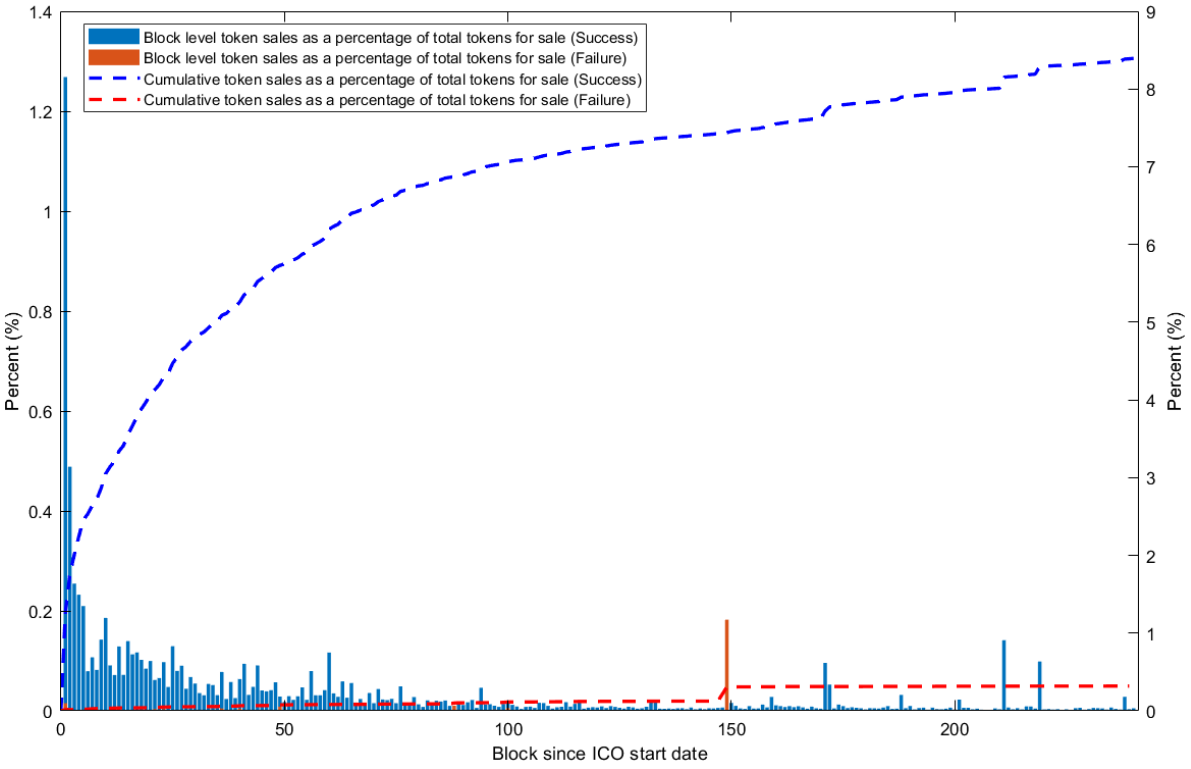


Table 2.1: ICOs by Country and Industry
This table provides descriptive statistics on ICOs from top 10 countries of incorporation in Panel A, and from top 10 industries in Panel B. We identify ICOs through ICObench, a data provider that specializes in ICO analytics. Our sample includes 1,549 ICOs that started between Q1 2016 and Q1 2018 and were completed as of May 31, 2018. In both panels, we report the number and proportion of ICOs within each country/industry, and the associated fundraising success rate. An ICO is considered successful if its soft cap was reached or the project raised more than $0.5 million in the absence of a soft cap.

Panel A. Most popular countries of incorporation

Country	No. of ICOs	Percent of total	Fundraising success rate
United States	278	18.0%	43.2%
Russia	176	11.4%	33.0%
Worldwide or multiple	160	10.3%	41.3%
United Kingdom	137	8.8%	43.1%
Singapore	113	7.3%	57.5%
Switzerland	62	4.0%	66.1%
China (includes Hong Kong)	52	3.4%	51.9%
Canada	50	3.2%	34.1%
Estonia	46	3.0%	47.8%
Australia	31	2.0%	41.9%
Sum	1,105	71.3%	44.2%
Total (all countries)	1,549	100%	45.4%

Panel B. Most popular industries

Industry	No. of ICOs	Percent of total	Fundraising success rate
Financial Services	159	10.3%	61.4%
Gaming and Virtual Reality	101	6.5%	59.3%
Investment	99	6.4%	50.0%
Exchanges and Wallets	99	6.4%	57.9%
Blockchain Infrastructure	82	5.3%	77.8%
Social Media and Communication	69	4.5%	36.1%
Trading	68	4.4%	54.1%
Business Services and Consulting	66	4.3%	51.4%
Marketing and Advertising	60	3.9%	58.8%
Commerce and Retail	58	3.8%	45.5%
Sum	863	55.7%	56.2%
Total (all industries)	1,549	100%	45.4%

Table 2.2: ICO Characteristics

This table reports characteristics of the 704 successful ICOs, and compares them with the 845 failed token sales. Our sample includes all ICOs on ICObench that started between Q1 2016 and Q1 2018 and were completed as of May 31, 2018. An ICO is considered successful if its soft cap was reached or the project raised more than $0.5 million in the absence of a soft cap. *Analyst rating* is the average rating (scale 1-5) for an ICO by independent experts on ICObench. *No. of analysts* is the number of analysts that rate an ICO on ICObench. *Soft cap* is the minimum amount of funds needed and aimed at by the startup to proceed as planned, and *Hard cap* is the maximum amount of capital that it aims to gather. *Presale* is an indicator equal to 1 if an ICO runs a token sale event before the official crowdsale goes live. *High bonus* equals 1 if an ICO offers a bonus over 20% (equivalent to a discount of 16.7%), and 0 otherwise. *Fraction of tokens for sale* is the number of tokens for sale divided by the total number of tokens generated. *Know Your Customer (KYC)* is an indicator equal to 1 if clients are required to provide information to confirm their identity. *Whitelist* is a dummy variable equal to 1 if customers have to register in advance to participate in an ICO. *Participation restriction* equals 1 if an ICO is restricted in certain countries, and 0 otherwise. *Multiple languages* is an indicator equal to 1 if the whitepaper or website for an ICO features more than one language. *Multiple currencies* equals 1 if an ICO accepts multiple currencies (digital or fiat), and 0 otherwise. *Gross proceeds* is the amount raised from investors in millions. *No. of subscribers* is the number of token buyers in an ICO. *Duration of offering* is the number of days between the ICO start and end dates. Column (1) reports the average, median and standard deviation of characteristics for successful ICOs. Column (2) shows the difference in average between successful and failed ICOs and its associated t-statistics. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Ex ante ICO characteristics

	Successful ICOs			Difference between successful and failed ICOs	
	Average	Median	Std. Dev.	Diff. in Avg.	t-stat.
	(1a)	(1b)	(1c)	(2a)	(2b)
Analyst rating	3.287	3.500	0.816	0.665***	15.49
No. of analysts	5.805	3.000	8.361	3.087***	9.45
Soft cap ($ million)	6.748	2.680	13.537	-0.046	-0.04
Hard cap ($ million)	88.204	21.755	777.613	48.385	1.02
Fraction of token for sale	0.572	0.600	0.224	-0.042***	-3.68
Presale	0.395	0	0.489	0.182***	8.26
High bonus	0.347	0	0.472	-0.045*	-1.76
Know Your Customer (KYC)	0.117	0	0.322	-0.048***	-2.67
Whitelist	0.082	0	0.274	-0.051***	-3.25
Participation Restriction	0.105	0	0.306	-0.037**	-2.20
Multiple languages	0.413	0	0.493	0.201***	8.79
Multiple currencies	0.320	0	0.467	0.040*	1.69

Panel B. Ex post ICO characteristics

	Successful ICOs			Difference between successful and failed ICOs	
	Average (1a)	Median (1b)	Std. Dev. (1c)	Diff. in Avg. (2a)	*t*-stat. (2b)
Gross proceeds ($ million)	18.733	9.500	39.735	16.07***	5.28
Gross proceeds/Hard cap	0.590	0.584	0.385	0.436***	11.23
No. of subscribers	2,198.468	658	4,321.582	2,159.710***	10.58
Duration of offering (days)	30.033	29.5	26.560	-7.752***	-5.17

Table 2.3: Determinants of Fundraising Success
This table examines the determinants of fundraising success for all ICOs that opened between Q1 2016 and Q1 2018 and were completed as of May 31, 2018. The sample includes a total of 1,461 ICOs that have all the required information. In Panel A, the dependent variable is an indicator equal to 1 if an ICO reaches its soft cap or the project raises more than $0.5 million in the absence of a soft cap, and 0 otherwise. In Panel B, the dependent variables are gross proceeds in millions of dollars and gross proceeds divided by the hard cap. Our sample is reduced to 727 (543) ICOs that have information on gross proceeds (both gross proceeds and hard cap). All independent variables are as defined in Table 2.2, and are measured immediately before the ICO start date. In each column, we report coefficient estimates, their heteroscedasticity-robust t-statistics and, when applicable, the corresponding marginal probability change induced by a one-unit change in the value of a specific covariate from its sample average. Standard errors are clustered at the quarter level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Fundraising success

Dependent variable:					
ICO success	Probit model			Linear probability model	
	Coefficient	t-stat.	Marg. Prob.	Coefficient	t-stat.
ICO characteristics	(1a)	(1b)	(1c)	(2a)	(2b)
Analyst rating	0.564***	5.17	22.2%	0.164***	4.75
No. of analysts	0.035***	5.16	1.4%	0.012***	8.22
Fraction of tokens for sale	-0.735***	-4.26	-29.0%	-0.160***	-3.24
Presale	0.385***	7.03	15.2%	0.119***	7.95
High bonus	-0.279***	-3.42	-10.9%	-0.083***	-3.12
Know Your Customer	-0.147***	-3.29	-5.7%	-0.057**	-2.27
Multiple languages	0.253*	1.88	10.0%	0.095**	2.35
Multiple currency	0.022	0.46	0.9%	0.015	1.20
Quarterly dummies		Yes		Yes	
Country fixed effects		No		Yes	
Observations		1,461		1,461	
Pseudo R-squared		0.26			
Adj. R-squared				0.32	
% (Dep variable = 1)		45.2%		45.2%	

Panel B. Gross proceeds

Dependent variable	Gross proceeds ($ million)		Gross proceeds/Hard cap	
	Coefficient	t-stat.	Coefficient	t-stat.
ICO characteristics	(1a)	(1b)	(2a)	(2b)
Analyst rating	4.680**	1.98	0.055***	3.23
No. of analysts	0.068	0.41	0.006***	3.25
Fraction of tokens for sale	-17.569***	-2.87	-0.238*	-1.79
Presale	-5.618*	-1.82	-0.023	-0.50
High bonus	-4.747**	-2.13	-0.108**	-2.18
Know Your Customer	-3.260	-1.14	-0.260***	-5.85
Multiple languages	5.450	1.04	-0.014	-0.92
Multiple currencies	0.830	0.46	-0.012	-0.62
Quarterly dummies	Yes		Yes	
Country fixed effects	Yes		Yes	
Observations	727		543	
Adj. R-squared	0.03		0.14	

Table 2.4: ICO Duration

This table relates token sale completion/resolution to ICO characteristics. The sample consists of a total of 1,115 ICOs from Q1 2016 to Q1 2018 that have all the required information. In column (1), the dependent variable is the logarithm of the number of days between ICO kickoff and completion. All independent variables are as defined in Table 2.2, and are measured immediately before the ICO start date. Column (1) reports results of an OLS model while column (2) applies a Cox (1972) proportional hazards model to estimate the hazard rate on a daily frequency for ICO completion. In both specifications, we report coefficient estimates and their heteroscedasticity-robust t-statistics. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable:					
# Days to resolution	OLS		Cox model		
	Coefficient	t stat.	Coefficient	t stat.	Hazard rate
	(1a)	(1b)	(2a)	(2b)	(2c)
Analyst rating	-0.111***	-3.50	0.101***	2.49	1.11
No. of analysts	0.001	0.07	-0.006	-1.20	0.99
Fraction of tokens for sale	0.088	0.68	-0.079	-0.56	0.92
Presale	-0.089	-0.80	0.229***	3.56	1.26
High bonus	0.112***	4.49	-0.017	-0.27	0.98
Know Your Customer	-0.122**	-2.33	0.032	0.28	1.03
Multiple languages	0.140*	1.75	-0.169***	-2.58	0.84
Multiple currencies	0.124***	3.39	-0.225***	-3.36	0.80
Quarterly dummies	Yes			Yes	
Country fixed effects	Yes			No	
Observations	1,115			39,850	
Adj. R-squared	0.07				
Wald Chi-squared				90.86	

Table 2.5: Primary Market Token Subscriptions

This table examines how first-day primary market subscriptions affect subsequent investor subscriptions for all Ethereum-based ICOs that opened between Q1 2016 and Q1 2018 and were completed as of May 31, 2018. The sample includes a total of 544 ICOs that have all the required information. *First-day subscription* is the number of tokens subscribed on the first day divided by the number of tokens for sale. *Second-day, Subscription between 2nd and 5th days, Subscriptions between 2nd and 15th days* are similarly defined. All other independent variables are as defined in Table 2.2, and are measured immediately before the ICO start date. In each column we report coefficient estimates and their heteroscedasticity-robust t-statistics. Standard errors are clustered at the country level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable		
	Second-day subscription (1a)	Subscription b/t 2nd and 5th days (1b)	Subscription b/t 2nd and 15th days (1c)
First-day subscription	0.114***	0.143***	0.166***
	[4.76]	[4.36]	[3.43]
Analyst rating	0.004	0.009*	0.015*
	[1.30]	[1.69]	[1.75]
No. of analysts	0.001	-0.001	0.001
	[0.25]	[-0.35]	[0.70]
Fraction of tokens for sale	0.003	0.014	-0.002
	[0.48]	[1.22]	[-0.11]
Presale	0.003	0.011	0.019
	[0.62]	[1.33]	[1.57]
High bonus	-0.004	-0.010*	-0.010
	[-1.41]	[-1.69]	[-0.91]
Know Your Customer	-0.005	-0.011	-0.034**
	[-1.43]	[-1.45]	[-2.25]
Multiple languages	0.005	0.006	0.001
	[1.09]	[0.74]	[0.04]
Multiple currencies	0.007*	0.016*	0.016
	[1.88]	[1.72]	[1.08]
Quarterly fixed effects	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes
Observations	544	544	544
Adj. R-squared	0.20	0.17	0.22

Table 2.6: Initial Investors' Investment Returns

This table reports results on the relationship between investor subscription timing and investment returns. The sample includes 298 listed Ethereum-based ICOs that opened between Q1 2016 and Q1 2018 and were completed as of May 31, 2018. In Panel A, the dependent variable is an investor's basis-adjusted return, computed as the percentage deviation of her aggregate cost basis from the current secondary market price (Frazzini, 2006). For any investor at any date t, the aggregate cost basis equals $\lambda^{-1}\sum_{n=0}^{t} Tokens_{t,t-n}Price_{t-n}$, in which $Tokens_{t,t-n}$ is the number of tokens an investor acquired on date $t-n$ that she still holds on date t, λ is a normalizing constant such that $\lambda = \sum_{n=0}^{t} Tokens_{t,t-n}$. $Price_{t-n}$ is the token price on $t-n$, taking into account discounts/bonuses offered at that time. The beginning date is the first day of primary sale. *# days waited until subscription* is the difference between the value-weighted date of purchase and ICO open time. In Panel B, the dependent variable is *Realized return*, which equals $\frac{Total\ proceeds\ in\ the\ second\ market}{Total\ primary\ market\ investment} - 1$. We calculate basis-adjusted returns and realized returns at the end of the first day, first month, three months, and six months, respectively., In each column we report coefficient estimates and their heteroscedasticity-robust t-statistics. Standard errors are clustered at the ICO level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Subscription timing and basis-adjusted returns

	Dependent variable: Basis-adjusted return (%)			
	First day	One month	Three months	Six months
	(1)	(2)	(3)	(4)
# days waited until subscription	-0.320***	-0.289**	-0.339**	-0.581**
	[-2.84]	[-2.37]	[-2.25]	[-2.28]
ICO fixed effects	Yes	Yes	Yes	Yes
Observations	482,126	426,552	345,567	195,936
Adj. R-squared	0.93	0.67	0.57	0.72

Panel B. Subscription timing and realized returns

	Dependent variable: Realized return (%)			
	First day	One month	Three months	Six months
	(1)	(2)	(3)	(4)
# days waited until subscription	-0.695**	-0.388**	-0.616***	-0.576***
	[-2.53]	[-2.39]	[-2.61]	[-2.78]
ICO fixed effects	Yes	Yes	Yes	Yes
Observations	22,608	70,791	115,113	123,150
Adj. R-squared	0.99	0.71	0.61	0.57

Table 2.7: Determinants of Primary Market Subscriptions

This table examines the determinants of primary market subscriptions for all Ethereum-based ICOs that opened between Q1 2016 and Q1 2018 and were completed as of May 31, 2018. The sample includes a total of 544 ICOs that have all the required information. We study the determinants of first-day token subscriptions, the concavity of investor subscriptions, and the number of days to sell 10% of token supply. *First-day subscription* is the number of tokens subscribed on the first day divided by the number of tokens for sale. Concavity of subscriptions equals the cumulative token sales on the 15th day minus 1/2 the cumulative token sales on the 30th day, with the latter date being the last day of an average ICO. # days to sell 10% of token supply is the number of days until 10% of token supply is sold. All other independent variables are as defined in Table 2.2, and are measured immediately before the ICO start date. In column (3), we apply a Tobit model with an upper limit of 30 days, which is approximately the average duration for ICOs. In each column we report coefficient estimates and their heteroscedasticity-robust t-statistics. Standard errors are clustered at the country level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	OLS		Tobit
Dependent variable	First-day subscription	Concavity of subscriptions	# days to sell 10% of token supply
	(1)	(2)	(3)
Analyst rating	0.026***	0.024**	-9.516***
	[3.08]	[2.27]	[-5.18]
No. of analysts	-0.001	-0.001	-0.490***
	[-0.62]	[-1.34]	[-5.78]
Fraction of tokens for sale	-0.004	-0.008	7.899
	[-0.17]	[-0.30]	[1.28]
Presale	0.012	0.024**	-8.113***
	[1.02]	[2.51]	[-2.81]
High bonus	-0.017**	-0.013	3.384
	[-2.16]	[-1.55]	[0.84]
Know Your Customer	-0.011	-0.028**	4.485
	[-1.06]	[-2.33]	[1.61]
Multiple languages	-0.008	0.002	-0.124
	[-0.82]	[0.13]	[-0.03]
Multiple currencies	-0.012	0.003	-0.525
	[-1.58]	[0.28]	[-0.71]
Quarterly fixed effects	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes
Observations	544	544	544
Adj. R-squared	0.18	0.13	
Pseudo R-squared			0.14

Table 2.8: Investor Subscriptions and ICO Success
This table replicates Table 2.3 except that column (1) replaces the covariates with *First-day subscription*, while the covariates in column (2) further include the residual from the regression in equation (1). In each column, we report coefficient estimates, their heteroscedasticity-robust t-statistics, and the marginal probability change induced by a one-unit change in the value of a specific covariate from its sample average. Standard errors are clustered at the quarter level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Fundraising success

Dependent variable: ICO success	Probit model					
	Coefficient	t-stat.	Marg. Prob.	Coefficient	t-stat.	Marg. Prob.
ICO characteristics	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
First-day subscription	18.959***	2.89	762.8%			
Residual of first-day subscription				11.306***	5.47	402.7%
Analyst rating				0.813***	6.65	29.0%
No. of analysts				0.053***	5.14	1.9%
Fraction of tokens for sale				-0.588***	-4.42	-20.9%
Presale				1.037***	13.42	37.8%
High bonus				-0.716***	-3.93	-24.6%
Know Your Customer				-0.456**	-2.32	-14.7%
Multiple languages				0.125	1.13	4.5%
Multiple currencies				-0.241	-0.70	-8.3%
Quarterly dummies		Yes			Yes	
Observations		635			540	
Pseudo R-squared		0.22			0.40	
% (Dep variable = 1)		31.2%			32.7%	

Panel B. Gross proceeds

	Dependent variable : Gross proceeds ($ million)			
	Coefficient	t-stat	Coefficient	t-stat
ICO characteristics	(1a)	(1b)	(2a)	(2b)
First-day subscription	17.752***	3.15		
Residual of first-day subscription			15.632***	2.34
Analyst rating			1.812*	1.65
No. of analysts			0.148***	3.74
Fraction of tokens for sale			-15.124***	-3.21
Presale			3.483***	2.09
High bonus			-4.379***	-2.86
Know Your Customer			-3.966***	-4.83
Multiple languages			3.020***	2.19
Multiple currencies			2.623***	3.30
Quarterly fixed effects	Yes		Yes	
Country fixed effects	Yes		Yes	
Observations	247		247	
Adj. R-squared	0.22		0.31	

Table 2.9: Token Returns and Turnover

This table reports statistics on returns and first-day turnover for all listed tokens that were sold through an ICO between Q1 2016 and Q1 2018. The number of observations varies depending on information availability. *First-day return* is measured from the token offer price to the first trading day closing price. *One-month, Three-month, Six-month, and One-year returns* are measured from the first after-market closing price to the 30th, 91th, 182th, and 365th trading closing prices, respectively. One-month, three-month, six-month, and one-year returns all exclude first-day returns. An excess return is calculated as the raw return minus the corresponding compounded daily return on the value-weighted index of Ethereum and Bitcoin. *Gross proceeds* is the amount raised from investors in millions. *Money left on the table* is calculated as the number of tokens issued times the change from the offer price to the first-day closing price. *First-day turnover* is the first-day volume divided by tokens issued. Columns (2), (4), and (6) report the average, median, and standard deviation of the variables of interest, while columns (3) and (5) report the p-value from the t-test and the Wilcoxon signed rank test, which is asymptotically normal, for testing whether the average and median are different from zero, respectively.

Variable	N	Average	p-value from t-test	Median	Wilcoxon p-value	Std. Dev.
	(1)	(2)	(3)	(4)	(5)	(6)
First-day return	432	158.2%	0.00	24.4%	0.00	485.0%
One-month return	430	63.3%	0.00	-24.5%	0.11	394.8%
Three-month return	420	71.6%	0.00	-30.0%	0.40	276.8%
Six-month return	370	143.6%	0.00	-43.3%	0.43	572.2%
One-year return	133	334.9%	0.00	-15.9%	0.02	1228.3%
First-day excess return	432	143.1%	0.00	24.1%	0.00	480.1%
One-month excess return	430	45.7%	0.03	-28.9%	0.00	390.5%
Three-month excess return	420	26.0%	0.04	-34.8%	0.00	257.1%
Six-month excess return	370	66.3%	0.02	-37.4%	0.00	529.1%
One-year excess return	133	21.7%	0.79	-125.3%	0.00	959.6%
Gross proceeds ($ million)	410	21.44	0.00	11.14	0.00	45.04
Money left on the table ($ million)	410	42.20	0.00	1.26	0.00	288.50
First-day turnover	405	6.7%	0.00	1.3%	0.00	27.1%
Days from ICO completion to listing	432	18.46	0.00	14.00	0.00	35.00

Table 2.10: Short-Run and Long-Run Token Returns

This table examines return patterns for all listed tokens that were sold through an ICO between Q1 2016 and Q1 2018. The number of observations varies depending on information availability. *First-day return, One-month, Three-month, Six-month, and One-year returns* are defined as in Table 2.9. *First-day excess return* is calculated as the raw first-day return minus the return on the value-weighted index of Ethereum and Bitcoin. *Market returns* are the value-weighted Ethereum and Bitcoin index return for the same return intervals as the dependent variables. All other independent variables are as defined in Table 2.2, and are measured immediately before the ICO start date. In each column we report coefficient estimates and their bootstrapped t-statistics. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

		Excluding first-day returns			
	First-day return	One-month return	Three-month return	Six-month return	One-year return
	(1)	(2)	(3)	(4)	(5)
First day market return	2.142**				
	[1.86]				
First-day excess return		-0.038*	-0.046**	-0.139***	-0.107
		[-1.93]	[-2.41]	[-3.16]	[-0.18]
One-month market return		1.159***			
		[2.85]			
Three-month market return			0.908***		
			[7.43]		
Six-month market return				2.020***	
				[5.85]	
One-year market return					1.102***
					[3.69]
Analyst rating	-0.002	0.108	0.603*	0.678	2.012**
	[-0.01]	[0.74]	[1.74]	[0.95]	[2.16]
No. of analysts	0.034	-0.011	-0.023	-0.020	2.286
	[0.56]	[-0.95]	[-0.89]	(-0.64]	[0.80]
Fraction of tokens for sale	1.130	-0.372	0.607	1.202	-3.661
	[1.35]	[-0.98]	[0.81]	[1.07]	[-0.38]
Presale	0.218	0.155	0.146	0.792	-1.989
	[0.450]	[0.81]	[0.67]	[1.46]	[-0.75]
High bonus	-1.335**	-0.452**	-0.136	-0.183	-2.577
	[-2.21]	[-2.26]	[-0.38]	[-0.19]	[-1.17]
Know Your Customer	-0.708	-0.490*	-0.761*	-0.320	-0.507
	[-0.77]	[-1.85]	[-1.83]	[-0.46]	[-0.15]
Multiple languages	0.582	-0.277	-0.613*	-0.820	-0.887
	[1.38]	[-1.15]	[-1.76]	[-1.28]	[-0.28]
Multiple currencies	0.323	-0.180	0.089	1.102	-0.855
	[0.91]	[-0.52]	[0.17]	[1.52]	[-0.22]
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	423	410	366	356	131
Adj. R-squared	0.06	0.06	0.25	0.34	0.42

Appendix

Figure A1: An Illustration of ICO Bonus Schedules

Panel A shows the bonus structure for Hash Rush, which ran an ICO from September 20 to October 20, 2017. HOQU's bonus structure is reported in Panel B. During HOQU's token sale from November 27, 2017 to February 26, 2018, bonuses were conditional on sales milestones (in HOQU's tokens).

