THE UNIVERSITY OF TULSA

THE GRADUATE SCHOOL

PRICE MANIPULATION IN THE CRYPTOCURRENCY ECOSYSTEM

by
JT Hamrick

A dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in the Discipline of Computer Science

The Graduate School
The University of Tulsa

2020

THE UNIVERSITY OF TULSA

THE GRADUATE SCHOOL

PRICE MANIPULATION IN THE

CRYPTOCURRENCY ECOSYSTEM

by
JT Hamrick

A DISSERTATION

APPROVED FOR THE DISCIPLINE OF

COMPUTER SCIENCE

By Dissertation Committee

Tyler Moore, Chair
Mauricio Papa
John Hale
Neil Gandal
Nicolas Christin

ABSTRACT

JT Hamrick (Doctor of Philosophy in Computer Science)

PRICE MANIPULATION IN THE CRYPTOCURRENCY ECOSYSTEM

Directed by Tyler Moore

164 pp., Chapter 10: Conclusions

(355 words)

In the recent past the cryptocurrency ecosystem has seen explosive growth. What was once a small market, made popular with the introduction of Bitcoin in 2009, has now grown to a pseudo-financial system with a market capitalization of over 230 billion USD. Many of the cryptocurrencies available have been built with security in mind, utilizing a public ledger as well as safe-guards against counterfeiting. One might be led to believe that the public nature of cryptocurrencies would make them an infrequent target for criminal activity. However, that is not entirely the case. Many of the services utilized by the cryptocurrency ecosystem are centralized utilities that store transactions in databases that, while partially made public through data endpoints, are not attached to the blockchain. This dissertation leverages public data, typically through off-chain utilities, to study cybercriminal activities. First, the impact that distributed denial-of-services attacks have on the Bitcoin ecosystem was investigated by examining trading activity on the Mt. Gox exchange. It was found that there are fewer extremely large trades on days following a shock. Next the suspicious trading activity of two actors found trading on the Mt. Gox exchange was analyzed. Between February 2013 to November 2013 these two actors were able to fraudulently purchase around 600,000 bitcoin worth approximately 112 million USD. The trading activity associated with one of the actors was found to be highly correlated with the impressive BTC/USD exchange rate increase at the end of 2013. Following the fall of Mt. Gox, finding out if cryptocurrencies fail in ways similar to the services they support was of interest. Examining 1,082 coins over nearly a five year period, it was found that 44% of publicly-traded coins are abandoned at least once with 18% of coin abandonments being permanent. In addition to the coins, 725 tokens were analyzed and it was found that 7% were abandoned at least once and 5% were abandoned permanently. Finally, the scope of cryptocurrency "pump-and-dump" schemes was quantified through groups found on Discord and Telegram, two popular messaging platforms. Nearly 5,000 different "pumps" were discovered over a six month period in early 2018, suggesting this phenomenon is widespread.

ACKNOWLEDGEMENTS

To those who were there for me, thank you.

TABLE OF CONTENTS

ABSTRACT . . . iii

ACKNOWLEDGEMENTS . . . v

TABLE OF CONTENTS . . . ix

LIST OF TABLES . . . xii

LIST OF FIGURES . . . xv

CHAPTER 1: **INTRODUCTION** 1
- 1.1 **Related Work** . . . 2
- 1.2 **Structure and Contribution of this Dissertation** . . . 7
  - 1.2.1 *Thesis Statement* . . . 7
  - 1.2.2 *Structure* . . . 7
  - 1.2.3 *Contributions* . . . 9

CHAPTER 2: **CRYPTOCURRENCY PRIMER** 11
- 2.1 **Coins vs Tokens** . . . 12
- 2.2 **The Cryptocurrency Exchange** . . . 13
- 2.3 **Conclusion** . . . 14

CHAPTER 3: **GATHERING DATASETS TO FIND EXCHANGE SHOCKS** 15
- 3.1 **Exchange Activity** . . . 15
  - 3.1.1 *Dataset Validation* . . . 16
- 3.2 **Ethical Considerations** . . . 19
- 3.3 **Shocks to Mt. Gox** . . . 19
  - 3.3.1 *D1: Reported DDoS attacks* . . . 19
  - 3.3.2 *D2: Additional security shocks* . . . 20
  - 3.3.3 *D3: Confirmed DDoS attacks* . . . 20
- 3.4 **Conclusion** . . . 21

CHAPTER 4: **IDENTIFYING SUSPICIOUS TRADES** 22
- 4.1 **Suspicious Trading Activity** . . . 22
  - 4.1.1 *Suspicious Trader 1 - the "Markus" Bot* . . . 22
  - 4.1.2 *Suspicious Trader 2 - the "Willy" Bot* . . . 25
  - 4.1.3 *What May Have Motivated The Operation of These Bots?* . . . 27

4.2 **Conclusion** . . . 28

CHAPTER 5: **ANALYZING SHOCKS ON THE MT. GOX CRYPTOCURRENCY EXCHANGE** 29
5.1 **Regression Analysis of Shocks** . . . 29
5.1.1 *Transaction Volume and Large Trades* . . . 30
5.1.2 *Skewness and Kurtosis* . . . 31
5.1.3 *Regression Analysis* . . . 33
Robustness Analysis . . . 36
5.2 **Price Manipulation** . . . 38
5.2.1 *Preliminary Analysis* . . . 38
Suspicious Purchases and Trade Volume . . . 39
Suspicious Purchases and Price Changes . . . 42
5.2.2 *Regression Analysis* . . . 45
Daily Price Change . . . 45
Daily Percentage Returns . . . 47
5.2.3 *Limitations* . . . 50
5.3 **Conclusion** . . . 50

CHAPTER 6: **MEASURING THE LIFESPAN OF A CRYPTOCURRENCY** 52
6.1 **Methodology** . . . 52
6.1.1 *Data Sources* . . . 52
6.1.2 *Identifying Peaks, Abandonments, and Resurrections* . . . 53
6.2 **Coin Results** . . . 55
6.2.1 *Peaks, Abandonments, and Resurrections* . . . 55
Peaks . . . 55
Abandonments, and Resurrections . . . 60
6.2.2 *Relationships Between Key Variables* . . . 63
6.2.3 *Bursting of Bubbles and the Changing of the Guard* . . . 67
6.2.4 *Further Analysis of The "Returns" From Top 80 Coins* . . . 70
6.3 **Conclusion** . . . 71

CHAPTER 7: **THE RISE AND FALL OF TOKENS** 72
7.1 **Methodology** . . . 73
7.1.1 *ICO Data* . . . 73
7.1.2 *CMC Data* . . . 75
7.1.3 *Evaluating and Combining ICO Data Sources* . . . 76
7.1.4 *Descriptive Analysis: ICO Data* . . . 79
7.2 **Measures of success** . . . 84
7.2.1 *Hardcap and Token Returns* . . . 84
7.2.2 *Pseudo-Trading* . . . 86
7.3 **Token Rise and Fall** . . . 88
7.3.1 *Abandonment, and Resurrection* . . . 92
7.3.2 *Relationships Between Key Variables* . . . 94
7.4 **Conclusion** . . . 98

CHAPTER 8: **MEASURING THE IMPACT OF CRYPTOCURRENCY "PUMP-AND-DUMP" SCHEMES** 99
8.1 **Methodology** . . . 100
8.1.1 *Pump Signals Data From Discord and Telegram* . . . 100
Transparent Pumps . . . 102
Obscured Pumps . . . 102
Differentiation between "countdown" vs. "price signal" . . . 103
Copied pumps . . . 103
Summary of Pump Signals . . . 104
8.1.2 *Pricing Data on Cryptocurrencies* . . . 104
Matching Discord/Telegram Information with Trading Data . . . 105
Identifying pump Timing and Success . . . 106
8.1.3 *Data Summary* . . . 106
8.2 **Informal Analysis of the Ecosystem** . . . 107
8.2.1 *A Bit of Theory...* . . . 107
8.2.2 *Dynamics over time in the ecosystem* . . . 107
8.2.3 Concentration in the Ecosystem . . . 108
8.3 **Analysis** . . . 110
8.3.1 Both Types of Pump Groups Examined Together . . . 110
8.3.2 *Independent Variables* . . . 110
8.3.3 *Analyzing the Two Types of Pump Groups Separately: Telegram Data Only* . . . 113
8.4 **Formal Regression Results** . . . 115
8.4.1 *All Pump and Dump Schemes Included Together* . . . 115
8.4.2 Transparent and Obscured Analyzed Separately: Telegram Only . . . 117
8.4.3 *What happens after the pump is over* . . . 119
8.4.4 Trading Volume Data . . . 120
8.4.5 *The number of exchanges mentioned by the pump* . . . 121
8.5 **Conclusion** . . . 121

CHAPTER 9: **MARKET MANIPULATION THROUGH ORGANIZED, TARGET-BASED TRADING** 123
9.1 **Methodology** . . . 123
9.1.1 *Extracting Targets from Pump Signals* . . . 123
9.1.2 *Pricing Data* . . . 126
9.1.3 *Matching Pump Information with Trade Data* . . . 127
9.2 **Analysis** . . . 128
9.2.1 *Identifying Target Cycles* . . . 128
9.2.2 *Success in Waves* . . . 133
9.2.3 *Price Movements within Cycles* . . . 136
9.3 **Conclusion** . . . 137

CHAPTER 10: **CONCLUDING REMARKS** 138
10.0.1 *Future Work* . . . 141
Anomalous Activity Detection . . . 141
Time Series Point-of-Interest Detection . . . 141

"Pump-and-Dump" Early Detection 141
Pump Signal Analysis Through Inferred Targets 142

BIBLIOGRAPHY 143

APPENDIX A: **DESCRIPTIVE STATISTICS AND OTHER TABLES** . . . 150
APPENDIX B: **SENSITIVITY ANALYSIS OF PEAK DETECTION** 155
APPENDIX C: **SENSITIVITY ANALYSIS OF ABANDONMENT AND RESURRECTION DETECTION** 157
APPENDIX D: **CRYPTOCURRENCY TRADING RISK AND REWARD** 159
APPENDIX E: **ADDITIONAL TOKEN STATISTICS** 162

LIST OF TABLES

1.1 Data, coverage and chapter(s) in which they appears. 8
3.1 Additional shocks, other than DDoS, affecting Mt. Gox. 21
4.1 Distribution of USD/BTC rates paid by "Markus" 23
4.2 Fraudulent transactions initiated by "Markus" (user ID in bold) 24
4.3 Duplicate "Markus" Transactions 24
4.4 Summary of "Markus" transactions 25
5.1 Transaction Volume and Large Trades 33
5.2 Skewness and Kurtosis 34
5.3 Skewness and Kurtosis With Trends Explanatory Variable 36
5.4 Robustness Analysis 37
5.5 Daily BTC purchased by "Markus" and "Willy" on days they were active. . . 40
5.6 Comparison of daily BTC volumes on days when suspicious trades occurred and did not. 41
5.7 Unauthorized activity and price changes on Mt. Gox 42
5.8 Suspicious trading activity: % of days active during each period 43
5.9 Average daily rate change (in $) and percentage rate change (in parentheses) in BTC-USD exchange rate by period 43
5.10 Examining Price Changes Within Mt. Gox and the other Exchanges 47
5.11 Examining Percent Price Changes Within Mt. Gox and the other platforms . 48
5.12 Skewness and Kurtosis Through Shocks and Insider Trading 49
6.1 Summary statistics on coin peaks, abandonment and resurrection, broken down by total trading volume per coin. 56

6.2 Monthly correlations between key variables in the ecosystem. 66

6.3 Currency movement during different influential time periods. 68

7.1 ICO data source summary statistics. 75

7.2 ICO Price representation by source. Statistically significant under and over-representations are indicated in bold with a (+/-). 77

7.3 Amount raised representation by source. Statistically significant under and over-representations are indicated in bold with a (+/-). 77

7.4 Start date representation by source. Statistically significant under and over-representations are indicated in bold with a (+/-). 77

7.5 End date representation by source. Statistically significant under and over-representations are indicated in bold with a (+/-). 77

7.6 Comparison of coverage before and after running data combination algorithm. 79

7.7 Summary statistics on token returns split by ICO success (binary). 85

7.8 Subset of tokens purchased 2015 for portfolio 86

7.9 Summary statistics for yearly portfolio percent returns. Token count at return calculation in parenthesis. 87

7.10 Summary statistics on token peaks, abandonment and resurrection, broken down by total trading volume per coin. 89

7.11 Monthly correlations between key variables in the ecosystem. 97

8.1 Pump (Median) Success/Profitability by Month in Percentage Terms 108

8.2 Descriptive Statistics: Discord, N=1,034 111

8.3 Descriptive Statistics: Telegram, N=3,767 111

8.4 Median Price Increases by Coin Rankings. 112

8.5 Correlations Among Variables: Discord, N=1,034 113

8.6 Correlations Among Variables: Telegram, N=3,767 113

8.7 Descriptive Statistics: (Telegram) Transparent Pumps, N=271 114

8.8 Descriptive Statistics: (Telegram) Obscured Pumps, N=2,198 114

8.9 Examining What Affects Success of Pump and Dump Schemes: 116

8.10 Examining What Affects Success of Pump and Dump Schemes: 118
8.11 Correlation Table: Pump Exchange and Independent Variables, N=546 . . . 121
9.1 First observed target for each pump signal. 129
9.2 Success based on inclusion or exclusion of trading cycles. 133
9.3 Stoploss crossed within X days of cycle start (first cycle only). 134
9.4 Percentage price movements between points of interest. 136
A.1 Summary statistics of independent and dependent variables 150
A.2 Correlation between daily rate changes and the independent variables 151
A.3 Correlation between daily percent rate changes and the independent variables 151
A.4 Correlation between independent variables 151
A.5 Suspicious trading activity and price changes on Bitstamp 152
A.6 "Willy": Volume activity (period 4) 152
A.7 "Markus": Volume activity (period 3) 153
A.8 Examining Skewness and Kurtosis Through Suspicious Trading Activity . . . 154
B.1 Sensitivity Analysis of Peak Definition Algorithm 155
C.1 Total Number of Abandonments (Sensitivity Analysis) 158
C.2 Median Number of Abandonments (Sensitivity Analysis) 158
C.3 Total Duration of Abandonments (Sensitivity Analysis) 158
E.1 Price movement through target values for Time New Bank. 162
E.2 Hours between pump signal and first target crossed (not necessarily a buy target). 162
E.3 Number of targets crossed per pump signal. 164

LIST OF FIGURES

3.1 Daily differences in transaction volume between leaked dataset and totals reported by `bitcoincharts.com`. Differences are normalized as a fraction of the leaked daily volume. Positive numbers indicate that the leaked data reported higher volume. 18

5.1 Distribution of transactions by amount in JPY on days following a reported DDoS attack (in red) and on all other days (in black) 30

5.2 Bitcoin-USD exchange rate with periods of suspicious activity shaded. 38

5.3 Percentage of total daily trade volume at Mt. Gox when "Willy" and "Markus" are active; shaded green if the BTC-to-USD exchange rate closed higher and red otherwise. 39

6.1 Volume plot showing currency abandonment. Red dots indicate peaks. 54

6.2 Percentage price and volume increase from a coin's launch to first peak, based on the year in which the coin was launched and its size. (Note: the vertical axis is logarithmic.) 57

6.3 Deciles of percent price and volume rises from the smallest value in the month prior to a peak (left graphs) and deciles of percent price and volume falls from the peak to the smallest value in the month following the peak (right graphs). 58

6.4 Abandonments (left) and resurrections (right) per coin, split by total trading volume. 60

6.5 Survival probabilities for the time to abandonment (left) and time to resurrection (right). 61

6.6 New currencies announced on the Bitcoin Forums each month (orange) compared to new currencies traded each month (black). 62
6.7 Average monthly market capitalization of all cryptocurrencies traded (in Billions of dollars). 63
6.8 Number (top graph) and percentage of active coins (bottom graph) experiencing price and volume peaks over time. 64
6.9 Cryptocurrency summary statistics including creation, resurrection, abandonment, and daily average trading volume. 65
7.1 Bitcoin price movement over time overlaid with the original dataset cutoff. . . 75
7.2 ICO creation and failure rates compared to the log scale BTC price movement. 80
7.3 Complete: Top Countries for ICOs. 81
7.4 Correct: Top Countries for ICOs. 81
7.5 Percent of ICO tokens that make it to CMC. 83
7.6 Time between ICO start date and first date on CMC. 83
7.7 Plot of the distribution of the % ICO success (Log10). 85
7.8 Distribution of returns over time split by periods of interest. 85
7.9 Daily values for yearly portfolios. Here the y-axis is log transformed. 88
7.10 Percentage price and volume increase from a token's launch to first peak, based on the year in which the token was launched and its size. The vertical axis is logarithmic. 90
7.11 Deciles of percent price and volume rises from the smallest value in the month prior to a peak (left graphs) and deciles of percent price and volume falls from the peak to the smallest value in the month following the peak (right graphs). 91
7.12 Abandonments (left) and resurrections (right) per token, split by total trading volume. 92
7.13 Survival probabilities for the time to abandonment (left) and time to resurrection (right). 93
7.14 Market capitalization of all tokens traded (in billions of dollars). 94

7.15 Number (top plot) and percentage of active tokens (bottom plot) experiencing price and volume peaks over time. 95
7.16 Token summary statistics including creation, resurrection, abandonment, and daily average trading volume. 96
8.1 Google Trends results from searching for pump and dump cryptocurrency. . . 109
9.1 Example pump signal from Whale Pump Group. 124
9.2 Example pump signal from Whale Pump Group. 124
9.3 Example target pump signal from Mega Pump Group. 124
9.4 Example pump signal from Big Pump Group. 124
9.5 Pump and dump using Genesis Vision as observed on Coinmarketcap. 127
9.6 Ark pump activity from November 2018. 129
9.7 Cycle identification within cryptocurrency time series data (iostoken). 131
9.8 Hours between pump signal and the price crossing the first buy target. 132
9.9 Difference between cryptocurrency price at the pump signal and the price at the first buy target. 133
9.10 Cycle timing summary statistics - median(mean) hours. 135
D.1 Scatterplot of token risk and reward. 159
D.2 Scatterplot of coin risk and reward. 160
D.3 Scatterplot of S&P 500 stock risk and reward. 160
E.1 Numerical value source categorization. 163
E.2 Categorical value source categorization. 163

CHAPTER 1
INTRODUCTION

Cryptocurrencies attract researchers, legitimate users, and criminals for some of the same reasons: they are a large, relatively unregulated market, and they are decentralized, meaning they are not controlled by a centralized governing body. Furthermore, many cryptocurrencies attempt to foster a culture of transparency by publishing every transaction to a public blockchain, or ledger. With this public ledger one might be led to believe that finding criminal activity was easy. However, that is not the case. Many of the available exchanges are centralized entities, with opaque governance, where transactions take place off-chain, meaning the transfer is never stored on the blockchain. Traditionally, to conduct financial research, a partnership with a financial institution is required. Although a database leak could also make some research easier [17], it is unpredictable and potentially unethical. Instead, through the use of public data endpoints offered by cryptocurrency exchanges, supplemented with data from the blockchain, the movement of money throughout the system can be observed.

The United States and other countries have similarly tried to apply broadly worded financial laws and regulations to cryptocurrencies. However, the ecosystem still operates as if it is the wild west. According to Autonomous Research, a London-based financial-services research firm, hacking related losses from cryptocurrency exchanges is said to total 1.63 billion USD as of July 2018 [48]. Additionally, there is no vetting process for individuals that wish to start their own cryptocurrency, exchange, wallet, or other blockchain tied service. Because of this, less than savory individuals have been able to steal at least 11 USD [45], and occasionally upwards of 660 million USD from seemingly legitimate ICOs, or initial coin offerings [4]. With little to no regulatory oversight or audit mechanisms in place, exchanges are regularly accused of employing and occasionally found to have employed, methods to artificially inflate the trade volume on their exchanges [46, 21]. Hacks, questionable behavior, and illegitimate cryptocurrencies add to the volatility of an ecosystem that has never truly been stable and create usability problems for those with legitimate use cases.

Utilizing publicly available data sources, such as API endpoints and cryptocurrency data aggregation websites, regressions and algorithms were developed to help analyze various cryptocurrency money-making endeavors. Greater detail is given in subsequent chapters to assist with the understanding of cryptocurrencies and their required parts, but it is worth noting, in most cases, the methods employed by the malicious actors are more important to the scheme's success than the cryptocurrencies they are used against.

1.1 Related Work

A large body of research exists within multiple disciplines that explores the viability of cryptocurrencies from various angles. The first popular cryptocurrency, Bitcoin, which surfaced in 2009, brought about much of the initial research. Bonneau et al. built a comprehensive description of Bitcoin and the related blockchain technologies by examining existing research and modifications to the cryptocurrency's codebase, and went on to propose changes that could have a positive, lasting impact on the security and efficiency of cryptocurrencies in general [14]. Böhme et al. offered an explanation of the design principles, regulatory and security risks, and available uses of cryptocurrencies for audiences outside of the discipline of computer science [11]. When viewed together, these two papers offered a baseline understanding of the inner workings of cryptcurrencies and pointed out key challenges facing them.

For many of the available cryptocurrencies an exchange, which is a central part of the operation, is the easiest, or only way for users to buy and sell currencies. Due to the nature of cryptocurrencies and the importance of the exchange it is a frequent target of nefarious activity. Although the exchange is a required service, it has already been established that, without a central agency offering financial assistance, exchanges can and will fail. Moore and

Christin found that by early 2013, 45% of all Bitcoin cryptocurrency exchanges had closed, and those that survived were plagued with service outages and hacks [39]. In a follow up study, Moore et al. found that of the 80 cryptopcurrency exchanges observed at the time almost half (38) had completely disappeared, confirming that the cryptocurrency market was extremely unstable [40].

Traditionally, networks like these only become valuable when they amass a large user base. However, within the cryptocurrency ecosystem many tokens solve the user issue by providing financial utility before providing application utility, or vice versa. This means tokens can be bought and sold, which attracts users, before they can be used within the application or platform being built. Furthermore, platform economics can easily be observed within cryptocurrency ecosystem: a large user base on one side of the business makes the platform more valuable to the other side. Mining or even exchanges, for example, are typically profitable when a cryptocurrency is actively traded. If a cryptocurrency is not traded, no miners collect block rewards and no exchanges collect transaction fees.

Additionally, a large body of empirical literature on the dynamics of entry and exit exists in Economics[1]. One particular focus in this literature is on the post-entry performance of firms. These studies typically examined the entry and exit rates over time, the number of firms in the industry over time, the survival rate of new firms, and the evolution of firm size over time.[2] One particularly robust finding in this literature was that entry into new markets generally occurs in waves. This seems to be the case in the cryptocurrency industry as well.

The collection of available cryptocurrencies has yet to revolutionize the financial sector in the ways its developers and users have envisioned. However, economists have taken advantage of these new financial markets to explore what drives a “value-less” currency. Li and Wang developed a model to explore the motivating factors behind the rise and fall of the bitcoin (BTC)-to-United States dollar (USD) exchange rate [32]. Corbet et al. expanded

[1]For a good summary of early work, see Audretsch and Mata on the post-entry performance of companies [8].