Panel A. Hash Rush's bonus structure



Panel B. HOQU's bonus structure

Bonus system during the main sale period

Bonus	Number of tokens
20%	13,777,764 HQX
15%	13,777,764 HQX
13%	13,777,764 HQX
11%	13,777,764 HQX
9%	13,777,764 HQX
7%	13,777,764 HQX
5%	13,777,764 HQX
3%	13,777,764 HQX
0%	27,555,528 HQX

Figure A2: An Illustration of the Concavity of Investor Subscriptions
We attempt to measure the concavity of the cumulative token demand curve. Our intuitive concavity metric equals the cumulative token sales on the 15th day minus 1/2the cumulative token sales on the 30th day, with the latter date being the last day of an average ICO. In the benchmark of steady daily fundraising volumes, this concavity measure equals zero (the orange line). However, it is strictly positive for a hot ICO that sell tokens faster in earlier periods (the red curve). The concavity measure is strictly negative for a cold ICO that sell tokens faster in later periods (the green curve).

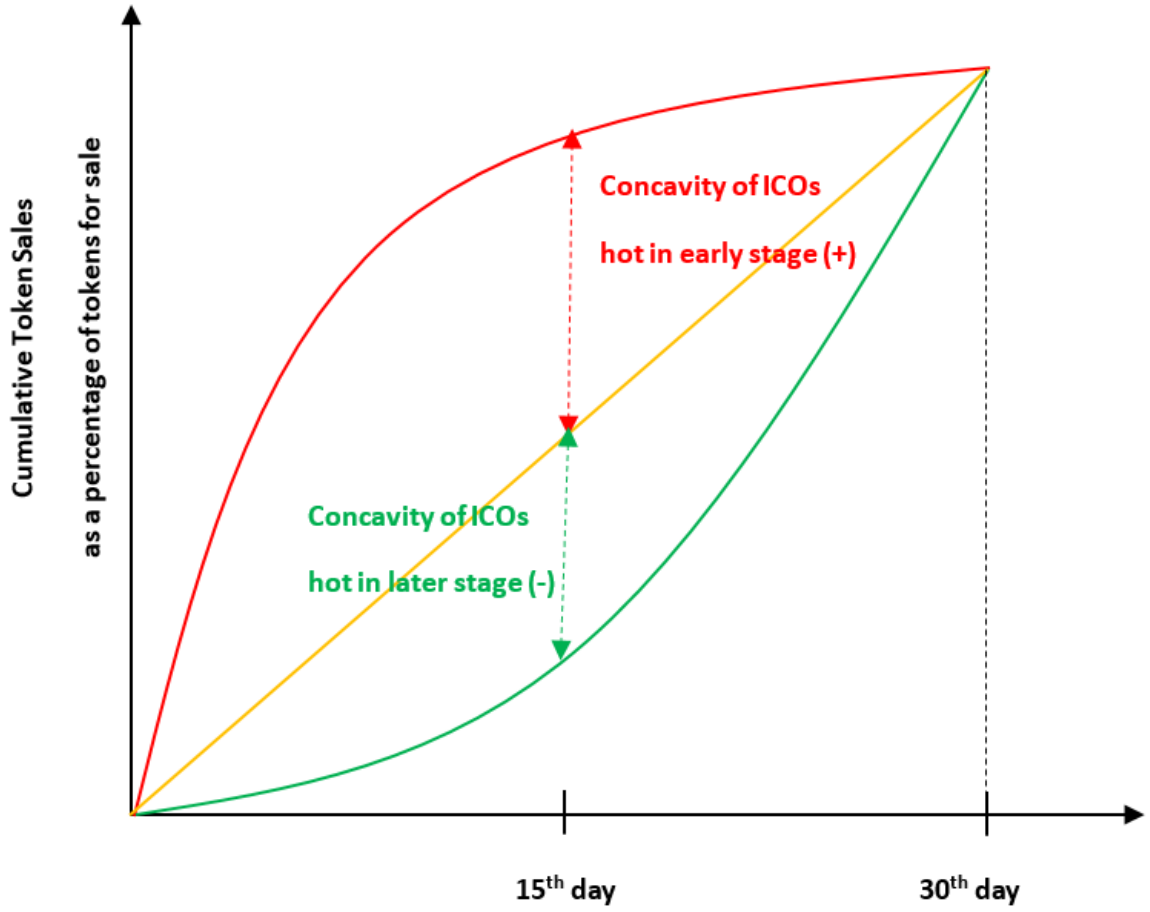


Table A1: Differences between ICOs and IPOs
This table describes the main differences between ICOs and IPOs at each stage of the fundraising process.

	ICO	IPO
1. Pre-announcement	Announce a summary featuring the idea and team to the cryptocurrency community to gather interest and feedback	Hire an investment bank to underwrite the IPO
2. Documentation	- Whitepaper - Website - Initial code	Filings with the securities regulator - Registration statement - Prospectus
3. Marketing	Public relations campaign - Crypto forums - Social network sites such as Medium, Steemit, Reddit, and Twitter	Road show - Meeting with potential investors -Bookbuilding by the underwriter - Offer price set
4. The sale	Subscribers send cryptocurrencies and/or fiat currencies to a digital address. Smart contracts issue tokens based on the exchange ratio.	Shares are allocated to investors
5. Listing	Tokens are listed on a cryptocurrency exchange	Shares are listed on a stock exchange

Table A2: Top 10 ICOs by Gross Proceeds

This table lists the top 10 ICOs by gross proceeds as of May 31, 2018. The ticker for each token is shown in the parenthesis next to the startup name. Information for each ICO is collected from ICObench, and the whitepaper and website for the token sale.

	Startup Name	ICO Start - End	Gross Proceeds ($)
	Dragon Coin (DRG)	2/15/2018 - 3/15/2018	320,000,000
	Huobi (HT)	1/24/2018 - 2/28/2018	300,000,000
	HADC (DAC)	11/27/2017 - 12/22/2017	258,000,000
	Filecoin (FIL)	8/10/2017 - 9/10/2017	257,000,000
	Tezos (XTZ)	7/1/2017 - 7/13/2017	232,000,000
	Sirin Labs (SRN)	12/12/2017 - 12/26/2017	157,886,000
	Styras (STY)	11/28/2017 - 12/30/2017	154,000,000
	The Bancor Protocol (BNT)	6/12/2017 - 6/12/2017	153,000,000
	Bankera (BNK)	11/27/2017 - 2/27/2018	151,800,000
	The DAO (DAO)	4/30/2016 - 5/28/2016	150,000,000

Table A3: Determinants of fundraising success

This table replicates Table 2.3 except that we replace the independent variable, *Know Your Customer*, with *Whitelist*. All independent variables are as defined in Table 2.2. In each column, we report coefficient estimates, their heteroscedasticity-robust t-statistics, and the marginal probability change induced by a one-unit change in the value of a specific covariate from its sample average. Standard errors are clustered at the quarter level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Fundraising success

Dependent variable:					
ICO success	Probit model			Linear probability model	
	Coefficient	t-stat.	Marg. Prob.	Coefficient	t-stat.
ICO characteristics	(1a)	(1b)	(1c)	(2a)	(2b)
Analyst rating	0.564***	4.96	22.2%	0.162***	4.66
No. of analysts	0.035***	5.34	1.4%	0.012***	9.04
Fraction of tokens for sale	-0.734***	-4.12	-28.9%	-0.159***	-3.24
Presale	0.383***	7.74	15.1%	0.118***	9.07
High bonus	-0.275***	-3.44	-10.7%	-0.082***	-3.05
Whitelist	-0.250**	-2.39	-9.6%	-0.091**	-2.21
Multiple languages	0.265**	2.06	10.5%	0.099**	2.49
Multiple currency	0.020	0.39	0.8%	0.014	1.26
Quarterly dummies		Yes		Yes	
Country fixed effects		No		Yes	
Observations		1,461		1,461	
Pseudo R-squared		0.26			
Adj. R-squared				0.32	
% (Dep variable = 1)		45.3%		45.3%	

Panel B. Gross proceeds

Dependent variable	Gross proceeds ($ million)		Gross proceeds/Hard cap	
	Coefficient	t-stat.	Coefficient	t-stat.
ICO characteristics	(1a)	(1b)	(2a)	(2b)
Analyst rating	4.834**	2.09	0.054***	3.35
No. of analysts	0.094	0.69	0.006***	3.03
Fraction of tokens for sale	-17.879***	-2.99	-0.227*	-1.80
Presale	-5.594*	-1.69	-0.019	-0.41
High bonus	-5.002**	-2.13	-0.107**	-2.25
Know Your Customer	-7.515***	-4.14	-0.224***	-3.00
Multiple languages	5.426	1.07	-0.007	-0.54
Multiple currencies	0.565	0.30	-0.021	-1.06
Quarterly dummies	Yes		Yes	
Country fixed effects	Yes		Yes	
Observations	727		543	
Adj. R-squared	0.03		0.13	

Table A4: Initial Investors' Investment Returns

This table replicates Table 2.6 except that the beginning date for both basis-adjusted returns and realized returns is the first day of presale. All variables are defined as in Table 2.6. In each column we report coefficient estimates and their heteroscedasticity-robust t-statistics. Standard errors are clustered at the ICO level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Subscription timing and basis-adjusted returns

	Dependent variable: Basis-adjusted return (%)			
	First day	One month	Three months	Six months
	(1)	(2)	(3)	(4)
# days waited until subscription	-0.435***	-0.356**	-0.378**	-0.530**
	[-2.71]	[-2.32]	[-2.20]	[-2.21]
ICO fixed effects	Yes	Yes	Yes	Yes
Observations	538,953	479,893	395,021	235,506
Adj. R-squared	0.93	0.69	0.58	0.73

Panel B. Subscription timing and realized returns

	Dependent variable: Realized return (%)			
	First day	One month	Three months	Six months
	(1)	(2)	(3)	(4)
# days waited until subscription	-1.081**	-0.537**	-0.848***	-0.840***
	[-2.43]	[-2.75]	[-2.81]	[-3.23]
ICO fixed effects	Yes	Yes	Yes	Yes
Observations	23,333	74,293	126,880	130,423
Adj. R-squared	0.99	0.73	0.60	0.57

Chapter 3

Cryptocurrency Pump-and-Dump Schemes

3.1 Introduction

Initial coin offerings (ICOs) have recently emerged as a popular method of financing blockchain-related startups. In an ICO, a startup creates and distributes its digital "tokens," typically in exchange for Bitcoin, Ethereum, or fiat currencies (e.g., U.S. dollars) to raise capital to fund their operations. A token gives its owner the right to use the firm's products or services once they are developed, and can be traded on the secondary market. ICOs raised nearly $11.4 billion across over 2,000 blockchain startup deals in 2018, exceeding traditional venture capital investments in funding blockchain-related innovative projects (Pozzi, 2019). The startling growth of this market, coupled with

This chapter is based on my paper, "Cryptocurrency Pump-and-Dump Schemes," which is coauthored with Tao Li (University of Florida) and Baolian Wang (University of Florida). This work was presented at Florida Blockchain Conference (2018), Georgia State University/Review of Financial Studies Conference on Current Research in FinTech (2019), 2nd Toronto FinTech Conference (2019)

rampant speculation and volatility, has both generated excitement and raised concern about potential exploitation or fraud.

This paper studies "pump-and-dump" schemes (P&Ds) in the cryptocurrency market. P&D is a form of price manipulation that involves artificially inflating an asset price before selling the cheaply purchased assets at a higher price. Once the assets are "dumped," the price falls and investors lose money. Such schemes are most common with microcap stocks and have recently become popular in the cryptocurrency market (Shifflet and Vigna, 2018). The U.S. Securities and Exchange Commission (SEC) deems P&Ds illegal in the stock market, but the regulation of P&Ds in the cryptocurrency market is weak or nonexistent. Many tokens are difficult to justify either as investment or consumer products, and do not fit neatly into existing securities or consumer-protection laws (Li and Mann, 2018). The regulation of cryptocurrencies also requires more global coordination than other assets, since tokens typically are traded globally.

In the cryptocurrency market, manipulators often organize "pump groups" using encrypted messaging apps such as Telegram. They create Telegram channels and invite other investors to join. They frequently advertise on social media platforms to attract investors. A Telegram channel operator can post messages for other members to read. For a planned pump, the operator announces the target date, time, and exchange, usually at least one day in advance. However, they do not disclose the identity of the target token until the scheduled time. Members also receive multiple reminder messages before the announcement of the token symbol. As we show in this paper, a typical cryptocurrency P&D lasts for only several minutes, leaving little time for non-members to participate. Therefore, it is reasonable to believe that Telegram channel members are dominant participants in P&Ds.

Relative to the stock market, on which most existing studies have focused, our setting provides several advantages for investigating P&Ds. First, in the cryptocurrency market,

a typical P&D episode lasts for only minutes, while such an episode frequently lasts for months in the stock market (Aggarwal and Wu, 2006; Leuz et al., 2017). Many other factors can cloud inferences when pumps last that long. Second, there is typically no false information release or firm actions associated with P&Ds in the cryptocurrency market, reducing the occurrence of information- or action-based manipulation. In fact, most of the Telegram channels have the word "pump" in their aliases and members understand that there is no fundamental news associated with P&Ds (see Table 3.3, Panel A for a list of the 10 largest channels). In the stock market, however, P&Ds are often associated with the release of false information or other actions (Aggarwal and Wu, 2006; Putniņš, 2012; Leuz et al., 2017).

Third, identifying P&Ds is much easier in the cryptocurrency market than in the stock market. The literature has focused on studying stock P&Ds that are ex post investigated by regulators (Aggarwal and Wu, 2006; Leuz et al., 2017) or the "stock pools" of the 1920s, which was perhaps the most famous case of alleged stock manipulation. The SEC anti-manipulation cases may not be representative, though, and existing studies show that "stock pools" engage in informed trading rather than manipulation (e.g., Mahoney, 1999; Jiang et al., 2005). In the cryptocurrency market, manipulators organize pump groups, advertise on social media to attract participants, and disclose their pump plans in real time. New encryption technologies allow them to do so without revealing their identities. This feature enables us to precisely identify a sample of P&Ds, pinpoint their timing, and conduct detailed analyses.

Using 500 hand-collected cryptocurrency P&Ds, we document several stylized facts. First, P&Ds have dramatic short-term impacts on the prices and volumes of most of the pumped tokens. In the first 70 seconds after the start of a P&D, the price increases by 25% on average, trading volume increases 148 times, and the average 10-second absolute return reaches 15%. A quick reversal begins 70 seconds after the start of the P&D. After

an hour, most of the initial effects disappear. The above findings hold for both liquid and illiquid tokens, although they are stronger for relatively illiquid tokens.

We also document that prices of pumped tokens begin rising five minutes before a P&D starts. The price run-up is around 5%, together with an abnormally high volume. These results are not surprising, as pump group organizers can buy the pumped tokens in advance. When we read related messages posted on social media, we find that some pump group organizers offer premium memberships to allow some investors to receive pump signals before others do. The investors who buy in advance realize great returns. We deem investors who know the identity of pumped tokens in advance as insiders and the rest as outsiders. Calculations suggest that insiders' average return can be as high as 18%, even after considering the time it may take to unwind positions. This far exceeds the round-trip trading costs that are between 20 and 50 basis points. For an average P&D, insiders make one Bitcoin (about $8,000) in profit, approximately one-third of a token's normal-period daily trading volume. The trading volume during the 10 minutes before a pump is 13% of the total volume during the 10 minutes after the pump starts. In the majority of cases, insiders are able to unwind their positions within 10 minutes. This implies that an average trade in the first 10 minutes after a pump has a 13% chance of trading against these insiders and on average the outsiders lose more than 2% (18%×13%).

The quick reversal means that an outside investor who is unaware of the P&D timing in advance needs to buy and sell very quickly to make a profit. Equally importantly, liquidity may disappear when the investor wants to sell. We conduct a performance analysis that accounts for real-time liquidity. Our analyses show that, on average, only investors who buy in the first 20 seconds after a P&D begins can make a profit, and they can do so only if they do not hold their tokens for very long. For example, if an investor buys tokens between 10 and 20 seconds after the P&D starts and begins selling

one minute later, he will lose 0.72% of his investment on average. An investor who buys target tokens one minute after a P&D starts will lose more than 1% even if he sells immediately after the purchase.

The inferred performance may overestimate the real performance one can achieve. First, in the analysis we assume that when investors decide to sell they can consume all the available liquidity in the market, but that is unlikely. Second, both the token price and the volume increase before a P&D starts, suggesting trading by pump insiders. If they start unwinding their positions immediately after the P&D starts, they will consume all the liquidity between the announcement and 140 seconds later, leaving scarcely any profit available even to the fastest outsiders. Third, we find that some outsiders receive pump signals systematically later than other outsiders do. All these findings make it puzzling why outsiders, especially those who receive signals systematically later, are willing to participate.

While the cryptocurrency market enables us to conduct detailed analyses of P&Ds, our data do not enable us to pin down the exact mechanism that explains why outsiders are willing to participate.[1] The price and volume patterns around P&Ds are analogous to asset bubbles that have been studied in the literature; however, these cryptocurrency bubbles develop within much shorter time intervals.[2] We conjecture that one plausible mechanism for this is that P&Ds attract overconfident investors who believe that they can time the market more accurately than others can (Scheinkman and Xiong, 2003;

[1]Tokens can also be traded off the exchanges. These trades are broadcast on a blockchain network and are viewable by the public. Blockchain trading data can potentially enable us to get investor identifier, as trading on blockchain can be linked to unique wallets. Research (2017) reports that 77% of cryptocurrencies issued before 2018 use the Ethereum platform and Mironov (2018) reports that this ratio increased to 95% in the first half of 2018. We therefore collect token trading data from Etherscan.io, a leading "block explorer" that allows users to view information about blocks and transactions on the Ethereum Blockchain. However, we find that virtually all P&D-related token trading occurs on the exchanges, likely due to lower trading costs and faster trading speed on exchanges.

[2]Asset bubbles have fascinated economists for centuries (Mackay, 1984; Tirole, 1985; Shiller, 2000; Scheinkman and Xiong, 2003; Fama, 2014). Many studies have documented abnormal asset price and volume behavior in bubble episodes (Hong and Stein, 2007; Xiong and Yu, 2011; Barberis et al., 2018; Greenwood et al., 2019).

Daniel and Hirshleifer, 2015). Another possible mechanism is that these investors have gambling preferences. The short-term returns on pumped tokens are very high and salient. Investors may overweight the possibility of these returns in their decision-making (Barberis and Huang, 2008; Barberis, 2013; Bordalo et al., 2012; Bordalo et al., 2013).

In trade-based price manipulation theories, where a manipulator cannot act in ways other than buying and selling assets, the manipulator needs to buy the target asset to "pump" its price first. As outside investors follow, the manipulator can sell at a higher price to make profits. Manipulators can make profits only if the price impact is greater when they buy than when they sell; otherwise this strategy is self-defeating (Friedman, 1953). Trade-based manipulation can be profitable if there is price momentum (Jarrow, 1992) or uninformed investors believe it is probable that the manipulator is informed (Allen and Gale, 1992). The price impact is abnormally high before a P&D starts, which is consistent with informed insider purchases. Interestingly, at the beginning of a P&D, the large price increase is not associated with an abnormally strong price impact, consistent with the fact that the P&D announcement leads to increased participation of Telegram channel members who are typically uninformed traders. Their purchases likely cause a large price increase. These findings contrast with the existing trade-based manipulation theories that typically assume that manipulators have to buy to increase the asset price.

P&Ds, or other forms of price manipulation, are generally considered undesirable and should be regulated. Besides illegitimate wealth transfers, are there any other detrimental consequences of P&Ds? We investigate this question using a natural experiment. On November 24, 2017, Bittrex, a U.S.-based cryptocurrency exchange, announced that it would ban P&Ds. Although it was not able to eliminate P&Ds altogether, it reduced the frequency of P&Ds on its exchange. This ban affected only the tokens traded on

Bittrex, not tokens on other exchanges. We employ a difference-in-differences approach and find that the ban causally increased token prices and liquidity.

Our paper contributes to the price manipulation literature. Price manipulation has been documented in the stock market (Aggarwal and Wu, 2006; Ben-David et al., 2013; Ni et al., 2005; Comerton-Forde and Putniņš, 2015), derivative markets including the equity and index options markets (Griffin and Shams, 2017), futures markets (Merrick Jr et al., 2005), and the LIBOR market (Mollenkamp, 2008; Snider and Youle, 2010). These assets are manipulated often because their prices are used for contract settlements. If manipulators have large positions in derivative contracts, they can make profits even if they lose in the underlying market. Manipulation has also been documented in emerging stock markets, such as in Pakistan and China (Khwaja and Mian, 2005; Chen et al., 2019). Gandal et al. (2018) document evidence of Bitcoin price manipulation in 2013 and Griffin and Shams (2018) document suspicious Bitcoin activities in 2017. Diverging from all these papers, we study cryptocurrency P&Ds. Special features of this market enable us to conduct detailed and clean empirical investigations.

Allen and Gale (1992) classify manipulation into information-, action-, or trade-based manipulation. The first type relies on spreading false information (Enron), the second on non-trade actions that can affect stock prices (such as a purported takeover bid), and the third on direct manipulation of stock prices by trading. Cryptocurrency P&Ds are not associated with company information releases or other actions, and therefore cannot be explained by information- or action-based manipulation theory.

There is also a theoretical debate over the profitability of trade-based manipulation. Friedman, 1953 concludes that arbitrage would make such a strategy self-defeating. More recent studies argue that such a strategy can be profitable, as it is when investors cannot distinguish between informed trading and a manipulator and believe it is likely that the manipulator is informed (Allen and Gale, 1992), when price momentum enables

a manipulator to establish a price trend and then to profit from trading against it, or when investors exhibit the disposition effect and are reluctant to sell plummeting assets (Mei et al., 2004).[3] The price impact behavior associated with cryptocurrency P&Ds is inconsistent with the implications of the existing trade-based manipulation theories. Instead, our results suggest that overconfident or gambling investors can "coordinate" to create short-term asset bubbles, followed by quick reversals, implying the presence of a new type of manipulation besides that considered by Allen and Gale (1992).