[2]See Geroski and the references cited within for a survey of the literature [22].

upon Li and Wang's model, finding that shocks to traditional financial assets did not affect cryptocurrencies [16]; but, the shocks to the price of the three cryptocurrencies they studied (Bitcoin, Ripple, and Litecoin) did affect each other. Bolt and van Oordt developed their own model to explore the economic factors that influenced cryptocurrency exchange rates [13]. Hayes developed a model that estimated the value of a cryptocurrency by taking into account its mining difficulty, its rate of unit production, and its level of competition between producers [25]. Xie et al. analyzed the effects the `bitcointalk.org` forum had on the price of bitcoin [56]. They found that disjoint discussion networks were a good indicator of price movement. Additionally, Ciaian et al. determined that market forces and coin attractiveness played a large role in the price movement of bitcoin [44][3].

An expanding area of research revolves around how cryptocurrencies have been and still can be manipulated for illicit financial gain. Because the cryptocurrency ecosystem is still a relatively unregulated market, often compared to the wild west, fraud runs rampant. Researchers have taken various approaches to establish the scope and determine the prevalence of undesirable activity such as Ponzi schemes [53], money laundering [41], mining botnets [26], and theft of "brain" wallets [52]. Meiklejohn et al. looked at the blockchain to explore the anonymity of Bitcoin payments. They were successfully able to determine wallet addresses for popular Bitcoin services such as currency exchanges [37]. Ron and Shamir utilized the blockchain to build a network of transactions, which they used to identify suspicious trading activity [47].

More recently Hougan et al. examined 83 of the top cryptocurrency exchanges by bitcoin trading pairs [36]. They found that only 10 of the exchanges are accurately reporting their trading volume; 95% of the trading volume reported for bitcoin related trades was most likely wash trading. Wash trading is a form of market manipulation in which bots are employed to trade cryptocurrencies with each other in an effort to create misleading, artificial activity in the marketplace.

[3] Gandal and Halaburda [19] examine competition among cryptocurrencies. They find that the data are consistent with strong network effects and winner-take-all dynamics.

Initial coin offerings (ICO) emerged in 2013 as a way to raise capital for a business without going through traditional methods of securing funding. Six billion dollars was raised in 2017 by ICOs, up from approximately 90 million USD in 2016. This phenomenal growth spurred new research into this area[4]. Huang et al. compared the potential profits from mining, and buying an altcoin to a more stable coin [27]. They found that, while some coins gave impressive returns, it was less risky and more profitable to mine coins than to buy them. Adhami et al. studied 253 ICO campaigns and found that, while most token offerings were successful, the success was tied to an openly accessible codebase [5]. Amsden and Schweizer studied features that caused tokens to trade on currency exchanges [7], and found that features like quality token operators increased the likelihood of trading. Bian et al. designed a system to identify scam ICOs by analyzing their whitepapers, and websites and other aspects of their associated tokens [10]. Huang et al. examine 917 ICOs to determine which country level offerings facilitate ICO growth [28]. They found that more ICOs are started in countries offering a clear regulatory framework as well as a well-developed financial market. Paul Momtaz finds that 40% of all ICOs destroy investor value on the first day of trading [38]. Additionally, he finds that highly visionary projects are abandoned at a higher rate than other projects making up a majority of the 21% of tokens that are delisted from exchanges. Lyandres et al. develop five indicators of ICO success and failure [33]. The authors also develop methods to determine source and record level data quality, which is very useful for combining ICO data sources. When examining ICO returns, Benedetti and Kostovetsky find that tokens are averaging returns of 179% after only 16 days [9].

In addition to exploring what gives "value-less" currencies value, researchers are analyzing how trade activity can alter the price of a cryptocurrency. Krafft et al. created bots that executed trades on 271 "penny cryptocurrencies," which are similar to penny stocks, using the Cryptsy cryptocurrency exchange [30]. They found that their bots caused a two percentage point increase in buying activity from others when the researchers executed a small trade.

[4] https://www.icodata.io

Similarly, outside of the cryptocurrency ecosystem, financially motivated individuals have manipulated low volume stocks through pump-and-dump schemes [6]. Aggarwal and Wu found that, during periods of manipulation, volatility, volume, and price increased. The authors suggest that, while these manipulative activities have declined on the main stock exchanges, pump-and-dump schemes continue to be an issue in the over-the-counter (OTC) market in the United States. And email spam has been used to promote and sell pump-and-dump stock schemes [12, 18, 24]. Massoud et al. examined the profit generating relationship between listed firms and the companies they pay to secretly, and illegally, promote their stocks and found that this activity increased the price and trade volume of the firm's stock and that the price increases reversed when regulators took action [35].

In considering larger manipulations, Griffin et al. investigated whether Tether, a digital cryptocurrency pegged to the United States dollar, contributed to the price movement of bitcoin and other cryptocurrencies [23]. They found that timed purchases of Tether resulted in significant increases in the price of the cryptocurrencies they studied and concluded that these increases could not be explained by investor demand but were consistent with the hypothesis that Tether was used to manipulate prices and provide price support.

Recently pump-and-dump schemes have gained media attention, which resulted in more published research. Four concurrent papers, produced by different groups with differing foci, were published at approximately the same time. Kamps and Kleinberg utilized pump-and-dump detection methods from traditional financial literature to propose a method for detecting that behavior in the cryptocurrency ecosystem [29]. They used manually identified pumps to test and verify the accuracy of their predictions.

Mirtaheri et al. took advantage of social media, specifically Twitter and Telegram, to detect pump signals which they then used verified by comparing it with market data. They also attempted to determine if the pump-and-dump would be successful. Xu and Livshits used data on 220 pump signals to build a model to predict which coins would be pumped [57]. Their model was able to distinguish between highly successful pumps and all other trading activity on the exchange. Li et al. used a difference-in-difference approach to show that pump-and-dump schemes lower the trading price of affected coins [31]. [5]

1.2 Structure and Contribution of this Dissertation

1.2.1 Thesis Statement

This dissertation documents the prevalence and impact of certain illicit financial schemes within the cryptocurrency ecosystem. It describes a wide range of unmistakably criminal techniques, including DDoS attacks, insider trading, pump and dump schemes. Additionally, because not all manipulations can be identified and not all fluctuations are intentional, it develops a general-purpose method for identifying when a cryptocurrency is likely to be abandoned and resurrected, which could be indicators of manipulation. The primary impact observed has been fluctuations in pricing, though volume is also considered. Again, in some cases these fluctuations are the result of clear manipulation, while in others the cause cannot be established.

1.2.2 Structure

Chapter 2 gives a short description of the cryptocurrency ecosystem.

Chapter 3 outlines the data gathered from the Mt. Gox cryptocurrency exchange for use in Chapter 5. The methods used to clean and validate the data are detailed, as are some of the problems faced throughout the process. Chapter 3 also explores how, and where reports of shocks to the Bitcoin ecosystem were collected.

Chapter 4 details the suspicious trading activity associated with two actors, which are referred to as “Markus” and “Willy.” The method of discovery is covered along with what appeared to be an incomplete attempt to cover up some of that trading activity. Finally, a hypothesis that explores a plausible reason for employing this type of suspicious trading activity is discussed.

[5] There have been media articles about the pump and dump phenomenon as well. Mac reported on pump and dump schemes in a Buzzfeed article published in January 2018 [34]. This was followed by work by Shifflet and Vigna in a Wall Street Journal article published in August 2018 [50].

Chapter 5 analyzes various shocks experienced on the Mt. Gox cryptocurrency exchange. The effects of planned and unplanned outages experienced by the cryptocurrency exchange are analyzed. The leaked Mt. Gox transaction database spanning April 2011 through November 2013 was utilized and the primary focus was on distributed denial-of-service (DDoS) attacks. Then, suspicious trading on Mt. Gox is examined as is its impact on the price of bitcoin between February 2013 and November 2013. The accounts involved were able to acquire 600,000 bitcoin valued at approximately 112 million USD and likely paid nothing to do so. During the two periods identified the bitcoin-to-United States dollar exchange rate stayed flat on the days with no suspicious trading activity and rose an average of 4% on the days when suspicious trading activity occurred.

Chapter 6 explores cryptocurrency competition and the identification of abandoned, or "dead" coins. By analyzing the trading activity of 1,082 coins over a five-year period it was possible to develop metrics to help identify when a cryptocurrency was likely to be abandoned.

Type of data	Date range	Chapter
Mt. Gox transactions	2011-04 - 2013-11	3, 4, 5
Exchange shocks	2011-02 - 2013-11	3, 4, 5
Coin price and volume	2013-02 - 2018-02	6
Token price and volume	2013-02 - 2018-02	7
Initial coin offerings	2012-07 - 2019-10	7
Token price and volume	2013-02 - 2019-10	7
Pump signals (all)	2018-01 - 2018-06	8
Cryptocurrency price and volume	2018-01 - 2018-06	8
Pump signals (obscured)	2017-07 - 2019-01	9
Cryptocurrency price and volume	2017-01 - 2019-11	9

Table 1.1: Data, coverage and chapter(s) in which they appears.

Chapter 7 examines token success through data reported on initial coin offerings (ICOs). By combining data from four ICO trackers, 8,305 unique ICOs were identified. The abandonment and resurrection analysis from Chapter 6 was repeated on this data to identify differences between coin and token abandonment and resurrection.

Chapter 8 attempts to determine the scope of "pump-and-dump" schemes within the cryptocurrency ecosystem as well as the rate of success. By combing through two prevalent messaging applications, Telegram and Discord, nearly 5,000 pump signals promoting more than 300 distinct cryptocurrencies over a six month period in 2018 were identified.

Chapter 9 focuses on analyzing a subset of pump-and-dump schemes that include target trade values. Longer term analysis of 3,683 pump signals explores success through the identification of trading cycles around target values.

Many chapters in this dissertation share data sources. Some of these data sources cover the same date ranges, while others are extended or modified. Table 1.1 shows that even though the sources may be identical, in many cases the date ranges have been updated for newer analysis.

1.2.3 Contributions

The contributions provided by this dissertation lie not only in the analysis of the collected data but also in the methodologies used to collect and cleanse the data. In Chapter 3, reports of DDoS events were gathered from the /r/bitcoin Reddit sub-forum, from social media accounts (Facebook, Google+, and Twitter), and finally from news reports. To ensure validity of the candidate events all forum posts were manually checked, and only those discussing DDoS events were used. The leaked Mt. Gox dataset was collected, cleansed, and normalized. In Chapter 4, a list of reported suspicious accounts on Mt. Gox was assembled and the existence of the accounts in the dataset established in Chapter 3 were manually verified. In Chapter 5, daily summary values including the open, high, low, and close prices, trading volume, exchanges, and market capitalization from `coinmarketcap.com` were scraped and stored locally. Five years worth of data ranging from February 2013 to February 2018 on 1,082 cryptocurrencies was collected and analyzed. Similarly, for Chapter 8, price data on close to 2,000 cryptocurrencies listed on 220 exchanges was collected from `coinmarketcap.com`. Unlike the data used in Chapter 5, this data was the finest granularity data offered by the site at the time and is roughly of a 5-minute interval. Additionally, Discord and Telegram were programmatically scraped for candidate pump signals, which, when found, were then filtered by keyword and subsequently manually verified. In Chapter 7, ICO trackers were evaluated to determine source reliability. Attributes associated with a token's ICO as well as pseudo-investment models are used in determining token and ICO success. In Chapter 9, regular expressions are developed to reliably separate pump and dump target data from the rest of the pump signal.

The work described above allowed for easier analysis of new datasets. Chapter 6 determines if cryptocurrencies fail in the same way as their exchange counterparts. To do this algorithms were developed to identify peaks and subsequent periods of abandonment in time series financial data. Furthermore, the scope of the abandonments was determined as were rules for resurrection. Chapter 9 examines long term pump success through cycles of price movements around pump target values.

CHAPTER 2

CRYPTOCURRENCY PRIMER

Using cryptography for anonymized digital payments was first developed and introduced by David Chaum's company, DigiCash, in 1990 under the name cyberbucks [15]. At the time e-commerce was still in its infancy and consumers still chose convenience over privacy, so Visa and Mastercard continued to be the preferred methods of sending and receiving payments. Even with the growth of online shopping, consumers continued to use products that offered convenience over privacy. DigiCash eventually died in 1999. Around the same time, other digital currencies such as E-Gold, which was backed by gold stored in a safety deposit box, and Liberty Reserve, which allowed unregulated money transfers came into existence. These services were not anonymous and they were run in a centralized manner which eventually led to their downfall when law enforcement shut them down for facilitating criminal activities. Ten years after the fall of DigiCash, Bitcoin[1] [42] was introduced as a new digital currency that was fully decentralized. Bitcoin was developed by the forever elusive Satoshi Nakamoto, was released as open source software, and became the world's first popular cryptocurrency.

From a single cryptocurrency in 2009, the world has been able to watch this ecosystem rapidly grow. At the time of this writing, `coinmarketcap.com` reports over 2,100 distinct coins and tokens offered by upwards of 250 exchanges.

Most of the available cryptocurrencies offer varying levels of four characteristics: they are decentralized; they are immutable; they are trustless; and, they provide a certain level of anonymity. The typical cryptocurrency has no central transaction storage mechanism

[1] The Bitcoin ecosystem includes the core network for propagating transactions, the blockchain, and many intermediaries such as currency exchanges, mining pools, and payment processors that facilitate trade. "Bitcoin" with a capital "B" will be used to refer to the ecosystem and "bitcoin" with a lowercase "b" will refer to the coin.

and no central governing body. With no central governing body these currencies operate without insurance or backing by a federal entity and it is the responsibility of the users running "full nodes" to verify that the agreed upon financial rules are being followed. In explanation, any computer that connects to the bitcoin network is called a "node," and computers that download all blocks and transactions and verify them against the consensus rules are called "full nodes." The users that run full nodes ensure that no one is double spending any fraction of a coin, that no one is creating coins out of thin air, and, most importantly, that payments received are verified as legitimate. Completion of these tasks ensures that each payment can be trusted even when there is no trust established between the parties involved in a transaction. Most cryptocurrencies attempt to provide anonymity by making sure there is nothing directly tying a specific person to an account/wallet address and some of the currencies go even further by obfuscating transaction details. However, a growing number of exchanges now require user registration for tax and legal purposes and this provides a link between a person and an account.

For a better understanding of the underlying mechanics of cryptocurrencies and blockchains the reader is referred to [43].

2.1 Coins vs Tokens

Within the cryptocurrency ecosystem there are two distinct entities: coins and tokens. Satoshi Nakamoto introduced the coin with the release of Bitcoin in 2009 and J. R. Willett introduced tokens with the release of Mastercoin [55] in 2013.

Coins have two main characteristics: they are tied to a public blockchain that anyone can use and they may be sent, received, and/or mined. They are not meant to act as anything other than a medium of exchange.

Tokens are not meant to be used as a medium of exchange or a store of value and they typically are not mined. Tokens are intended to be a form of tech start-up funding consumers can purchase. Tokens also differ from coins in that they give users the ability to participate in a specific network. There are many different types of tokens available to consumers, and upon exploring the ecosystem it was found that the community largely agrees on the following five token types: utility, security, currency, reward, and asset. Regardless of type, most tokens operate like a company scrip in an old coal mining town. The tokens can only be used within the organizer's project. One exception to this type of use is the currency token which works much in the same way a typical cryptocurrency does. The average token has a very specific use case within a certain project similar to tickets to an event which cannot be used to pay bills or purchase goods, but can be used a single time at a certain venue. Additionally, tokens do not reside on their own blockchain, but are built on top of an existing cryptocurrency's infrastructure. For example, Tether is built on top of Omni, while Binance Coin, OmiseGo, and 0x are built on top of Ethereum, and Everipedia is built on top of EOS.

Tokens are increasingly popular cryptocurrency offerings. According to `icodata.io`, a website that tracks the sale of tokens through initial coin offerings (ICOs), 1,255 ICOs raised just over 7.8 billion USD from token sales in 2018. This year is slated to also be an impressive one with ICO sales having raised over 262 million USD as of mid-May.

2.2 The Cryptocurrency Exchange

The main reason users join the cryptocurrency ecosystem is to buy a cryptocurrency through an exchange. To buy and sell coins and/or tokens users typically maintain balances of both fiat currency and cryptocurrency on the exchange without having direct access.

Exchanges are a necessary evil in the world of cryptocurrency and they effectively nullify the four cryptocurrencies characteristics mentioned earlier in this chapter (anonymity, immutability, decentralization, and trustless operation). They typically employ a centralized structure for ease of governance, security, and speed. Furthermore, because these exchanges use a centralized database for off-chain transactions, they are able to offer refunds in the event of a theft; which cannot be done when transacting directly on the blockchain because of its immutability. The anonymity or pseudo-anonymity offered by cryptocurrencies is removed by the identification requirements of some exchanges, which differ from exchange to exchange and from country to country because of the differences in the laws and perceptions of digital money. Some countries, such as Egypt and Morocco, have completely outlawed cryptocurrencies while others, like the United States, in an effort to curb money laundering require that substantial cryptocurrency businesses know their customers. This means that exchanges require identification when opening an account. Finally, because these exchange operations mirror those of a bank, the consumer's position requires that he/she trust exchanges for service and account stability.

An important difference between cryptocurrency exchanges and banks is that, even though most transactions happen off-chain, meaning they are only saved to the blockchain in the event of a withdrawal from, or deposit to, the exchange, users still have access to some forms of anonymized trade data through publicly accessible endpoints. Many exchanges give access to open trade requests, giving the amount of a cryptocurrency a trader wants to buy or sell along with a price he/she are willing to pay or receive. Typically exchanges also give access to trades completed on the exchange which usually includes a trade identification (ID), the amount traded, the price paid, a time stamp, and other miscellaneous information. Because this data is anonymized, exchanges never willingly give out user IDs or any form of identifiable information. The type of financial information available to the public through a cryptocurrency exchange is not made available through traditional banks.

2.3 Conclusion

Cryptocurrencies and exchanges, although growing in number and size, still have problems that need to be addressed. The centralization, lack of regulation, and irregular security implementations make the cryptocurrency ecosystem a near perfect environment for theft and manipulation.

CHAPTER 3

GATHERING DATASETS TO FIND EXCHANGE SHOCKS

For the purpose of this dissertation simply looking at exchange activity through the records stored in the blockchain did not provide sufficient information because most exchange activity is stored off-chain in a centralized database controlled by the exchange with only the exchange user deposits and withdrawals being written to the blockchain.

The following sections provide an outline of how rarely seen exchange transaction data were collected and verified. Furthermore, the methodologies used to collect reported shocks to the Bitcoin ecosystem are documented.

3.1 Exchange Activity

Shortly after Mt. Gox filed for bankruptcy in early 2014, a trade history of its transactions was publicly leaked. The leaked data include transaction times, currency conversions to/from bitcoins, transaction amounts and exchange rates. And importantly the user ID, the internal number associated with each Mt. Gox user, was included. The user ID was a crucial piece of information because it linked different transactions to the same user. This data offered much finer granularity than was typically available because most buy and sell transactions are recorded only by the exchange and never appear on the blockchain. And the data could be leveraged to monitor changes in user participation as well as overall transaction volume at times surrounding shocks. In total, nearly 18 million matching buy and sell transactions were reported between April 2011 and November 2013.

The Mt. Gox data were supplemented with data with daily transaction volumes reported by the website `bitcoincharts.com` for all other monitored Bitcoin exchanges. Some entries obtained from `bitcoincharts.com` included missing values, making it necessary to gather weekly transaction data from `bitcoinity.org` to verify trading volumes, to compare Mt. Gox exchange rates to other leading platforms, and to verify daily BTC-to-USD exchange rates.

3.1.1 Dataset Validation

While it is impossible to directly ascertain the validity of the Mt. Gox transaction data, it was possible to conduct checks to ensure that the data is consistent. As a first check, it was verified that the total buy transactions matched the total sell transactions in number and aggregate value.

Through further study of the Mt. Gox leaked data, it was found that there were many duplicate entries in the dumped files. It was also found that the Mt. Gox registry occasionally contained multiple entries for transactions with the same user ID, transaction time, transaction type (buy/sell) and transaction amount. In order to clear this problem two forms of de-duplication were considered. The more conservative approach was to treat each (user ID, timestamp, transaction type, amount in bitcoin, amount in Japanese Yen) tuple as unique (de-duplication strategy 1). Removing such duplication narrowed the data from approximately 18 million rows to 14 million rows[1]. The more aggressive de-duplication strategy was to consider (user ID, timestamp, transaction type, amount in bitcoin) tuples as unique (de-duplication strategy 2). Using this strategy, transactions that are reported at the same time but at different exchange rates are treated as duplicates. The difference between the two strategies is roughly 1 million rows; strategy 1 results in a dataset with approximately 14 million rows, and strategy 2 results in a dataset with approximately 13 million rows. De-duplication strategy 1 was chosen for Section 5.1 and de-duplication strategy 2 was chosen for Section 5.2.

As a further check of data consistency the Mt. Gox de-duplicated data was compared with other reported data. For example, the Mt. Gox transaction volumes were compared to the daily totals reported on `bitcoincharts.com`. Both de-duplicated datasets were more

[1]Note that each completed transaction has both a buy and sell record, which means that the total number of unique completed transactions is 7 million.

consistent with the daily totals found on `bitcoincharts.com` than those found in the original leaked data.

Figure 3.1 plots the daily differences in transactions between the leaked dataset and the totals reported by `bitcoincharts.com`. Differences are normalized as a fraction of the leaked daily volume. Positive numbers indicate that the leaked data reported higher volume. Note that some difference is expected, particularly if the time zones used in the leaked data and on `bitcoincharts.com` differ. Also, note that there were a few gaps in when data was reported by `bitcoincharts.com` (e.g., in mid-2012 and January 2013). These gaps only affect the comparisons between datasets, not the subsequent analysis.

Overlaid on the graph is a red dotted line on days where DDoS attacks are reported at Mt. Gox, and a blue dashed line for other shocks. From this it can be seen that data are available during the shocks, and there does not appear to be any increase in the disparity between sources on days when shocks occurred.

The top graph reports on de-duplication strategy 1. It can be seen that the transaction volume is always the same or higher in the leaked data. The difference, while volatile, increases somewhat as time passes. The bottom graph reports on de-duplication strategy 2. During 2011, `bitcoincharts.com` reported higher volumes than Mt. Gox tracked internally, but that changed over time and the overall trend lines are similar in both graphs.

Finally, As a final check there were communications made with multiple Mt. Gox users, who confirmed that their own transactions were accurately reported in the leaked data.

From this analysis, it was concluded that the de-duplicated leaked data appears robust enough to provide a reliable picture of the actual levels of trade activity at Mt. Gox. Both de-duplication strategies were employed in the following analysis. Utilizing strategy 1 for the Section 5.1, the impact of shocks to the Bitcoin ecosystem was evaluated and using strategy 2 for Section 5.2 price movements caused by spurious trading activity were examined. It is important to note that the results for these sections remain consistent regardless of the de-duplication strategy used or whether one was used at all.