Our findings also contribute to the burgeoning literature on the application of blockchain technology in finance. These studies have examined how blockchain technology may reshape transaction costs and economic exchange (Biais et al., 2018; Ciaian et al., 2016; Malinova and Park, 2017; Cong and He, 2018; Yermack, 2017); ICOs (Catalini and Gans, 2018b; Chen et al., 2018; Cong et al., 2018; Cong and Xiao, 2018; Howell et al., 2018; Lee et al., 2018; Li and Mann, 2018; Sockin and Xiong, 2018); the ecosystem of Bitcoin (Böhme et al., 2015), including mining and transaction fees (Cong et al., 2019; Easley et al., 2019; Huberman et al., 2017); and Bitcoin adoption and usage, including how Bitcoin facilitates illegal activity (Athey et al., 2016; Foley et al., 2018). Hu et al. (2018) document stylized facts pertaining to cryptocurrency prices. Liu and Tsyvinski (2018) study how cryptocurrency prices are related to other traditional financial assets. Makarov and Schoar (2019) study cryptocurrency mispricing and arbitrages across exchanges. Of greatest relevance to our study, Gandal et al. (2018) and Griffin and Shams (2018) provide evidence consistent with Bitcoin price manipulation. They do not, however, study P&Ds.[4]

[3]Jarrow (1992) and Huberman and Stanzl (2004) derive conditions for the nonexistence of price manipulation, and both studies show that market manipulation strategies can exist under reasonable conditions.

[4]Two concurrent papers also study cryptocurrency P&Ds. Xu and Livshits (2018) study how to predict P&D events. Feder et al. (2018) examine which factors affect the returns of pumped tokens. Their main finding is that a token's market capitalization is the most important factor, consistent with our finding that P&Ds increase illiquid tokens' prices more than those of the liquid ones. Our token trading data are trade-by-trade from the target exchanges, while Feder et al. (2018) use data at five-minute intervals and Xu and Livshits (2018) use data at one-hour intervals. Both of their data sets

Our study also informs the recent debate over cryptocurrency regulation. While there are many aspects of cryptocurrency regulation, our findings suggest that P&Ds should be regulated. First, the existence of price run-ups before P&Ds implies that an average outsider loses money by trading in P&Ds. Standard market forces suggest that outsiders would be unlikely to participate in P&Ds. Our evidence suggests that this market mechanism is not working. Second, our analysis based on the Bittrex ban provides causal evidence that P&Ds are detrimental to the health of the cryptocurrency market. Third, the fact that Bittrex was able to sharply reduce P&Ds on its exchange suggests that banning P&Ds is technologically feasible. An exchange may be able to distinguish itself by preventing manipulation.

3.2 Institutional Background

In this section, we briefly discuss what cryptocurrencies are, how they are traded, and how cryptocurrency P&D schemes are operated. We conclude with a discussion of the current regulatory environment.

3.2.1 Cryptocurrencies

Cryptocurrencies are digital tokens typically issued through ICOs. In an ICO, a technology startup creates and distributes its token in exchange for Bitcoin, Ethereum, or fiat currencies (e.g., U.S. dollars) to raise capital to fund their operations. A token typically provides a specific set of rights to its holders, including access to a platform or network, the right to create or develop features for an ecosystem, and the right to vote, among others (Lee et al., 2018; Sockin and Xiong, 2018).

are aggregated across *all* exchanges. The granularity of our data enables us to pin down the effects of P&Ds more accurately.

Most secondary market trading in cryptocurrencies occurs on online exchanges. Tokens are typically listed on one or more exchanges. Exchanges operate 24 hours a day, 365 days a year. On most cryptocurrency exchanges, traders can place market or limit orders. Limit orders provide liquidity and are rewarded by the exchanges with lower trading commissions than market orders. However, Binance, Bittrex, and Yobit, the three exchanges we study, charge flat fees for both market and limit orders of 0.1%, 0.2% and 0.25%, respectively.[5]

Token prices are quoted in ticker pairs that function in a way that is similar to exchange rates. For example, NXS/BTC is a pair that indicates that the trade is an exchange between NXS (Nexus) and BTC (Bitcoin). In this pair, Nexus is quoted in the number of Bitcoins, which is the base currency. Besides Bitcoin, other cryptocurrencies such as Ethereum, Tether, Binance Coin, and Litecoin are also commonly used as numeraires. Bitcoin is by far the most widely used base currency and all the P&Ds in our sample target trading pairs involving Bitcoin. In seven P&Ds, both Bitcoin pairs and Ethereum pairs are targeted. We therefore focus on Bitcoin pairs in our main analysis. In additional analysis, we also examine Ethereum pairs.

3.2.2 Pump-and-Dump Schemes

In the cryptocurrency market, operators of P&Ds often organize "pump groups" using encrypted messaging apps such as Telegram. Multiple groups also coordinate to pump cryptocurrencies together. Telegram is a cloud-based instant messaging service developed by Telegram Messenger LLP. Telegram uses end-to-end encryption, which

[5]In addition to trading on online exchanges, investors can also trade cryptocurrencies over the counter (OTC) or on decentralized exchanges. In the OTC market, traders typically discuss trade terms via e-mail or messaging apps such as Skype. After a trade is agreed upon, fiat money is sent to a third party who handles the settlement. On a decentralized exchange, trades occur directly between users through an automated process, without a third-party service that holds the customers' funds. Off-exchange trading is subject to the same blockchain-based mining process, which can take several minutes to several days to confirm a transaction.

means that nobody but the sender and recipient can read the messages. Chat messages cannot be forwarded to others outside of the conversation, and no trace is left on the servers. It is considered as one of the most heavily-encrypted messaging platforms available today. Telegram users can also use aliases. Both of these features make it difficult to track who is involved in a particular P&D scheme.

Another unique feature of Telegram is that it allows operators to create Telegram channels to broadcast messages.[6] Members can join a channel free of charge and access the entire message history. Each message, timestamped to the second, has its own view counter that shows how many users have viewed it. To attract participants, P&D operators often advertise their pump groups on social media platforms such as Reddit and BitcoinTalk, and they often urge their existing participants to do the same. Some operators offer monetary incentives to members who invite new members to participate.

Figure 3.1 illustrates how a typical P&D works. On the morning of July 4, 2018, Big Pump Signal, one of the biggest pump groups in our sample, announced that they would pump a token at 7 p.m. Coordinated Universal Time or UTC (3 p.m. EDT) on Binance. One-and-a-half hours before the pump, the operator of the Telegram channel reminded its 70,000 followers of the event, and encouraged them to reach out to other investors: "Bigpumpsignal has always been focused on reaching outsiders. Because, a good pump is decided by the amount of interest of the outsiders ... You are free to help us with spreading news about the coin we pump on social media!" About 30 minutes before the pump, the operator urged members to get ready, and they were reminded again 20, 10, five, and two minutes before the pump. Although a target return was not set in this P&D, such a target is often announced in other pump events.

[Insert Figure 3.1 here.]

[6]Users can send private messages to the operator, but are not able to leave feedback in the public domain.

At 6:59:57 p.m. UTC, three seconds before the scheduled time, the ticker of the target token, NXS (Nexus), was announced. Telegram's record shows that 28,800 followers viewed the message. The price of Nexus jumped 52.0% immediately, to 0.0003045 Bitcoin ($2.01), before plummeting 24.5% from the peak after 10 minutes. During this time, 4,153 trades worth 222.68 Bitcoin ($1.5 million) were executed, compared with virtually no trading during the hour before the announcement.

Not all group members are treated equally by the operators. Tiered access to pump signals appears to be common with many groups. High-ranking members may be sent a signal several seconds earlier than others receive it. Members can pay a fee to become VIP members. The operators may also incentivize members to invite new members to join by offering premium memberships. For example, Mega Pump Group, a pump group with more than 15,000 followers on Telegram, sends pump signals 0.5 (3.5) seconds earlier to members who successfully invited at least four (50) members to join.

3.2.3 The current regulatory environment

P&Ds in stocks are illegal in many countries and considered fraudulent by U.S. regulators. Stock exchanges that do not take adequate measures to prevent P&D schemes can also face legal penalties. The SEC regularly targets P&D scams (Mei et al., 2004). Section 9 of the Securities Exchange Act of 1934 specifically makes it unlawful to manipulate security prices. P&D groups active on cryptocurrency exchanges, however, have been operating with relative impunity because cryptocurrencies are not necessarily considered securities and the exchanges currently are unregulated markets.

In December 2017, the SEC issued a statement on cryptocurrencies and ICOs, warning investors to beware of scams and criminal activities in the sector. The statement explains that "excessive touting in thinly traded and volatile markets can be an indicator of 'scalping,' 'pump and dump' and other manipulations and frauds." On February 15,

2018, the U.S. Commodity Futures Trading Commission (CFTC) published a customer advisory, warning investors to avoid cryptocurrency P&D schemes. It offers eligible whistleblowers between 10% and 30% of the value of enforcement actions that involve $1 million or more in P&D investing.

Despite the regulatory scrutiny, P&D groups in the cryptocurrency market have not ceased operating. One potential reason for this is that most cryptocurrencies are traded on multiple exchanges globally, which could lead to regulatory arbitrage. Encrypted messaging apps such as Telegram can also help P&D organizers hide their identities.

3.3 Data

We obtain a list of P&D events, cryptocurrency trading data, and additional token-level variables from various data sources. We discuss them one by one. Separate data sources may have distinct token identifiers and we manually match them by token symbol and name.[7] We discuss potential problems with our data at the end of this section.

3.3.1 P&D events

There exists no central database of P&D events. We therefore manually construct a sample. We first collect a list of pump groups from Reddit and BitcoinTalk, two popular cryptocurrency message boards. Pump groups typically advertise on these boards to attract participants. For completeness, we also conduct an Internet search for additional advertisements posted by such groups. We initially identify 210 Telegram groups that engage in P&Ds, 129 of which were still accessible as of September 2018. Forty-nine

[7]We discuss the details of matching tokens across multiple data sets in Part A1 of the Appendix.

of the 129 groups did not initiate any pumps.[8] All these Telegram groups are publicly accessible.

We read all messages posted in these Telegram channels. For each P&D event, we collect the target token (name and symbol), target exchange, announcement time, initially scheduled time (timestamped to the second), number of viewers of the announcement, and target price or return if available. All timestamps in this paper are based on the UTC scale. In cases where a token is targeted by multiple channels at the same scheduled time, we use the earliest announcement time for our analysis. As we document later in the paper, the sharp changes in token price and trading volume occur exactly at the announcement, providing justification that our data on P&D announcement times are accurate.

We identify 3,412 P&D announcements from 80 active Telegram channels in the period between May 15, 2017 and August 26, 2018. Many P&Ds are coordinated by more than one channel. The total number of unique P&Ds is 1,747. These P&Ds target tokens traded on Binance, Bittrex, Yobit, Cryptopia, HitBTC, Poloniex, and CoinExchange. We restrict our attention to the 1,040 P&Ds targeting tokens on Binance, Bittrex, and Yobit. This choice is motivated by the small number of P&Ds on HitBTC, Poloniex, and CoinExchange, and the unavailability of trading data from Cryptopia.[9] Binance, Bittrex, and Yobit are among the largest cryptocurrency exchanges in the world, and are

[8]We also identify a small number of Discord pump groups. We cannot access them, however, because an invitation is needed to join a Discord group. Discord groups are much smaller and less popular than Telegram groups (with up to 100,000 members) as Discord imposes a limit cap of 5,000 concurrent online users for each group. Groups reaching 5,000 simultaneous online members need to apply for hardware supporting larger servers. At that moment, members start getting "Server Unavailable" errors. In fact, the Big Pump Signal group switched to Telegram on December 28, 2017 due to "reaching maximum amount of users online on discord!"

[9]In our sample, only four announcements target HitBTC, while the number of announcements for Poloniex and CoinExchange are 54 and 44, respectively. We collected information on 690 P&Ds on Cryptipia. However, Cryptopia does not allow any third party to download tick-level trading data that are older than one week. To the best of our knowledge, no data provider collects its trading data either, potentially due to Cryptopia's small size. Cryptopia is currently ranked 77th among all cryptocurrencies, according to CoinMarketCap.

ranked first, 36th, and 27th by trading volume, respectively, on CoinMarketCap.com, a popular data provider.

The large number of P&D events suggests that P&Ds are pervasive in the cryptocurrency market. Binance is located in Asia, Bittrex in the U.S., and Yobit in Russia. Therefore, our sample covers major regions around the globe. Not all pumped cryptocurrencies are covered by our trading data. After merging with trading data, our sample shrinks to 507 P&Ds. We further require that a target token be traded at least once in the period from 37 days before an announcement to eight days before the announcement. Our final sample includes 500 distinct P&Ds involving 239 unique tokens.

Figure 3.2, Panel A displays the distribution of our sample P&Ds by scheduled hour (UTC). We report the distribution for each of the three exchanges. It is evident that although there is some difference across the three exchanges, the majority of pumps are scheduled from 12:00 to 22:00 UTC (daytime in the U.S.).

[Insert Figure 3.2 here.]

3.3.2 Exchange trading data

We obtain trade-by-trade data on tokens listed on Binance, Bittrex, and Yobit from two sources. For Binance tokens, we download data using its public application programing interface (API), a web interface for data retrieval. Binance's API enables us to retrieve the entire history of token trading data. However, APIs provided by Bittrex and Yobit allow access only to the past 24 hours' worth of data. We thus purchase these trading data from Kaiko Data, a data provider that has been collecting token trading data since 2014. Kaiko obtains the data by querying the APIs on a daily basis.

The data structure is similar to trade files of the NYSE Trade and Quote (TAQ) database. For each transaction, we have the ticker symbol pair (e.g., NXS/BTC, in which BTC is the base currency), execution price, quantity, UTC-based timestamp, and

an indicator that specifies whether the trade is buyer- or seller-initiated. Therefore, we do not need to infer trading direction based on any algorithm such as the one proposed by Lee and Ready (1991). Trades are time-stamped to the millisecond on Binance and Bittrex, and to the second on Yobit. All three exchanges cover both active coins and delisted coins and therefore are not subject to survivorship bias.

Figure 3.2, Panel B displays the distribution of trading volume by UTC hour. Although the three exchanges are located on three continents (again, Binance in Asia, Bittrex in the U.S., and Yobit in Russia), the trading volume distributions by hour look remarkably similar. First, trading is active at almost all hours, reflecting the global nature of token trading. Second, trading volume is higher between 12:00 and 20:00, which is similar to the distribution of P&Ds shown in Figure 3.2, Panel A.

3.3.3 CoinMarketCap data

Our exchange trading data may not cover all the trading activities of a token if it is cross-listed on exchanges other than Binance, Bittrex, and Yobit. We therefore also obtain data from CoinMarketCap.com. CoinMarketCap is widely considered the top source for trading information on cryptocurrencies. CoinMarketCap aggregates pricing and volume data for nearly 2,000 tokens from hundreds of exchanges. The data CoinMarketCap provides are not trade-by-trade records but aggregated at the daily level. For each day, CoinMarketCap reports price (volume-weighted across exchanges), total volume, and market capitalization. CoinMarketCap does not cover all tokens, but tends to feature larger and more liquid ones.

3.3.4 Other data

We also collect token-level data for all cryptocurrencies traded on Binance, Bittrex, and Yobit. The initial listing date for each token is obtained from CoinMarketCap.

In cases where CoinMarketCap does not feature a token or its initial listing date is more recent than the initial trading date on Binance, Bittrex, or Yobit, we use the latter. CryptoCompare.com compiles a social media activity index for cryptocurrencies, which aggregates the number of users on Reddit, Twitter, Facebook, and its own site. We manually search each token name on CryptoCompare and obtain its social media activity index.

3.3.5 Summary statistics

In Table 3.1, Panel A we report the characteristics of our P&D events. In the full sample, on average, 1.6 Telegram channels coordinate one P&D, suggesting that coordination is popular. The total number of viewers across all channels in a pump is 5,942, with Binance pumps attracting the most viewers, and Yobit pumps the least. Nearly 90% of the pumps are scheduled at the usual time, which is defined as the most frequently scheduled time for a given group. On average, P&Ds are announced 24.5 seconds (the median is 5.2 seconds) after the scheduled time. The earliest channel receives the signal about 3 seconds earlier than the average. In 229 P&Ds, pump groups specify the returns they target. If a target return is specified as a range running, for example, from 200% to 300%, or it differs across channels, we use the lower bound. The target returns are 212% on average. Target returns for Yobit tokens (233%) are much higher than those for Binance tokens (69%), likely because Yobit tokens are less liquid.

[Insert Table 3.1 here.]

In Table 3.1, Panel B we report summary statistics on the target tokens as well as on the non-target tokens listed on Binance, Bittrex, and Yobit. For each token characteristic on each day on which at least one pump occurs, we calculate the average value for target and non-target tokens. We then compute the time-series average. The differences between target and non-target tokens and their associated t-values are calculated using

these time-series averages. The token characteristics we report are log trading volume (in Bitcoin), log market capitalization (in dollars), return volatility, coin age, social media index, and the frequency of being covered by CoinMarketCap. Trading volume and volatility are calculated over the period running from day -37 to day -8. Market capitalization is available only if a token is covered by CoinMarketCap. We also report the average percentile of pumped tokens among all tokens.

On average, 2.02 tokens are pumped per day in our sample, and 684 are not. The likelihood that a token will be pumped on a given day is 0.3%, which is equivalent to more than 107% a year. The characteristics of non-target tokens inform us of the average token's characteristics, as the majority of tokens are not pumped on any given day. The results presented in Table 3.1, Panel B show that the average log trading volume of non-target tokens over the previous 30 days is 3.345 (28.36 Bitcoin or around $0.23 million) and the average daily return standard deviation is 14.1%, suggesting that the average token is illiquid and its price is highly volatile. CoinMarketCap covers 52.4% of tokens on the three exchanges. In unreported analysis, we find that covered tokens generally exhibit much higher trading volumes than uncovered ones, suggesting that CoinMarketCap is more likely to cover larger tokens. For covered tokens, the average log market capitalization is 16.375 ($12.70 million). All these results show that the average token is small, illiquid, and volatile.

In Table 3.1, Panel B we also show that target tokens exhibit higher trading volumes, lower volatility, and higher social media indexes than non-target tokens, and they are older and more likely to be covered by CoinMarketCap. Conditional on being covered by CoinMarketCap, their market capitalization is similar to that of other covered tokens. The average market capitalization of target tokens is comparable to that of the 10th

percentile of U.S. common stocks in the 10th percentile, and their volatility is comparable to that of stocks in the 99th percentile (Hou and Loh, 2016).[10]

3.3.6 Potential problems with the data

There are a few potential sources of selection bias in our sample. The first is that 81 of our 210 initially identified Telegram pump channels are closed. Telegram closes channels to save disk space. When closing a channel, Telegram will delete all messages, contacts and data stored in the Telegram cloud. This explains why in those circumstances we can no longer access the messages. The most prominent cause of channel closures is that operators do not log onto their channels within a six-month period. Many channels in our sample had been inactive for months (but for fewer than six months) but were not closed. We assess the importance of this potential bias in Part A2 of the Appendix and find that P&Ds that occurred within the past six months before we started this study, which are free of such bias, perform similarly to other P&Ds.

The second potential problem arises from the fact that the trading data we collected from Bittrex and Yobit do not cover all the listed tokens. This is because our data provider occasionally had technical issues in querying the two exchanges for some token-days. We believe this is unlikely to cause any bias in our results. Nevertheless, to mitigate any such concern, we replicate our main results by focusing only on P&Ds that target Binance. We download trading data from Binance directly and are able to collect

[10]In addition to using CryptoCompare's social media index, we also create an indicator variable equal to 1 if a cryptocurrency is rated by analysts from ICObench.com and ICORating.com, two popular websites that provide investment ratings of cryptocurrencies before they were listed. ICObench uses a rating scale that ranges from 1 to 5, with 5 being the highest rating. Their assessment algorithm evaluates four separate aspects: team, token sale information, product representation, and marketing and social media. ICORating provides three ratings—a hype score, a risk score, and an overall investment rating. The hype score indicates the level of interest on the part of potential investors, the risk score is aimed at assessing the risk of potentially fraudulent activities, and the investment rating "demonstrate[s] maximum openness to potential investors." See Lee et al. (2018) for more information. We find that pumped tokens are more likely to be rated by these two websites.

data on all of their tokens. In Part A3 of the Appendix, we report similar results after analyzing targeted Binance tokens only.

Third, our sample Telegram channels may over-represent the channels that actively advertise on Reddit and BitcoinTalk or are more successful. After reading many social media posts, however, we conclude that most Telegram channels are active on Reddit and BitcoinTalk and our list of Telegram pump groups is comprehensive. Of the 129 pump groups that were still open as of September 2018, 49 did not initiate any pumps while 37 initiated ten or fewer P&Ds. This suggests that our sample covers a wide range of pump groups, including unsuccessful ones.

3.4 Empirical Results Pertaining to P&Ds

In this section, we first report the effects of P&Ds on token prices and trading patterns, using the standard event study method. We report all the results without adjusting for the effect of the market. Adjusting for the market effect has little influence on our results. At the end of the section, we discuss why P&Ds are puzzling and offer several potential interpretations.

3.4.1 Can P&Ds move token prices?

Figure 3.3 displays the distribution of maximum returns (Panel A) and time to maximum returns (Panel B) across P&D events. A P&D's maximum return is defined as the ratio of the highest price achieved within 10 minutes after the pump announcement to the price ten minutes before the announcement minus 1. The time to a maximum return is defined as the number of seconds it takes from the announcement to reach the maximum price. There are a few extreme returns. For ease of illustration, we cap maximum returns at 500% for the results displayed in Figure 3.3, Panel A. In all other analyses, returns are not capped. A maximum return can be negative if the

maximum price of a pumped token during the ten minutes after the announcement is lower than its price ten minutes before the announcement. This can happen if the scheme is unsuccessful and the price drops.

[Insert Figure 3.3 here.]

In Figure 3.3, Panel A we show clearly that in the majority of P&Ds, token prices increase. On average, the mean (median) maximum return is 68.94% (24.39%). For the first and third quartiles the returns are 9.27% and 70.73%, respectively. There are only seven P&Ds in which the maximum return is negative, suggesting that these P&Ds are unable to pump the token prices. The token price reductions in these cases are much smaller relative to price increases in successful P&Ds. On average, it takes only 2.6 minutes for prices to reach the maximum level. This short window indicates that one needs to be fast to make profits from P&Ds. Later in this section, we conduct more detailed analyses on trading performance by time of purchase.

In Figure 3.3, Panel A we report cumulative returns, abnormal volume, and the volatility of the tokens targeted by P&D groups for each 10-second interval. The X-axis indicates the time from 600 seconds before to 600 seconds after the announcement of a P&D. Time 0 indicates the 10-second interval running from 0 to 10 seconds where 0 is the announcement time. The solid lines show the means across all P&Ds, and the dashed lines indicate 95% confidence intervals. Cumulative returns are calculated as the log change in price from 600 seconds before the announcement of P&Ds. We use log returns to mitigate the effects of extreme returns. Abnormal volume is calculated as $\log(1 + 10\text{-second volume/average 10-second volume over day } t - 37 \text{ to day } t - 8)$. Volume is measured by the number of tokens traded. Volatility is measured as the absolute value of returns in each 10-second interval. We measure volatility as the absolute value of returns rather than squared returns to minimize the effects of extremely large returns, as we observe in Figure 3.3.

[Insert Figure 3.3 here.]

It is evident from the results displayed in Figure 3.3, Panel A that P&Ds move token prices and generate significant abnormal trading volumes. It is interesting that the changes start before announcements, suggesting potential leakages. Such leakages are likely driven by VIP members of pump groups who tend to receive signals in advance, or by pump group operators who may trade directly, or both.

We use Figure 3.3, Panel A to infer investor performance. An "insider," who knows the token identity and timing of the pump in advance and buys target tokens ten minutes before the announcement and holds until 70 seconds after the announcement, achieves a nearly 25% return. For an "outsider" who does not know about a P&D in advance but buys immediately after the announcement and sells 70 seconds later, his return of 15% is still eye-popping. These are the maximum returns an insider or an outsider can potentially achieve. In these analyses, we assume that the insider and the outsider are small investors and their trades do not affect the price. In reality, this assumption is unlikely to hold. In later analyses (Table 3.2), we examine this assumption in greater detail.