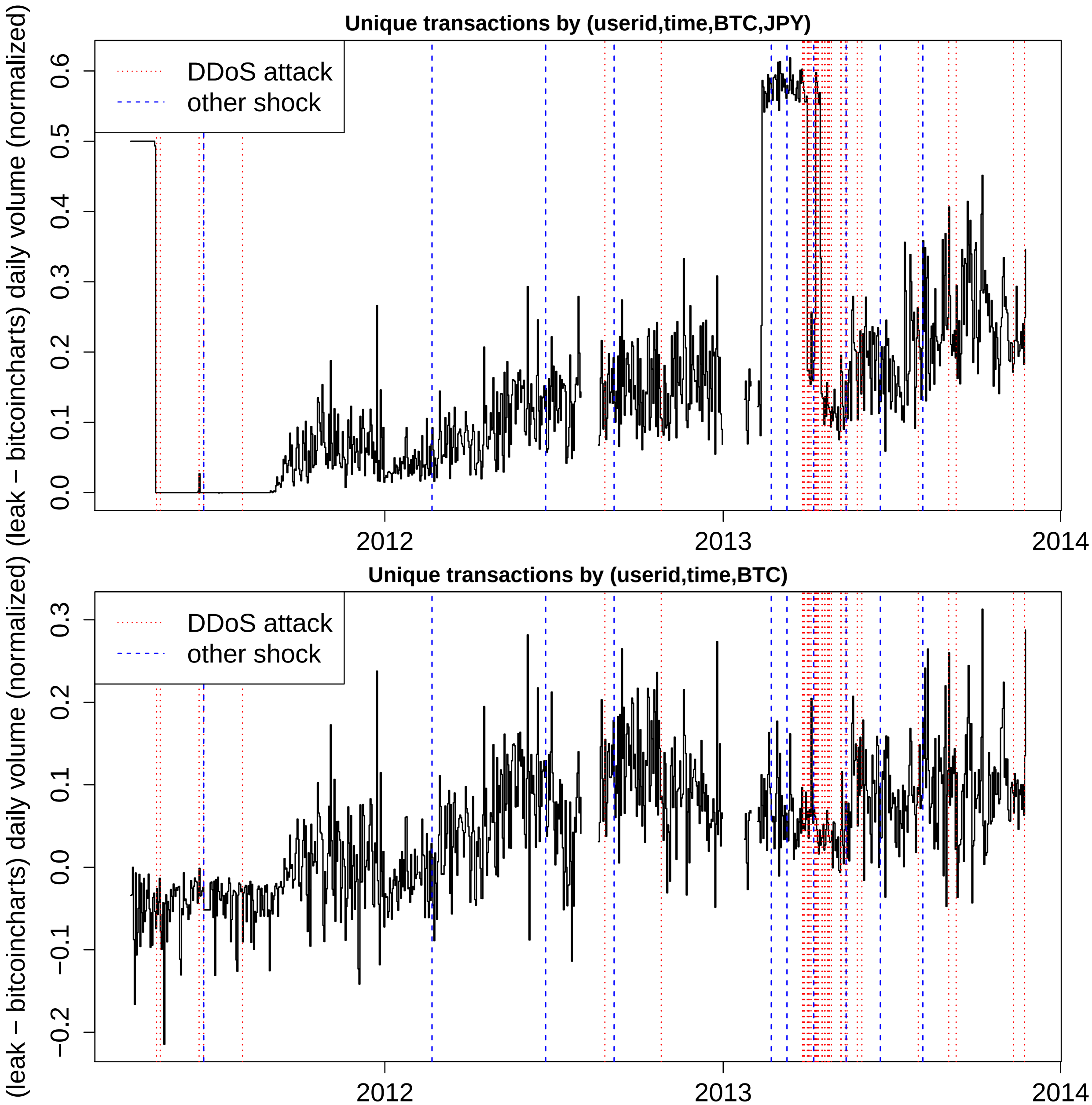


Figure 3.1: Daily differences in transaction volume between leaked dataset and totals reported by `bitcoincharts.com`. Differences are normalized as a fraction of the leaked daily volume. Positive numbers indicate that the leaked data reported higher volume.

3.2 Ethical Considerations

The decision was made to use the leaked Mt. Gox data in this research because the data had already been publicly disclosed. Consequently, any subsequent examination of the data would not add to any existing harms experienced because of the datasets initial publication. In fact, by analyzing the transactions of a prominent dead exchange, it is hoped that light is shed on how denial-of-service attacks could impact todays exchanges.

3.3 Shocks to Mt. Gox

Measuring the impact of denial-of-service (DDoS) attacks targeting the Mt. Gox exchange was of primary interest. It was expected that the attacks would affect the different types of traders on Mt. Gox in different ways. In particular, the expectation was that an attack, for one of two reasons, would lead to a subsequent temporary reduction in "large-volume" trades on Mt. Gox following the attacks. There are two reasons for this. First, large traders probably had better and more up-to-date information than small traders. And second, large traders would struggle to find sufficient depth in the market to complete large-volume trades immediately following a DDoS attack.

3.3.1 D1: Reported DDoS attacks

Three sources of reported outages affecting Mt. Gox were combined: user reports in the `bitcointalk.org` forum, user reports in the /r/bitcoin Reddit sub-forum, and public announcements by Mt. Gox in the press and on social media.

In [54], Vasek et al. measure the prevalence of DDoS attacks on a range of Bitcoin services by inspecting posts on the popular `bitcointalk.org` discussion forum. The information used herein was data published by the authors (available from `doi:10.7910/DVN/25541`), which reported the day a thread describing an alleged DDoS attack on Mt. Gox started. The authors in [54] used a keyword-based classifier to identify candidate threads discussing DDoS attacks, then manually inspected all threads to ensure that a purported DDoS attack was in fact being discussed as opposed to a general discussion of DDoS attacks or their hypothetical impact. Reports were gathered from between February 2011 and October 2013, with 34 attacks on Mt. Gox reported.

The /r/bitcoin forum on Reddit is another popular discussion forum. Following the same procedure as the authors in [54] historical posts using the Reddit API were inspected. In all, we found 8 reported DDoS attacks on Mt. Gox discussed on Reddit, reported between April and November 2013. Three of those attacks were also reported on `bitcointalk.org`.

What's being measured with these data are *reported* DDoS attacks, not confirmed events. It is possible that some of the outages experienced by users were caused by reasons other than a DDoS attack.

Mt. Gox frequently issued press releases via its website and social media when outages occurred. Sometimes the outages were directly attributed to DDoS attacks. Unfortunately, after Mt. Gox collapsed, most of these pages were deleted. As a result, these public statements have been lost forever[2]. However, in a few cases reports could be obtained from third-party websites or Gox's Google+ page which was seemingly forgotten when the other social media accounts were deleted. In total, direct acknowledgment of DDoS attacks by Mt. Gox were found on 9 occasions and some of the attacks were reported by more than one source

Across all three data sources, DDoS attacks were reported on 37 days.

3.3.2 D2: Additional security shocks

DDoS attacks were far from the only adverse event afflicting Mt. Gox during its operational life. The exchange faced pressure from regulators, thefts from users, and self-inflicted IT outages. Ten publicly reported shocks were documented by examining statements made by Mt. Gox that were gleaned from news reports, press releases and social media. Those events are described in Table 3.1.

3.3.3 D3: Confirmed DDoS attacks

Because it cannot be proven that all DDoS attacks reported on the discussion forums

[2] `archive.org` did not preserve the Mt. Gox pages containing public statements.

Date	Description
2011-06-19	Security breach causes BTC fall to 0.01 USD
2012-02-21	Kernel panic triggers outage
2012-06-23	Invalid trading causes outage
2012-09-05	Unplanned trading outage
2013-02-22	Dwolla anti-money laundering efforts cancel USD transfers
2013-03-11	Blockchain fork glitch
2013-04-09	Outage reportedly caused by high trade volume
2013-05-14	Department of Homeland Security seizes cash in court action
2013-06-20	Suspends USD withdrawals
2013-08-05	Announces significant losses due to early crediting

Table 3.1: Additional shocks, other than DDoS, affecting Mt. Gox.

actually happened, a narrow subset of 9 DDoS attacks that Mt. Gox directly acknowledged was also examined.

While the possibility of false negatives such as shock events that transpired but were not verified cannot be eliminated, there is confidence that most events affecting Mt. Gox are included. Herein the scouring of public reports from the two most popular discussion forums and direct acknowledgments by the company indicates that the number of missing events does not affect results in a meaningful way.

3.4 Conclusion

Due to the unfortunate failure of the first popular bitcoin exchange it was possible to gain access to the internal database of Mt. Gox. This dataset of approximately 14 million rows, provided a behind the scenes look at how an early exchange operated, the problems it faced, and the mistakes that were made. Coupling this with publicly available reports on shocks to the ecosystem, a compelling story of how malicious actors might alter the price of a cryptocurrency was available.

CHAPTER 4

IDENTIFYING SUSPICIOUS TRADES

In this chapter, the existence of two suspicious actors reportedly found within the leaked Mt. Gox dataset is verified. Section 4.1.1 covers the first actor with one account referred to as "Markus." Section 4.1.2 goes into great detail about the trading activity within the 49 accounts of the second actor, referred to as "Willy." The information presented in this chapter is a precursor to the analysis preformed in Section 5.2.

4.1 Suspicious Trading Activity

As mentioned in Section 3.1, in the midst of theft allegations in early 2014, Mt. Gox transactional history was leaked. During initial data exploration it was found that a group of users with attributes that differed from the rest of the users existed in the dataset. In particular, every transaction for these users had "??" as an entry for the user country and user state fields. This appeared suspicious as these fields normally contain FIPS location codes, a NULL value, or "!!". When compared to other accounts, one containing the abnormal location values stood out because that account bought and sold bitcoins, while the other accounts only bought. This anomalous trading behavior had been observed before [3]. Therefore, the naming convention found on blogs where individuals are discussing this trading activity and refer to the first account as "Markus" and the remaining accounts as "Willy" is continued herein.[1]

4.1.1 Suspicious Trader 1 - the "Markus" Bot

[1]Although "Markus" sold bitcoin on a few occasions, most of this account's activity involved buying bitcoin.

"Markus" began "buying" bitcoin on 2013-02-14 and was active until 2013-09-27. This user traded on 33 of the 225 days the account was active, and was able to acquire 335,898 bitcoins with a value around 76 million USD. Upon closer inspection the trades made by this account raised many red flags. "Markus" never paid transaction fees and reportedly paid seemingly random prices for bitcoins. Most curious of all, many duplicate transactions were identified in which the amount paid was changed from an implausibly random price to one that was consistent with other trades that day.

"Markus" seemingly paid random rates for the bitcoins acquired. For example, in two transactions that took place the same hour on 2013-04-03, "Markus" paid 0.000374 USD per bitcoin for one transaction and 925,489.67 USD per bitcoin for another.

Table 4.1 shows the wide range of rates that "Markus" paid. The table reports the number of purchases that "Markus" made for different ranges of rates. During the time "Markus" traded, published exchange rates ranged from 20 USD to 229 USD. Hence, any transactions with rates outside this range raise suspicion. In fact, only one quarter of "Markus's" trades fell within this range. Thirteen percent of the time, "Markus" paid less than one USD, while in 821 transactions (3% of the total), "Markus" reportedly paid a rate exceeding 100,000 USD per bitcoin!

Table 4.1: Distribution of USD/BTC rates paid by "Markus"

	$\leq \$0.10$	$> \$0.10$, $\leq \$1$	$> \$1$, $\leq \$20$	$> \$20$, $\leq \$229$	$> \$229$, $\leq \$1K$	$> \$1K$, $\leq \$10K$	$> \$10K$, $\leq \$100K$	$> \$100K$
#	1,050	2,586	6,320	7,009	3,658	4,604	2,311	821
%	3.7%	9.2%	22.3%	24.7%	12.9%	16.2%	8.1%	2.9%

After further investigation, the random exchange rates appear to come from transactions posted before "Markus's" transactions. Table 4.2 illustrates the pattern. Transaction 1362466144485228 was posted with user 238168 buying ≈0.398 bitcoin for 15.13 USD. Every "Markus" transaction that followed (indicated in bold) "borrowed" the Money amounts, and Money_JPY values from the previous transaction. This pattern of behavior was confirmed throughout. Whenever "Markus" bought, the amount this account paid came from a previous unrelated transaction while the number of bitcoins acquired appears random.

Table 4.2: Fraudulent transactions initiated by "Markus" (user ID in bold)

Trade_Id	Date	User_Id	Type	Bitcoins	Money	Money_JPY
1362466099116388	2013-03-05 6:48	238168	buy	0.58932091	22.39419	2094.796
1362466099116388	2013-03-05 6:48	109955	sell	0.58932091	22.39419	2094.796
1362466144485228	2013-03-05 06:49	238168	buy	0.3982007	15.13163	1415.442
1362466144485228	2013-03-05 06:49	132909	sell	0.3982007	15.13163	1415.442
1362466154623847	2013-03-05 06:49	**698630**	buy	1.70382	15.13163	1415.442
1362466154623847	2013-03-05 06:49	96376	sell	1.70382	15.13163	1415.442
1362466154714939	2013-03-05 06:49	**698630**	buy	1	15.13163	1415.442
1362466154714939	2013-03-05 06:49	201601	sell	1	15.13163	1415.442

Occasionally "Markus" would also sell bitcoin, and the bitcoin-to-fiat currency exchange rate for those transactions appears to be correct. For example, on 2013-06-02 "Markus" sold 31.5 bitcoins for 3,757.95 USD, or 119.3 USD per bitcoin which is similar to the average rate paid by users that day. In total, "Markus" sold 35,867.18 bitcoin worth approximately 4,018,681.65 USD in 2,927 transactions on six different days.

As stated in Section 3.1.1, closer attention was paid to the choice of what records needed to be removed when de-duplicating the data. This was due to the fact that several transactions contained duplicate buy and sell rows; see Table 4.3 for an example of those transactions. It is evident that user 201601 sold one bitcoin twice at the same exact time, first to user 698630 for 15.13 USD and second to user 634 for 38.11 USD.

Table 4.3: Duplicate "Markus" Transactions

Trade_Id	Date	User_Id	Type	Bitcoins	Money	Money_JPY
1362466154714939	2013-03-05 06:49	201601	sell	1	15.13163	1415.442
1362466154714939	2013-03-05 06:49	**698630**	buy	**1**	**15.13163**	1415.442
1362466154714939	2013-03-05 06:49	201601	sell	1	38.11000	3564.883
1362466154714939	2013-03-05 06:49	**634**	buy	**1**	**38.11000**	3564.883

Upon closer inspection, the rows containing 15.13163 in the Money column are the original rows for this transaction. In every instance where duplicates were discovered they involved user 698630 and user 634; 634 appeared to "correct" for 698630. There are multiple oddities involving user 634. First, the numeric user ID is extremely low which strongly suggests that it could be an internal account. Second, prior to issuing the "corrected" transactions, user 634 bought and sold a total of 824,297.7 bitcoins between 2011-04-07 and 2012-08-01. The 634 account was inactive for 197 days before being used again in the duplicate transactions involving "Markus."

Table 4.4 summarizes the discrepancies between "Markus"' identities. 2,966 buy transactions made by 698630 were later duplicated as originating from user 634 at market prices. In total, as user 698630, "Markus" reportedly paid 1,080,617 USD for 67,452 bitcoin. When acting as user 634 instead, "Markus" "paid" 2,000,729 USD for the same transactions. This only includes the corrected transactions involving user 634. Any trading activity that occurred before "Markus" was active was ignored. It is worth noting that only the amounts paid for bitcoins were altered, never the bitcoin amount. Additionally, for the 196 transactions where user 698630 sold bitcoin and a duplicate row was found with user 634, no monetary amounts were altered. Only the user ID had changed.

Finally, the majority of transactions by user 698630 were never changed, despite the frequent presence of wild exchange rates. User 698630 only operated between February and September of 2013, and during that time the account purchased 268,446.09 BTC, reportedly at prices totaling 76.4 million USD. Note that this total USD amount should be viewed with caution because it is based on seemingly random exchange rates.

Table 4.4: Summary of "Markus" transactions

	User ID	# Transactions	Total BTC	Total USD
Manipulated Buy	698630	2,966	67,451.61	$1.1M
Manipulated Buy	634	2,966	67,451.61	$2.0M
Unchanged Buy	698630	25,407	268,446.09	$76.4M
Manipulated Sell	698630	196	5,049.86	$0.2M
Manipulated Sell	634	196	5,049.86	$0.2M
Unchanged Sell	698630	2,927	35,867.18	$4.0M

4.1.2 Suspicious Trader 2 - the "Willy" Bot

In the case of "Willy," in addition to the circumstantial evidence of sequential use and proximity to “Markus,” the most solid evidence that foul play was involved can be traced to the internal user ID. “Willy” did not use a single ID; instead, it was a collection of 49 separate accounts that each rapidly bought exactly 2.5 million USD worth of bitcoin and never sold the acquired bitcoin. Additionally, the IDs associated with the accounts were abnormally high for the time period in which they were used [3]. The typical account for that time period had IDs that capped around 650000 while the users at the center of “Willy’s” activity had IDs in the 658152-832432 range.

The first “Willy” account became active on 2013-09-27, a mere 7 hours and 25 minutes after “Markus” became permanently inactive, and “Willy’s” activity was tracked until the research data cut-off date of 2013-11-30. After activation, each “Willy” account proceeded to spend exactly 2.5 million USD before becoming inactive. Then the next account would become active and the process would repeat. Unlike “Markus,” “Willy” was apparently interacting with real users. While accounts of those users were “nominally” credited with fiat currency, there is no evidence Willy paid for the bitcoins.

“Willy” traded on 50 of the 65 days before the data cutoff. In total, the collection of accounts acquired 268,132 bitcoins for around 112 million USD. While “Willy” acquired slightly fewer bitcoins than “Markus,” the “Markus” activity occurred on 33 days spread over a 225-day period. Thus, the “Willy” activity was much more intense. By November 2013, these bogus users had acquired around 600,000 bitcoins.

In addition to the outlined suspicious activity, several user reports can be found detailing Mt. Gox trading API outages during various periods of time in which almost no trading activity was being processed. “Willy’s” brazen trading activity was the exception, and it was continually processed throughout those outages [1]. On 2014-01-07 the trading API was offline for 90 minutes. During that time the only activity being processed followed the exact buying pattern of “Willy” when he was active: 10-19 bitcoins purchased every 6-20 minutes.

Recently, in a trial in Japan, the former Mt. Gox owner, Mark Karpeles, confirmed that the exchange itself operated the “Willy” accounts and that the trades were issued automatically [51].[2]

4.1.3 What May Have Motivated The Operation of These Bots?

The publicly reported trading volume at Mt. Gox included the fraudulent transactions and signaled to the market that heavy trading activity was taking place. It is shown later that, even when the fraudulent activity is set aside, average trading volume on all major exchanges trading bitcoin and USD was much higher on the days the "Markus" and "Willy" bots were active. The associated increase in "non-bot" trading was profitable for Mt. Gox, because it collected transaction fees.

The "Willy" bot could have served another purpose as well. A theory, initially espoused in a Reddit post shortly after Mt. Gox's collapse [2], was that hackers stole a huge number (approximately 650,000) of bitcoin from Mt. Gox in June 2011 and the exchange took extraordinary steps for several years to cover up the loss.[3]

Bitcoin currency exchanges function in many ways like banks. Customers buy and sell bitcoins, but typically those customers maintain balances of both fiat currencies and bitcoins on the exchange without retaining direct access to that currency. If an exchang was trying to hide a huge number of BTC missing from its coffers, it could succeed only so long as customers maintained confidence in the exchange. By offering to buy large numbers of bitcoins, a bot could prop up the trading volume at an exchange and "convert" customer bitcoin balances to fiat money. That could stave off collapse of the exchange only as long as customers who sold bitcoin had enough confidence to leave the bulk of their fiat balance at the exchange. If customers wanted to take out bitcoins, an exchange would immediately have to supply them. On the other hand, if customers wanted to redeem the fiat cash in their accounts, the exchange could delay the withdrawal by saying that its bank was placing limits on how much fiat cash could be withdrawn in a particular time period. This seems

[2]It appears that Karpeles operated the "Markus" bot as well, and this is where most of the prosecutor's evidence against Karpales has focused.

[3]When Mt. Gox folded, it initially announced that around 850,000 bitcoins belonging to customers and the company were missing and likely stolen. Shortly thereafter, Mt. Gox found an additional 200,000 bitcoins. Hence, the total loss was 650,000 bitcoins.

to have happened at Mt. Gox. Some (but not all) customers could not withdraw cash from their fiat accounts during the last couple of months before it ceased operations. By using this strategy, the "Willy" bot could turn the Mt. Gox "bitcoin deficit" into a fiat currency deficit which may have delayed the crash of the exchance.

4.2 Conclusion

Thanks, in part, to misconfigured database fields and suspicious transaction patterns it was possible to identify two distinct "traders" responsible for the sharp rise in the price of bitcoin. Together these bots were able to falsely acquire approximately 600,000 bitcoins worth around 112 million USD over a period of ten months.

CHAPTER 5

ANALYZING SHOCKS ON THE MT. GOX CRYPTOCURRENCY EXCHANGE

Although the developers of Bitcoin implemented safeguards against counterfeiting the cryptocurrency, malicious actors frequently mount attacks against this ecosystem through intermediaries such as exchanges. Section 5.1 details how one such attack, a distributed denial-of-service (DDoS) attack, effected the Bitcoin ecosystem. Then, Section 5.2, examines how the trading activity of two suspicious actors may have driven the price of Bitcoin up more than 800 USD in late 2013.

These analyses are important because exchanges are critical parts of the Bitcoin ecosystem. Exchanges continue to be the main way through which users enter the cryptocurrency market. In exchanges, sellers benefit from a larger number of buyers and buyers benefit from a larger number of sellers. For an exchange to succeed, it must build up confidence among its users. A loss of trust in an exchange can quickly lead to its downward spiral during which buyers and sellers quickly move to another exchange.

5.1 Regression Analysis of Shocks

The regressions used in this study are discussed in the following sections. In Section 5.1.1 a first attempt at explaining how shocks effect trades is examined, utilizing transaction volume and large trades as the dependent variable. Section 5.1.2 describes a more robust model which uses skewness and kurtosis of the daily transaction volumes as the dependent variable.

Before running any regressions the data was examined to see if there were any clear relationships between transaction volume and shocks. By examining figure 5.1 one can see

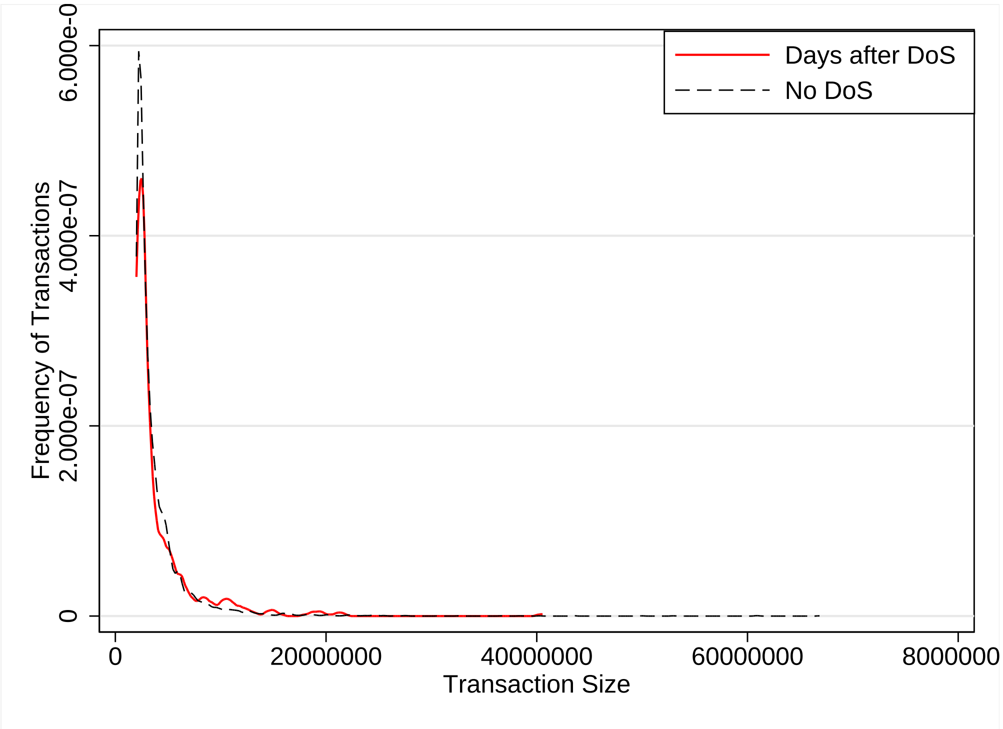


Figure 5.1: Distribution of transactions by amount in JPY on days following a reported DDoS attack (in red) and on all other days (in black)

that there are fewer large trades on days immediately following a DDoS attack. Now that a relationship exists between fewer large trades and DDoS events in the data regressions can be built to measure its significance.