The second and third graphs on Figure 3.3, Panel A display the abnormal volumes and volatility, respectively. At the maximum, the change in the 10-second volume is around 148 times (the exponential of 5) higher than the average 10-second interval in the period running from 37 days to eight days before the announcement of a P&D. The volatility (measured as the absolute value of 10-second returns) is 15% in the first 10-second interval after the announcement.

The results for returns, trading volume, and volatility all demonstrate that P&Ds have a significant short-term impact on token prices and trading. Ten minutes after an announcement, the price, trading volume, and volatility are still significantly higher

than their pre-announcement levels, although the reversal for volume and volatility is quicker than that for price.

In Figure 3.3, Panel B we present the patterns of returns, trading volume, and volatility over a longer period—running from seven days before an announcement to seven days after. Outside of the short period around a P&D, trading decreases sharply, preventing us from calculating these variables accurately based on 10-second data. Therefore, for this longer-term analysis we calculate price, volume, and volatility over one-hour intervals. We find that the effect of P&Ds on volatility disappears in less than a day, and the effects on price and volume last for two to three days. However, as trading volume quickly goes back to the normal level after minutes, there is hardly any liquidity for trading against the continual reversal.

In Figure 3.5, we show the results of replicating the above analysis using two groups of P&Ds sorted by the liquidity of pumped tokens. Liquidity is defined as the total trading volume (in Bitcoin) over the period running from 37 days to eight days before the pump. The results show that the effects of P&D on price and volume are much stronger for illiquid tokens than for liquid ones.[11]

[Insert Figure 3.5 here.]

3.4.2 Spillovers to Ethereum pairs and other exchanges

In previous reported analyses, we focus only on Bitcoin pairs traded on the target exchanges. We focus on Bitcoin pairs because P&Ds specifically target these pairs. In this section, we investigate the effects of P&D spillovers to Ethereum (ETH) pairs, which are the second most commonly traded pairs. We also investigate whether there is any spillover effect on the same tokens traded on other exchanges.

[11] An alternative measure of token liquidity is market capitalization. We do not use market capitalization because it is available for only approximately half of our tokens (see Table 3.1, Panel B). In untabulated results, we find that the correlation between log trading volume and log market capitalization is 0.74.

We expect a spillover effect on the ETH pair because the limits to arbitrage between a BTC pair and an ETH pair are low. Investors can easily buy BTC pairs and sell them as ETH pairs and vice versa. BTC/ETH pairs are listed on all three of the exchanges we study and are very liquid, facilitating exchange between these two base currencies. We also expect that the spillover to other exchanges would be low in volume, due to stronger limits to arbitrage. Trades happen instantly within an exchange, but once a cryptocurrency must be transferred to another exchange the settling of that transaction takes much longer, partly due to the mining process. For example, it takes nearly an hour to settle a Bitcoin transaction, which makes exploiting mispricing opportunities across exchanges difficult. This is consistent with Makarov and Schoar (2019), who find that there are many recurrent arbitrage opportunities in Bitcoin and Ethereum across exchanges.

There are 92 pumped tokens that have also traded ETH pairs and 89 pumped tokens that are cross-listed on one or two other exchanges. Figure 3.6 displays the results pertaining to these ETH pairs and cross-listed tokens. If a token is cross-listed on two other exchanges besides the target exchange, we take the average of maximum returns, abnormal volume, and volatility across the two exchanges. The results show that the average return, abnormal volume, and volatility of the ETH pairs are comparable to those of the BTC pairs (Figure 3.6, Panel A), but the average return, abnormal volume, and volatility of the pumped tokens traded on other exchanges are all much smaller than those of the tokens traded on the target exchanges (Figure 3.6, Panel B). The results reported Figure 3.6 and in Figure 3.3, Panel A are not directly comparable as the sample compositions are different. In Part A4 of the Appendix, we conduct an analysis by restricting our data to the same sample and find similar results.

[Insert Figure 3.6 here.]

Overall, the results reported in Figure 3.6 and Part A4 of the Appendix confirm our expectation that there is a substantial spillover to ETH prices, but the spillover to other exchanges is less significant. This is consistent with our previous discussion of the differences between the arbitrage limits on BTC prices and ETH prices, respectively, as opposed to similar limits across exchanges. The quick reversal shown in Figure 3.3 and weaker spillover to other exchanges shown in Figure 3.6, Panel B provide strong evidence that the price increases are not supported by changes in token fundamentals.

3.4.3 Performance by purchase and sale time

We can infer how investor performance depends on the timing of buying and selling based on the results reported in Figure 3.3, Panel A. This inference reflects the perspective of a small investor whose trade does not, however, move the market. In this section, we investigate this question from an average real-world investor's perspective. Specifically, we account not only for real-time returns but also for real-time liquidity. Note that the returns are calculated before exchange commissions, which range from 0.2% to 0.5% per round trade. Incorporating commissions does not materially affect our results.

In Table 3.2, Panel A we report the mean "achievable" returns by purchase time. Given the time of purchase, we also check whether the achievable returns vary with the time over which an investor delays before he begins selling. We can illustrate how to read the numbers presented in Table 3.2 with the following example. Suppose the total volume for token ABC in the first 10-second interval after the P&D announcement is N and the volume-weighted average price is P. Assume that all these investors follow the same strategy by waiting for D seconds before they start to sell. When they start to sell, we track how long it will take them to unwind all of their purchased tokens, of which we assume there are N. We also assume that when they sell, they are the only

sellers on the market and can trade on the volume-weighted average price until they sell all of their purchased tokens. We then calculate their returns and gauge how long it takes to unwind their purchases. Given that it is unlikely that these investors are the only sellers when they sell, our measure of the time taken to unwind purchased tokens will underestimate the actual time required. The inferred achievable returns therefore will overestimate investor performance.

[Insert Table 3.2 here.]

In Panel A we report the mean achievable returns for each purchase time-delay combination across P&Ds. We highlight in bold the returns that are significantly positive at the 5% level. In Panel B we show the corresponding $t-$values, while in Panels C and D we report the mean and median times to unwind the initial purchase. In addition to the 12 10-second intervals after a P&D, we also include a "Before P&D" interval, which is the 10-minute window before the P&D announcement. We use this interval to evaluate the performance of "insiders."

Not surprisingly, returns depend critically on when an investor buys and sells. Investors who buy before a P&D realize great returns. If they start to sell immediately after the announcement, their return will be 13.93%. The median time to unwind their portfolio is 20 seconds, while the mean is 210 seconds, suggesting that some positions take some time to unwind. These investors' average return, at 18.44%, is highest if they sell after a delay of 30 seconds. On average, it takes 263 (median=30) seconds to unwind a portfolio. Returns start declining if sales are delayed further. However, even if sales are delayed by 180 seconds, the average return is still 8.21%. It will, however, take an average of 733 (median 145) seconds to unwind the position.

Other groups' returns are much lower. Investors who buy in the first 10 seconds after a P&D will realize a positive return if they do not delay selling by more than 140 seconds. However, they cannot delay selling by more than 80 seconds to make

statistically significant positive returns. With optimal timing, their return is 6.53%. The next group who buy in the second 10-second interval cannot delay selling by more than 10 seconds to make statistically significant profits. Other groups, on average, cannot make money at all.

The results reported in Table 3.2 show that, to make positive returns from P&Ds, an outsider has to buy tokens in the first 20 seconds and make sure that he can sell before it is too late. Most others lose money. Compared with the results reported in Figure 3.3, Panel A, these results suggest that the performance of outsiders is much less impressive, highlighting the importance of real-time liquidity. As discussed above, these estimates are likely to overestimate investors' real performance, as we assume investors use the fastest possible method to unwind their positions.

The consistently positive performance achieved by investors in the "Before P&D" interval (i.e., 10 minutes before a P&D) suggests the occurrence of wealth transfers from later participants to these early investors. We deem the former insiders and the latter outsiders. The average volume in the "Before P&D" interval is 13% of that during the 10 minutes after P&Ds. If insiders time the market optimally, they can achieve a 18.44% return. This 18.44% return implies a 2.40% loss for an average outsider.

3.4.4 Other potential sources of unfairness

As noted above, not all pump outsiders are treated equally. Some receive signals earlier than others. We do not, however, observe the times at which VIP members receive their signals, and thus we cannot evaluate the advantages VIP members enjoy. In this section, we investigate whether participants in some Telegram channels receive signals systematically later than others.

The results are reported in Table 3.3. Panel A reports delays in announcements, which are measured by the number of seconds that elapse after a scheduled pump time.

A negative number indicates that the announcement time is ahead of schedule. We report the largest ten pump groups (measured by number of all P&Ds in which they participate) and an aggregate number for all other channels. The delay is capped at ±30 seconds, and the Table 3.2 statistics show that the marginal effect of further delays beyond 30 seconds weakens considerably. This also helps us avoid undue influence by a small number of large deviations from the scheduled time. We also report the fractions of early announcements.

[Insert Table 3.3 here.]

We conduct the analyses for two samples. In the first four columns we report results for all P&Ds across the three exchanges. In the last four columns we report results based on P&Ds in our final sample, in which we require the availability of trading data. The results reported in the first four columns indicate that only about 20% of announcements are made ahead of schedule and most are delayed. The mean (median) delay for PumpZone is only 0.73 (0) seconds, while it is 11.29 (7) seconds for Premium Yobit Pump. The results based on our final sample are similar. This suggests the presence of significant heterogeneity in receiving pump signals across groups. In untabulated results, we analyze delays in a regression specification by including P&D fixed effects. The fixed-effect model tests whether separate Telegram channels for the same P&D receive the signal at the same time. The null hypothesis that all channels receive the signal at the same time is strongly rejected.

In Panel B, we present the results of investigating whether there is any persistence in delays. We regress a channel's delay in receiving a P&D signal on its lagged delay. We do this for both the Early dummy, which equals 1 if an announcement time occurs is before the scheduled time, and a continuous measure of delay. We use one specification without P&D fixed effects and another with fixed effects. All our results show strong persistence in the delay in receiving signals. Overall, although we cannot judge whether

the heterogeneity in signal reception is necessarily an intentional outcome, the results reported in Table 3.3 provide strong evidence of additional unfairness besides that which occurs between insiders and outsiders.

3.4.5 Maximum returns, target returns, and viewership

In this section, we analyze factors that affect P&D maximum returns and target returns, and determine whether the target returns are achieved. This analysis sheds light on the extent to which pump group operators behave strategically. We also study the dynamics of Telegram viewership, focusing on whether investors learn from past experience. We study the former at the P&D level and the latter at the Telegram channel level.

At the P&D level, we examine factors that affect the maximum returns. For Figure 3.3, we define maximum return as the ratio of the highest price achieved within 10 minutes after a pump announcement and the price ten minutes before the announcement minus 1. Because we study the price run-up, which is mechanically related to the maximum return defined in this way, we redefine the maximum return as the ratio of the highest price achieved within 10 minutes after the pump announcement to the price at the announcement minus 1. At the channel level, it is interesting to study whether past pump performance, such as recent pump returns and abnormal volume, affects Telegram viewership, which we use as a proxy for investor participation.[12]

The results reported in column (1) of Table 3.4 confirm the findings displayed in Figure 3.5, which demonstrate that maximum returns are lower in liquid tokens than in illiquid ones. Potential leakages, proxied by price run-ups, negatively predict the maximum returns achieved. Pumps launched by multiple groups achieve a maximum

[12]One potential concern is that Telegram's post view count system may overestimate the number of real-time viewers, as the view count is updated over time. However, random checks several months after we collected the data indicate that few people view old posts retrospectively.

return that is substantially higher than returns involving a single channel, suggesting that P&Ds rely on investor participantion. As shown in column (2), pump operators set target returns at a lower level for liquid tokens. In column (3) we further show that in 31.4% of the cases the target returns among the channels are achieved. However, liquidity does not affect whether or not target returns are reached. The finding that target returns are not achieved in most cases suggests that either the pump operators intentionally set the target returns too high to mislead participants or coordination between pump members is difficult to accomplish.

[Insert Table 3.4 here.]

Given that the average P&D lasts for only several minutes, it is reasonable to believe that Telegram channel members are the dominant outsiders because there is little time for non-members to react to sudden jumps in token price and volume. Does the performance of channel members affect the popularity of P&Ds? In Table 3.5, we report the results of examining factors that affect the dynamics of Telegram channel viewership. Our dependent variable is the change in the natural logarithm of Telegram channel viewers. The results suggest that past P&D performance does not affect Telegram channel viewership. Although viewership does not necessarily lead to participation in P&Ds, our finding that viewership is not sensitive to past P&D performance provides indirect evidence that learning by investors may be limited.

[Insert Table 3.5 here.]

3.4.6 Potential mechanisms

To summarize, we have reported three main findings. First, an average P&D moves token prices and trading volume significantly. Second, investor performance depends critically on when they obtain their signals. Insiders could make abnormally high returns,

but it is difficult for outsiders to achieve this. Third, some outsiders are disadvantaged and receive signals systematically later than others. All these findings make it puzzling that an outsider, especially a disadvantaged outsider, is willing to participate in a P&D.

In trade-based price manipulation theories in which a manipulator cannot act in ways other than buying and selling assets, the manipulator needs to buy the target asset to "pump" its price first. As outside investors follow, the manipulator can sell at a higher price to make profits. The manipulator can make profits only if the price impact is greater when they buy than when they sell; otherwise this strategy will be self-defeating (Friedman, 1953). Trade-based manipulation can be profitable if there is price momentum (Jarrow, 1992) or uninformed investors believe that it is likely that the manipulator is informed (Allen and Gale, 1992).

In Figure 3.7, we report the results pertaining to price impact. We estimate this impact for every minute instead of every ten seconds because estimating price impacts requires more data. Specifically, price impact λ_t is estimated based on the following equation,

$$\Delta p_{i,t} = \alpha_t + \lambda_t q_{i,t} + \epsilon_{i,t} \tag{3.1}$$

where $q_{i,t}$ and $\Delta p_{i,t}$ are order flow and return over interval t for token i, respectively. Order flow is defined as the volume of buy orders minus the volume of sell orders in Bitcoin units. $\epsilon_{i,t}$ is the error term. λ_t is our measure of price impacts. This method is similar to those adopted in Glosten and Harris (1988), Brennan and Subrahmanyam (1996), and more recently by Makarov and Schoar (2019).[13]

[Insert Figure 3.7 here.]

The results show that the price impact immediately after a P&D starts is weaker than that immediately before and is similar to that in later intervals. The abnormally high

[13]Using a similar model, Makarov and Schoar (2019) study the relationship between Bitcoin order flow and prices. They consider the fact that Bitcoin is listed on many exchanges. We consider only the order flow from the target exchange in light of our finding that spillover to other exchanges is small.

price impact immediately before a P&D is consistent with insiders' purchase causing the price run-ups we report in Figure 3.3, Panel A. The weaker price impact after a P&D starts is consistent with the fact that the P&D announcement leads to increased participation of uninformed speculative traders. Their purchases likely cause the large price increase. Our results suggest that outside investors crowd into the market and effectively "coordinate" to create short-term price bubbles, followed by quick reversals. This contrasts with the existing trade-based manipulation theories which typically assume that the manipulators need to buy to increase the asset price.

These results do not, however, explain why outsiders are willing to participate. Since information on investor identification is not available, we cannot pin down the exact mechanism. We conjecture that one plausible mechanism is that P&Ds attract overconfident investors who believe that they can time the market better than others can (Daniel et al., 1998; Scheinkman and Xiong, 2003; Daniel and Hirshleifer, 2015). As we show in Table 3.2, buyers in the first 20 seconds can potentially make money, and there are significant variations in average returns across P&Ds even if one buys tokens later. Overconfidence has been found to be useful in explaining many phenomena in the stock market (Barber and Odean, 2000; Barber and Odean, 2001; Gervais and Odean, 2001; Daniel and Hirshleifer, 2015). Barber and Odean (2001) find that on average men are more overconfident than women. Survey evidence shows cryptocurrency traders are predominantly men (Leinz, 2018). It is likely that cryptocurrency investors may also be overconfident.

Another possible mechanism encouraging outsiders to participate is that these investors have gambling preferences. The short-term return on pumped tokens is very high and salient. Investors may overweight these returns in their decision-making or overestimate the skewness of the tokens, consistent with salient thinking or prospect theory (Barberis and Huang, 2008; Bordalo et al., 2012; Bordalo et al., 2013; Wang,

2018). Evidence of gambling preferences on the part of stock market investors abounds (Kumar, 2009; Barberis et al., 2016).

3.5 The Bittrex Ban

In this section, we examine the economic consequences of P&Ds. We find that P&Ds lead to short-term cryptocurrency bubbles featuring dramatic increases in prices, volume, and volatility, which suggests that P&Ds lower the informativeness of cryptocurrency prices. Evidence suggests that exposures to scandals lower investor trust, and reduce stock market participation and the use of financial intermediation services (Giannetti and Wang, 2016; Gurun et al., 2017). Do P&Ds play a similar role by lowering investors' willingness to invest in the cryptocurrency market? We study this question using a quasi-experiment.

3.5.1 The background

On November 24, 2017, Bittrex, then the world's third-largest cryptocurrency exchange by volume, sent a notice to customers warning them about market manipulation tactics. Customers could be banned or have their accounts frozen for artificially manipulating the prices of tokens trading on its platform. Bittrex stated that it "actively discourages any type of market manipulation, including pump groups. Consistent with our terms of service, we will suspend and close any accounts engaging in this type of activity and notify the appropriate authorities." Bittrex's ban followed an investigation by Business Insider ten days earlier that found traders were colluding in groups on Telegram to inflate the price of cryptocurrencies on platforms such as Bittrex and Yobit (Williams-Grut, 2017).

Messages about pump groups on Telegram show that traders involved in P&D tactics had taken notice of the warning. Many scheduled P&D events were immediately

canceled. For example, one prominent pump channel, "Trading signals for crypto," canceled its P&D event on November 26, 2017 (see Figure 3.8, Panel A). Some message groups solicited feedback from group members regarding whether to switch to other exchanges. Many message groups eventually ceased to pump tokens traded on Bittrex and/or switched to alternative exchanges such as Yobit.

[Insert Figure 3.8 here.]

In Figure 3.8, Panel B we plot the number of P&Ds on Bittrex and the other two exchanges over the period running from five months before to eight months after the ban. This period spans our sample period. Although the ban was not able to eliminate P&Ds altogether, it is evident that there was a sharp decrease in the frequency of P&Ds on Bittrex. Before the ban, the monthly average number of P&Ds targeting Bittrex in our sample is 44. We have 225 tokens listed on Bittrex with trading data, implying a significant hazard of being targeted before the ban. This suggests that Bittrex was able to adhere to its announced policy and worked to reduce the occurrence of P&Ds on its exchange.

3.5.2 Consequences of the ban

We adopt a difference-in-differences approach to examine the effects of the Bittrex ban on the prices and trading volumes of tokens. All tokens listed on Bittrex comprise the treatment group, and other tokens comprise the control group. We conduct our analysis based on CoinMarketCap data. We use CoinMarketCap data rather than exchange data because the tokens on the other two exchanges differ in important ways from Bittrex tokens (see Table 3.1).

To control for differences between Bittrex tokens and others, we conduct a matched-sample exercise. Specifically, we match each token listed on Bittrex at the time of the ban with a token that is not listed on Bittrex. We match them based on the average

trading volume (in Bitcoin) in the period running from 60 days to 31 days before the ban. We use the non-Bittrex token with the closest trading volume as the matched token. The matching is performed with replacement. We further require that the log difference in volume between a Bittrex token and the matched token not be greater than 20%.

In Figure 3.9, we present the average trading volumes and prices over the period running from 14 days before the ban to 30 days after. Trading volume is measured as the natural logarithm of the ratio between the trading volume on a given day and the average volume in the period running from 60 days to 31 days before the ban. Price is calculated as the log price change from the price 30 days before the ban to the end of a given day. We conduct these transformations to mitigate the effects of cross-token heterogeneity in volume and price. Tokens with no volume in the period running from 60 days to 31 days before the ban are excluded. Our final sample includes 190 treatment tokens and 190 matched tokens. Figure 3.9 shows that in this period there was a sharp increase in both volume and price for both the treatment and control groups, suggesting a booming market. The results presented in Figure 3.9 make it evident that both the volumes and prices of the Bittrex tokens increased relative to those of the controls. The increases started almost exactly on the event day. In the pre-ban period, although the treatment and control groups do not move perfectly in tandem, their volumes and prices fluctuate mostly in the same direction, suggesting that the control group represents a reasonable counterfactual.[14]

[Insert Figure 3.9 here.]

[14]As we discussed, the ban caused some pump groups to switch from Bittrex to other exchanges. Therefore, the ban reduced the frequency of P&Ds on Bittrex, and at the same time, it increased the frequency of P&Ds on other exchanges. The difference-in-differences estimate reflects both the effect of reduced frequency of P&Ds on Bittrex and the increased frequency of our controls. However, we think this "spillover" effect is likely weak, because the majority of control tokens in our sample are not listed on the exchanges that the pump groups switch to.

We conduct formal difference-in-differences tests and report the results in Table 3.6. Specifically, we estimate model (2) below:

$$DV_{i,t} = \alpha + \beta_1 Bittrex_i + \beta_2 Post_t + \beta_3 Bittrex_i * Post_t + \epsilon_{i,t} \tag{3.2}$$

where DV is either volume or price, $Bittrex$ is a dummy equal to 1 for Bittrex tokens and 0 for control tokens, and $Post$ is a dummy equal to 1 if it is after November 24, 2017 and 0 otherwise. β_3 is the difference-in-differences estimator. Instead of running a pooled regression with observations at the token-day level, we aggregate all the data and produce two observations for each token, one for the pre-ban period and one for the post-ban period (Bertrand et al., 2004). Specifically, for trading volume we calculate the average for the pre-ban and post-ban periods for each token. For price, we use the price at the beginning of the pre-ban period and the price at the end of the post-ban period.

[Insert Table 3.6 here.]

Our results are robust to variations in the lengths of the pre- and post-ban periods. To obtain the results we report in Figure 3.9, we define the pre-ban period as running from day -14 to day -1, and the post-ban period as running from day 1 to day 30. We exclude day 0 as it is ambiguous whether to include it in one or the other period. In Figure 3.9, we report the results of examining two specifications, one with token fixed effects and another without such effects. In the specification with token fixed effects, *Bittrex* is subsumed. We cluster the standard errors by treatment-control pair. The results show that in both the volume and price regressions the coefficient on the interaction term is both economically and statistically significant. The estimates suggest that the Bittrex ban increased the prices and volumes of its tokens by 40.1% and 63.9%, respectively.

Overall, the results we report in Figure 3.9 and Table 3.6 provide casual evidence that P&Ds are detrimental to the health of the cryptocurrency market. Kyle and

Viswanathan (2008) propose two necessary conditions to classify a scheme as "illegal price manipulation:" (1) the scheme makes prices less accurate as signals for efficient resource allocation, and (2) it makes markets less liquid for risk transfer. Our results suggest that cryptocurrency P&Ds satisfy both conditions and should be considered illegal.

3.6 Conclusion

This paper studies pump-and-dump schemes (P&Ds) in the cryptocurrency market. We find that most P&Ds lead to short-term bubbles where prices, volume, and volatility increase dramatically, followed by a quick reversal. We also find evidence of significant wealth transfers from potential outsiders who do not know the identities of pumped tokens in advance to insiders. The quick reversals imply that it is difficult for any outsiders except the fastest movers to make profits. Standard trade-based price manipulation theories assume that manipulators need to buy to pump the asset price. In contrast to this, we find that pumps are mainly driven by uninformed speculative traders.

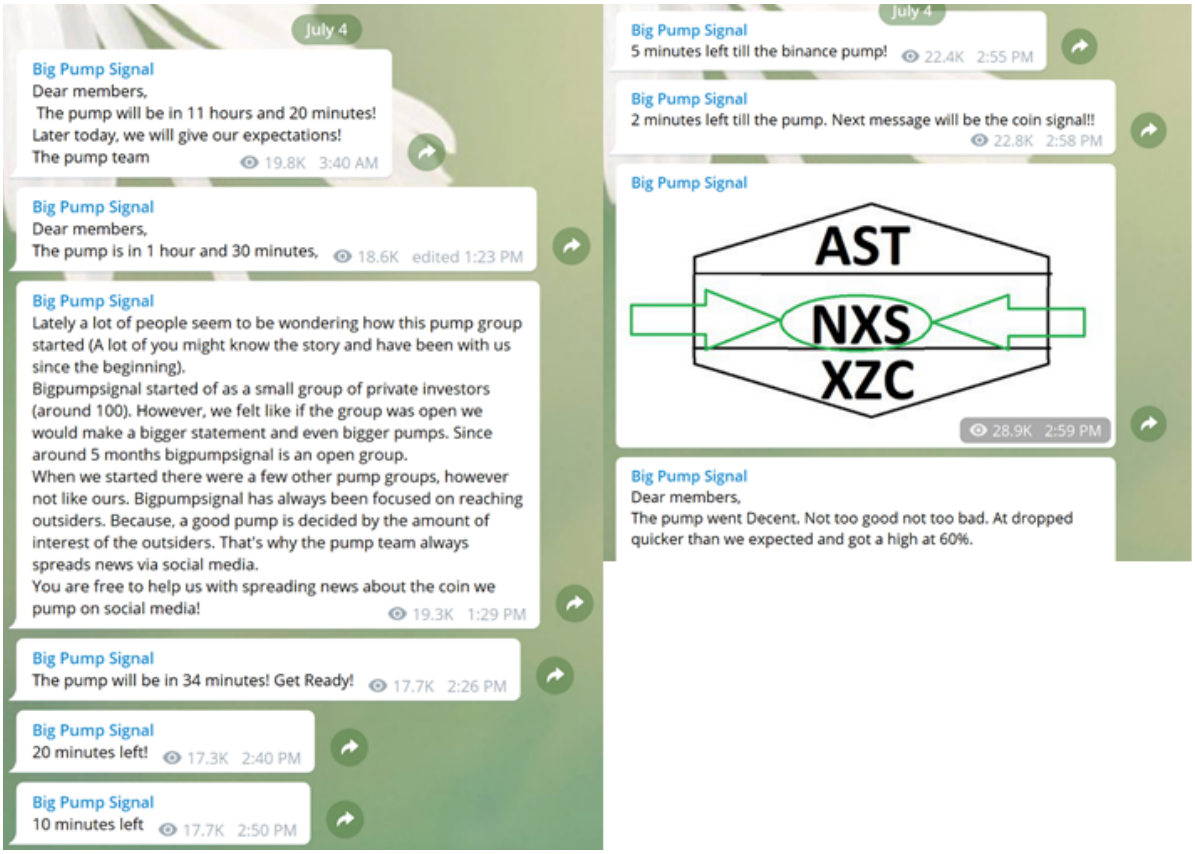
Our findings make it puzzling why these outsiders are willing to participate in P&Ds. We conjecture that one plausible mechanism is that P&Ds attract overconfident investors who believe that they can time the market more accurately than others can (Scheinkman and Xiong, 2003; Daniel and Hirshleifer, 2015). Another possible mechanism is that these investors are affected by the salience of short-term extreme returns, and overweight the possibility that they will realize similar returns in their decision-making or overestimate the skewness of token returns (Barberis and Huang, 2008; Barberis, 2013; Bordalo et al., 2012; Bordalo et al., 2013).

We also conduct a difference-in-differences test of the Bittrex ban of P&Ds on its exchange to shed light on the equilibrium effects of P&Ds. Using tokens on Bittrex as the treatment group and a matched sample of tokens as the control group, we find

strong evidence that P&Ds are detrimental to the health of the cryptocurrency market. Specifically, banning P&Ds increased the prices and volumes of tokens listed on Bittrex to a greater extent than those of other tokens.

Figure 3.1: An Example of Pump-and-Dump Schemes on Telegram
Panel A displays a series of Telegram announcements by Big Pump Signal, one of the biggest pump groups, regarding their July 4, 2018 pump targeting Nexus (NXS). Panel B plots Nexuss prices and volumes before and after the pump announcement (t =0).

Panel A. Telegram pump announcements on Nexus



Panel B. Nexuss price and volume before and after the P&D announcement (t =0)

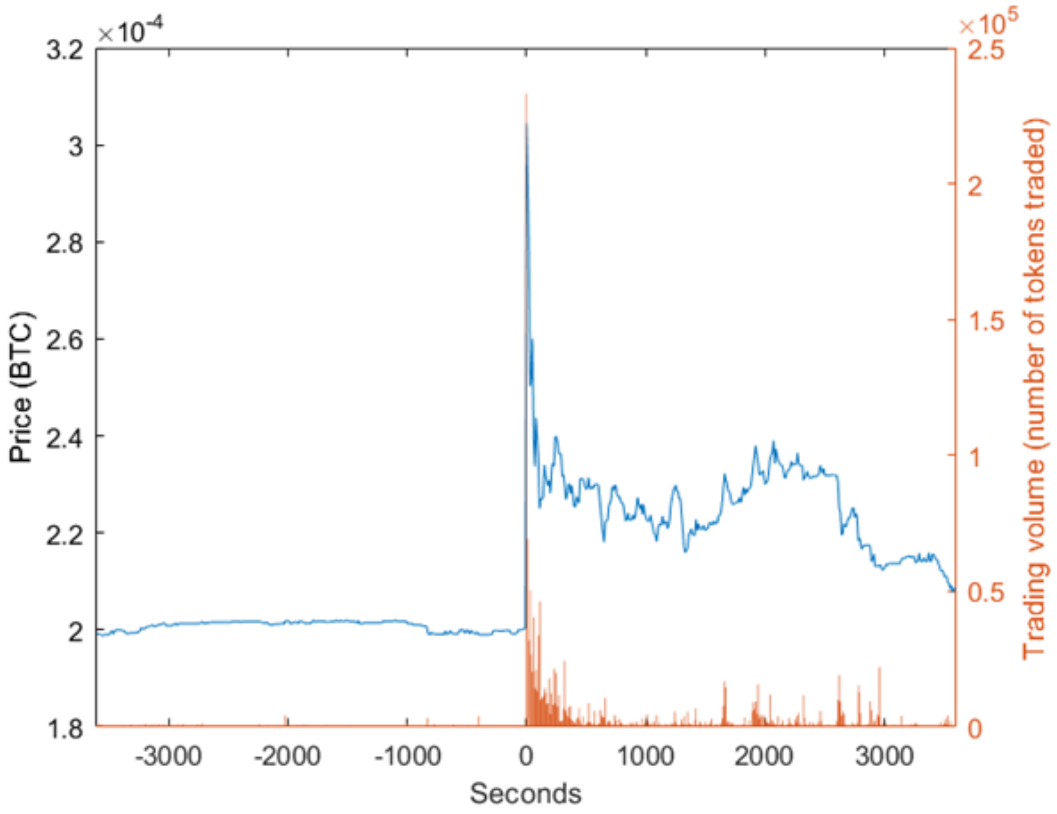
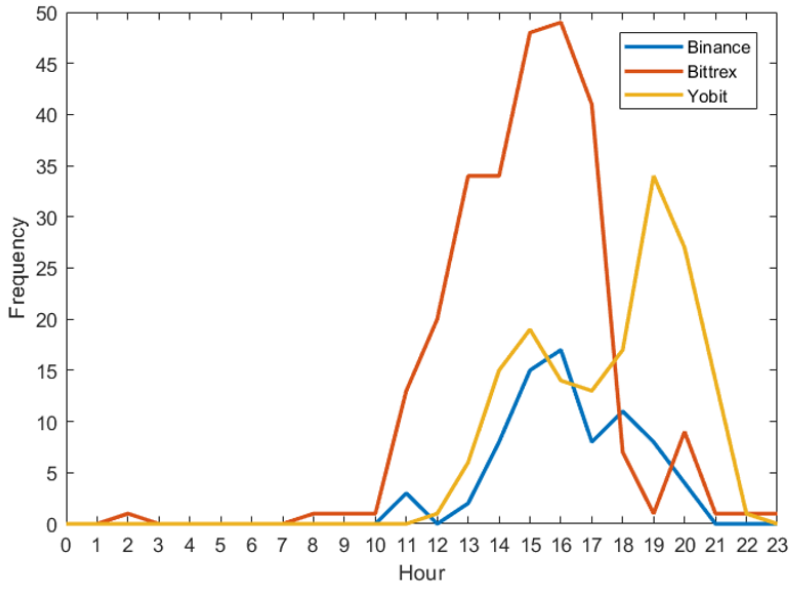


Figure 3.2: Distribution of Pump and Dumps and Trading Volume by Hour
In Panel A we present the distribution of P&D events by scheduled hour and exchange. Panel B displays the fraction of trading volume by hour and exchange. For each exchange, we first calculate the fraction of the trading volume (in Bitcoin) occurring in each hourly interval, and then compute the average volume across all dates running from May 15, 2017 to August 26, 2018. The X-axis represents the hour, where 0 indicates the interval between 00:00:00 and 00:59:59 (Coordinated Universal Time or UTC). Other hourly intervals are similarly defined.

Panel A. Distribution of P&Ds by hour and exchange



Panel B. Distribution of trading volume by hour and exchange

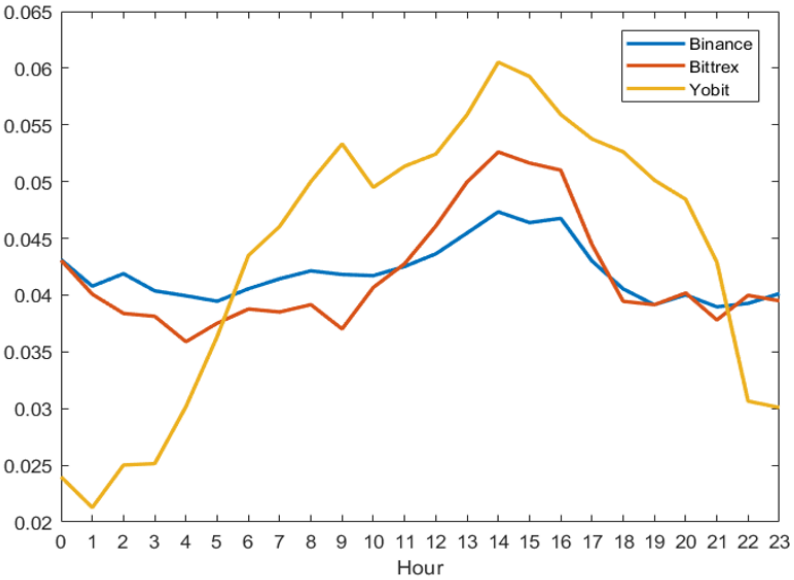
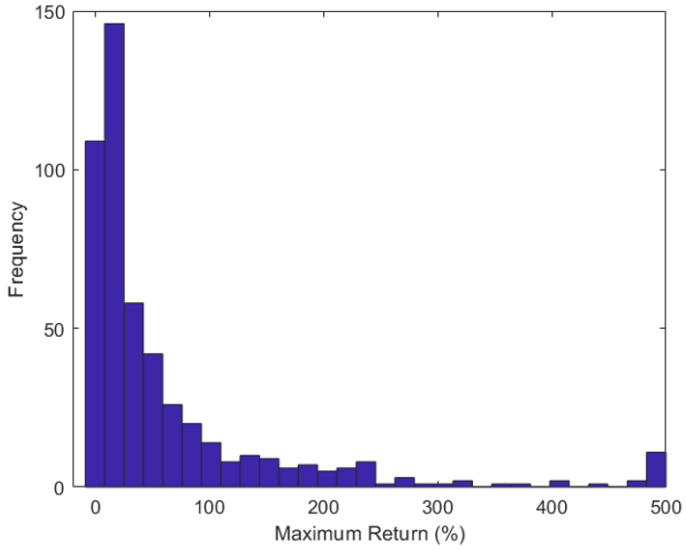


Figure 3.3: Distribution of Maximum Returns and Time to Maximum Returns

Panel A displays the distribution of P&D maximum returns, while Panel B plots the distribution of the number of seconds it takes to reach a maximum return for our sample of 500 P&Ds. A P&D's maximum return is defined as the ratio of the highest price achieved within 10 minutes after the pump announcement and the price ten minutes before the announcement minus 1. The time to a maximum return is defined as the number of seconds it takes from the announcement to reach the maximum return. For ease of illustration, maximum returns are capped at 500% in Panel A.

Panel A. Distribution of P&D maximum returns



Panel B. Distribution of time to maximum returns

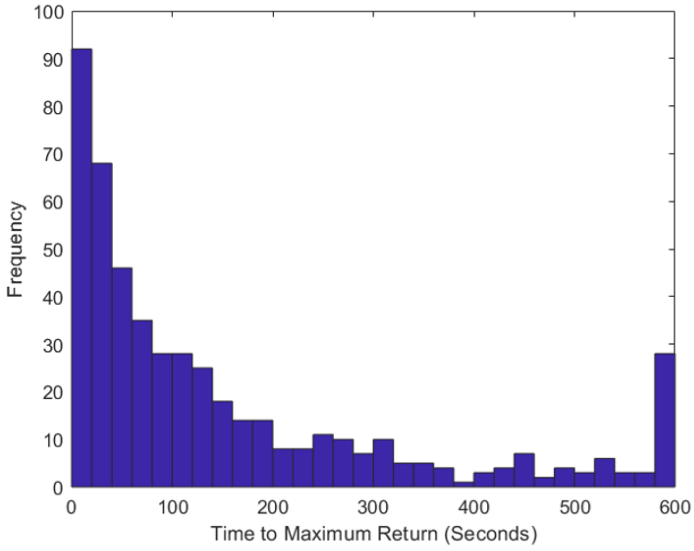
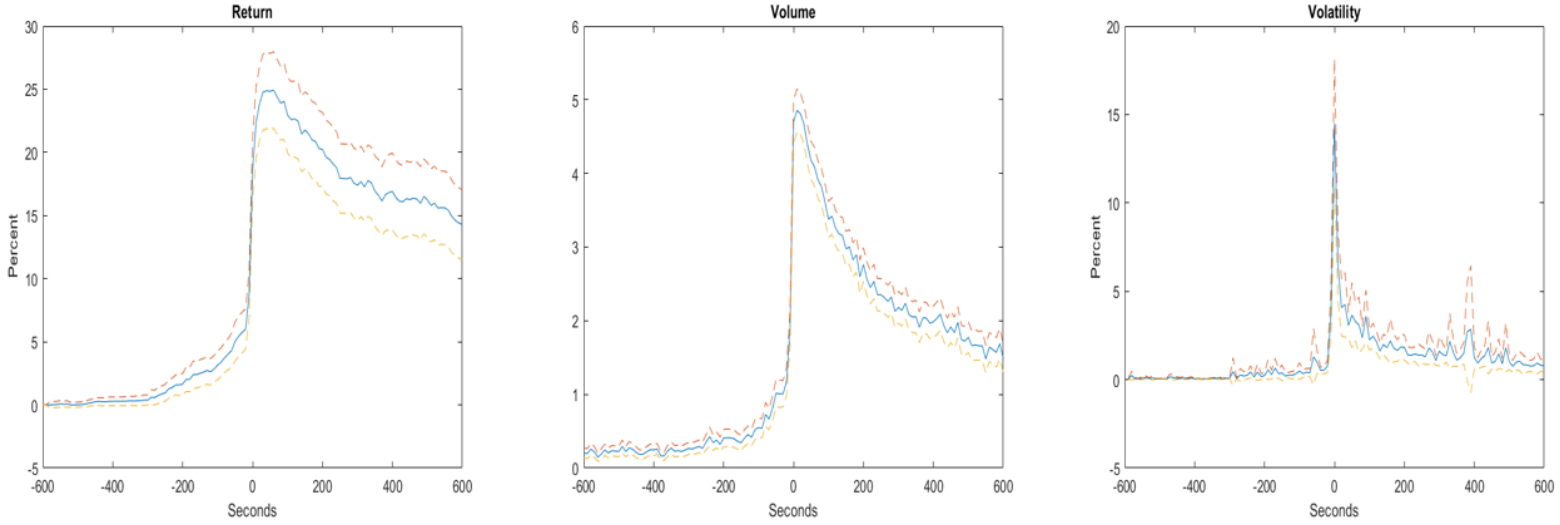


Figure 3.4: Return, Volume and Volatility Around Pump and Dumps
This figure displays cumulative returns, abnormal volume, and the volatility of pumped tokens. Panel A reports the pattern for each 10-second interval from 600 seconds before to 600 seconds after P&D announcements. Time 0 indicates the 10-second interval between 0 and 10, in which 0 is the announcement time. Panel B shows the analysis for each one-hour interval from 7 days before to 7 days after P&D announcements. The X-axis indicates time. The solid lines represent the averages across all target tokens, and the dashed lines show the 95% confidence intervals. Cumulative returns are calculated as the logarithm of price changes from 600 seconds before an announcement in Panel A and from seven days before an announcement in Panel B. Abnormal volume equals log(1 + volume over an interval/average volume in the interval over day -37 to day -8). Volume is measured by the number of tokens traded. Volatility is measured as the absolute value of the return during each 10-second interval in Panel A and each one-hour interval in Panel B.

Panel A. Short-term analysis



Panel B. Long-term analysis

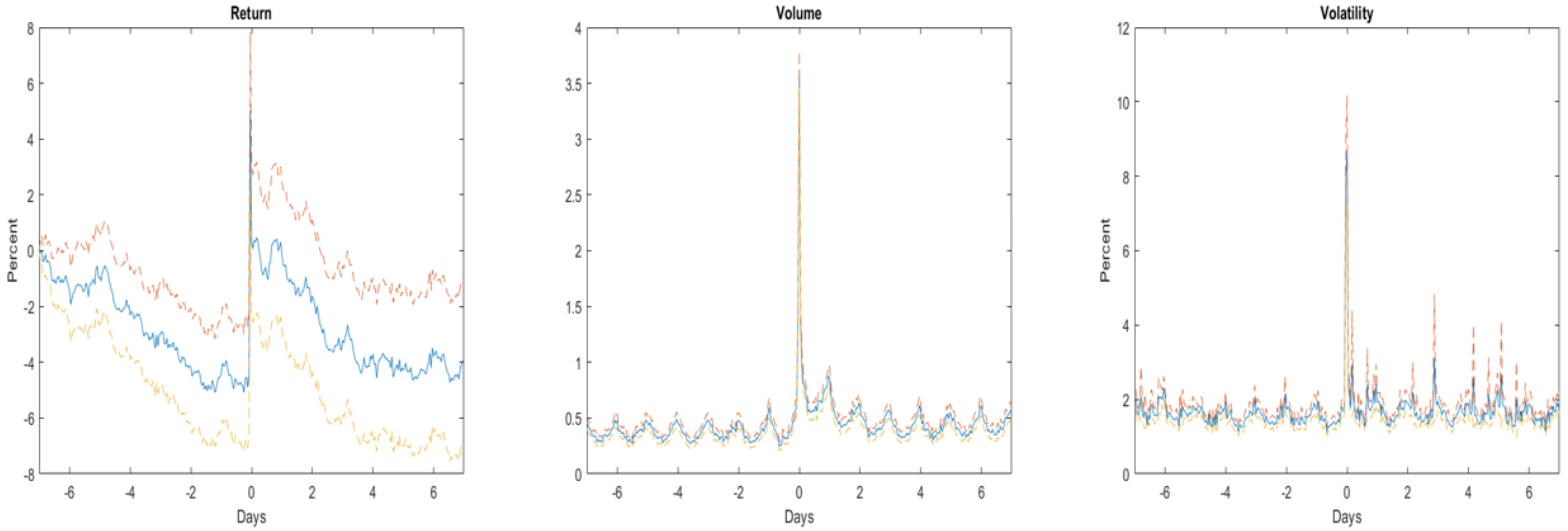
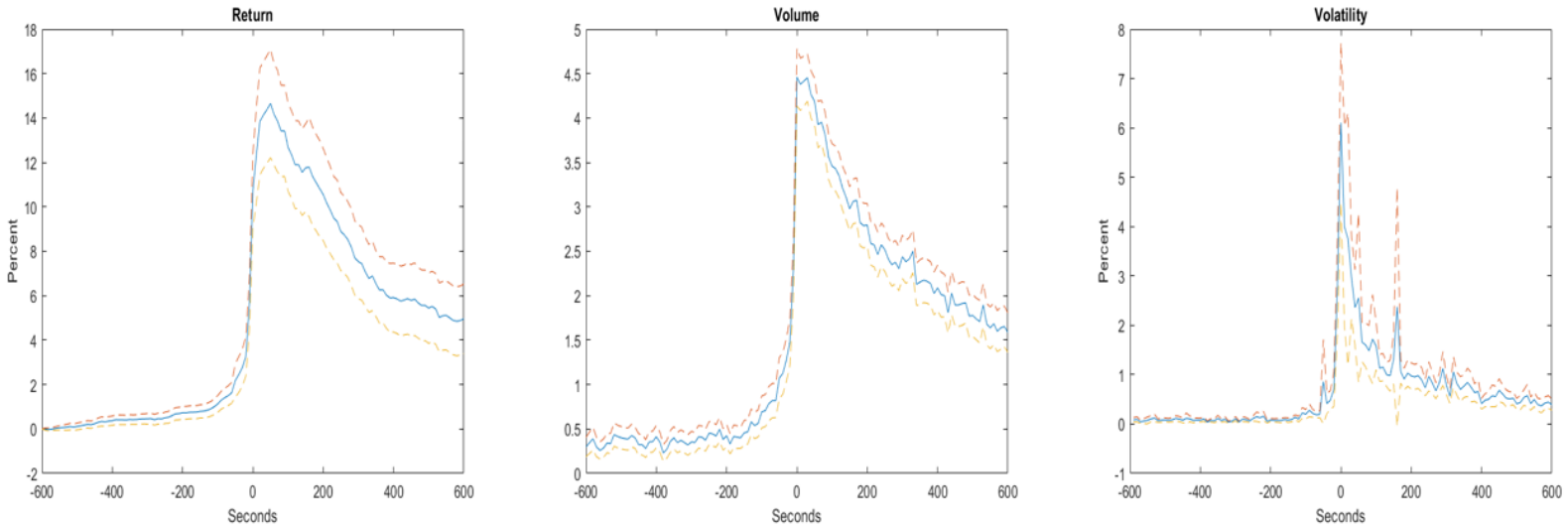


Figure 3.5: Return, Volume and Volatility Sorted by Liquidity

This figure displays short-term cumulative returns, abnormal volume, and the volatility of pumped tokens sorted by liquidity. Panel A features liquid tokens while Panel B shows plots for illiquid tokens. Liquidity is defined as the total trading volume (in Bitcoin) over the period running from 37 days to eight days before a P&D announcement. Liquid (illiquid) tokens are target tokens in our sample where the targets liquidity is above (below) the median value. We plot cumulative returns, abnormal volume, and the volatility for each 10-second interval from 600 seconds before to 600 seconds after P&D announcements. Time 0 indicates the 10-second interval between 0 and 10, in which 0 is the announcement time. The X-axis indicates time. The solid lines represent the averages across all target tokens, and the dashed lines show the 95% confidence intervals. Cumulative returns are calculated as the logarithm of price changes from 600 seconds before an announcement. Abnormal volume equals log(1 + volume over a 600-second interval/average volume in the interval over day -37 to day -8). Volume is measured by the number of tokens traded. Volatility is measured as the absolute value of return in each interval.