5.1.1 Transaction Volume and Large Trades

Shocks, malicious or otherwise, can increase the probability of a failed trade on an exchange and, in some cases, a shock to the system can erase the value of a transaction completely. Because of this, it is not unreasonable for informed users to temporarily halt trading on an exchange after an attack occurs.

Beginning with the reported events in the $D1$ and $D2$ groups, the effects on the recorded transaction volume were examined. Assuming a linear time trend, the following regression is estimated:

$$TransactionVolume_t = \beta_0 + \beta_1 D1_t + \beta_2 D2_t + \beta_3 Time_t + \epsilon_t \tag{5.1}$$

Transaction volume is the daily trade volume in Japanese Yen (JPY). $D1$ and $D2$ are dummy variables. $D1$ is equal to one (1) on the day following a DDoS attack and zero (0) otherwise. $D2$ is equal to one (1) on the day following the other 10 shocks described in Section 3.3.2. The $Time$ variable is a time trend, and ϵ is the error term. Subscript t indicates that the data used consists of daily observations.

Since a drop in large transactions in the days following a DDoS attack is hypothesized the largest daily transaction (denoted Max. Transaction) is used as an independent variable. As in equation 5.1, a time trend is employed, and the following regression is estimated:

$$Max.Transaction_t = \beta_0 + \beta_1 D1_t + \beta_2 D2_t + \beta_3 Time_t + \epsilon_t \tag{5.2}$$

To gain a broader understanding of how shocks effect trades the daily number of large transactions was computed and used as an independent variable. Although the thresholds used can be debated, similar results were found with all that were tried. Taking into account the exchange rate, as Mt. Gox used JPY for its internal storage, large transactions are defined as those exceeding 1,000 USD. Again, a regression with the same dependent variables was employed:

$$LargeTransactions_t = \beta_0 + \beta_1 D1_t + \beta_2 D2_t + \beta_3 Time_t + \epsilon_t \tag{5.3}$$

5.1.2 Skewness and Kurtosis

Since the data used consists of daily aggregates listed in chronological order, certain problems associated with using time series data needed to be dealt with. Prior work has shown that attacks are more likely to occur during periods with higher transaction volume and high liquidity [54]. Because of this, regressions described in equations 5.1, 5.2, and 5.3 would all be susceptible to endogeneity bias. To solve this issue, skewness and kurtosis of the daily transaction volume was used as independent variables. Including these two new measures is important for the following reasons.[1] First, while trade volume grows over time,

[1]Results from equations 5.1, 5.2, and 5.3 are reported in Table 5.1; but, because of the potential endo-

the skewness and kurtosis of the daily trades does not change. Second, changes in skewness and kurtosis are not likely to lead to an increase in the likelyhood of a DDoS attack.

Heavy-tailed distributions have higher skewness and kurtosis values. On the Mt. Gox exchange, a high number of small transactions and a low number of large transactions are generally seen. Therefore, if a DDoS attack leads to fewer large trades, the skewness and kurtosis will also fall. In these regressions the natural log of both the skewness and kurtosis are used; however, the results remain robust. Although skewness and kurtosis can be negative, the distribution of trades on Mt. Gox is highly skewed towards smaller transactions giving the distribution a long right tail. Therefore, in this data, skewness and kurtosis are always positive[2] and no problems arise when using the natural log of skewness and kurtosis in the following analysis.

Similar to the regressions in 5.1, $D1$ and $D2$ are included; $D1$ is, again, the key independent variable. Mirroring the earlier regressions, $D1$ is a dummy variable equal to one (1) on the day following a DDoS attack and zero (0) otherwise. In the event that a DDoS attack lasted for more than one day two alternatives are considered: (1) definition $D1$ as the day after the end of the ongoing attack, or (2) extend the define of $D1$ to include the second, third, etc. day of an attack as "days after an attack." Results are robust to either of these definitions.[3]

Also included in these regressions are independent variables such as the number of users on the exchange denoted by $Users$, the total volume of the exchange denoted by $TransactionVolume$, and a time trend denoted by $Time$. Even though $Users$ and $TransactionVolume$ are co-determined in the data, there should be no correlation between these variables and the error term while using skewness or kurtosis as the dependent variable. Therefore, no bias introduced and ordinary least squares (OLS) regressions were utilized. However, due to the possibility that errors are not identically and independently distributed

geneity bias, the parameter estimates from these OLS regressions are also potentially biased.

[2]Summary statistics are reported in Appendix A

[3]When a dummy variable for the day the attack is taking place is added, the results are qualitatively unchanged. That is, there is reduced volume the day following the attacks and the coefficients on the lagged variables are essentially the same.

the regressions were run using robust standard errors. The main results are derived from the following regressions:

$$ln(skewness)_t = \beta_0 + \beta_1 D1_t + \beta_2 D2_t + \beta_3 ln(TransactionVolume)_t + \beta_4 Users_t + \beta_5 Time_t + \epsilon_t \tag{5.4}$$

$$ln(kurtosis)_t = \beta_0 + \beta_1 D1_t + \beta_2 D2_t + \beta_3 ln(TransactionVolume)_t + \beta_4 Users_t + \beta_5 Time_t + \epsilon_t \tag{5.5}$$

5.1.3 Regression Analysis

Table 5.1: Transaction Volume and Large Trades

VARIABLES	(1) Transaction Volume	(2) Max. Transaction	(3) Large Transactions
D1	-2.826e+07	-700,953	-104.6
	(1.306e+08)	(1.265e+06)	(277.3)
D2	1.588e+08	1.559e+06	311.4
	(1.963e+08)	(1.901e+06)	(416.8)
Time	1.053e+06***	13,140***	2.246***
	(76,263)	(738.5)	(0.162)
Constant	-2.334e+08***	-2.215e+06***	-537.5***
	(4.064e+07)	(393,531)	(86.28)
Observations	924	924	924
Adjusted R-squared	0.171	0.255	0.172

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The regression results from looking at the effects of D1 and D2 events on transaction volume and large trades on Mt. Gox are, unfortunately, inconclusive. Table 5.1 shows the estimated coefficient for D1 is negative, as hypothesized, however, this result is not significant. This may be due to the previously discussed endogeneity bias, which would lead to upward biased estimates. The estimated coefficient for D2 events is positive, but also insignificant. These estimates may also suffer from an upward bias[4]. This shows that the endogeneity bias is a severe handicap when it comes to identifying what really happens after users realize that a DDoS attack has occurrred.

Table 5.2: Skewness and Kurtosis

VARIABLES	(1) ln(Skewness)	(2) ln(Kurtosis)	(1.1) ln(Skewness)	(2.1) ln(Kurtosis)
D1	-0.276**	-0.560***		
	(0.094)	(0.184)		
D2	-0.0766	-0.160		
	(0.146)	(0.289)		
Users	-0.000144***	-0.000247***	-0.000129***	-0.000218***
	(1.97e-05)	(3.84e-05)	(2.41e-05)	(4.62e-05)
ln(Transaction Volume)	0.327***	0.640***	0.329***	0.643***
	(0.0280)	(0.0538)	(0.0276)	(0.0529)
Time	-0.000889***	-0.00167***	-0.00089***	-0.00167***
	(0.000113)	(0.000214)	(1.07e-04)	(2.05e-04)
Constant	-2.358***	-4.192***	-2.414***	-4.280***
	(0.435)	(0.834)	(0.428)	(0.820)
DDoS			-0.2298**	-0.4390**
			(0.112)	(0.214)
Lagged DDoS			-0.1155	-0.2406
			(0.111)	(0.212)
Other			-0.3806*	-0.7337*
			(0.218)	(0.417)
Observations	924	924	925	925
Adjusted R-squared	0.17	0.20	0.18	0.20

Standard errors in parentheses
Robust Standard errors are employed
*** p<0.01, ** p<0.05, * p<0.1

In Table 5.2, the results for the preferred models containing skewness and kurtosis as the dependent variables are reported. This table shows that a DDoS attack causes both skewness and kurtosis to change in the days following the attack. Kurtosis drops by a staggering 56 percent following a DDoS attack, and similarly, skewness drops by approximately

[4] The high values of adjusted R-squares are due to the extremely significant time trend in the data.

28 percent following an attack. The sign of the estimated coefficient for D2 events is now negative, as expected. However, it still lacks significance. This result suggests that DDoS attacks had more serious effects than any other type of shock experienced by the Mt. Gox exchange[5].

In both regressions the natural log of the transaction volume was included as a control variable. The estimated effect, as expected, is both positive and significant in both cases. Excluding this variable had no effect on the main results.

Due to the fact that different specifications for the DDoS variables were used in the following section, Section 5.2, those variables were included as 1.1 and 2.1 in Table 5.2. Replacing D1, and D2 in these models are three dummy variables named DDoS, Lagged DDoS, and Other. DDoS takes on the value one (1) on days where an attack occurred, and zero (0) otherwise. Lagged DDoS takes on the value one (1) on the day immediately following an attack, and zero (0) otherwise. And, Other takes on the value one (1) on days where a non-DDoS event occurred, and zero (0) otherwise.

The results between the two specifications are very similar with the main difference being the significance of the Other/D2 group. The difference between these two variables comes from the specification; D2 flags the day after the event and Other flags the day of the event. Since these results are relatively unchanged, the following sections utilize the original variable definitions.

In addition to examining transaction distribution fluctuations through negative shocks to the ecosystem, a regression was run that included a positive measure that might explain some of the movement. Weekly Google trends data was included for the "blockchain" keyword. The results for this regression can be seen in Table 5.3. The addition of the trends variable did not change the results for the other variables in the regression, and trends are not a significant driver for differences in skewness and kurtosis.

[5] Regressions with a variable that is the interaction between "D1" and time were also run. The main results are qualitatively unchanged, namely that there are fewer large trades following DDoS attacks. Interestingly, the coefficient on the interaction term is positive and "borderline" significant at the 10 percent level. This suggests that, over time, large traders became slightly less sensitive to the attacks.

Table 5.3: Skewness and Kurtosis With Trends Explanatory Variable

VARIABLES	(1.2) ln(Skewness)	(2.2) ln(Kurtosis)
D1	-0.266**	-0.540**
	(0.113)	(0.218)
D2	-0.068	-0.1496
	(0.170)	(0.326)
Users	-0.0001468***	-0.0002523***
	(2.38e-5)	(4.57e-5)
ln(Transaction Volume)	0.3298983***	0.6468517***
	(0.0285)	(0.055)
Time	-0.000893***	-0.0016941***
	(1.08e-4)	(2.08e-4)
Constant	-2.440***	-4.357973***
	(0.446)	(0.856)
Trends	.001047	0.0021941
	(0.001)	(0.001)
Observations	924	924
Adjusted R-squared	0.16	0.19

Standard errors in parentheses
Robust Standard errors are employed
*** p<0.01, ** p<0.05, * p<0.1

Robustness Analysis: It is desirous to know whether or not the regression results reported in table 5.2 are robust. Therefore, in this section, the results from four robustness regressions are reported in Table 5.4. In regressions 1 and 2 regressions 5.4 and 5.5 are rerun including this time the variable D3. Here, D3 takes on a value one (1) when Mt. Gox directly acknowledged a DDoS attack, and zero (0) otherwise. The variable D1-without-D3 only includes DDoS attacks not acknowledged by Mt. Gox. The regressions show that non-acknowledged DDoS attacks against Mt. Gox resulted in a 36.5% reduction in skewness and a 74.2% reduction in kurtosis. Additionally, attacks acknowledged by Mt. Gox also resulted in a reduction of skewness and kurtosis, but this effect is not significant[6].

Regressions 3 and 4 use an alternative specifications for D1. Here, D1-alt-without-D3 is a dummy variable that takes on the value one (1) for every day of a DDoS attack,

[6] This may be because a very small number of attacks acknowledged by the Mt. Gox exchange was found.

excluding the first day, and zero (0) otherwise. The results for these regressions show that the original findings are robust to this alternative definition as well.

In general, regressions 5.4 and 5.5 were found to be robust to the following:

- Including or excluding a time trend.
- Including or excluding transaction volumes and/or the number of users.
- Estimating 5.4, and 5.5 in levels and not logarithms.
- All combinations of the above.

Table 5.4: Robustness Analysis

VARIABLES	(1) ln(Skewness)	(2) ln(Kurtosis)	(3) ln(Skewness)	(4) ln(Kurtosis)
D1-without-D3	-0.365***	-0.742***		
	(0.086)	(0.165)		
D1-alt-without-D3			-0.241**	-0.497**
			(0.092)	(0.177)
D2	-0.0663	-0.140	-0.0789	-0.165
	(0.148)	(0.292)	(0.146)	(0.288)
D3	-0.0535	-0.150	-0.0208	-0.0825
	(0.243)	(0.453)	(0.246)	(0.460)
Users	-0.000147***	-0.000252***	-0.000145***	-0.000248***
	(2.0e-05)	(3.9e-05)	(2.0e-05)	(3.9e-05)
ln(TransactionVolume)	0.328***	0.644***	0.327***	0.641***
	(0.0282)	(0.0540)	(0.0282)	(0.0539)
Time	-0.000890***	-0.00167***	-0.000885***	-0.00166***
	(0.000113)	(0.000214)	(0.000113)	(0.000214)
Constant	-2.383***	-4.242***	-2.363***	-4.202***
	(0.436)	(0.836)	(0.436)	(0.835)
Observations	924	924	924	924
Adjusted R-squared	0.17	0.20	0.17	0.20

Standard errors in parentheses
Robust Standard errors was employed
*** p<0.01, ** p<0.05, * p<0.1

5.2 Price Manipulation

Being the first popular cryptocurrency, Bitcoin has seen impressive growth since its introduction in 2009. Due to its popularity and market capitalization of 28 billion USD at the time of writing, the market has been flooded with "me too" cryptocurrencies, usually referred to as altcoins. Although these cryptocurrencies often promise to disrupt portions of the existing banking infrastructure through technical innovation, they continue to be the target of attacks by financially motivated criminals.

In this section, a dataset that gives an inside look at the operations of the largest bitcoin exchange at the time, Mt. Gox, is used in order to examine suspicious trading activity on that exchange. The extent of the suspicious trading activity was examined first and showed that it makes up a large percentage of the daily transaction volume on Mt. Gox. Finally, the effects this trading activity had on the bitcoin price on the top four exchanges during the time period of this study was examined.

5.2.1 Preliminary Analysis

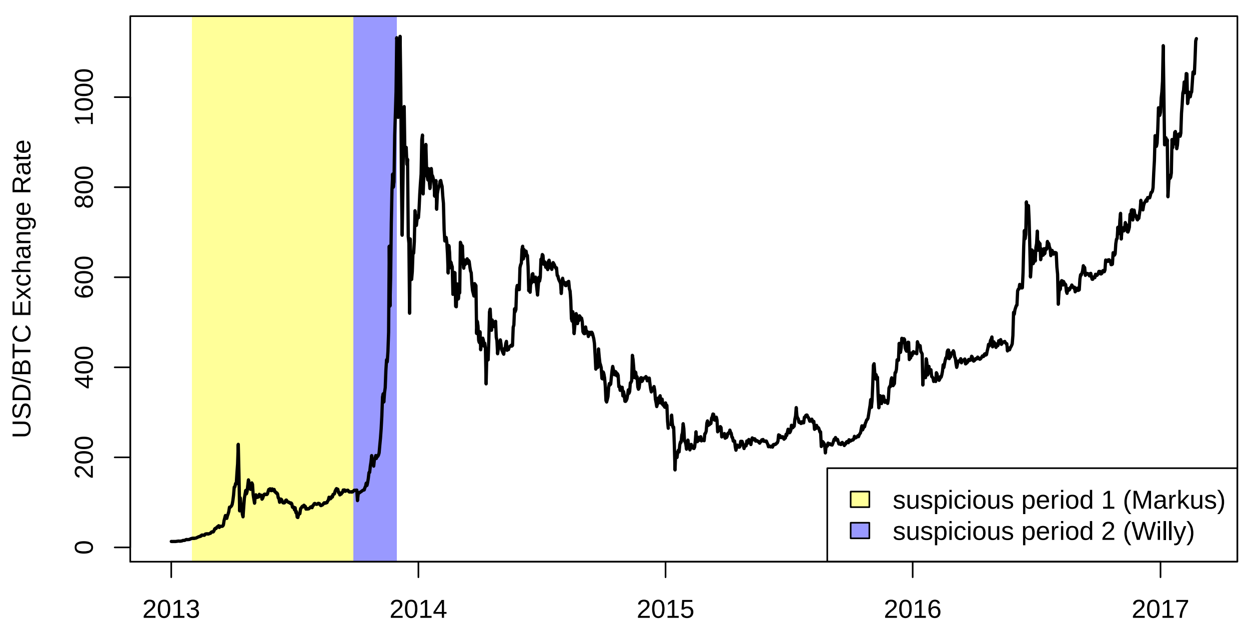


Figure 5.2: Bitcoin-USD exchange rate with periods of suspicious activity shaded.

In this section the effects of the suspicious trading activity on Mt. Gox discussed in detail in Chapter 4 are outlined. When examining figure 5.2, it is clear there was a sharp appreciation in the BTC-to-USD exchange rate while "Willy" was active and trading. However, the overlap does not necessarily mean that "Willy's" activity *caused* the increase in the price. In the following section compelling evidence that Willy's fraudulent trading activity was a driving force behind the dramatic increase in the exchange rate is provided. First the impact on trading volume is outlined and then the impact on the bitcoin is investigated.

Suspicious Purchases and Trade Volume: Although the "Markus" and "Willy" bots had active accounts for 225 days and 65 days respectively before this study's data cutoff, these traders were not always buying bitcoins. However, on the days when they were trading they were big players in the Bitcoin market. On average, "Markus" purchased 9,302 bitcoin a day, which accounted for 21% of Mt. Gox's daily trade volume. When "Willy" was active this bot purchased, on average, 4,962 bitcoins a day. This amount, although lower than the "Markus" mean, still accounted for an impressive 18% of the daily trade volume on Mt. Gox. Figure 5.3 shows a detailed breakdown of daily percent of bitcoin bought by both "Markus" and "Willy." Although there are times during which the price fell when these bots were trading, there are more instances of a price increase. Refer to Table 5.7 for a detailed break down of daily rate movements and price changes.

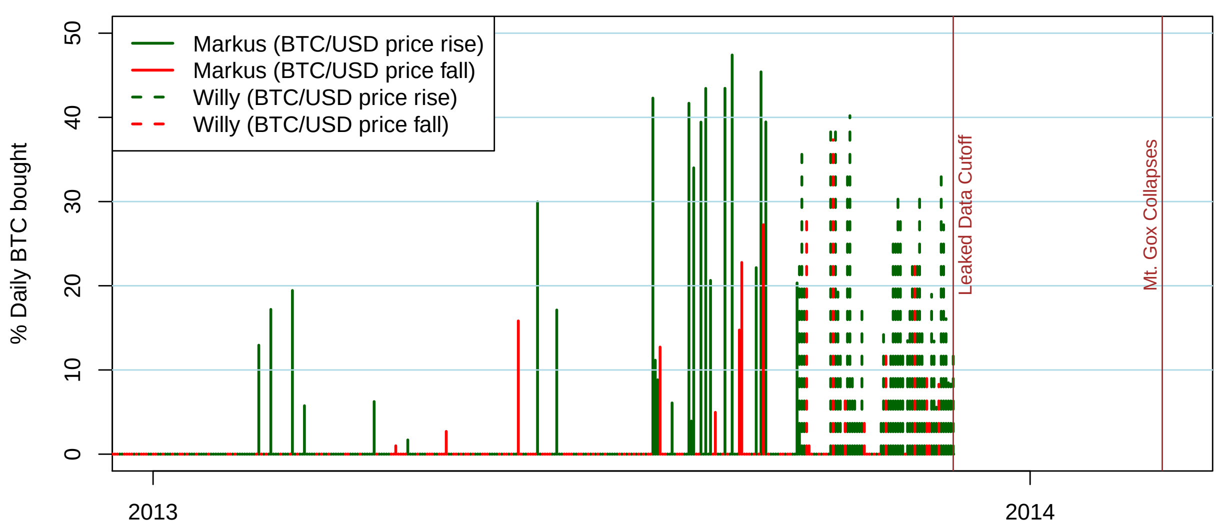


Figure 5.3: Percentage of total daily trade volume at Mt. Gox when "Willy" and "Markus" are active; shaded green if the BTC-to-USD exchange rate closed higher and red otherwise.

Table 5.5: Daily BTC purchased by "Markus" and "Willy" on days they were active.

	Mean	SD	Median	N
Markus:				
BTC purchased	9,302	7,310	5,874	33
% of Mt.Gox daily trade	21		17	
% of total trade at 4 main exchanges	12		10	
Willy:				
BTC purchased	4,962	4,462	3,881	50
% of Mt.Gox daily trade	18		15	
% of total trade at 4 main exchanges	6		5	

In addition to Mt. Gox, the top exchanges at the time of this study's analysis were Bitstamp, BTC-e, and Bitfinex. Even when comparing the "Markus" and "Willy" volume against the aggregate volume for these four exchanges the percentages are still significant. On average, when "Markus" and "Willy" were actively trading, they accounted for 12% and 6%, respectively, of the total trading volume. These four exchanges, Mt. Gox, Bitstamp, BTC-e, and Bitfinex, accounted for over 80% of the total BTC-to-USD trading activity during the period used in this study's analysis.

For this analysis, the data was divided into four equal three-month time periods, beginning December 2012 and ending where the leaked dataset cuts off at the end of November 2013. This allowed for 2.5 months before any activity is seen from "Markus," whose trades mainly occur in Period 3. All of "Willy's" trading activity resides in Period 4.

The substantial increase in trading volume in those periods cannot solely be accounted for by the trades of these two bots. Both "Markus"' and "Willy's" trading activity was associated with a total trading volume that was substantially higher than their own contributions. Table 5.6 shows that during Period 3, when "Markus" was most active, the account was responsible for "purchasing" 8,900 bitcoins per day on Mt. Gox. The median daily volume on Mt. Gox when "Markus" was active during Period 3 was just north of 42,000 bitcoins but it was only 17,421 bitcoins on the days the account made no trades. Similar volume increases can be seen across the other three exchanges as well. Median volume on the four exchanges was 67,691 on the days where Markus was actively trading, but only 31,173

Table 5.6: Comparison of daily BTC volumes on days when suspicious trades occurred and did not.

Buyer	Period	Bot?	Exchange	Daily BTC Volume Mean	Median	N
"Markus"	3	Active	Mt. Gox	10,056	8,901	17
Everyone	3	Active	Mt. Gox	39,619	42,022	17
Everyone	3	Inactive	Mt. Gox	27,672	17,421	75
Everyone	3	Active	Overall	63,984	67,691	17
Everyone	3	Inactive	Overall	46,962	31,173	75
"Willy"	4	Active	Mt. Gox	4,962	3,881	50
Everyone	4	Active	Mt. Gox	30,854	25,939	50
Everyone	4	Inactive	Mt. Gox	17,472	10,444	41
Everyone	4	Active	Overall	90,611	82,779	50
Everyone	4	Inactive	Overall	46,263	29,476	41

on the days with no "Markus" trading activity. For a detailed break down of volumes on individual exchanges, refer to the tables in Appendix A.

The same can be said about "Willy's" activity in Period 4. When "Willy" was actively trading, the account was responsible for "purchasing" a median number of approximately 3,900 bitcoins per day. Furthermore, on the 50 days "Willy" was trading, the total daily median volume on Mt. Gox was approximately 26,000 bitcoins. On the 41 days with no "Willy" trades the median was a more modest 10,444 bitcoins. Similarly, when looking at the volume for the top four exchanges at the time, a median number of just under 83,000 bitcoins a day was purchased when "Willy" was trading, and the median number was only 29,476 bitcoins when no "Willy" trades were observed.

Therefore, although there are differences in the methods employed by each of these bots, the days during either are active are associated with trading volumes much higher than the account's direct contributions. It is likely their high volume trading activity sent a signal to the market and encouraged others to enter and/or purchase bitcoin. This could be one of the reasons why their activity had such a profound effect on the price of bitcoin. In the following section preliminary analysis is conducted on "Markus"' and "Willy's" effect on the price of bitcoin.

Suspicious Purchases and Price Changes: After examining the volume movements on days with and without trading activity associated with "Markus" and "Willy," a relationship between these trades and an overall price rise of bitcoin was expected; not just on Mt. Gox. This expectation was simply due to the fact that an increase in the demand should lead to an increase in the price. Even though this suspicious trading activity only took place on Mt. Gox, how it affected the other exchanges in the Bitcoin ecosystem, namely the top four exchanges mentioned earlier, is of interest.

Table 5.7: Unauthorized activity and price changes on Mt. Gox

		Days with no bots		Days with bots	
		Days	%	Days	%
"Markus"	Daily rate decrease	84	44	7	21
	Daily rate increase	109	56	26	79
"Willy"	Daily rate decrease	9	60	10	20
	Daily rate increase	6	40	40	80
Total	Daily rate decrease	93	45	17	21
	Daily rate increase	115	55	65	79

The descriptive evidence presented here suggests that exchange rates increased on days with suspicious trading activity on Mt. Gox. In this research, the interests lay in whether or not there were differences between price changes on days with suspicious trading activity. For each of the days when "Markus" and "Willy" were active, the exchange rate movement was observed, whether it be positive or negative. In Table 5.7, it can be seen that on days without suspicious trading activity the exchange rate rose 55% of the time. However, on 79% of the 82 days with suspicious trading activity the exchange rate rose. According to a chi-squared test of proportions, it is unlikely this difference was due to randomness ($p < 0.05$).

During the same time period, similar patterns of price appreciation at other cryptocurrency exchanges were observed. Table A.5 shows on days without suspicious trading activity the exchange rate rose on the Bitstamp exchange 55% of the time. However, on the

82 days when “Markus” and “Willy” were trading, the exchange rate increased more than 80% of the time. These increases indicate that the suspicious trading had spillover effects on other exchanges which makes sense as all these exchanges traded the same BTC/USD currency pair.

Table 5.8: Suspicious trading activity: % of days active during each period

	Period 1 2012-12-01 – 2013-02-28	Period 2 2013-03-01 – 2013-05-31	Period 3 2013-06-01 – 2013-08-31	Period 4 2013-09-01 – 2013-11-30
“Markus”	3%	5%	19%	9%
“Willy”	0	0	0	55%
N	90	92	92	91

Table 5.8 shows the percentage of days each of the bots was active for the four periods defined. The “Markus” account was active during at least part of all four of the periods, but was most active in Period 3. In total, “Markus” was active for over 8 months of the dataset. Unlike “Markus,” “Willy's” account activity spans less than three months and is completely contained in Period 4. There is no data on suspicious trading activity more recent than the end of Period 4 because Mt. Gox shut down shortly after.

Table 5.9: Average daily rate change (in $) and percentage rate change (in parentheses) in BTC-USD exchange rate by period

	Period 1	Period 2	Period 3			Period 4		
			All	“Markus” active	“Markus” not active	All	“Willy” active	“Willy” not active
Rate change Mt.Gox	0.21 (1%)	1.00 (1.8%)	0.16 (0.2%)	3.15 (2.9%)	-0.51 (-0.4%)	11.61 (2.6%)	21.85 (5%)	-0.88 (-0.2%)
Rate change Bitstamp	0.23 (1.1%)	1.02 (2.1%)	0.02 (0.1%)	2.35 (2.3%)	-0.51 (-0.4%)	10.99 (2.6%)	20.37 (4.9%)	-0.45 (-0.05%)
Rate change Bitfinex	. .	0.92 (1.3%)	0.04 (0.1%)	2.14 (2.2%)	-0.44 (-0.3%)	10.75 (2.7%)	19.54 (5%)	0.03 (-0.07%)
Rate change BTC-e	0.22 (1%)	1.05 (2.1%)	-0.1 (0.01%)	1.81 (1.9%)	-0.53 (-0.4%)	10.30 (2.6%)	19.22 (4.8%)	-0.58 (-0.07%)
N	90	92	92	17	75	91	50	41

In table 5.9, how the daily exchange rate (closing price minus opening price) fluctuated, on average, is shown for each of the four main cryptocurrency exchanges[7]. Since a majority of the suspicious trading activity is clustered in Periods 3 and 4 those two periods are the focus of this analysis. Periods 1 and 2 can be seen as a benchmark.

In Period 3, where “Markus” activity is concentrated, the price of bitcoin remains relatively flat. However, if the days in which “Markus” bought bitcoin versus the days in which the account was inactive are reviewed the average daily price increase is higher on the days when “Markus” was active. This holds true across all of the top four exchanges.

In Period 4, when “Willy” was active, much more impressive increases in the price of bitcoin are seen. Separating Period 4 into days that “Willy” was trading and days that “Willy” was not trading shows that the difference is much more dramatic. On three of the top four exchanges the price of bitcoin fell, on average, when “Willy” was not active. On the 50 days “Willy” was active, the price of bitcoin increased on the top four exchanges, on average, by a minimum of 19.22 USD a day.

Also included in parenthesis in Table 5.9 is the daily return, which is the typical for measuring the performance of assets. Instead of dollar values, these are the percentage change in the daily exchange rate.[8]

Table 5.9 shows that, across all four top exchanges, the average daily returns ranged from 1.9 to 2.9 percent when “Markus” was active. On the days lacking “Markus” activity the average daily return was slightly negative for those same four exchanges. Similarly, when “Willy” was active the daily average return was between 4.8 and 5 percent across the top exchanges, and, again, was slightly negative for all four exchanges when “Willy” was not active.

The results in Table 5.9 show that the suspicious trading activity of “Markus” and “Willy” could have caused the impressive increase in the exchange rate of bitcoin. In the following section, regressions to control for other possible effects on the exchange rate are

[7]Each of these exchanges allows for 24 hour trading, so the closing rate on one day is the opening rate on the following day. Bitfinex lacks Period 1 as it was not active until April 2013.

[8]This measure is calculated by subtracting the opening price from the closing price and then dividing by the opening price.

run.

5.2.2 Regression Analysis

The analysis in the previous section provides convincing evidence that the suspicious trading activity on the Mt. Gox exchange may have affected prices on all exchanges. In this section, regression analysis to control for other events, like distributed denial of service (DDoS) attacks that may have caused the exchange rate to fluctuate are used. Regressions with the dependent variables being (1) the absolute daily price changes and (2) the daily returns on all four exchanges are run.

Daily Price Change: The following regressions are run:

$$RateChange_t = \beta_0 + \beta_1 Markus_t + \beta_2 Willy_t + \beta_3 DDoS_t + \beta_4 DayAfterDDoS_t + \beta_5 Other_t + \epsilon_t \tag{5.6}$$

$$Returns_t = \beta_0 + \beta_1 Markus_t + \beta_2 Willy_t + \beta_3 DDoS_t + \beta_4 DayAfterDDoS_t + \beta_5 Other_t + \epsilon_t \tag{5.7}$$

In equation 5.6, the dependent variable, *RateChange*, is the normalized daily difference in the BTC/USD exchange rate, i.e., the closing price minus the opening price, divided by the opening price[9]. In equation 5.7, *Returns* is the dependent variable, which in this case is the daily difference in the BTC/USD exchange rate, i.e., the daily difference between the opening and closing price.

Both regression equations employ the same independent variables. *Markus* and *Willy* are both dummy variables that take on the value of zero (0) on days with no activity, and a value of one (1) when each of the respective bots are actively trading. *DDoS* is

[9] The closing price on day t is equal to the opening price on day $t+1$ and since these exchanges employ 24 hour trading, the opening and closing prices are at 24:00 GMT.

a dummy variable that takes on the value one (1) on the day a DDoS attacks occurs on Mt. Gox and zero (0) otherwise. $DayAfterDDoS$ is a dummy variable that takes on the value one (1) on the day immediately following DDoS attacks on Mt. Gox and zero (0) otherwise.[10] Similar to $DDoS$, $Other$ is a dummy variable that takes on the value one on the days where a non-DDoS event occurred and zero otherwise. ϵ is a white noise error term[11], and the subscript t refers to time. With the exception of Bitfinex, which was not in operation during the entirety of Period 1, there are a total of 356 observations for each of the top four exchanges.

The intention for regressions 5.6, and 5.7 was not to estimate the supply or demand for bitcoin. Instead, interest was in estimating the effect of changes in the right-hand-side variables on the left-hand-side variables (daily rate change and percent daily returns). Measuring the coefficients in both of these reduced-form regressions was where the interest lay. Summary statistics for the variables used in these regressions can be found in Appendix A.

It can be seen in Table 5.10 that the coefficient representing "Willy" is both positive and significant. Therefore, there is a strong correlation between the price increase of bitcoin on the Mt. Gox exchange and "Willy's" trading activity. Confirming the results from Section 5.2.1, the estimated coefficient for "Willy's" activity was 21.65 USD, compared to 21.86 USD in the earlier section. The results for the other top exchanges at the time are similar to those for Mt. Gox; the BTC/USD price rose on average 19 USD - 20 USD a day when "Willy" was actively trading. The estimated coefficients are consistent with the "estimates" from the summary statistics in Section 5.2.1[12].

None of the other estimated coefficients from this analysis appear to be significant. The estimated coefficient associated with "Markus's" activity is positive on all four ex-

[10] Most of the DDoS attacks during the time period studied were targeting the Mt. Gox exchange. This is most likely due to the exchange's popularity at the time.

[11] Auto-correlation of errors was checked for by calculating the Durbin Watson (DW) statistic for each regression. The value of DW is not statistically different from two in any of the four cases, which strongly suggests that there is no auto-correlation and a white noise error term is appropriate.

[12] Controlling for other factors, the estimates of the constant show the price change on days when the bots were not active was essentially zero.

Table 5.10: Examining Price Changes Within Mt. Gox and the other Exchanges

Independent Variables / Dependent Variable	Mt.Gox Rate Change	Bitstamp Rate Change	Bitfinex Rate Change	BTC-e Rate Change
"Markus"	2.79	3.24	2.06	2.37
	(0.72)	(0.96)	(0.31)	(0.71)
"Willy"	21.65***	20.21***	19.23***	19.04***
	(6.66)	(7.18)	(3.63)	(6.81)
DDoS	-2.38	-1.67	-1.87	-2.01
	(-0.55)	(-0.44)	(-0.26)	(-0.54)
Day After DDoS	-3.50	-3.25	-2.9	-2.68
	(-0.80)	(-0.86)	(-0.41)	(-0.72)
Other Attacks	7.16	5.70	7.35	5.61
	(0.82)	(0.75)	(0.44)	(0.75)
Constant	0.37	0.30	0.45	0.32
	(0.28)	(0.26)	(0.17)	(0.28)
N	365	365	244	365
adj. R^2	0.10	0.12	0.037	0.11

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

changes, however, it lacks significance. This suggests that "Markus's" spread out trading activity was not associated with a large increase in the daily BTC/USD exchange rate. Interestingly, none of the shocks appear to influence the BTC/USD exchange rate. While this does not prove "Willy's" activity drove the price of bitcoin up, it suggests that it was likely the cause of the sharp bitcoin price increase.

Daily Percentage Returns: It is common for researchers in finance to examine daily returns to currencies in percentage returns, i.e., closing price minus opening price divided by opening price. Repeating the previous analysis, daily percentage returns were used as the dependent variable.

Table 5.11 shows that activities of "Markus" and "Willy" led to returns that were

Table 5.11: Examining Percent Price Changes Within Mt. Gox and the other platforms

Independent Variables / Dependent Variable	Mt.Gox % Rate Change	Bitstamp % Rate Change	Bitfinex % Rate Change	BTC-e % Rate Change
"Markus"	0.0371**	0.0434***	0.0272*	0.0348**
	(3.18)	(3.55)	(1.66)	(2.90)
"Willy"	0.0433***	0.0423***	0.0469***	0.0413***
	(4.45)	(4.14)	(3.54)	(4.12)
DDoS	-0.0182	-0.00758	-0.00391	-0.00903
	(-1.40)	(-0.55)	(-0.22)	(-0.67)
Day After DDoS	-0.0144	-0.0128	-0.0167	-0.0111
	(-1.10)	(-0.94)	(-0.94)	(-0.83)
Other Attacks	0.0374	0.0234	0.0239	0.0235
	(1.43)	(0.85)	(0.57)	(0.87)
Constant	0.0071	0.0065	0.0032	0.0069
	(1.77)	(1.57)	(0.46)	(1.68)
N	365	365	244	365
adj. R^2	0.075	0.064	0.044	0.054

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

significantly higher than those earned without the bots. On the days with no "Markus" or "Willy" activity the average rate of return, as the estimates of the constant show, was less than one percent. The estimated daily rates of return for "Willy" on the top four exchanges are contained in a fairly tight range, from 4.1% to 4.7% over the average returns when "Willy" was not trading. On days without "Markus" and "Willy" trading activity the percentage rate change remained flat; in other words, there was no percentage change in the exchange. All of the coefficient estimates for "Willy's" activity are significant at the 99% level of confidence.

In the case of "Markus," the results are significant across all four of the top exchanges at the time. However, the range for the daily rate of return isn't as tight as it was for "Willy."

Table 5.12: Skewness and Kurtosis Through Shocks and Insider Trading

VARIABLES	(1) ln(Skewness)	(2) ln(Kurtosis)	(1.3) ln(Skewness)	(2.3) ln(Kurtosis)
D1	-0.276**	-0.560***		
	(0.094)	(0.184)		
D2	-0.0766	-0.160		
	(0.146)	(0.289)		
Users	-0.000144***	-0.000247***	-0.0001218***	-0.0002***
	(1.97E-05)	(3.84E-05)	(2.4E-5)	(4.61E-5)
ln(Transaction Volume)	0.327***	0.640***	0.3385***	0.6616***
	(0.0280)	(0.0538)	(0.0275)	(.0527)
Time	-0.000889***	-0.00167***	-0.0008***	-0.0015***
	(0.0001)	(0.0002)	(0.0001)	(0.0002)
Constant	-2.358***	-4.192***	-2.6079***	-4.649***
	(0.435)	(0.834)	(0.4274)	(0.8189)
DDoS			-0.2719**	-0.5152**
			(0.1117)	(0.2140)
Lagged DDoS			-0.1527	-0.3081
			(0.1104)	(0.2116)
Other			-0.4334**	-0.8284**
			(0.2166)	(0.4149)
Willy Buy Day			-0.3491***	-0.6445***
			(0.0856)	(0.1641)
Markus Buy Day			-0.1615*	-0.3316*
			(0.0965)	(0.1849)
Observations	924	924	925	925
Adjusted R-squared	0.17	0.20	0.19	0.22

Standard errors in parentheses
Robust Standard errors are employed
*** p<0.01, ** p<0.05, * p<0.1

The daily returns when "Markus" was trading range from 2.7% to 4.3% more than when "Markus" was not trading. With the exception of Bitfinex, which, again, was not in operation throughout all four periods, the "Markus" coefficients are significant at the 99% confidence level.

In an effort to understand the price movements of Bitcoin, the analysis found in Table 5.10 and Table 5.11 was repeated with the skewness and kurtosis measures from Section 5.1.2. The results from this analysis were insignificant, and are therefore included in Table A.8 of

Appendix A.

Additionally, skewness and kurtosis measures from Table 5.2 in Section 5.1.3 were re-examined to see if the buying activity associated with Willy and Markus had any impact on the distribution of trades. Dummy variables were added to the regressions for each of the bots buying activity. Similar to previous regressions, on days with bot trading activity the dummy variables took on the value of one (1), and zero (0) otherwise. The results in Table 5.12 show skewness and kurtosis dropped when Markus and Willy were purchasing bitcoin. Therefore, bot activity and DDoS events targeting the Mt. Gox exchange have the same result; fewer large trades.

5.2.3 Limitations

The research presented in this section was made possible through the use of the data leaked from the Mt. Gox cryptocurrency exchange. Without this data it would have been extremely difficult to group transactions by a single user or actor. As mentioned in Section 3.1 this data dump included a unique, numeric user ID. This user ID is not generally shared through exchange data APIs and blockchain data proved to be insufficient for this analysis. The blockchain, on the other hand, gives access to an identifier, in this case a wallet address, that can be used to group transactions by an actor. However, due to the centralized nature of cryptocurrency exchanges only account deposits and withdrawals are stored to the blockchain.

Because of these limitations found within other datasets, generalizing the analysis found within Section 5.2 will require additional methods to group transactions by the same actor.

5.3 Conclusion

The research presented in this chapter shows that, although Bitcoin is a maturing market, its ecosystem offers opportunities for fraudulent attempts to sway the price either up or down. In Section 5.1, the first econometric study measuring the impact of distributed denial-of-service attacks on a Bitcoin currency exchange was conducted. In Section 5.2, off-chain trade data was used to conclude that the suspicious trading activity on Mt. Gox was highly correlated with the sharp rise in the price of Bitcoin during the period studied.

A series of regressions to measure the effect of shocks on transaction volume in were constructed in Section 5.1. Unfortunately, due to endogeneity issues and the rising trend in transaction volume over time, using the transaction volume directly as the dependent variable in the regressions was problematic. Because skewness and kurtosis of the daily transaction volume do not suffer from the same problems as measuring transaction volume directly these measures were employed. With these it was found that on days where DDoS attacks or other shocks occurred, both the skewness and kurtosis decreased. In other words, the distribution of daily transaction volume shifted so that fewer extremely large transactions took place when shocks occurred.

By examining the fraudulent activity of two actors in Section 5.2, it was shown that manipulations can have important real effects. The suspicious trading activities of the bots was associated with a 4% daily rise in the price of bitcoin which, in the case of the second actor, combined to result in a massive spike in the BTC/USD exchange rate in late 2013 from around 150 USD to over 1,000 USD. The fall in value was even more dramatic and rapid and it has taken more than three years for bitcoin to match the rise during the previous period.

CHAPTER 6

MEASURING THE LIFESPAN OF A CRYPTOCURRENCY

In this chapter, the dynamics of the rapidly growing cryptocurrency industry are analyzed. Here, attention is restricted to coins, that is, entities with their own distributed ledger. Tokens, which are entities built on top of coins, are not considered.

A methodology was developed that allowed the definition of volume peaks, price peaks, coin abandonment, and coin "resurrections." The association between entry and exit and other key variables such as price, volume, and market capitalization was examined in order to analyze and understand the intuition underpinning the fundamentals of this market. The after-effects of two periods in which bitcoin prices experienced a sharp appreciation and then an equally impressive depreciation were also examined.

6.1 Methodology

First the data sources used to investigate cryptocurrencies are described. Next, the methodology used to identify peaks in trading volume and price, as well as when coins are abandoned and resurrected is described.

6.1.1 Data Sources

The publicly available data on coins from `coinmarketcap.com` was gathered and used to examine the dynamics in the cryptocurrency industry. This website lists all cryptocurrencies that report pricing and 24-hour trading volume via a public API.[1] Such transparent and easy-to-access information has enabled the website to become the most comprehensive public repository of cryptocurrency trading information. The available data for each cryptocurrency includes daily summary values for the opening, high, low, and closing prices,

[1] This is true as long as at least one such API reports positive trade volume.

trading volume, exchanges, and market capitalization. All monetary values reported by `coinmarketcap.com` are given in USD. Data was collected on 1,082 currencies on 2018-02-07 and yielded 662,837 daily observations, starting from February 2013 up to February 2018.

Because currencies appearing on `coinmarketcap.com` are already being traded, this data excludes coins that have been announced but not yet traded. In order to identify when coins fail prior to public launch, supplemental data was also gathered from the altcoin announcements forum on `bitcointalk.org`. The forum was scraped and all announcements which had the term "coin" in them and did not reference a token platform such as Bitcoin, Waves, or Ethereum were considered. Posts referring to coins that appeared on the token section of coinmarketcap were eliminated. The name of the coin was semi-automatically parsed out and the timestamp of the first post for a given coin was considered the announcement date.

6.1.2 Identifying Peaks, Abandonments, and Resurrections

In order to address exits, peaks in volume need to be identified. Because trade in marginal cryptocurrencies can be dormant for many months only to increase again when investment surges in the industry. Identifying price peaks which indicate the potential profits or losses that may result from trading was also of interest.

"Candidate" price and volume peaks for each cryptocurrency were identified first. A "candidate peak" is defined as a day in which the 7-day rolling average value is greater than any value 30 days before or after. In order to identify only those peaks with sudden jumps in value, a "candidate peak" is further defined as a peak that satisfies two additional criteria:

- The candidate peak value must be greater than or equal to 50% of the minimum value in the 30 days prior to the candidate peak.
- The candidate peak value must be at least 5% as large as the currency's maximum peak.