Panel A. Liquid tokens



Panel B. Illiquid tokens

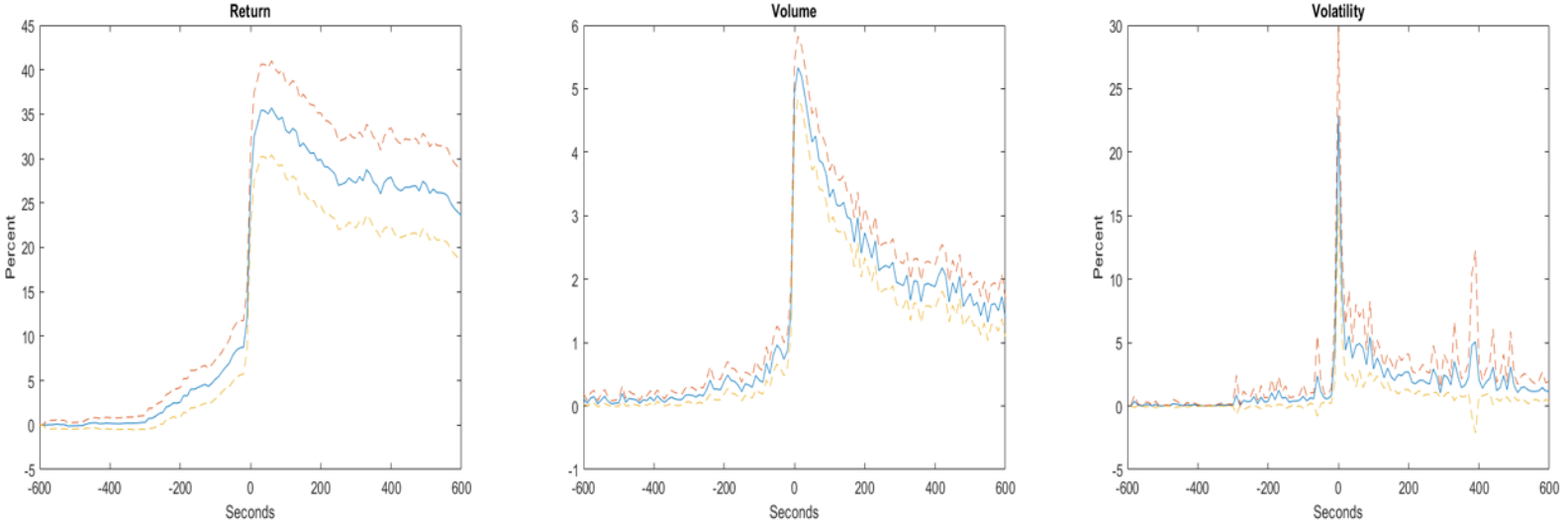
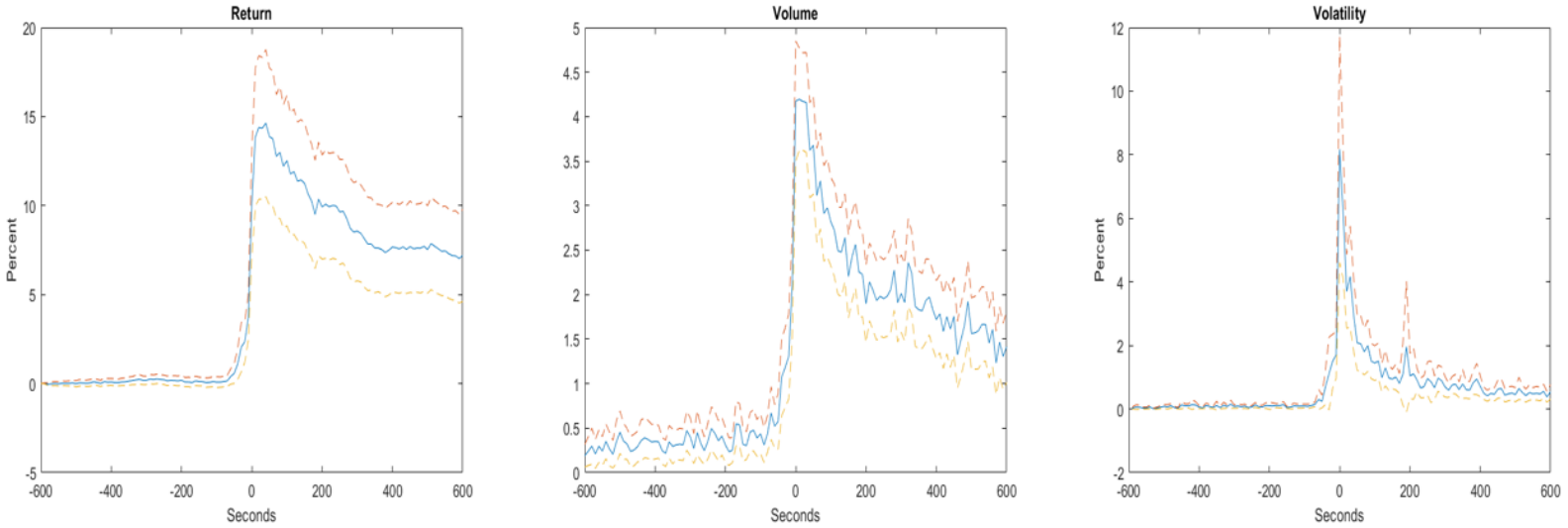
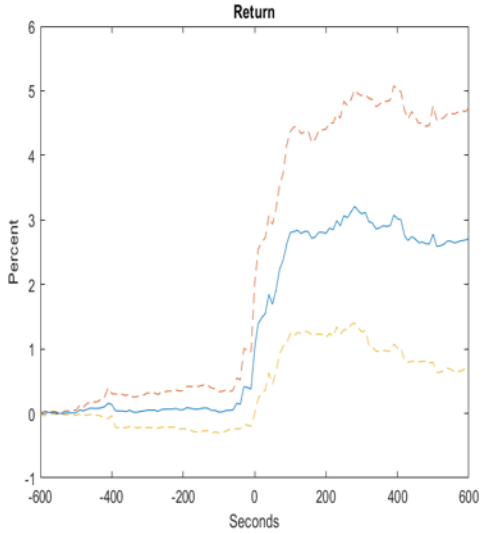


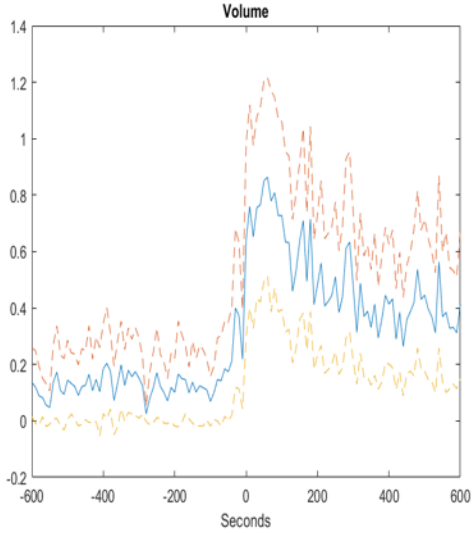
Figure 3.6: Spillovers to Ethereum Trading Pairs and Other Exchanges
This figure shows whether any spillover to prices and volume quoted in Ethereum occurs during P&Ds (Panel A) and prices and volumes quoted in Bitcoin on other exchanges (Panel B). In Panel A, for pumped tokens with trading data on Ethereum pairs, we report cumulative returns, abnormal volume, and the volatility of the Ethereum pairs. In Panel B, for pumped tokens with trading data from other exchange(s), we report cumulative returns, abnormal volume, and the volatility of these cross-listed tokens on the non-targeted exchange(s). If a token is cross-listed on two other exchanges besides the target exchange, we take the average of maximum return, abnormal volume, and volatility across the two exchanges.

Panel A. Ethereum pairs



Panel B. Bitcoin pairs in other exchanges





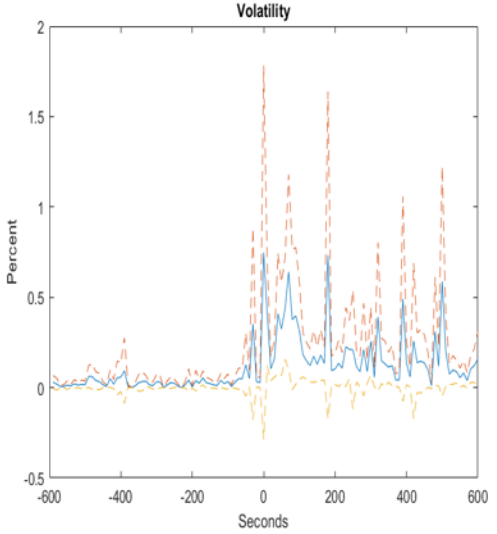


Figure 3.7: Price Impact
This figure displays the price impact. The price impact is estimated from the following regression for each 1-minute interval across all the P&Ds: $\Delta p_{i,t} = \alpha_t + \lambda_t q_{i,t} + \epsilon_{i,t}$ in which $q_{i,t}$ and $\Delta p_{i,t}$ are order flow and return at time t for token i, respectively. Order flow is defined as the volume of buy orders minus the volume of sell orders in Bitcoin units. $\epsilon_{i,t}$ is the error term. $\lambda_{i,t}$ is our measure of price impact and α_t is an intercept.

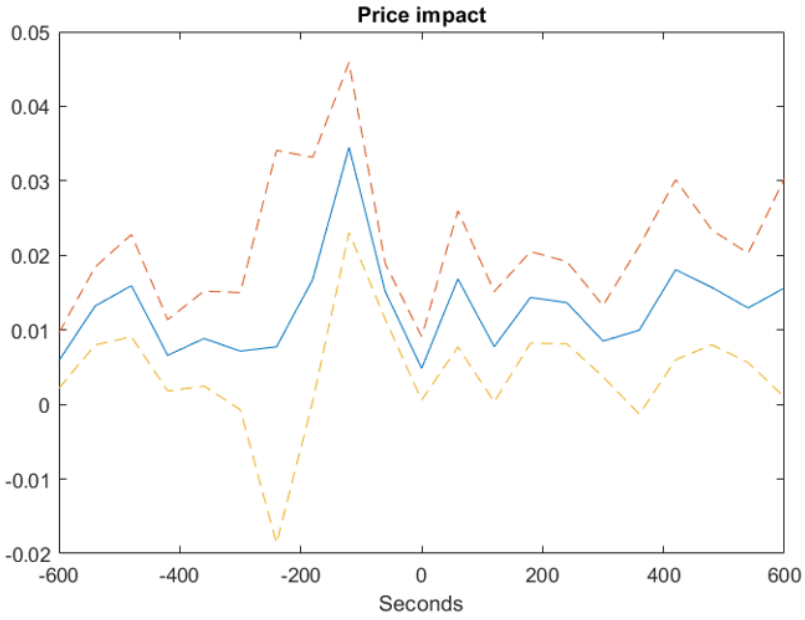
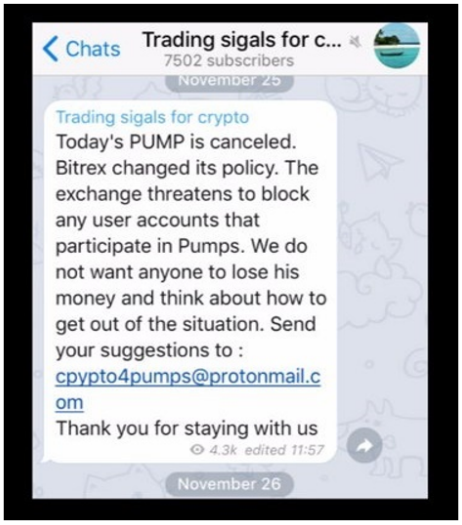


Figure 3.8: A Sharp Decrease in P&Ds on Bittrex after the Ban
Figure A displays a screenshot of the pump cancellation after Bittrexs ban of P&Ds. Panel B displays the frequency of P&Ds before and after the Bittrex ban. The X-axis shows the months relative to the ban that occurred on November 24, 2017. Our sample period runs from five months before to eight months after the ban, corresponding to the sample period for P&Ds.

Panel A. An example of pump cancellation



Panel B. Frequency of P&Ds around the ban

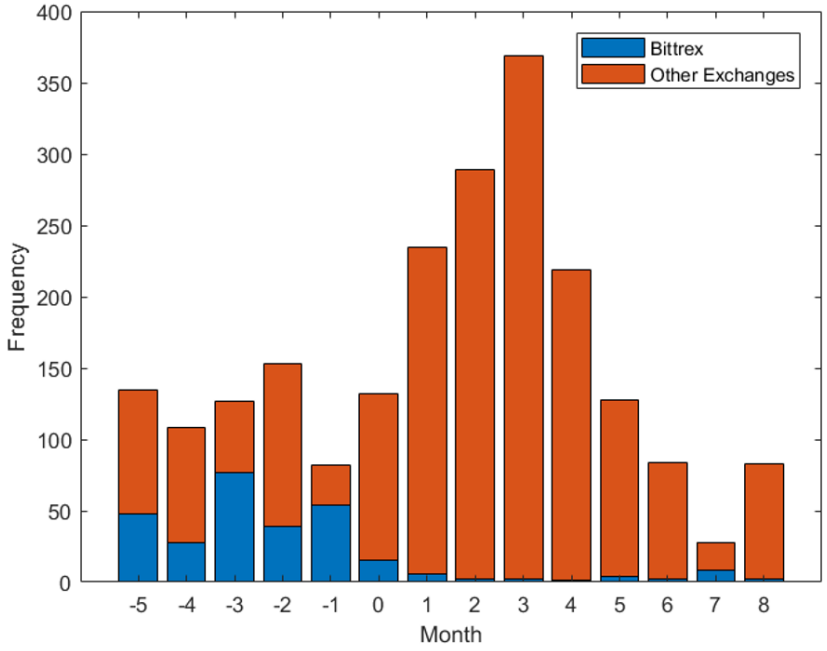
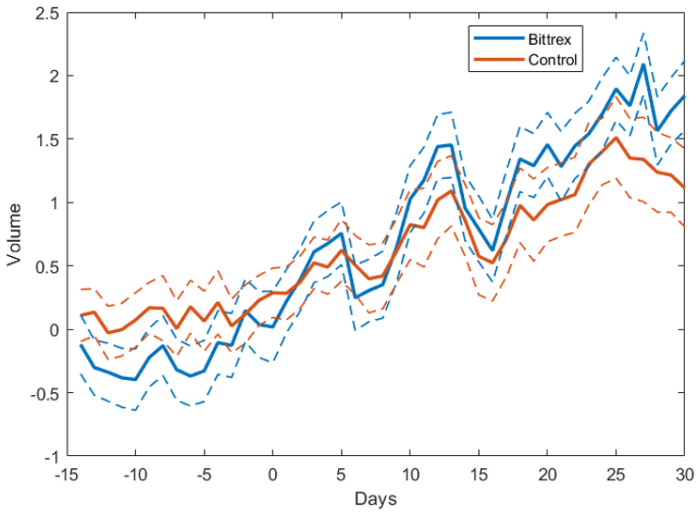


Figure 3.9: The Effects of the Bittrex Ban on Price and Volume
This figure displays volumes (Panel A) and prices (Panel B) of Bittrex tokens and the control tokens. The treatment group comprises tokens listed on Bittrex on November 24, 2017, and the control group includes the matched tokens not listed on Bittrex at that time. We use the non-Bittrex token with the closest trading volume (in Bitcoin) as the matched token. Matching is based on the trading volume over the period running from 60 days to 31 days before the ban. The X-axis is the day relative to November 24, 2017. Volume is measured as the natural logarithm of the ratio between volume on day t and the average volume over the period from 60 days to 31 days before the ban. Price is defined as the natural logarithm of the ratio between the end-of-day price and the price at the end of day -31. Each point on the solid lines is the average across either the Bittrex tokens or other control tokens. The dashed line displays the 95% confidence intervals.

Panel A. Volume



Panel B. Price

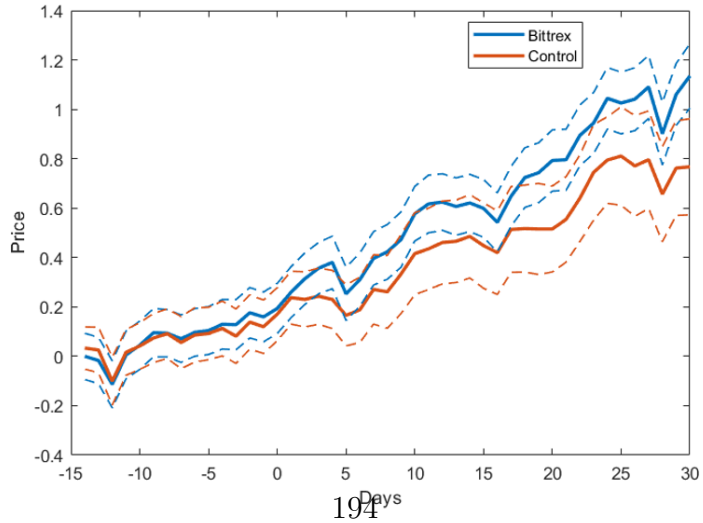


Table 3.1: Summary Statistics

In Panel A we report characteristics of P&D events between May 15, 2017 and August 26, 2018. *Earliest channels delay time* is the difference between the earliest pump announcement among all participating channels and the scheduled time. *Average delay time* is the average difference between the announcement and the scheduled time among all participating channels. *Total number of viewers* is the sum of viewers across all channels in a P&D event. *Target return* is the pre-specified return before a P&D announcement. *Regular scheduled time* is a dummy variable equal to 1 if a channel's P&D announcement is scheduled at the regular pump time and 0 otherwise. *A channels regular pump time* is the time most frequently used by its operator. Panel B displays the characteristics of pumped tokens on Binance, Bittrex and Yobit, and compares them with non-target tokens listed on the three exchanges. For each token characteristic, on each day with at least one pump we calculate the average value for target and non-target tokens. We then compute the time-series average. The characteristics we report are log market capitalization (U.S. dollar), log trading volume (Bitcoin), volatility, the frequency at which a token is covered by CoinMarketCap.com, token age, and a social media index. Volatility is defined as the standard deviation of daily log returns. Trading volume and volatility are calculated over the period running from day -37 to day -8. *Token age* is the number of years since the cryptocurrency was first listed. *Social media index* is the logarithm of social media activity points computed by CryptoCompare.com. sym*, sym**, and sym*** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Pump-and-dump events

	All (N=500)		Binance (N=76)		Bittrex (N=263)		Yobit (N=161)	
Event-level	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Earliest channels delay time (seconds)	21.7	4	20.1	5	14.0	4	34.9	3
Average delay time (seconds)	24.5	5.2	31.2	8	14.3	4.5	38.0	5.5
Average number of channels	1.58	1	1.42	1	1.08	1	2.45	1
Total number of viewers	5,941.5	2,094	22,445.0	3,049	3,887.1	2,382	1,507.0	1,038
Average number of viewers per channel	4,295.6	1,633	13,915.1	2,973	3,649.7	2,373	787.8	523.5

	All (N=788)		Binance (N=108)		Bittrex (N=285)		Yobit (N=395)	
Channel-level	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Target return	211.5%	200%	69.0%	51.2%	246.9%	100%	233.0%	200%
Regular scheduled time	89.5%	100%	86.1%	100%	84.2%	100%	94.2%	100%

Panel B. Characteristics of tokens

	Target tokens	Non-target tokens	Difference	t-stat. of Diff.	Percentile
ln(Market capitalization (dollar))	16.078	16.375	-0.219*	-1.72	49.63
ln(Volume (Bitcoin))	4.538	3.345	1.193***	5.96	58.00
Volatility	12.830%	14.072%	-1.242%***	-3.13	45.29
Covered by CoinMarketCap	0.590	0.524	0.066***	3.10	56.11
Token age (in years)	1.667	0.950	0.717***	11.53	60.47
Social media index	8.102	7.518	0.585***	3.62	59.56
Average number of tokens per day	2.02	684.0			

Table 3.2: Trade Performance by Purchase Time

In this table we report the "achievable" returns for each purchase timedelay combination. All the returns are calculated before taking into account exchange commissions. We consider 12 10-second intervals after the P&D starts and one "Before P&D" interval which is the 10-minute interval before a P&D. These are shown in each column. Each row displays the results by delay time. Time 0 indicates sales immediately after the 10-second interval after purchase. For a combination of purchase timedelay, we assume that an investor buys tokens at the volume-weighted average price during the purchase time interval, and starts to sell after the delay. We further assume that when the investor decides to sell, he is the only seller and can trade at the volume-weighted average price until he fully unwinds his initial purchase. In Panels AD we report the mean returns, t-values of the average returns, and the average and median times to unwind the purchase, respectively. The average returns that are significantly positive at 5% are highlighted in bold.

Panel A. Pump-and-dump events

	Purchase time, in seconds												
Delay time	Before P&D	10	20	30	40	50	60	70	80	90	100	110	120
0	**13.926**	**5.619**	**1.309**	-0.063	-0.921	-1.272	-1.184	-1.051	-1.589	-0.015	-1.375	-0.703	0.048
10	**18.387**	**6.505**	**1.594**	-0.579	-1.516	-1.849	-1.050	-2.005	-1.959	-0.708	-1.890	-0.557	-0.186
20	**18.395**	**6.534**	1.180	-0.995	-1.494	-2.683	-2.143	-2.180	-2.736	-1.066	-1.606	-0.863	-0.917
30	**18.436**	**5.405**	0.709	-1.408	-1.837	-3.638	-2.541	-2.936	-3.030	-1.021	-1.969	-1.525	-1.411
40	**17.806**	**4.659**	0.200	-2.008	-2.643	-3.719	-2.347	-3.449	-2.931	-1.558	-2.702	-2.100	-1.783
50	**16.989**	**4.195**	-0.717	-2.931	-2.903	-4.267	-2.741	-3.339	-3.509	-2.128	-3.129	-2.346	-2.430
60	**15.883**	**4.222**	-1.258	-2.775	-3.383	-4.680	-2.704	-4.028	-3.979	-2.324	-3.437	-2.824	-2.884
70	**14.964**	**3.538**	-1.368	-3.105	-3.658	-4.751	-3.266	-4.361	-4.437	-3.267	-4.111	-3.540	-3.804
80	**13.894**	**3.516**	-1.698	-3.378	-3.683	-5.253	-3.723	-5.003	-4.598	-3.790	-4.863	-4.133	-4.035
90	**12.626**	2.876	-2.119	-3.603	-4.318	-5.786	-4.651	-5.732	-5.042	-4.470	-5.255	-4.498	-4.553
100	**11.534**	2.268	-2.453	-4.099	-4.910	-6.505	-5.824	-6.222	-5.864	-4.743	-5.530	-4.932	-4.714
110	**11.087**	1.963	-3.136	-4.612	-5.535	-5.999	-6.351	-6.941	-6.211	-4.981	-6.091	-5.108	-4.778
120	**11.077**	1.320	-3.352	-5.178	-6.388	-6.467	-7.082	-7.181	-6.437	-4.893	-6.204	-5.307	-5.346
130	**10.596**	0.887	-4.130	-6.146	-5.951	-6.838	-7.229	-6.706	-6.898	-4.711	-6.466	-5.743	-5.383
140	**10.045**	0.403	-3.511	-6.416	-6.648	-6.727	-7.471	-6.726	-6.028	-5.153	-6.304	-5.899	-5.682
150	**9.974**	-0.745	-3.935	-6.974	-7.024	-6.613	-8.021	-6.689	-6.399	-5.710	-6.588	-5.644	-5.621
160	**9.368**	-1.216	-4.492	-7.394	-7.251	-6.802	-7.930	-6.912	-6.547	-5.994	-7.055	-5.558	-5.988
170	**8.724**	-1.221	-4.363	-7.542	-7.580	-6.963	-8.238	-7.496	-7.008	-5.892	-7.023	-6.076	-6.181
180	**8.213**	-1.290	-4.615	-7.653	-7.747	-7.105	-8.753	-7.910	-7.458	-5.817	-7.463	-6.429	-6.584