The resulting peak data was then used to define cryptocurrency abandonment. Each

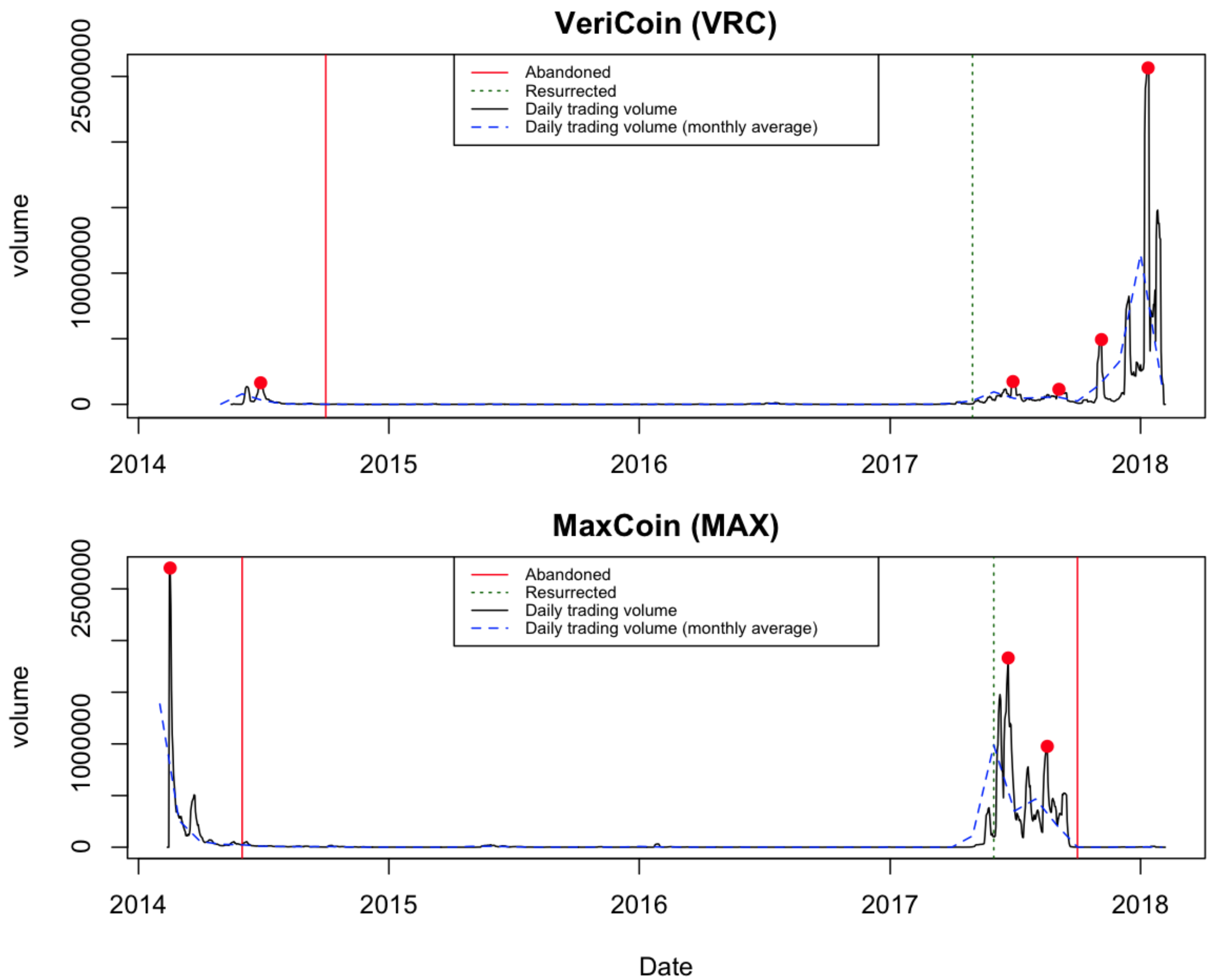


Figure 6.1: Volume plot showing currency abandonment. Red dots indicate peaks.

of the peak values was compared to all of the succeeding daily volume values for each cryptocurrency. Abandonment is defined as follows:

- If the daily average volume for a given month is less than or equal to 1% of the peak volume, the currency is considered *abandoned*.

Unlike other industries, where exit is a "one-way street," currencies don't necessarily stay "dead" when they are abandoned. If the average daily trading volume for a month following a peak is greater than ten percent of the peak value and that currency is currently abandoned, then its status changes to *resurrected*.

Two examples of currency abandonment and subsequent resurrection are shown in figure 6.1. VeriCoin was established in mid-2014, reached an early peak volume of 1.5 million

USD, but was abandoned within a few months. Nearly two years later, in mid-2016, volume jumped slightly, but to less than 10% of the prior peak value. Then, in the spring of 2017, the currency was resurrected, eventually reaching a trading volume more than 15 times greater than its first peak volume of 1.6 million USD.

MaxCoin began trading in early 2014 and quickly reached a peak volume of 2.7 million USD before becoming abandoned less than four months later. The cryptocurrency was resurrected during the 2017 period of massive growth before once again becoming abandoned in October of 2017. During this period peak trading volume did not reach its initial peak value; however, it came close with 1.8 million USD being its highest point. The last abandonment of this currency appears to be a permanent abandonment as it has not yet been resurrected in 2018.

Note that these algorithms work well for the time period covered in this study. However, for real world applications or for more general use, the definitions utilized by these algorithms need to be revisited.

6.2 Coin Results

Beginning with Section 6.2.1, summary measures pertaining to the life of a cryptocurrency, namely the peaks, abandonments, and resurrections are presented and discussed. In Section 6.2.2, the correlation between bitcoin's popularity and altcoin creation and abandonment are analyzed. And finally, in Section 6.2.3, by examining two sharp drops in the price of cryptocurrencies, it is shown that the advantages of being first to market diminish over time.

6.2.1 Peaks, Abandonments, and Resurrections

Peaks: Based on the algorithm developed, almost every coin in the data used had at least one price peak during the time period studied. Out of 1,082 total coins in the data, 1,068 coins had a price peak, yielding a total of 3,508 peaks across all cryptocurrencies. Additionally, almost every coin in the data experienced a volume peak at least once during the time period studied. 1,076 out of 1,082 cryptocurrencies peaked at least once, with a total of 3,828 volume peaks across all cryptocurrencies. Due to the volatility associated with cryptocurrency price, volume is used for a majority of the following analysis.

Not all cryptocurrencies generate the traffic or trading volume that Bitcoin was able to during its peak popularity. A majority of the approximately 2,100 cryptocurrencies available for trade at the time of writing only trade in small volumes. To better study the peak, abandonment, and resurrection characteristics of these coins, the data was split into different groups based on aggregate lifetime trade volume.

	overall	<$1M	$1–10M	$10–100M	$100M-1B	>$1B
# coins	1 082	374	344	183	124	57
# price peaks (total)	3 508	1 426	1 022	531	376	153
# price peaks (median)	3	4	3	2	3	3
% price increase 1st peak (median)	749	418	583	999	1 936	3 441
# volume peaks (total)	3 828	1 734	1 064	468	406	156
# volume peaks (median)	3	4	2	2	3	3
% volume increase 1st peak (median)	3 714	917	1 561	6 915	24 992	90 530
# coins abandoned	475	239	154	50	32	0
% coins abandoned	44	64	45	27	26	0
# abandonments	642	347	192	62	41	0
days abandoned (median)	182	153	184	242	426	—
# coins resurrected	336	183	103	25	25	—
% coins resurrected	71	38	27	13	19	—
# resurrections	452	261	135	30	26	—
months to resurrection (median)	6	5	6	10	19	—
# coins permanently abandoned	190	86	57	32	15	0
% coins permanently abandoned	18	23	17	17	12	0

Table 6.1: Summary statistics on coin peaks, abandonment and resurrection, broken down by total trading volume per coin.

Table 6.1 shows summary statistics for peaks, abandonments, and resurrections for all of the coins in the data used as an overall measure and split into the different volume groups. By considering the number of coins per group, it can be seen that this data consists mainly of coins with low trading volume with some of these coins being unpopular and some of these coins being new offerings. Only 57 cryptocurrencies have a total trading volume that exceeds 1 billion USD, while most, just under two-thirds, of the dataset reported less than 10 million USD in total trading volume. Coins with a total trading volume under 1 million USD, comprised of 344 cryptocurrencies, make up the largest group.

It is not surprising that the group with the most coins also reports the largest number of price and volume peaks. However, the number of peaks per coin is relatively consistent across all size groups in Table 6.1, with the median only deviating by 1 on either side of the overall median of 3.

One important measure is the length of time to the first peak, as it provides an indication of how much profit the early cryptocurrency backers potentially made. Utilizing trade volume for peak discovery, the median time to the first peak is only 40 days and the median percent increase between the first day of trading and the first peak is an impressive 3,714%. Although price data is more volatile than volume data, the price values do not increase at the same rate as the volumes do. Therefore, the price percent increases are lower than the volume percent increases. The median percent increase in price from the first day of trading to the first peak was still an impressive 749%. This means that half of the early coin backers saw at least a seven-fold rise in price by the time the first peak was reached.

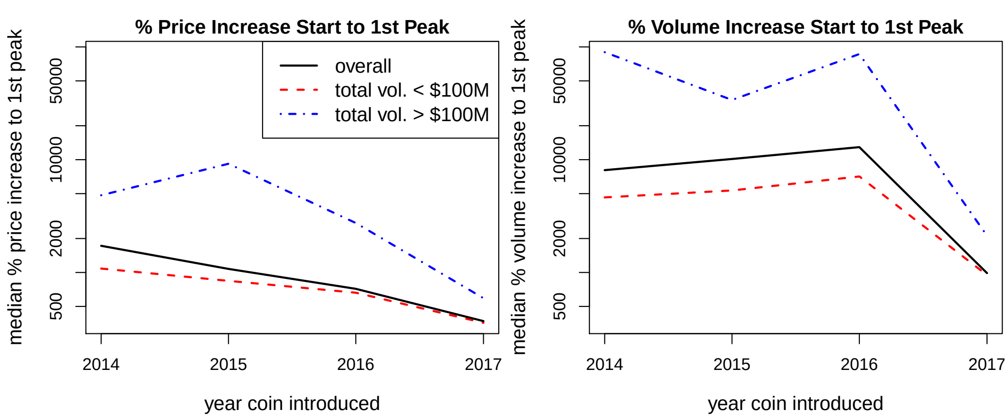


Figure 6.2: Percentage price and volume increase from a coin's launch to first peak, based on the year in which the coin was launched and its size. (Note: the vertical axis is logarithmic.)

Figure 6.2 examines the relationship between when a coin is launched and the magnitude of the initial peak after launch. With the left graph plotting the median percent price rise based on the coin's launch year. Overall, coins launched in 2015 enjoyed a median initial price jump of over 1,700%. This fell steadily, to 1,075% in 2016 and 370% for coins launched in 2017. Coins with higher transaction volume fared even better with the median initial price rise peaking at over 9,000% in 2015.

The initial volume jumps show a slightly different story. Median percentage jumps for the first volume peak were consistently higher than for prices but stayed relatively level for coins launched in 2014 through 2016. However, the median initial volume rise fell sharply in 2017. Taken together, these figures indicate that jumps in trading volume are very high while initial price peaks have become less extreme over time.

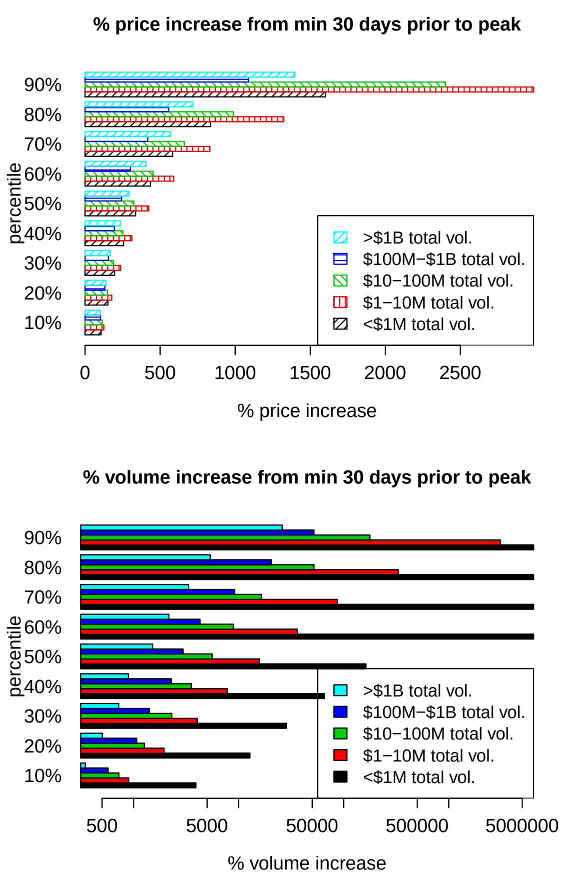


Figure 6.3: Deciles of percent price and volume rises from the smallest value in the month prior to a peak (left graphs) and deciles of percent price and volume falls from the peak to the smallest value in the month following the peak (right graphs).

The distribution of the size of the rise and the subsequent fall surrounding all of the peaks discovered is examined here. Recall from Section 6.1.2 that for a point to be a peak, the value must be, at a minimum, 50% greater than the lowest value in the previous 30 days. Figure 6.3 quantifies how large those increases were and how far the value fell afterwards.

The top two plots in Figure 6.3 show the distribution of the rise and fall associated with peaks in trading price. Beginning with the top left plot, the 10th to 90th percentiles of the peak's percent increase relative to the smallest value in the previous 30 days is shown[2]. To better understand the movements of coins with different popularity, data was further divided by the same coin size used in Table 6.1. The top 10% of price increases range from 1,100% for coins with a total trading volume between 100 million USD to 1 billion USD to a whopping 3,000% for coins with a total trading volume beteen 1 million USD to 10 million USD. Throughout these plots it can be seen that the second smallest volume group rises the fastest, and the second largest volume group rises the slowest.

The top right plot in Figure 6.3 illustrates the falls that occur following a peak value. It can be seen that smaller coins fall farther than the other groups across every percentile, it is important to note how far the falls are across the board. Nine out of 10 coins lost at least 40 to 50% of their value in the 30 days following a peak. Furthermore, half of all coins lost at least 60 to 70% of their value shortly after a peak.

The bottom row of plots in Figure 6.3 shows the movements around a peak associated with trading volume instead of price. As seen in the plot on the left, currencies in the group with the smallest trading volume consistently experience the greatest volume increase leading up to a peak. These plots use a logarithmic scale, so the median increase shown ranges from approximately 1,500% for the most frequently traded coins, to over 100 times that for the coins with the lowest aggregate trading volume. For coins with less than 1 million USD in total trading volume, for more than 30% of the time, there were days with no trading volume within 30 days of hitting a peak.

[2]For this analysis any price or volume rises from peaks occurring in the first week of a coin's operation were excluded, as were any falls within the last week of its operation. This is to accommodate edge effects from the 7-day rolling average used to compute peaks.

Finally, similar to the percentage price decreases following a peak, volume decreases after a peak are extreme. For almost every coin group, trading volume falls more than 90% in the 30 days following a peak.

Abandonments, and Resurrections: Although a majority of the coins saw impressive increases in price and volume leading up to a peak, coin interest was not sustained. From Table 6.1, 44% of all coins were abandoned at least once according to the definition established for abandonment. Out of those 475 abandoned coins, only 336 were "resurrected," that is, the coin's trading volume rose to 10% of the previous peak value following the abandonment.

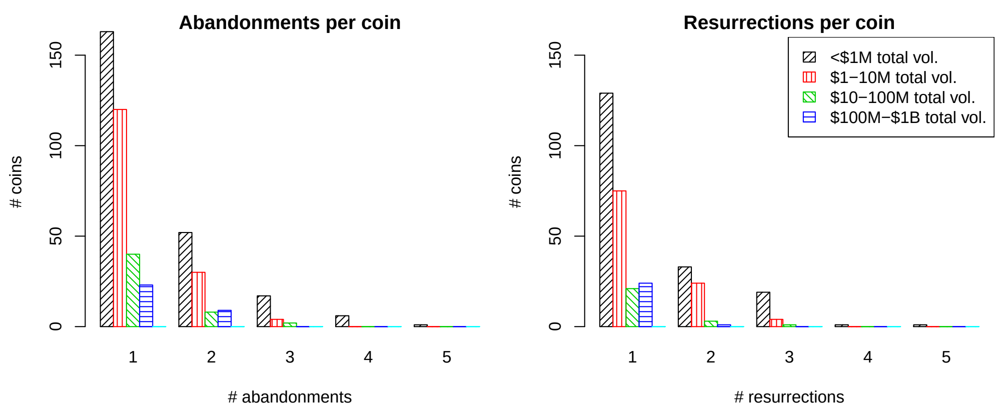


Figure 6.4: Abandonments (left) and resurrections (right) per coin, split by total trading volume.

Knowing that a cryptocurrency can experience multiple price and volume peaks, it is not surprising that, according to the definitions of abandonment and resurrection herein, a coin can experience abandonment and resurrection multiple times as well. A total of 642 abandonments and 452 resurrections were found. Figure 6.4 shows the number, per coin, of abandonments in the plot on the left, and the number of resurrections in the plot on the right based on trading volume. It can be seen that most coins were abandoned just once but some coins were abandoned up to five times. Most of the cases of multiple abandonments and resurrections occurred in in the group with the lowest aggregate trading volume.

A coin's trading volume is a good measure of how popular the cryptocurrency is and it can also be used as an indicator for its abandonment potential. As shown in Table 6.1, 65% of coins with less than 1 million USD in trading volume were subsequently abandoned, compared to just 26% for coins with trading volume between 100 million USD and 1 billion USD. Note that, according to the definition of abandonment established herein, no coins with total trading volume in excess of 1 billion USD have been abandoned. Similar trends follow for resurrection; lower-volume coins are more likely to be resurrected than higher volume coins.

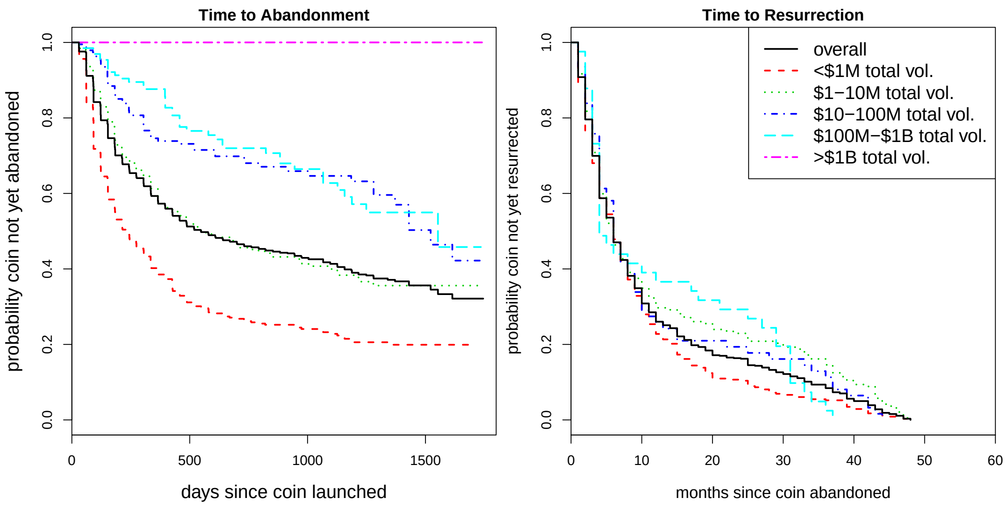


Figure 6.5: Survival probabilities for the time to abandonment (left) and time to resurrection (right).

On average, coins are abandoned within 7.5 months of reaching their peak value with a 4 month median, showing that when coins fail, it can happen quickly. Coin resurrection is a slower process, with a 6 month median overall. In addition to being less likely to resurrect, coins in the higher volume categories take longer to reach a point of resurrection. The median time to resurrection for coins with more than 100 million USD in total trading volume is 19 months.

For a more comprehensive look at the time to abandonment and resurrection survival probabilities were computed, as shown in Figure 6.5, using Kaplan Meier estimators. This enabled the empirical estimation of the time from launch to abandonment using the duration of all coins, even those that had not been abandoned. Overall, the median time to abandonment for coins is 547 days. The time to abandonment varies considerably with the coin's total trading volume. For lightly-traded coins under 1 million USD, the median time from launch to abandonment is just 242 days. By contrast, for coins traded between 100 million USD and 1 billion USD, the median time to abandonment is 1,249 days. Note, once again, that no coins with trading volume in excess of 1 billion USD have been abandoned.

The right graph in Figure 6.5 shows the estimated time to resurrection. The time to resurrection is shorter than the time to abandonment. Overall, the median time from abandonment to resurrection is six months. While there is variation between coin groups, these differences are smaller in magnitude and not statistically significant.

Looking back at Table 6.1, 190 coins remain abandoned at the end of the gathered data. This 18% rate of permanent abandonment is estimated to be lower than the true rate of abandonment. This is due to the fact that some cryptocurrencies fail before being listed on an exchange, and because `coinmarketcap.com` does not list all coins. Failures before being listed can range from the cryptocurrency being a known scam to a lack of resources available to the coin's founders.

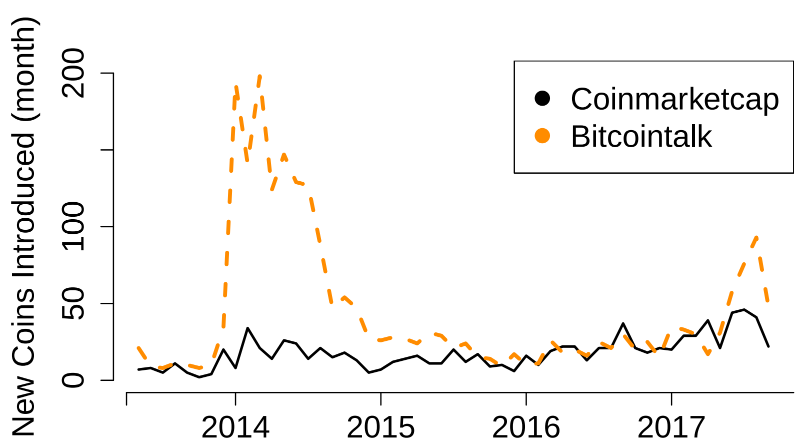


Figure 6.6: New currencies announced on the Bitcoin Forums each month (orange) compared to new currencies traded each month (black).

After a careful review of 12,794 posts on `bitcointalk.org`, 2,361 different cryptocurrencies were found to have been listed on the altcoin announcements section from January 2014 through September 2017. Of these currencies, only 346 appeared later on `coinmarketcap.com`. While 18% of listed currencies become permanently abandoned later, 85% of announced currencies on `bitcointalk.org` failed before ever becoming publicly traded. Figure 6.6 shows that trend over time. In 2014, it was easy to create an alternative currency using the now-defunct `coingen.io`. This service, which at the time cost less than 100 USD, created clones of Bitcoin's code with a few altered parameters. However this fad died off because many of these currencies failed to trade publicly.

6.2.2 Relationships Between Key Variables

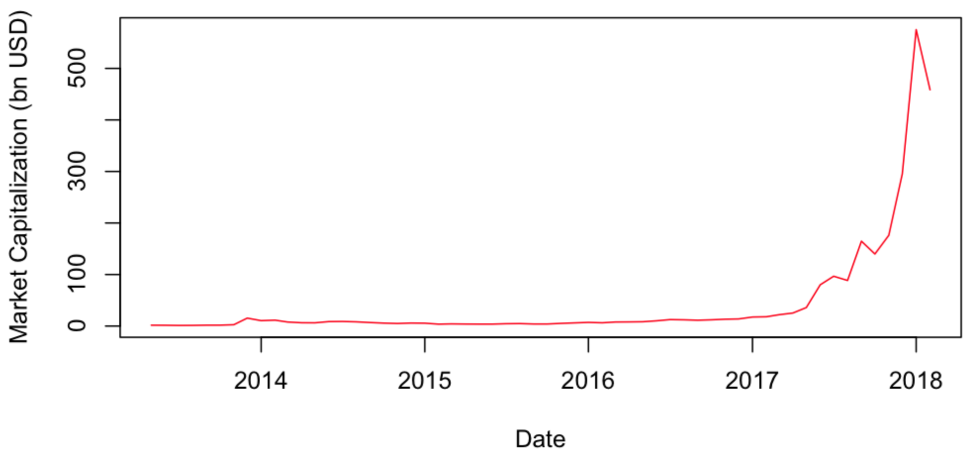


Figure 6.7: Average monthly market capitalization of all cryptocurrencies traded (in Billions of dollars).

As shown in Figure 6.7, throughout 2017 unprecedented growth in the market cap of all cryptocurrencies was seen. This was due, in part, to the massive increase in the price of all cryptocurrencies during that same time period, which was driven by the popularity of Bitcoin. During that period, people tried to profit off of the popularity of cryptocurrencies by issuing their own coin. Some of those coins, such as ZCash and Ethereum, improved existing technologies while others merely tried to "ride the wave" of cryptocurrency success and provided a coin with features no different from one of the popular offerings. This led to an explosion of newly offered cryptocurrencies.

The trend of cryptocurrency issuance follows the fluctuations in the price of Bitcoin and other major cryptocurrencies: when the price is high, more new coins are released with the opposite being true when the price falls. As shown in Figure 6.6, this is most visible during the bitcoin price rise of 2013 which was the first time bitcoin prices over 1,000 USD were seen. When the price fell after the peak the number of new coins being released went down with it. The same was true of the price increase in 2017 when bitcoin peaked at a value of more than 19,000 USD per coin.

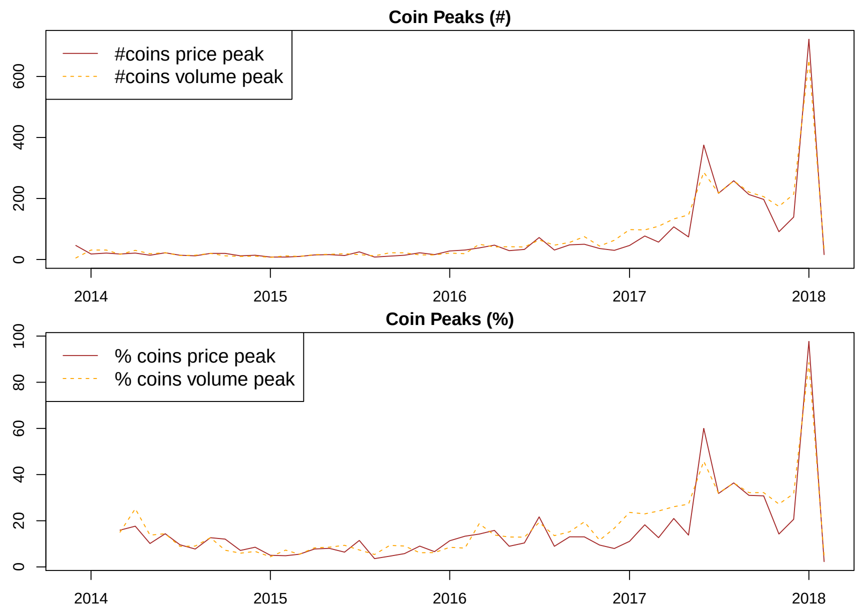


Figure 6.8: Number (top graph) and percentage of active coins (bottom graph) experiencing price and volume peaks over time.

The two plots in Figure 6.8 show a significant correlation between the number of price and volume peaks each month. Until the end of 2016, between 10 and 20% of all coins reported a price or volume peak every month. In 2017, that number increased to 60% of all coins reporting a peak in June of that year. Finally, with the massive spike in cryptocurrency prices at the end of 2017, over 90% of coins reported a peak in January 2018.

The relationships between bitcoin price, coin creation, abandonment, and resurrection are visible in the plots found in Figure 6.9.

Benefiting from the fear of missing the next rapidly appreciating cryptocurrency, when new coins enter the market during periods of high prices, they do so with substantial trading volumes. However, when markets correct and prices fall, some of these cryptocurrencies are abandoned. Looking for this in the data, altcoin abandonment is expected to lag behind the price trend. Well established coins, such as bitcoin, can be seen to endure the volatility

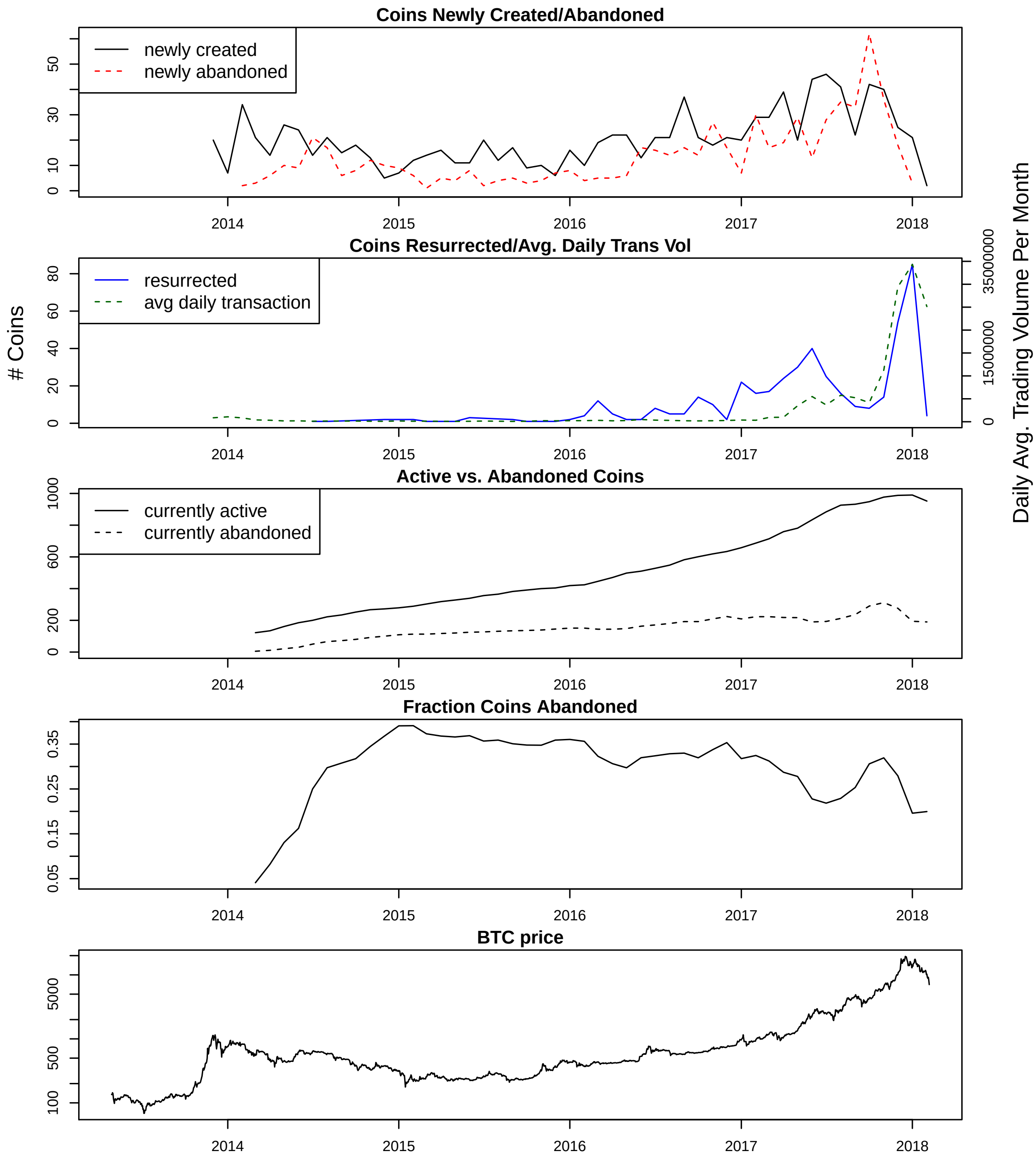


Figure 6.9: Cryptocurrency summary statistics including creation, resurrection, abandonment, and daily average trading volume.

of the cryptocurrency markets because no coins with a total trading volume greater than 1 billion USD have ever been abandoned. However, a new cryptocurrency without a sizable user base will likely suffer from the network effects that push activity towards larger, more established, cryptocurrencies.

The top plot in Figure 6.9 shows the differences over time within the number of newly created and abandoned coins per month and illustrates the spike in new coin creation in early 2014 followed by the spike in coin abandonment later that same year. The rate of creation and abandonment stayed relatively constant through 2015, before accelerating again in 2016, and continuing to rise through 2017.

The second plot from the top examines the rate of coin resurrection per month and the daily transaction volume over time. There is a substantial correlation between the two: coin resurrection remained flat through 2015, slowly grew through 2016, and finally exploded in 2017 which indicates that, as more people trade cryptocurrencies, trade opportunities in little used coins with substantial upward potential are sought.

The next two plots in Figure 6.9 show how the rate of abandonment effects the overall number of coin offerings over time. The first of these shows a steadily increasing number of active coins with the number of abandoned coins also increasing, but at a much slower rate. Additionally, a small rise in abandonment in late 2017 just before the spike in the price of bitcoin can be seen. The following plot better illustrates the impact of that increase in abandonment. In early 2015, almost 40% of all coins were flagged as abandoned. That percentage has slowly declined since then and by January 2018 approximately 20% of coins were considered to be abandoned.

The final plot shows the price of bitcoin, in USD, over time, on a logarithmic scale. Spikes on this plot regularly coincide with points of interest on the other plots in this figure.

	# Coins Abandoned	# Coins Resurrected	# Coins Created	Trade Volume	log_{10}(Average BTC Price)	# Price Peaks	# Volume Peaks
# Abandoned	1						
# Resurrected	0.2080	1					
# Created	0.6107	0.3858	1				
Trade Volume	0.0695	0.7512	0.0959	1			
log_{10}(Average BTC Price)	0.5321	0.7078	0.5053	0.7996	1		
# Price Peaks	0.2756	0.8504	0.4515	0.6524	0.6798	1	
# Volume Peaks	0.3795	0.9007	0.5013	0.7072	0.7756	0.9721	1

Table 6.2: Monthly correlations between key variables in the ecosystem.

To further examine the relationships between these variables, the correlations between them, as seen in Table 6.2, were calculated and these relationships revealed two key trends in the market.

- Resurrection is highly correlated with the number of price and volume peaks, 0.85 and 0.90 respectively which suggests that many of the resurrected coins are riding "the wave" created by the huge increase in the cryptocurrency market. Additionally, trade volume (0.75) and the log-transformed bitcoin price (0.71) are both positively correlated with resurrection (0.75)[3].

- And there is a high positive correlation (0.61) between the number of coins abandoned and the number of new coins created, which suggests that new coins are created to fill gaps left by coin abandonment. Thus, despite the general upward trend in prices and volume, there appears to be some competition between coins. This also suggests that there is substitutability among some of the coins so it is not the case that a "rising tide" is lifting all cryptocurrencies.

These trends can be seen in Figure 6.9. An explanation of these cryptocurrency trends is shown in the bottom plot - the BTC/USD exchange rate over time. Bitcoin continues to be the market leader and to set the trend for other coins as shown by the high correlation values among all variables and the log transformed BTC/USD price.

6.2.3 Bursting of Bubbles and the Changing of the Guard

During the steep decline in bitcoin prices in 2014, Gandal and Halaburda found that trading prices of other cryptocurrencies fell when the price of bitcoin fell [20]. For example, when bitcoin fell from 1,151 USD on December 4, 2013 to 448 USD on April 30, 2014, Litecoin, the second most popular cryptocurrency at the time, also fell from 44.73 USD to 10.90 USD. Even though the drop in bitcoin was steep (-61%), Litecoin fell by even more (-76%). Between April 2014 and February 2016, the price of bitcoin stayed virtually

[3] See Figure 6.9 for a graphic representation of the latter correlation.

Coin	Percent change during each time period 10/16 - 10/17	10/16-12/17	10/16-2/18	Percent change in 12/17 bubble 52 days prior	52 days following	all 104 days
Bitcoin (BTC)	774	2 861	972	239	-64	23
Ethereum (ETH)	2 519	6 018	6 119	134	2	137
Ripple (XRP)	2 201	8 389	7 837	269	-6	245
Bitcoin Cash (BCH)				446	-51	168
Cardano (ADA)				1 311	-14	1 108
Litecoin (LTC)	1 344	7 618	3 161	434	-58	126
NEO (NEO)	20 887	36 554	61 853	75	69	195
Stellar (XLM)	1 395	9 852	14 181	566	43	855
NEM (XEM)	5 162	16 692	11 564	219	-31	122
IOTA (MIOTA)				723	-59	234
Dash (DASH)	2 978	10 704	4 737	251	-55	57
Monero (XMR)	1 348	5 287	2 870	272	-45	105
Lisk (LSK)	2 686	5 553	8 075	103	45	193
Ethereum Classic (ETC)	941	3 310	1 542	227	-52	58
Qtum (QTUM)				174	-28	96
Bitcoin Gold (BTG)				117	-71	-38
Nano (XRB)				2 079	269	7 947
Zcash (ZEC)				117	-35	41
Steem (STEEM)	470	1 106	1 803	111	58	234
Bytecoin (BCN)	2 402	6 566	6 139	166	-6	149
Verge (XVG)	20 215	160 350	166 215	690	4	719
Siacoin (SC)	761	2 792	3 948	236	40	370
Stratis (STRAT)	12 465	26 702	15 328	113	-42	23
BitShares (BTS)	1 072	9 959	4 398	758	-55	284
Waves (WAVES)	948	3 673	1 179	260	-66	22
Dogecoin (DOGE)	382	2 602	1 518	461	-40	236
Decred (DCR)	3 941	9 609	6 779	140	-29	70
Hshare (HSR)				130	-64	-17
Ardor (ARDR)	1 518	7 350	2 908	360	-60	86
Komodo (KMD)				138	-27	74
Ark (ARK)				94	-36	24
DigiByte (DGB)	2 673	9 210	7 869	236	-14	187
PIVX (PIVX)	96 865	196 640	133 964	103	-32	38
ZClassic (ZCL)				137	1 683	4 134
Bitcore (BTX)				136	1	137
Syscoin (SYS)	2 451	5 217	4 165	108	-20	67
GXShares (GXS)				156	-16	116
MonaCoin (MONA)	12 268	44 501	10 345	261	-77	-16
Factom (FCT)	491	1 110	724	105	-32	39
ZCoin (XZC)	259	1 677	1 092	395	-33	232
ReddCoin (RDD)	2 152	5 824	10 889	163	86	388
Nxt (NXT)	747	9 446	2 075	1 027	-77	157
Neblio (NEBL)				-18	110	72
Vertcoin (VTC)	8 140	22 740	6 100	177	-73	-25
DigitalNote (XDN)	569	2 078	5 549	226	159	745
ZenCash (ZEN)				34	-10	20
Achain (ACT)				376	-14	307
Asch (XAS)				58	-35	4
Einsteinium (EMC2)	4 919	135 317	19 688	2 598	-85	294
Metaverse ETP (ETP)				-16	-64	-70
LBRY Credits (LBC)	223	783	527	174	-29	94
BitConnect (BCC)				122	-99	-98
Voxels (VOX)	73	1 484	486	814	-63	238
Steem Dollars (SBD)	9	1 069	198	972	-75	173
Elastic (XEL)				50	-45	-18
Rise (RISE)	4854	13 004	2 816	165	-78	-41
ATBCoin (ATB)				-44	-58	-76
Internet of People (IOP)				120	-65	-24
Regalcoin (REC)				-74	-97	-99
ATMCoin (ATMC)				0	27	27
Tezos (Pre-Launch) (XTZ)				237		
SegWit2x (B2X)				-80	-96	-99
InfChain (INF)				172	-59	12

Table 6.3: Currency movement during different influential time periods.

constant, only falling by 2%, while the USD prices of all other top cryptocurrencies declined significantly with the declines ranging between 69% and 94%.

In the more recent rise and fall of Bitcoin, the currency reached a peak value of 19,498 USD on December 17, 2017. In the fifty-two days preceding that peak (period I), bitcoin rose from 5,905 USD to 19,498 USD. In the fifty-two days following the peak (from December 17, 2017 to February 6, 2018), bitcoin declined to 6,955 USD, a loss of 64% of its value.

Some other currencies saw a significant decline in value similar in magnitude to bitcoin. Table 6.3 shows that eight of the top 14 coins, including bitcoin, experienced steep declines after bitcoin's peak. Another three coins declined slightly. Ethereum did not fall in value at all, Ripple fell by just 6%, and Cardana declined by less than 20%. NEO and XLM, two coins found in the top 10, continued to rise even after Bitcoin's price peak.

This behavior differs from the "rise and fall" at the end of 2013/beginning of 2014. In large part, this appears to be due to innovations by late-entrant cryptocurrencies which led to changes in the cryptocurrency ecosystem. The changes show that Bitcoin's network effect and first-mover advantage may not be able to compensate for the fact that Ripple and Ethereum's platforms have included useful complementary products. Ethereum, for example, has applications outside of simple financial transactions, something that Bitcoin does not really have. Ethereum uses its own token, Ether, to create a decentralized marketplace for computing power and other services. Ripple focuses on sending global payments quickly and cheaply.

These two platforms have cut deeply into Bitcoin's market share. At the beginning of 2017, Bitcoin's market share was above 80%. As of February 2018, Bitcoin's share of the total cryptocurrency market had fallen to just 34%. Ethereum's market share is now 20%, while Ripple's is now 10%. And, it is not just Ethereum and Ripple who are challenging Bitcoin. Many other late entrants are creating platforms for the exchange of digital goods. A changing of the guard may be in process.

6.2.4 Further Analysis of The "Returns" From Top 80 Coins

The returns from the top 80 coins produced during the following three, 52 day periods were examined in more detail. The returns were measured by price changes in percentage terms.[4] Coin rank here is based on trading volume.

- Period I: From October 26, 2017 – Dec 17, 2017
- Period II: From Dec 17, 2017 – Feb 6, 2018
- Period III: From Feb 6, 2018 – March 31, 2018

In Period I, it was found that the median return was 174%. Regardless of the impressive median return, 25% of the top 80 coins lost 18% or more during this period. Conversely, 25% of the coins earned a median return greater than 376%. The variance of the returns was extremely large. The highest return during this period was 2,600 percent!

In Period II, when bitcoin declined significantly, the median return was -32% and more than 75% of the top 80 coins also declined in value. Further, 25% of the coins lost more than 60% of their value. Nevertheless, 10% of the coins increased in value by 58% or more. The variance of the returns was an order of magnitude smaller than Period I. The highest return during this period was 1,683%.

In Period III, when bitcoin remained virtually unchanged, the median return was -36%. More than 75% of the coins declined in value. 25% of the coins lost more than 60% of their value. More than 95% of the valuations fell. The variance was two orders of magnitude smaller than the variance in Period II. The highest return during this period was only 17%.

There is virtually no correlation between Period I and Period II returns, while the correlation between Periods II and III is -0.41. Suggesting that the coins that did not decline in Period II did so in Period III, and vice versa.

Total volume and market capitalization of the coins are uncorrelated with Period I and Period II returns, but total volume and market capitalization are positively correlated

[4] 52 day periods were chosen in order to have three time periods for data analysis: the rise, the fall, and the aftermath.

with Period III returns.

When the top 80 coins were split into two groups: large trade volume vs. small trade volume and large market capitalization vs. small market capitalization the correlation between Period III returns and volume/market cap is found to be much higher for the more important coins, i.e., those with higher trading volume and market capitalization. The analysis suggests that investors/speculators became somewhat more selective in Period III.

6.3 Conclusion

In this chapter, a preliminary analysis of the dynamics in the cryptocurrency market is provided. Methods were devised to identify peaks in trading volume and coin prices for over 1,000 coins. Lower-volume coins were found to face greater risk of abandonment but they were also more likely to rise again. Many of the entrants and resurrected coins were found to be riding "the wave" created by the huge increase in the cryptocurrency market. Nevertheless, the high correlation between resurrection and exit suggests that there is increasing competition among coins.

Evidence consistent with increasing competition is that, unlike the bursting of bitcoin's first bubble in early 2014 when nearly all altcoins followed bitcoin down, some coins were able to increase in value during the 2018 bubble. Although, some coins' fates are indeed tied to bitcoin, there are now clear exceptions to that.

CHAPTER 7

THE RISE AND FALL OF TOKENS

In this chapter, the dynamics of cryptocurrencies are examined through tokens. A majority of this analysis has already been performed on coins and can be found in Chapter 6. Although coins and tokens share surface level similarities, tokens were not included in this original analysis because there are several functional differences that exist between the two entities. Separate analysis of coins and tokens allows the ability to see if these differences provide additional price support or merely stifle growth.

On the surface both coins and tokens rely on cryptocurrency exchanges as the main method of transfer and some form of a wallet for storage. Likewise, with the exception of a few "stablecoins" that tie their value to a fiat currency, tokens prices are just as variable as their coin equivalents.

Although some names like "filecoin"[1] may suggest otherwise, coins typically have a single purpose, which is to be a store of value or a means of transfer. Coins are meant to act in a manner similar to traditional forms of currency. Tokens, on the other hand, provide additional utility on top of use as a means of transfer. Tokens can act as a security or a form of equity providing additional value on top of the tokens' trading price. Utility tokens provide people with access to a product or service, and payment tokens can only be used to pay for goods. Additionally, many cryptocurrency coins rely on mining to distribute new coins to the ecosystem, whereas many tokens are pre-mined and use initial coin offerings (ICO) to distribute the lifetime pool of available tokens. Many users of these tokens theorize that this mass distribution of the entire pool should lead to an increase in value as the supply is slowly used. Another key difference between coins and tokens is the blockchain

[1]Filecoin is in fact a token.

they reside on. Coins typically utilize their own native blockchain. In contrast, tokens are usually built on top of these existing blockchains where developers can take advantage of existing infrastructure to build out apps.

Taking advantage of some of these new cryptocurrency attributes tokens are analyzed to determine the rates of success for ICOs. The use of new ICO data sources required the development of methods for the reliable combination of varying levels of numerical and categorical variables into a single unified dataset. This ICO data is then used to determine token success through amounts raised as well as pseudo-investments.

Finally, adapting the methodology and analysis used in Chapter 6, periods of abandonment and resurrection were identified for each of the tokens observed. Furthermore, relationships between key variables and market entry and exit are explored in an effort to understand the fundamentals of this market.

7.1 Methodology

In this section the data collection methods are outlined. Great detail is given on the issues that arose as well as the methods utilized while massaging the data into a usable state. This section also includes preliminary analysis that explores two different data filtering methods: one for correctness and one for completeness.

7.1.1 ICO Data

The analysis in this chapter relies heavily on ICO-related data. Unfortunately, `coinmarketcap.com`[2] does not record or report much more than daily cryptocurrency price and volume data. Instead of attempting to collect the required data from individual token websites, where it exists, the data used in this analysis was collected from ICO-tracking websites. To locate potential ICO data sources a simple web search was performed and each candidate source was manually inspected for relevant token level data as well as the total number of ICOs tracked. The sources giving access to the most relevant data were then selected for this analysis: ICORating (`icorating.com`), ICObench (`icobench.com`), Foundico

[2]Moving forward `coinmarketcap.com` will be referred to as CMC.

(`foundico.com`), and TokenData (`tokendata.io`). Each source provided a different number of records with varying levels of token overlap. Consequently a substantial effort, outlined in future sections, was put into comparing values, removing outliers, and combining records where possible.

Each of the four sources outlined above provided useful data to varying degrees. All sources provide a token name, a token ICO price, start and end dates for the ICO, accepted currency types, as well as the amount raised[3]. This list is essentially where the TokenData reporting ends; looking at the other sources, they provide more data in the form of exchange listings, countries of origin, location restrictions, blockchain technologies, and token types. A few sources go above and beyond this base list of variables and present users with KYC (know your customer) scores, hype scores, scam scores, and their own ICO/token ratings.

Because cryptocurrencies lack a central body allocating ticker symbol for coins and tokens, collisions frequently arise in which different cryptocurrencies utilize the same ticker symbol. In addition, not all data collected contains this ticker value. Therefore, the main identifier used to combine similar columns was a normalized version of the full token name.