Panel B. t values

	Purchase time, in seconds												
Delay time	Before P&D	10	20	30	40	50	60	70	80	90	100	110	120
0	8.10	7.57	2.77	-0.11	-2.13	-2.75	-3.13	-1.72	-3.72	-0.02	-3.06	-1.39	0.11
10	8.77	7.02	2.42	-1.01	-2.71	-3.63	-0.96	-2.98	-3.58	-0.84	-3.79	-1.05	-0.36
20	8.48	6.48	1.57	-1.46	-2.26	-4.32	-1.87	-3.00	-4.43	-1.22	-2.99	-1.60	-1.70
30	8.37	5.45	0.90	-1.91	-2.04	-5.22	-2.09	-3.68	-4.52	-1.16	-3.41	-3.14	-2.57
40	8.19	4.50	0.24	-2.34	-2.76	-4.80	-1.55	-4.03	-4.13	-1.77	-4.45	-4.02	-3.22
50	7.93	3.88	-0.79	-3.23	-2.92	-5.09	-1.81	-3.92	-4.95	-2.45	-4.99	-4.45	-4.15
60	7.65	3.00	-1.30	-2.55	-3.27	-5.16	-1.80	-4.59	-5.40	-2.63	-5.37	-4.99	-4.40
70	7.16	2.43	-1.35	-2.71	-3.47	-5.08	-2.19	-4.87	-6.05	-5.22	-6.11	-5.53	-6.05
80	6.62	2.22	-1.54	-2.83	-3.51	-5.73	-2.51	-6.31	-6.18	-5.54	-6.66	-6.47	-6.11
90	6.52	1.79	-1.86	-2.99	-4.08	-6.12	-4.28	-7.59	-6.77	-6.13	-7.20	-6.82	-6.74
100	6.08	1.42	-2.13	-3.42	-4.52	-7.36	-6.64	-8.12	-7.35	-6.40	-7.36	-7.32	-6.77
110	5.80	1.24	-2.70	-3.77	-5.76	-4.64	-7.28	-8.55	-7.87	-6.58	-7.91	-7.46	-6.90
120	5.64	0.83	-2.87	-4.27	-6.71	-4.98	-8.17	-8.81	-8.04	-5.34	-7.91	-7.75	-7.70
130	5.33	0.55	-3.73	-5.86	-5.98	-5.23	-8.35	-7.90	-8.49	-4.83	-8.04	-8.15	-7.86
140	5.00	0.29	-2.35	-6.10	-6.54	-5.11	-8.76	-7.69	-6.79	-5.49	-7.58	-8.27	-8.48
150	5.03	-0.58	-2.63	-6.49	-6.95	-4.95	-9.26	-7.85	-7.30	-6.06	-8.11	-7.91	-8.18
160	4.76	-0.94	-2.97	-7.04	-7.05	-4.89	-8.79	-8.14	-7.49	-6.17	-8.73	-7.55	-8.31
170	4.56	-0.94	-2.86	-7.09	-7.42	-5.01	-9.12	-8.79	-8.11	-5.46	-8.63	-7.79	-8.57
180	4.28	-1.01	-3.05	-7.13	-7.64	-5.12	-9.74	-9.35	-8.85	-5.28	-8.94	-5.25	-8.94

Panel C. Average time taken to unwind the purchase

	Purchase time, in seconds												
Delay time	Before P&D	10	20	30	40	50	60	70	80	90	100	110	120
0	20.97	4.64	3.57	2.33	3.60	5.57	2.30	2.41	3.47	5.05	2.57	2.19	2.12
10	22.15	7.32	4.70	3.21	4.25	7.56	2.93	3.01	4.23	5.90	2.86	2.51	2.25
20	24.05	9.27	5.71	3.91	4.87	8.03	4.84	3.43	4.84	7.84	3.12	2.96	2.43
30	26.26	12.52	6.63	4.61	5.92	8.49	7.61	4.01	6.56	7.93	3.96	3.26	3.16
40	29.84	15.01	12.71	8.00	6.56	8.84	8.39	4.28	6.69	8.52	4.70	3.87	3.69
50	32.05	19.84	13.94	8.50	8.00	9.49	8.63	4.95	7.39	8.88	5.44	4.06	4.61
60	36.35	22.76	15.03	9.41	9.40	9.81	8.88	7.02	7.60	9.21	5.76	4.34	6.82
70	39.59	25.71	17.18	10.66	9.60	10.06	9.51	7.15	8.08	9.39	7.46	6.15	7.83
80	42.67	33.25	18.56	11.11	10.06	11.00	9.85	7.42	8.25	9.77	7.68	6.34	8.75
90	47.81	36.50	19.76	11.76	10.54	11.33	10.44	7.76	8.62	10.09	7.78	6.86	9.06
100	49.38	39.04	22.28	12.63	11.09	12.21	10.52	8.67	10.77	11.26	7.77	7.50	9.08
110	50.71	41.64	27.67	13.10	11.85	12.56	11.48	9.06	11.02	11.45	8.42	8.17	9.09
120	53.35	43.16	29.17	13.82	12.04	12.85	12.03	9.26	11.25	11.85	8.37	8.38	9.18
130	56.41	44.95	29.93	14.90	15.25	13.42	12.28	9.59	11.90	12.08	8.43	8.47	9.44
140	59.21	47.39	30.58	15.80	16.20	13.56	12.62	10.18	11.88	12.08	8.49	8.79	9.53
150	62.35	49.85	32.53	16.47	16.57	14.46	13.16	10.19	11.95	12.27	8.86	8.94	9.52
160	63.96	51.11	35.06	16.93	17.61	14.98	13.40	10.20	12.42	12.69	8.95	11.28	11.27
170	66.60	54.24	36.86	19.41	18.21	15.11	13.47	11.01	12.57	13.22	9.35	12.25	12.11
180	73.31	55.32	37.97	21.43	18.85	15.26	13.93	11.39	12.93	13.64	10.84	12.54	12.12

Panel D. Median time taken to unwind the purchase

	Purchase time, in seconds												
Delay time	Before P&D	10	20	30	40	50	60	70	80	90	100	110	120
0	2	1	1	1	1	1	1	1	1	1	1	1	1
10	2	1	1	1	1	1	1	1	1	1	1	1	1
20	2	2	1	1	1	1	1	1	1	1	1	1	1
30	3	2	2	1	1	1	1	1	1	1	1	1	1
40	3	2	2	2	2	2	2	1	2	1	1	1	1
50	4	3	2	2	2	2	2	2	2	1	2	1	1
60	5	4	3	3	2	2	2	2	2	2	2	2	1
70	6	4	3	3	3	2	2	2	2	2	2	2	2
80	6	5	4	3	3	3	2	2	2	2	2	2	2
90	7	5	4	4	3	3	3	2	2	2	2	2	2
100	8	5	4.5	4	4	3	3	3	3	2	2	2	2
110	8	6	5	4	4	4	3	3	3	2	3	2	2
120	9	7	6	5	4	4	3	3	3	3	3	2	2
130	10	7	6	5	4	4	4	3.5	3	3	3	2	2
140	11	8	6	5	5	4	4	4	3	3	3	3	2
150	12	8	6.5	6	5	4	4	4	3	3	3	2	2
160	13	9	7	6	5	5	4	4	3	3	3	2	2
170	14	10	7	6	6	5	4	4	4	3	3	3	3
180	14.5	10	7	7	7	5	5	4	4	3	3	3	3

Table 3.3: Persistent Advantages for Certain Telegram Channels
In this table we indicate the persistent advantages certain Telegram channels enjoy. The analysis is carried out at the channel level. In both panels, the reported results are based on analyses using two samples: one with all the P&D channels targeting Binance, Bittrex and Yobit, and the other with the channels for which trading data are available. In Panel A we report the announcement delay times for each channel, measured in seconds. A positive number indicates that the announcement occurred after the scheduled time. We winsorize the variable at ± 30 seconds. In Panel A, we report the average and median delays for each top-ten Telegram channel based on P&D events initiated. *Early* is a dummy variable equal to 1 if the announcement time is before the scheduled time and 0 otherwise. *% Early* is the fraction of P&Ds in which the announcement occurs before the scheduled time. In Panel B we report the results of our regression analysis, in which the dependent variable is either delay time or the *Early* dummy, and the independent variable is lagged delay time or lagged *Early* dummy. We include only P&Ds that are covered by more than one channel when using P&D fixed effects. In each column, we report coefficient estimates and their heteroscedasticity-robust t-statistics. *, **, and sym*** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Delay in announcement (seconds)

	All P&D channels				P&D channels with trading data			
Pump group name	N	Average	Median	% Early	N	Average	Median	% Early
Alt the Way	303	9.79	7	12.5%	62	16.21	14	4.8%
Crypto Mega Pumps Yobit	244	6.90	5	25.0%	44	13.70	9	15.9%
The Pumpers	220	7.33	4	21.8%	37	15.43	16	8.1%
Superb Pumps	180	7.05	6	20.6%	40	11.05	12	25.0%
YoBit Pumpers	159	4.45	1	39.0%	31	14.74	30	22.6%
Bittrex Signals	148	5.83	2	38.5%	25	16.56	19	16.0%
World Pumps	130	6.65	3	16.9%	7	1.86	3	14.3%
Crypto VIP Signals	116	4.99	1	38.8%	23	14.61	25	21.7%
Premium Yobit Pump	107	11.29	9	11.2%	53	10.13	6	7.5%
PumpZone	89	0.73	0	13.5%	27	0.96	1	0.0%
Other channels	788	7.15	4	27.9%	439	6.38	4	25.1%

Panel B. Persistent advantages

Dependent variable	All P&D channels				P&Ds with trading data available			
	Delay time		Early dummy		Delay time		Early dummy	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Lagged delay time	0.293***	0.152***			0.341***	0.109***		
	(15.17)	(8.62)			(9.85)	(2.99)		
Lagged Early dummy			0.302***	0.301***			0.290***	0.348***
			(15.64)	(13.46)			(8.24)	(7.09)
Constant	4.962***		0.171***		6.052***		0.136***	
	(16.62)		(17.82)		(9.74)		(8.86)	
P&D fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,433	1,901	2,433	1,901	746	377	746	377
Adj. R-squared	0.09	0.87	0.09	0.59	0.11	0.87	0.08	0.52

Table 3.4: Maximum Returns and Target Returns

In this table we report results pertaining to an analysis of factors that predict P&D returns at the event level. The dependent variable for column (1) is the ratio of the highest price achieved within 10 minutes after the pump announcement and the price at the announcement minus 1. The dependent variable for column (2) is the minimum target return among participating groups. The dependent variable for columns (3) is an indicator equal to 1 if the maximum return is greater or equal to the minimum target return among participating groups and 0 otherwise. *Pre-pump liquidity* is the logarithm of trading volume (in Bitcoin) over the period running from 37 to eight days before a pump. *Price run-up* is the token return over the period between 10 minutes before a P&D announcement and the announcement. *Multiple Telegram channels* is an indicator equal to 1 if multiple channels participate in the same event and 0 otherwise. *Delay in announcement* is the difference between an announcement and the scheduled time (winsorized at ±30 seconds). All other control variables are as defined in Table 3.1. In each column, we report coefficient estimates and their heteroscedasticity-robust t-statistics. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively

	Ordinary least squares						Linear probability model		
Dependent variable	Maximum return			Minimum target return among channels			Dummy for achieving minimum target return		
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)
Pre-pump liquidity	-0.114***	-0.123***	-0.107***	-0.073***	-0.073***	-0.061*	-0.025**	-0.016	-0.009
	(-7.04)	(-7.19)	(-5.61)	(-2.85)	(-2.61)	(-1.86)	(-2.27)	(-1.54)	(-0.58)
Price run-up	-0.903***	-0.767***	0.023	-0.584	0.497***	0.460**			
		(-4.24)	(-3.95)		(0.05)	(-1.16)		(2.66)	(2.02)
Multiple Telegram channels			0.824***			1.038**			0.159
			(5.00)			(2.31)			(1.14)
Token age			-0.062**			-0.130			0.076***
			(-2.17)			(-1.53)			(3.09)
Social media index			0.076*			-0.018			-0.015
			(1.69)			(-0.17)			(-0.30)
Delay in announcement			0.003			0.018***			0.003
			(1.06)			(2.58)			(0.78)
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	500	500	500	229	229	229	229	229	229
Adj. R-squared	0.19	0.22	0.30	0.45	0.45	0.47	0.23	0.27	0.31
% (Dep variable = 1)							31.4%	31.4%	31.4%

Table 3.5: Viewership

In this table we report results pertaining to factors that predict changes in the number of Telegram viewers at the channel level. *Average recent pump return* is average maximum returns after a channel's three most recent P&D announcements. *Average recent abnormal volume* is the average abnormal volume in the three most recent P&Ds, in which abnormal volume equals log(1 + volume over a [-10 minutes, +10 minutes]/average volume in the interval running from day -37 to day -8). All other control variables are as defined in Table 3.4. In each column, we report coefficient estimates and their t-statistics. Standard errors are clustered at the channel and month levels. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable	ΔLog (No. of channel viewers)				
	(1)	(2)	(3)	(4)	(5)
Average recent pump return	-0.010	-0.009	-0.009	-0.009	-0.010
	(-1.05)	(-0.92)	(-0.84)	(-0.84)	(-1.04)
Average recent abnormal volume		0.011	0.011	0.011	0.013
		(0.98)	(1.00)	(1.00)	(1.11)
Pre-pump liquidity			0.002	0.002	-0.004
			(0.29)	(0.29)	(-0.32)
Price run-up				0.007	-0.012
				(0.09)	(-0.15)
Multiple Telegram channels					-0.071**
					(-2.46)
Token age					0.012
					(0.75)
Social media index					0.009
					(0.60)
Delay in announcement					0.002
					(1.36)
Month fixed effects	Yes	Yes	Yes	Yes	Yes
Channel fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	720	720	720	720	720
Adj. R-squared	0.05	0.06	0.06	0.06	0.07

Table 3.6: The Bittrex Ban

In this table we report the results of a difference-in-differences estimation of the Bittrex ban. The treatment group comprises tokens listed on Bittrex on November 24, 2017, and the control group includes matched tokens not listed on Bittrex at that time. We use the non-Bittrex token with the closest trading volume (in Bitcoin) as the matched token. Matching is based on the trading volume over the period running from 60 days to 31 days before the ban. In the regression, each token has two observations: one for the pre-ban period and one for the post-ban period. For volume, we take the average volume over the period running from day -14 to day -1 for the pre-ban period, and from day 1 to day 30 for the post-ban period. The pre-ban period price is measured at day -14, and the post-ban period price is measured at day 30. Volume is the natural logarithm of the ratio between volume on day t and the average volume over the period running from 60 days to 31 days before the ban. Price is defined as the natural logarithm of the ratio of the end-of-day price to the price at the end of day -31. All standard errors are clustered at the treatmentcontrol pair level.

Dependent variable	Volume		Price	
	(1)	(2)	(3)	(4)
Bittrex × Post	0.639***	0.639***	0.401***	0.401***
	(5.25)	(5.26)	(3.88)	(3.89)
Post	0.880***	0.880***	0.734***	0.734***
	(8.34)	(8.35)	(7.65)	(7.66)
Bittrex		-0.311**		-0.033
		(-2.15)		(-0.49)
Constant	0.096	0.033		
	(0.93)	(0.69)		
Token Fixed Effects	No	Yes	No	Yes
Observations	754	754	750	750
Adj. R-squared	0.14	0.78	0.23	0.60

Appendix

Part A1. Matching Cryptocurrencies across Data Sets

Both our exchange data and the CoinMarketCap data use token symbols to identify unique tokens. In most P&Ds, the Telegram channels provide token symbols and often token names as well. In most cases, a token has the same symbol on the exchange (or exchanges if it is cross-listed) and on CoinMarketCap. In some cases, however, a token can have a different symbol on the exchanges than it has on CoinMarketCap. It is also possible that two tokens have the same ticker on different exchanges. We manually matched these tokens.

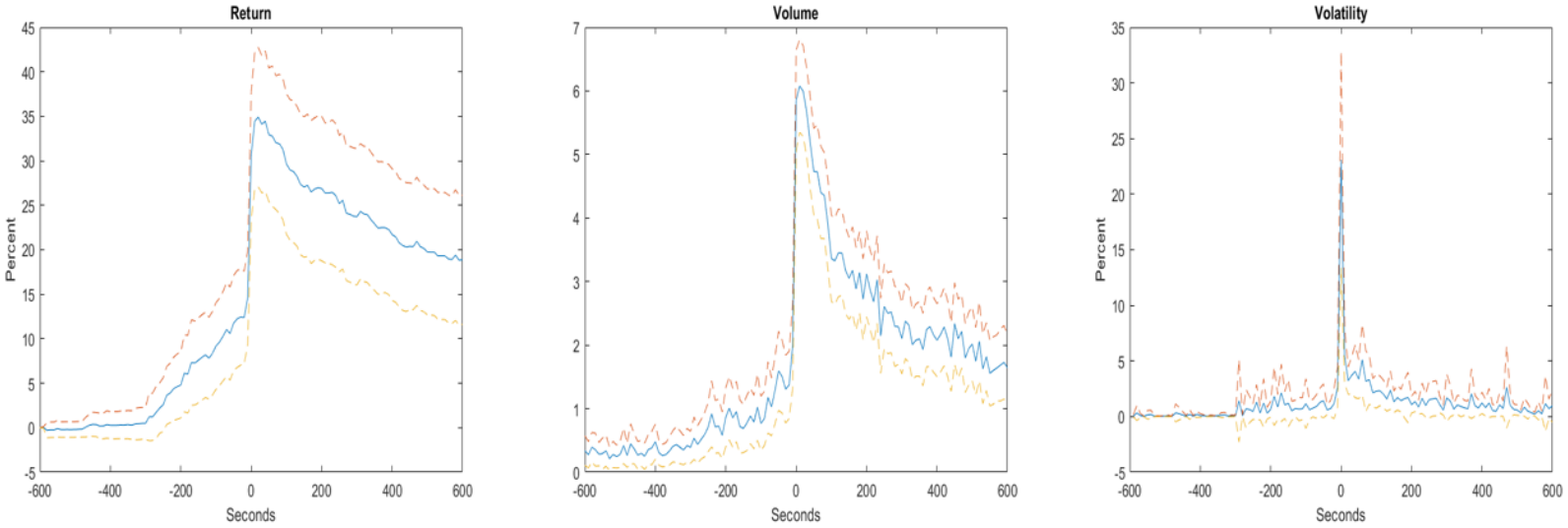
CoinMarketCap provides both token symbols and names. However, the exchange data include only token symbols but not their names. We search for token names using their symbols on the websites of the exchanges. Unlike the application programming interfaces (APIs) that we use to download trading data for both active and inactive tokens, the websites of the three exchanges (Binance, Bittrex, and Yobit) do not return names of delisted tokens. We therefore rely on Binances Announcements page (https://support.binance.com/hc/en-us/categories/115000056351-Announcements), Bittrexs Coin Removals page (https://support.bittrex.com/hc/en-us/sections/200560334-Coin-Removals), and Yobits press releases to identify such names.

Token prices are quoted in ticker pairs that function much like exchange rates. For example, NXS/BTC is a pair that indicates that the trade is an exchange between NXS (Nexus) and BTC (Bitcoin). In this pair, Nexus is quoted in the volume in Bitcoin, which is the base currency. Bitcoin is by far the dominant base currency. The second most widely used base currency is Ethereum (ETH). The symbols in our exchange data

contain token tickers and base currencies. Based on this, we filter out pairs with base currencies other than BTC and ETH.

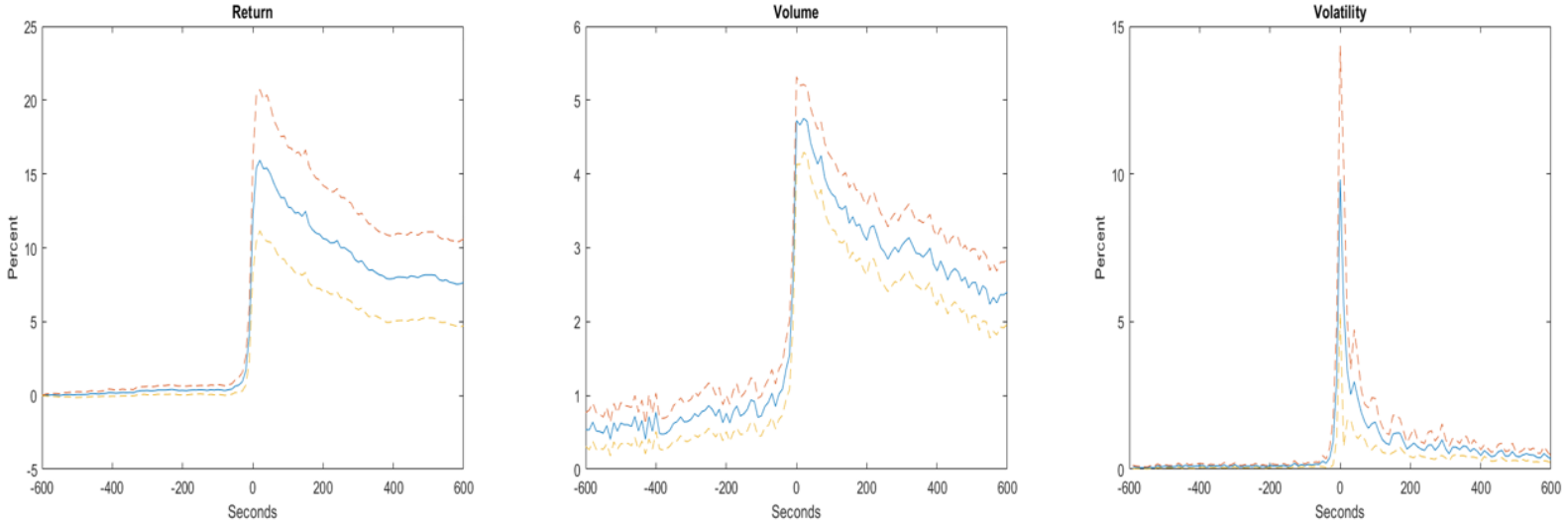
Part A2: Pump-and-Dump Schemes in the Past Six Months

This figure displays cumulative returns, abnormal volume, and volatility for the P&Ds that occurred in the half-year period from February 27, 2018 to August 26, 2018. The last P&D event in our sample took place on August 26, 2018. Our sample includes 146 P&D events in total.



Part A3: Pump-and-Dumps on Binance

This figure presents abnormal returns, abnormal volume, and volatility for the P&Ds that targeted Binance. In total, we have 76 P&D events.



Part A4. Spillovers to Ethereum Trading Pairs and Other Exchanges

This table reports results pertaining to potential spillovers. The figures in Panel A show whether there is any spillover to trading pairs in Ethereum for the same pumped tokens, while those in Panel B represent trading pairs in Bitcoin on other exchanges. In Panel A, for pumped tokens with trading data on Ethereum pairs, we report maximum returns, abnormal volume, and the volatility of the Ethereum pairs. In Panel B, for pumped tokens with trading data from other exchange(s), we report maximum returns, abnormal volume, and the volatility of these cross-listed tokens on the non-targeted exchange(s). If a token is cross-listed on two other exchanges besides the target exchange, we take the average of maximum returns, abnormal volume, and volatility across the two exchanges. The t-values and the Wilcoxon p-values are calculated based on the paired sample.

Panel A. Spillover to Etherium pairs

	Bitcoin pair	Ethereum pair	Difference	t-value	Wilcoxon p-value
Maximum return	22.79	23.07	-0.28	-0.14	0.86
Abnormal volume	1.92	1.32	0.61	9.58	<0.01
Volatility	0.66 0.69	-0.03	-1.06	0.66	
Number of P&Ds	92	92			

Panel B. Spillovers to other exchanges

	Targeted exchange	Other exchange(s)	Difference	t-value	Wilcoxon p-value
Maximum return	26.84	5.15	21.69	6.01	<0.01
Abnormal volume	1.62	0.31	1.31	12.59	<0.01
Volatility	0.79	0.12	0.67	5.60	<0.01
Number of P&Ds	89	89			

Bibliography

Aggarwal, Rajesh K and Guojun Wu, "Stock Market Manipulations," *Journal of Business*, 2006, *79* (4), 1915–1953.

Aït-Sahalia, Yacine and Andrew W Lo, "Nonparametric Estimation of State-price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 1998, *53* (2), 499–547.

Akerlof, George A, "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Allen, Franklin and Douglas Gale, "Stock-Price Manipulation," *Review of Financial Studies*, 1992, *5* (3), 503–529.

_ **and** _ , *Financial Innovation and Risk Sharing*, MIT press, 1994.

Amromin, Gene and Steven A Sharpe, "Expectations of Risk and Return Among Household Investors: Are Their Sharpe Ratios Countercyclical?," *Working paper*, 2009.

Amsden, Ryan and Denis Schweizer, "Are Blockchain Crowdsales the New 'Gold Rush'? Success Determinants of Initial Coin Offerings," *Working paper*, 2018.

Arnold, Marc, Dustin Robert Schuette, and Alexander F Wagner, "Neglected Risk in Financial Innovation: Evidence from Structured Product Counterparty Exposure," *Working paper*, 2018.

Athey, Susan, Ivo Parashkevov, Vishnu Sarukkai, and Jing Xia, "Bitcoin Pricing, Adoption, and Usage: Theory and Evidence," *Working paper*, 2016.

Atkins, Ralph, "Switzerland Sets Out Guidelines to Support Initial Coin Offerings," *Financial Times*, 2018.

Balakrishnan, Karthik and Daniel A Cohen, "Competition and Financial Accounting Misreporting," *Working paper*, 2013.

Barber, Brad M and Terrance Odean, "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance*, 2000, *55* (2), 773–806.

_ **and** _ , "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics*, 2001, *116* (1), 261–292.

Barberis, Nicholas, Abhiroop Mukherjee, and Baolian Wang, "Prospect Theory and Stock Returns: An Empirical Test," *Review of Financial Studies*, 2016, *29* (11), 3068–3107.

_ **and Ming Huang**, "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices," *American Economic Review*, 2008, *98* (5), 2066–2100.

Barberis, Nicholas C, "Thirty Years of Prospect Theory in Economics: A Review and Assessment," *Journal of Economic Perspectives*, 2013, *27* (1), 173–96.

Barberis, Nicholas C., "Psychology-Based Models of Asset Prices and Trading Volume," *Working paper*, 2018.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, "Extrapolation and Bubbles," *Journal of Financial Economics*, 2018.

Bauer, Janis, Holger Fink, and Eva Stoller, "Are Gross Margins of Structured Products Priced in a Market-Consistent Way? Evidence From the New Issuer Estimated Value," *Working paper*, 2016.

Beatty, Randolph P and Jay R Ritter, "Investment Banking, Reputation, and the Underpricing of Initial Public Offerings," *Journal of Financial Economics*, 1986, *15* (1-2), 213–232.

Becker, Bo and Victoria Ivashina, "Reaching for Yield in the Bond Market," *Journal of Finance*, 2015, *70* (5), 1863–1902.

Ben-David, ITZHAK, Francesco Franzoni, Augustin Landier, and Rabih Moussawi, "Do Hedge Funds Manipulate Stock Prices?," *Journal of Finance*, 2013, *68* (6), 2383–2434.

Benveniste, Lawrence M and Paul A Spindt, "How Investment Bankers Determine the Offer Price and Allocation of New Issues," *Journal of Financial Economics*, 1989, *24* (2), 343–361.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Biais, Bruno, Christophe Bisiere, Matthieu Bouvard, and Catherine Casamatta, "The Blockchain Folk Theorem," *Working paper*, 2018.

_ **, Peter Bossaerts, and Jean-Charles Rochet**, "An Optimal IPO Mechanism," *Review of Economic Studies*, 2002, *69* (1), 117–146.

Binsbergen, Jules Van, Michael Brandt, and Ralph Koijen, "On the Timing and Pricing of Dividends," *American Economic Review*, 2012, *102* (4), 1596–1618.

Black, Fischer and Myron Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 1973, *81* (3), 637–654.

Böhme, Rainer, Nicolas Christin, Benjamin Edelman, and Tyler Moore, "Bitcoin: Economics, Technology, and Governance," *Journal of Economic Perspectives*, 2015, *29* (2), 213–38.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience Theory of Choice under Risk," *Quarterly Journal of Economics*, 2012, *127* (3), 1243–1285.

_ , _ , and _ , "Salience and Asset Prices," *American Economic Review Papers & Proceedings*, 2013, *103* (3), 623–28.

_ , _ , and _ , "Competition for Attention," *Review of Economic Studies*, 2015, *83* (2), 481–513.

Brav, Alon, Wei Jiang, and Tao Li, "Picking friends before picking (proxy) fights: How mutual fund voting shapes proxy contests," *Working paper*, 2018.

Brennan, Michael J and Avanidhar Subrahmanyam, "Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns," *Journal of Financial Economics*, 1996, *41* (3), 441–464.

Brunnermeier, Markus and Martin Oehmke, "Complexity in Financial Markets," *Working paper*, 2009.

Calvet, Laurent, Claire Célérier, Paolo Sodini, and Boris Vallée, "Can Financial Innovation Solve Household Reluctance to Take Risk," *Working paper*, 2017.

Cameron, Matt, "Uridashi Losses Put at 500 Million Dollars after Nikkei Rebounds," *Risk.Net*, 2013.

Canidio, Andrea, "Financial Incentives for Open Source Development: the Case of Blockchain," *Working paper*, 2018.

Carlin, Bruce I, "Strategic Price Complexity in Retail Financial Markets," *Journal of Financial Economics*, 2009, *91* (3), 278–287.

_ and Gustavo Manso, "Obfuscation, Learning, and the Evolution of Investor Sophistication," *Review of Financial Studies*, 2010, *24* (3), 754–785.

Catalini, Christian and Joshua S. Gans, "Initial Coin Offerings and the Value of Crypto Tokens," *Working paper*, 2018.

_ and Joshua S Gans, "Initial Coin Offerings and the Value of Crypto Tokens," Technical Report, National Bureau of Economic Research 2018.

Célérier, Claire and Boris Vallée, "Catering to Investors through Security Design: Headline Rate and Complexity," *Quarterly Journal of Economics*, 2017, *132* (3), 1469–1508.

Cha, Kyung-Wook and Joo hyun Park, "Analysis of ELS Investment Behavior and ELS Types: Principal Guaranteed vs. Non-Guaranteed ELS (in Korean)," *Journal of Consumer Studies,*, 2015, *28* (2), 227–250.

Chen, Mark A, Qinxi Wu, and Baozhong Yang, "How Valuable is FinTech Innovation?," *Review of Financial Studies, Forthcoming*, 2018.

Chen, Ting, Zhenyu Gao, Jibao He, Wenxi Jiang, and Wei Xiong, "Daily Price Limits and Destructive Market Behavior," *Journal of Econometrics*, 2019, *208* (1), 249–264.

Chevalier, Judith and Glenn Ellison, "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, 1997, *105* (6), 1167–1200.

Chod, Jiri and Evgeny Lyandres, "A theory of ICOs: Diversification, Agency, and Information Asymmetry," *Working paper*, 2018.

Choi, Jaewon and Mathias Kronlund, "Reaching for Yield in Corporate Bond Mutual Funds," *Review of Financial Studies*, 2017, *31* (5), 1930–1965.

Ciaian, Pavel, Miroslava Rajcaniova, and dArtis Kancs, "The Economics of BitCoin Price Formation," *Applied Economics*, 2016, *48* (19), 1799–1815.

Comerton-Forde, Carole and Tālis J Putniņš, "Dark Trading and Price Discovery," *Journal of Financial Economics*, 2015, *118* (1), 70–92.

Cong, Lin William and Yizhou Xiao, "Up-Cascaded Wisdom of the Crowd," *Working paper*, 2018.

_ **and Zhiguo He**, "Blockchain Disruption and Smart Contracts," *Review of Financial Studies, Forthcoming*, 2018.

_ **, Ye Li, and Neng Wang**, "Tokenomics: Dynamic Adoption and Valuation," *Working paper*, 2018.

_ **, Zhiguo He, and Jiasun Li**, "Decentralized Mining in Centralized Pools," *Working paper*, 2019.

Coval, Joshua, Jakub Jurek, and Erik Stafford, "The Economics of Structured Finance," *Journal of Economic Perspectives*, 2009, *23* (1), 3–25.

Cox, David R., "Regression Models and Life-Tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1972, *34* (2), 187–202.

Credit Suisse Equity Derivative Strategy, "Systematic Trading Strategies: Extracting The Dividend Risk Premium," *Credit Suisse*, 2013.

_ , "European Equity Volatility Outlook: The Mighty European Autocall Market," *Credit Suisse*, 2014.

Da, Zhi and Xing Huang, "Harnessing the Wisdom of Crowds," *Working paper*, 2016.

Damodaran, Aswath, "Equity Risk Premiums (ERP): Determinants, Estimation and Implications - The 2018 Edition," *Working paper*, 2018.

Daniel, Kent and David Hirshleifer, "Overconfident Investors, Predictable Returns, and Excessive Trading," *Journal of Economic Perspectives*, 2015, *29* (4), 61–88.

_ , _ , and Avanidhar Subrahmanyam, "Investor Psychology and Security Market Under-and Overreactions," *Journal of Finance*, 1998, *53* (6), 1839–1885.

Derman, Emanuel and Iraj Kani, "The Ins and Outs of Barrier Options: Part 2," *Derivatives Quarterly*, 1997, *3*, 73–80.

Deutscher Derivate Verband, "Are Structured Products Too Expensive? : Facts and Figures," *Working paper*, 2015.

Diamond, Douglas W, "Financial Intermediation and Delegated Monitoring," *The Review of Economic Studies*, 1984, *51* (3), 393–414.

Dindo, Pietro and Filippo Massari, "The Wisdom of the Crowd Revisited," *Working paper*, 2017.

Döhrer, Björn, Lutz Johanning, Nils Steiner, and Arndt Völkle, "Issuer Margins for Structured Products in Germany," *Working paper*, 2013.

Downes, David H and Robert Heinkel, "Signaling and the Valuation of Unseasoned New Issues," *Journal of Finance*, 1982, *37* (1), 1–10.

Duffie, Darrell and Rohit Rahi, "Financial Market Innovation and Security Design: An Introduction," *Journal of Economic Theory*, 1995, *65* (1), 1–42.

Easley, David, Maureen O'Hara, and Soumya Basu, "From Mining to Markets: The Evolution of Bitcoin Transaction Fees," *Journal of Financial Economics*, 2019.

Egan, Mark, "Brokers vs Retail Investors : Conflicting Interests and Dominated Products," *Working paper*, 2018.

_ , Gregor Matvos, and Amit Seru, "The Market for Financial Adviser Misconduct," *Working paper*, 2016.

Eyal, Ittay and Emin Gün Sirer, "Majority is Not Enough: Bitcoin Mining is Vulnerable," *Communications of the ACM*, 2018, *61* (7), 95–102.

Fama, Eugene F, "Two Pillars of Asset Pricing," *American Economic Review*, 2014, *104* (6), 1467–85.

Feder, Amir, Neil Gandal, JT Hamrick, Tyler Moore, Arghya Mukherjee, Farhang Rouhi, Marie Vasek et al., "The Economics of Cryptocurrency Pump and Dump Schemes," *Working paper*, 2018.

Fernández-Villaverde, Jesús and Daniel Sanches, "Can Currency Competition Work?," Technical Report, National Bureau of Economic Research 2016.

Financial Supervisory Service, "A Survey on ELS Investors," 2010.

Foley, Sean, Jonathan R Karlsen, and Tālis J Putniņš, "Sex, Drugs, and Bitcoin: How Much Illegal Activity is Financed through Cryptocurrencies?," *Review of Financial Studies, Forthcoming*, 2018.

Frazzini, Andrea, "The Disposition Effect and Underreaction to News," *Journal of Finance*, 2006, *61* (4), 2017–2046.

Friedman, Milton, "Essays in Positive Economics," 1953.

Gabaix, Xavier and David Laibson, "Shrouded Attributes, Consumer myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

Gandal, Neil and Hanna Halaburda, "Competition in the Cryptocurrency Market," *Working paper*, 2014.

_ , JT Hamrick, Tyler Moore, and Tali Oberman, "Price Manipulation in the Bitcoin Ecosystem," *Journal of Monetary Economics*, 2018, *95*, 86–96.

Gans, Joshua S. and Hanna Halaburda, "Some Economics of Private Digital Currency," in "Economic Analysis of the Digital Economy," University of Chicago Press, 2015, pp. 257–276.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, "Neglected Risks, Financial Innovation, and Financial Fragility," *Journal of Financial Economics*, 2012, *104* (3), 452–468.

Gervais, Simon and Terrance Odean, "Learning to be Overconfident," *Review of financial studies*, 2001, *14* (1), 1–27.

Ghent, Andra C, Walter N Torous, and Rossen Valkanov, "Complexity in Structured Finance," *Review of Economic Studies*, 2017, pp. 1–29.

Giannetti, Mariassunta and Tracy Yue Wang, "Corporate Scandals and Household Stock Market Participation," *Journal of Finance*, 2016, *71* (6), 2591–2636.

Glosten, Lawrence R and Lawrence E Harris, "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics*, 1988, *21* (1), 123–142.

Greenwood, Robin and Andrei Shleifer, "Expectations of Returns and Expected Returns," *Review of Financial Studies*, 2014, *27* (3), 714–746.

_ , _ , and Yang You, "Bubbles for Fama," *Journal of Financial Economics*, 2019, *131* (1), 20–43.

Griffin, John M and Amin Shams, "Manipulation in the VIX?," *Review of Financial Studies*, 2017, *31* (4), 1377–1417.

_ and _ , "Is Bitcoin Really Un-Tethered?," *Working paper*, 2018.

Gurun, Umit G, Noah Stoffman, and Scott E Yonker, "Trust Busting: The Effect of Fraud on Investor Behavior," *Review of Financial Studies*, 2017, *31* (4), 1341–1376.

Habib, Michel A. and Alexander P. Ljungqvist, "Underpricing and Entrepreneurial Wealth Losses in IPOs: Theory and Evidence," *Review of Financial Studies*, 2001, *14* (2), 433–458.

Hansson, Fredrik, "A Pricing and Performance Study on Auto-callable Structured Products," 2012.

Harvey, Campbell R., "Cryptofinance," *Working paper*, 2016.

Henderson, Brian J and Neil D Pearson, "The Dark Side of Financial Innovation: A Case Study of the Pricing of a Retail Financial Product," *Journal of Financial Economics*, 2011, *100* (2), 227–247.

Heston, Steven L, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 1993, *6* (2), 327–343.

Hong, Harrison and Jeremy C Stein, "Disagreement and the Stock ¡arket," *Journal of Economic Perspectives*, 2007, *21* (2), 109–128.

_ and Marcin Kacperczyk, "Competition and Bias," *Quarterly Journal of Economics*, 2010, *125* (4), 1683–1725.

Hou, Kewei and Roger K Loh, "Have We Solved the Idiosyncratic Volatility Puzzle?," *Journal of Financial Economics*, 2016, *121* (1), 167–194.

Howell, Sabrina T., Marina Niessner, and David Yermack, "Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales," *Working paper*, 2018.

Hu, Albert, Christine A Parlour, and Uday Rajan, "Cryptocurrencies: Stylized Facts on a New Investible Instrument," *Working paper*, 2018.

Huberman, Gur and Werner Stanzl, "Price Manipulation and Quasi-Arbitrage," *Econometrica*, 2004, *72* (4), 1247–1275.

_ , Jacob Leshno, and Ciamac C. Moallemi, "Monopoly Without a Monopolist: An Economic Analysis of the Bitcoin Payment System," *Working paper*, 2017.

Hull, John C., *Option, Futures and Other Derivatives, 8th edition*, Prentice-Hall, 2011.

Hunt, Stefan, Neil Stewart, and Redis Zaliauskas, "Two Plus Two Makes Five? Survey Evidence that Investors Overvalue Structured Deposits," *Working paper*, 2015.

Ibbotson, Roger G. and Jeffrey F. Jaffe, ""Hot issue" markets," *Journal of Finance*, 1975, *30* (4), 1027–1042.

Jarrow, Robert A, "Market Manipulation, Bubbles, Corners, and Short Squeezes," *Journal of Financial and Quantitative Analysis*, 1992, *27* (3), 311–336.

Jia, Chunxin, Jay R. Ritter, Zhen Xie, and Donghang Zhang, "Pre-IPO Analyst Coverage: Hype or Information Production?," *Working paper*, 2018.

Jiang, Guolin, Paul G Mahoney, and Jianping Mei, "Market Manipulation: A Comprehensive Study of Stock Pools," *Journal of Financial Economics*, 2005, *77* (1), 147–170.

Jr, John J Merrick, Narayan Y Naik, and Pradeep K Yadav, "Strategic Trading Behavior and Price Distortion in a Manipulated Market: Anatomy of a Squeeze," *Journal of Financial Economics*, 2005, *77* (1), 171–218.

Khwaja, Asim Ijaz and Atif Mian, "Unchecked Intermediaries: Price Manipulation in an Emerging Stock Market," *Journal of Financial Economics*, 2005, *78* (1), 203–241.

Kovbasyuk, Sergei, "Wisdom of the Crowd," *Working paper*, 2011.

Kremer, Ilan, Yishay Mansour, and Motty Perry, "Implementing the "Wisdom of the Crowd"," *Journal of Political Economy*, 2014, *122* (5), 988–1012.

Kumar, Alok, "Who Gambles in the Stock Market?," *Journal of Finance*, 2009, *64* (4), 1889–1933.

Kyle, Albert S and S Viswanathan, "How to Define Illegal Price Manipulation," *American Economic Review*, 2008, *98* (2), 274–79.

Lee, Charles MC and Mark J Ready, "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 1991, *46* (2), 733–746.

Lee, Jongsub, Tao Li, and Donghwa Shin, "The Wisdom of Crowds in FinTech: Evidence from Initial Coin Offerings," *Working paper*, 2018.

Leinz, Kailey, "A Look at Who Owns Bitcoin (Young Men), and Why (Lack of Trust)," *Bloomberg*, 2018.

Leland, Hayne E. and David H. Pyle, "Informational Asymmetries, Financial Structure, and Financial Intermediation," *Journal of Finance*, 1977, *32* (2), 371–387.

Leuz, Christian, Steffen Meyer, Maximilian Muhn, Eugene Soltes, and Andreas Hackethal, "Who Falls Prey to the Wolf of Wall Street? Investor Participation in Market Manipulation," *Working paper*, 2017.

Li, Jiasun and William Mann, "Initial Coin Offering and Platform Building," *Working paper*, 2018.

Liu, Yukun and Aleh Tsyvinski, "Risks and Returns of Cryptocurrency," *Working paper*, 2018.

Ljungqvist, Alexander, "IPO underpricing," in "Handbook of Empirical Corporate Finance," Elsevier, 2007, pp. 375–422.

Mackay, Charles, "Extraordinary Popular Delusions and the Madness of Crowds. Harmony Books," 1984.

Mahoney, Paul G, "The Stock Pools and the Securities Exchange Act," *Journal of Financial Economics*, 1999, *51* (3), 343–369.

Makarov, Igor and Antoinette Schoar, "Trading and Arbitrage in Cryptocurrency Markets," *Journal of Financial Economics, Forthcoming*, 2019.

Malinova, Katya and Andreas Park, "Market Design with Blockchain Technology," *Working paper*, 2017.

Mei, Jianping, Guojun Wu, and Chunsheng Zhou, "Behavior Based Manipulation: Theory and Prosecution Evidence," *Working paper*, 2004.

Merton, Robert C, "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 1973, pp. 141–183.

Mironov, Mikhail, "ICO Market Research Q1 2018," *ICORating.com*, 2018.

_ **and Steven Campbell**, "ICO Market Research Q1 2018," *ICORating.com*, 2018.

Mollenkamp, Carrick, "Bankers Cast Doubt on Key Rate Amid Crisis," 2008.

Momtaz, Paul P., "Initial Coin Offerings," *Working paper*, 2018.

Muller, Sigrid, Lutz Johanning, Christian Koziol, Dirk Schiereck, and Bernd Rudolph, "Total Costs and Cost Components of Investment in Retail Structured Products," *Working paper*, 2017.

Ni, Sophie Xiaoyan, Neil D Pearson, and Allen M Poteshman, "Stock Price Clustering on Option Expiration Dates," *Journal of Financial Economics*, 2005, *78* (1), 49–87.

Pagnotta, Emiliano and Andrea Buraschi, "An Equilibrium Valuation of Bitcoin and Decentralized Network Assets," *Working paper*, 2018.

Pozzi, Daniele, "ICO Market 2018 vs 2017: Trends, Capitalization, Localization, Industries, Success Rate," *Cointelegraph*, 2019.

Putniņš, Tālis J, "Market Manipulation: A Survey," *Journal of Economic Surveys*, 2012, *26* (5), 952–967.

PwC, "Initial Coin Offerings: A Strategic Perspective," 2018.

Research, EY, "Initial Coin Offerings (ICOs)," 2017.

Ritter, Jay R., "The "Hot Issue" Market of 1980," *Journal of Business*, 1984, pp. 215–240.

_ , "The Long-Run Performance of Initial Public Offerings," *Journal of Dinance*, 1991, *46* (1), 3–27.

Ritter, Jay R, "Initial Public Offerings: Updated Statistics," *Working paper*, 2018.

_ **and Ivo Welch**, "A Review of IPO Activity, Pricing, and Allocations," *Journal of Finance*, 2002, *57* (4), 1795–1828.

Rock, Kevin, "Why New Issues Are Underpriced," *Journal of Financial Economics*, 1986, *15* (1-2), 187–212.

Sagar, Jayanand, "Why Generous ICO Bonuses May Indicate Scam, November 23, 2017," *NEWSBTC*, 2017.

Scheinkman, Jose A and Wei Xiong, "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 2003, *111* (6), 1183–1220.

Securities and Futures Commission, "Structured Product Investor Survey," 2006.

Sehra, Avtar, Philip Smith, and Phil Gomes, "Economics of Initial Coin Offerings," *Working paper*, 2017.

Shifflet, Shane and Paul Vigna, "Traders are Talking Up Cryptocurrencies, then Dumping Them, Costing Others Millions," *Wall Street Journal*, 2018.

Shiller, Robert, "Irrational Exuberance," *Princeton University Press*, 2000.

Simsek, Alp, "Speculation and Risk Sharing with New Financial Assets," *Quarterly Journal of Economics*, 2013, *128* (3), 1365–1396.

Sirri, Erik R and Peter Tufano, "Costly Search and Mutual Fund Flows," *Journal of Finance*, 1998, *53* (5), 1589–1622.

Snider, Connan Andrew and Thomas Youle, "Does the LIBOR Reflect Banks' Borrowing Costs?," *Working paper*, 2010.

Sockin, Michael and Wei Xiong, "A Model of Cryptocurrencies," *Working paper*, 2018.

Starr, Amy M., "Structured Products - Complexity and Disclosure - Do Retail Investors Really Understand What They are Buying and What the Risks Are?," *U.S. Security and Exchange Commission*, 2015.

Stiglitz, Joseph E. and Andrew Weiss, "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, 1981, *71* (3), 393–410.

Stoimenov, Pavel A and Sascha Wilkens, "Are Structured Products 'Fairly' Priced? An Analysis of the German Market for Equity-Linked Instruments," *Journal of Banking and Finance*, 2005, *29* (12), 2971–2993.

Strausz, Roland, "A Theory of Crowdfunding: A Mechanism Design Approach with Demand Uncertainty and Moral Hazard," *American Economic Review*, 2017, *107* (6), 1430–76.

Surowiecki, James, "The Wisdom of Crowds," 2005.

Szymanowska, Marta, Jenke Ter Horst, and Chris Veld, "Reverse Convertible Bonds Analyzed," *Journal of Futures Markets*, 2009, *29* (10), 895–919.

Tirole, Jean, "Asset Bubbles and Overlapping Generations," *Econometrica*, 1985, pp. 1499–1528.

Vissing-Jorgensen, Annette, "Perspectives on Behavioral Finance: Does Irrationality Disappear with Wealth? Evidence from Expectations and Actions," *NBER Macroeconomics Annual*, 2003, *18*, 139–194.

Wallmeier, Martin and Martin Diethelm, "Market Pricing of Exotic Structured Products: The Case of Multi-Asset Barrier Reverse Convertibles in Switzerland," *Working paper*, 2008.

Welch, Ivo, "Sequential Sales, Learning, and Cascades," *Journal of Finance*, 1992, *47* (2), 695–732.

Williams-Grut, Oscar, ""Market Manipulation 101": "Wolf of Wall Street"-Style "Pump and Dump" Scams Plague Cryptocurrency Markets," *Business Insider*, 2017.

Xiong, Wei and Jialin Yu, "The Chinese Warrants Bubble," *American Economic Review*, 2011, *101* (6), 2723–53.

Xu, Jiahua and Benjamin Livshits, "The Anatomy of a Cryptocurrency Pump-and-Dump Scheme," *Working paper*, 2018.

Xu, Ting, "The Informational Role of Crowdfunding," *Working paper*, 2016.

Yermack, David, "Corporate Governance and Blockchains," *Review of Finance*, 2017, *21* (1), 7–31.