This new token name was then used for a first round of duplicate removal from the combined dataset. If duplicates were encountered after removing ICO-specific tags from the token name, those duplicates were removed from the dataset. Duplicate token listings were also encountered when a service provider offered their token on multiple blockchains. For example, EncryptoTel trades under the same name on the Ethereum blockchain as well as the WAVES blockchain. These duplicate listings were also removed as it complicates the process of calculating returns and token growth. It was discovered that the ICO trackers also follow the progress of coin based ICOs, not just token based ICOs. In total, 191 coin based ICO records were removed bringing the total number of records down from 8,553 to 8,305.

Table 7.1 gives the earliest and latest dates seen in each set of source data as well

[3]Foundico changes this ICO price to the current token value when the ICO is complete so the ICO price from this source has been omitted from the analysis. The remainder of the data from Foundico will remain part of this analysis.

Source	Start	End	Record Count
Foundico	2017-05-19	2028-08-20	3,027
ICORating	2012-07-04	2023-01-24	4,551
ICObench	2015-08-04	2019-11-10	5,672
TokenData	2014-07-08	2019-02-21	2,240

Table 7.1: ICO data source summary statistics.

as the total record count for each. In total each source provided the following number of clean rows: Foundico (3,027), ICObench (5,672), ICORating (4,551), and TokenData (2,240). Combined into a single group the four token sources provided data on 8,305 unique ICOs for cryptocurrency tokens with ICO start dates as early as 2012-07-04 and end dates as late as 2028-08-20.

7.1.2 CMC Data

The original dataset used in Chapter 6 ended on 2018-02-06. If the date range remains the same the new data consists of 765 tokens going as far back as 2014-03-20. If instead, the date range is extended this new CMC data includes 1,894 tokens with data ending on 2019-10-13. This shows that after the price of cryptocurrencies stabilized following the crash at the end of 2017 the number of tokens listed on exchanges rose significantly.

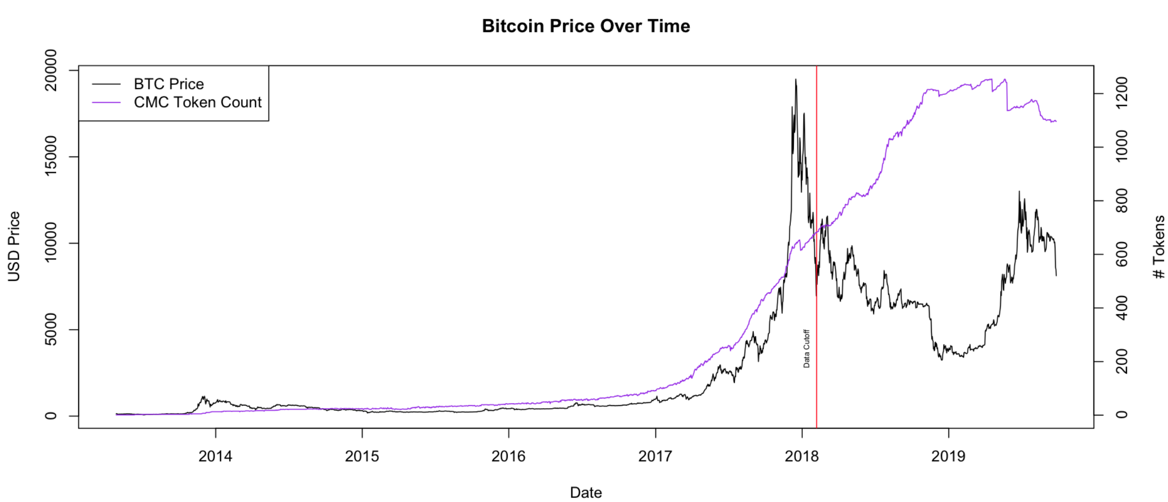


Figure 7.1: Bitcoin price movement over time overlaid with the original dataset cutoff.

The full dataset ending in October 2019 contains roughly 737k daily records, and the dataset ending in February 2018 contains just under 121k daily records. Again, this data contains daily values for open, close, high, and low price along with daily volume and market capitalization values for each of the tokens tracked.

7.1.3 Evaluating and Combining ICO Data Sources

Before the ICO data from the four sources could be used, a method needed to be developed to combine identical ICO attributes into a single, unified ICO record. Although many of the values reported between the sources match, some differences exist and these differences need to be reconciled[4].

Many ICO trackers rely on self-reported values from organizers so the reliability of each of the sources outlined in Section 7.1.1 needed to be decided before the data could be combined. An algorithm was developed that grouped agreement into three different levels: complete, majority, and none. If all reported values for an attribute were equal then the row was marked as “complete agreement,” if a majority ($> 50\%$) of the values reported were equal then the row was marked “majority agreement,” and if no values were equal then the group was marked “no agreement.” In the event that agreement was split 50/50 between four values the group was marked no agreement as it was no better than two non-matching values. During the attribute level agreement calculation each of the sources was also tagged with its position in the result. If a source provided an agreed upon value it was tagged with the resulting complete, or majority flag. Conversely, if the value was not equal to any of the others in the set it was tagged no agreement, even if the other values were tagged complete or majority. These source level agreement values were then combined and can be seen in Table 7.2 for ICO price agreement, Table 7.3 for amount raised agreement, and Table 7.4 and Table 7.5 for start date and end date agreement respectively. Visual representations of these tables can be seen in the figures in Appendix E.

Because of its relatively small record count, coupled with its over-representation in the no agreement category, TokenData was determined to be the least reliable ICO source. Using

[4]Unfortunately, there is no way of knowing if these inconsistencies are simply human error, or misrepresentations intended to attract more token sales.

	Complete Agreement			Majority Agreement			No Agreement		
Data Source	#	%	Sig.	#	%	Sig.	#	%	Sig.
Foundico	0			0			0		
ICObench	718	49.52		67	47.18		1,205	45.10	
ICORating	715	49.31		69	48.59		1,201	44.95	
TokenData	17	**0.01**	(-)	6	4.23		266	**9.96**	(+)

Table 7.2: ICO Price representation by source. Statistically significant under and over-representations are indicated in bold with a (+/-).

	Complete Agreement			Majority Agreement			No Agreement		
Data Source	#	%	Sig.	#	%	Sig.	#	%	Sig.
Foundico	408	37.67		235	39.17		440	27.08	
ICObench	321	29.64		191	31.83		429	26.40	
ICORating	150	13.85		63	10.50		480	29.54	
TokenData	204	**18.84**	(-)	111	**18.50**	(-)	276	**16.98**	(+)

Table 7.3: Amount raised representation by source. Statistically significant under and over-representations are indicated in bold with a (+/-).

	Complete Agreement			Majority Agreement			No Agreement		
Data Source	#	%	Sig.	#	%	Sig.	#	%	Sig.
Foundico	487	**24.06**	(+)	543	**28.91**	(+)	600	**12.18**	(-)
ICObench	681	33.65		581	30.94		1,320	26.79	
ICORating	551	27.22		320	**17.04**	(-)	1,722	**34.95**	(+)
TokenData	305	15.07		434	**23.11**	(+)	685	13.90	

Table 7.4: Start date representation by source. Statistically significant under and over-representations are indicated in bold with a (+/-).

	Complete Agreement			Majority Agreement			No Agreement		
Data Source	#	%	Sig.	#	%	Sig.	#	%	Sig.
Foundico	391	24.11		464	**28.48**	(+)	772	**15.59**	(-)
ICObench	519	32.00		387	**23.76**	(-)	1,671	33.75	
ICORating	482	29.72		422	**25.91**	(-)	1,682	33.97	
TokenData	230	14.18		356	**21.85**	(+)	826	16.68	

Table 7.5: End date representation by source. Statistically significant under and over-representations are indicated in bold with a (+/-).

these source level results the agreement algorithm was updated to adhere to the following rules:

a) If only one value is reported between all sources take that value.

- If the only value is reported by TokenData then treat that group as if no data was reported.

b) If all sources report the same value take that value.

c) If a majority of the sources report the same value (2/3, 3/4, or in some cases 2/4) take the majority value.

- In some cases all four sources reported a value and within those values two sets of 50% agreement existed. These cases were flagged as no agreement because there was no clear majority, and the median was reported.

d) If no majority or complete agreement exist take the median of all values reported.

- If columns consist of string values instead of numerical values then do nothing when no agreement is encountered.
- If columns consist of date values take the median value in the event of no agreement.

Two subsets of the combined data were created using some, or all of the agreement rules outlined above. The first dataset, referred to as "complete," employs all four of the rules for greater coverage. The second dataset, referred to as "correct," only uses rules B and C. Coverage for both datasets can be seen in Table 7.6.

The main driver of the differences between the two datasets comes from the inclusion or exclusion of the single value attribute groups. Of the six attributes observed in Table 7.6, the percentage of groups only containing a single value ranges from 52% to 81% depending on the attribute being observed. When those groups are combined with the groups where

	Complete Result Count	Original Count	% Coverage	Correct Result Count	Original Count	% Coverage
Hardcap	4,184	4,184	100.00	380	4,184	9.08
Amount Raised	2,486	2,677	92.87	642	2,677	23.98
ICO Price	5,627	5,666	99.31	789	5,666	13.93
Location	6,527	6,695	97.49	2,165	6,695	32.34
Start Date	6,409	6,850	93.56	1,741	6,850	25.42
End Date	6,402	6,825	93.80	1,479	6,825	21.67

Table 7.6: Comparison of coverage before and after running data combination algorithm.

no agreement exists it is easy to see that the correct dataset omits a significant number of records and has a much lower percent coverage.

7.1.4 Descriptive Analysis: ICO Data

While examining the combined ICO tracker data, patterns were discovered that give a glimpse into ICOs and the tokens that use them. Figure 7.2 explores the relationship between the price of bitcoin and the three monetary values from Table 7.6. The top two figures utilize the complete dataset, the bottom two use the correct dataset, and the price of bitcoin over time separates the two groups.

Beginning with ICO creation, the two plots follow the movement of bitcoin during the largest bubble to date. The number of ICOs slowly rise, with a sharp decrease in creation at the beginning of 2018, when the bubble burst. This market wide scare was short-lived, however, and ICO creation resumed with rates that were nearly the same as pre-dip creation. By mid 2018 creation had peaked and started to taper off and by the second week of August 2019 weekly ICO creation was back in the single digits for good. Although the two creation plots roughly follow the same trends the maximum ICOs created per week are vastly different. The complete dataset peaked in April 2018 with 153 ICOs created and the correct dataset a month earlier with 64 ICOs.

The only source that reported ICO failure was also decidedly the least reliable source (TokenData), which partially explains why weekly failure rates appear flat. It is also worth noting that this does not take into account subsequent token failure rates, it only examines ICO failure. Failure rates remain flat, but are centered around the bitcoin bubble.

Average USD raised per week examines the reported amount raised against the ICO creation date. Looking past the few spikes these plots also follow the price movements of bitcoin. Traders were looking for promising, low cost bitcoin alternatives that might also give the same returns in the long run. Compared to the price of bitcoin, even after the bubble, the prices of tokens are relatively low.

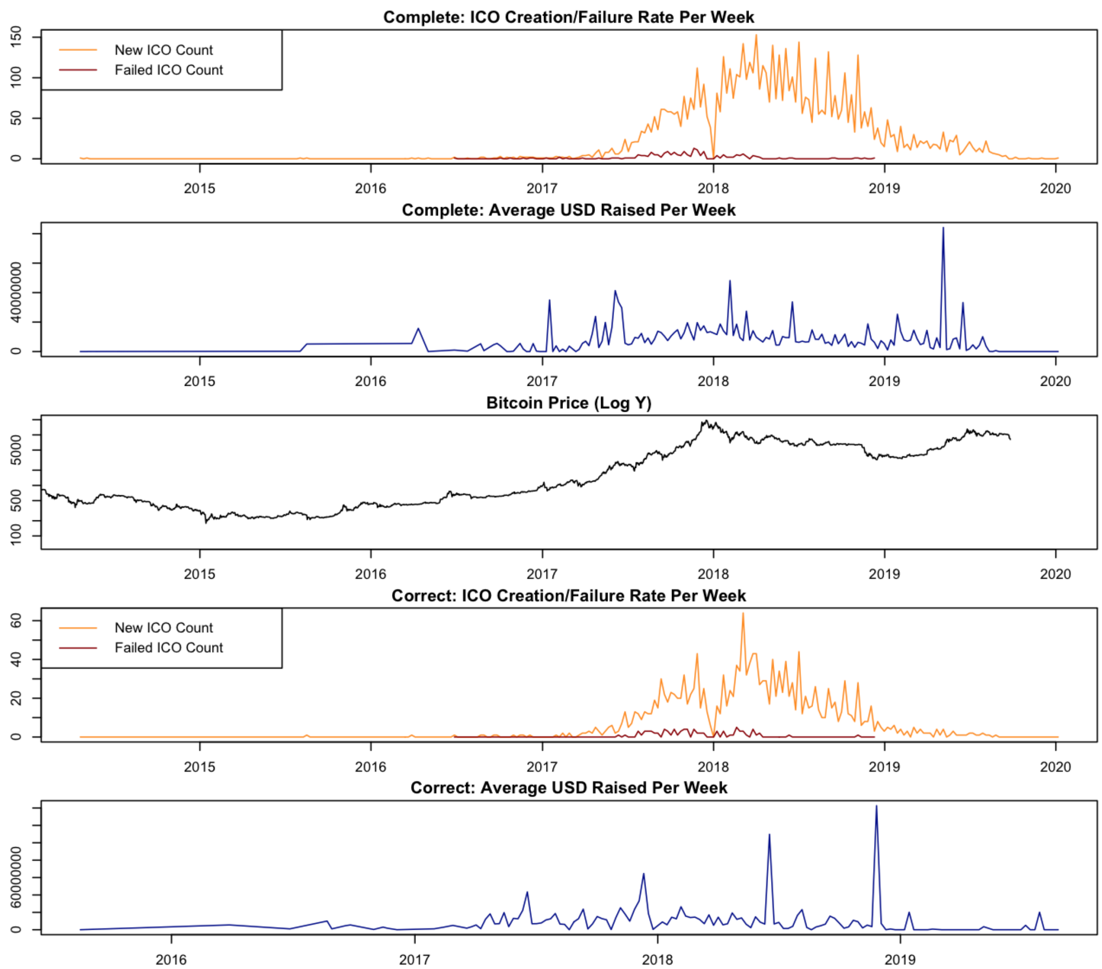
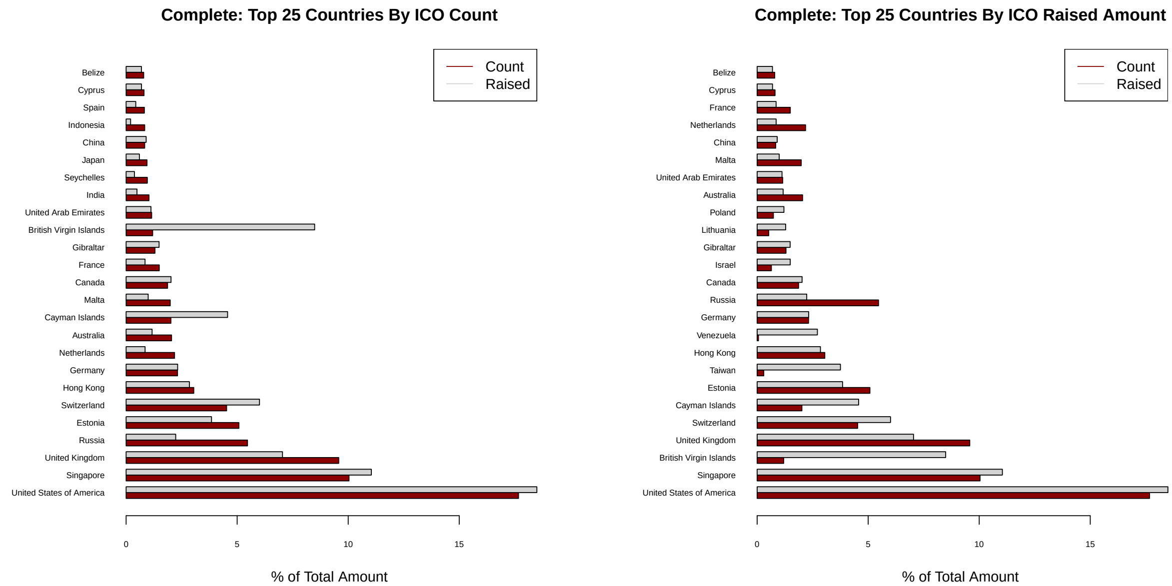


Figure 7.2: ICO creation and failure rates compared to the log scale BTC price movement.

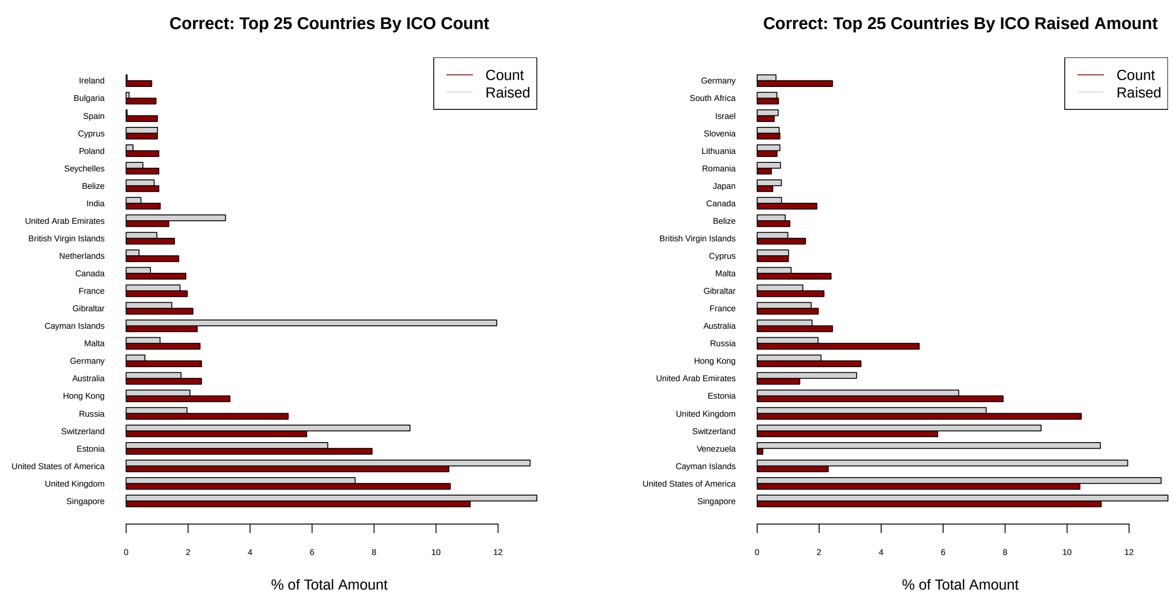
With overall creation rates identified, it is of interest to figure out what countries are responsible for most of these ICOs and their associated tokens. By grouping ICOs by attributes like amount raised, and location the most popular and the most profitable countries for ICOs can be found. Figure 7.3a through Figure 7.4b plot the relationship beetween these variables for both the complete and the correct data.



(a) (b)

Figure 7.3: Complete: Top Countries for ICOs.

In the complete data, Figure 7.3a and Figure 7.4a show the United States as a popular choice with 1,183 ICOs raising 5,011,092,493 USD. Singapore, the second most popular ICO destination raised 2,991,745,897 USD through 672 ICOs.



(a) (b)

Figure 7.4: Correct: Top Countries for ICOs.

Conversely the correct data shows Singapore to be the top location by count and amount raised, with 242 ICOs raising 879,661,769 USD. The U.S. is a close third by count with 227 ICOs raising 865,179,858 USD; when sorting by amount raised the U.S. is ranked second.

An interesting difference between the complete and the correct datasets is the ranking between the top countries. Venezuela's ranking is 10th in the complete dataset and 4th in the correct dataset when sorting by amount raised. According to the data this country only had 4 ICOs, but raised 735,000,000 USD. There are a few countries that exhibit the opposite behavior, instead of a few ICOs raising a significant amount of money they are host to many ICOs that regularly do not meet set hardcap goals. Germany, for instance, was home for 53 ICOs raising a paltry 40,333,913 USD. As mentioned at the end of Section 7.1.3, these differences are a result of single value attributes being excluded from the complete data.

The analysis in the following section is highly dependent on the number of tokens that make it to `coinmarketcap.com` following their ICO. Through discovery and data comparisons, it was found that an increasing number of tokens are never listed on `coinmarketcap.com`. The number of ICO tokens that make it to `coinmarketcap.com` over time can be seen in Figure 7.5.

Initially the number of ICOs being created and the number of tokens that made it to `coinmarketcap.com` was very low. Starting in mid-2016 more and more tokens began to be listed on the data aggregation service, but the acceptance lagged far behind the rates of ICO creation. By early 2018 the percentage of ICO tokens listed on `coinmarketcap.com` plateaued and by the end of the data only 11% of ICO tokens had been listed.

One unexpected issue was the gap that exists between the ICO start time and the time at which the token appears on `coinmarketcap.com`. There is an mean time delay of 100 days (69 day median) for the complete data and 91 days (61 day median) for the correct data.

This means if anything of interest occurs within that rough 3.3 month window whether it be high, low, or flat it is missing from this analysis. Not all tokens trade within that

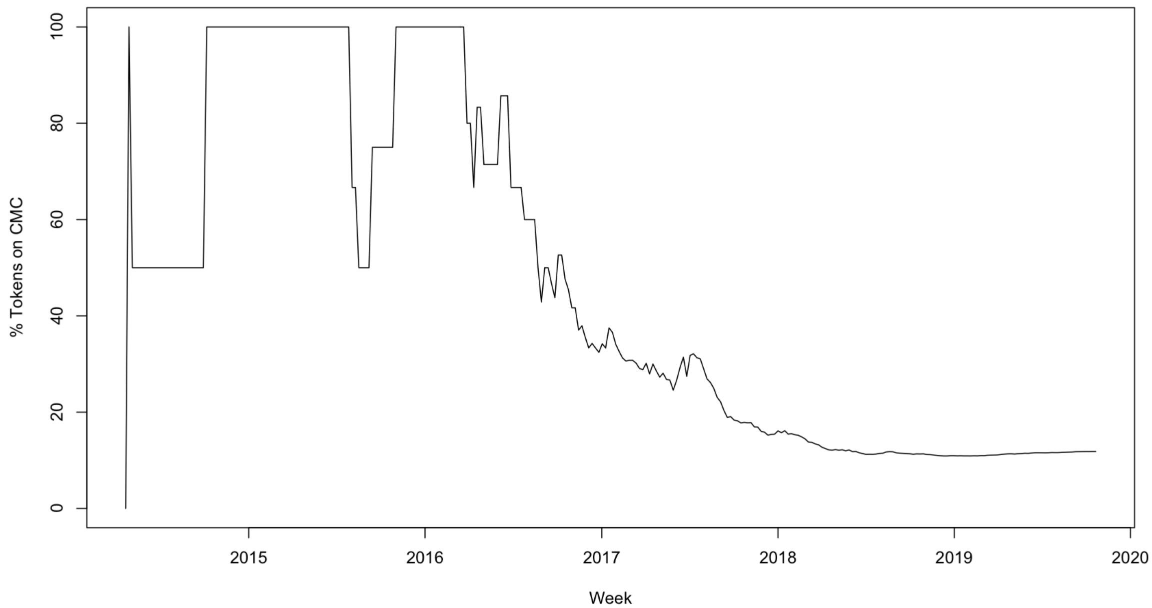


Figure 7.5: Percent of ICO tokens that make it to CMC.

window, though some gaps in the data do exist. Figure 7.6 shows the distribution of the differences between the reported ICO start date and the first date the token was observed on `coinmarketcap.com`.

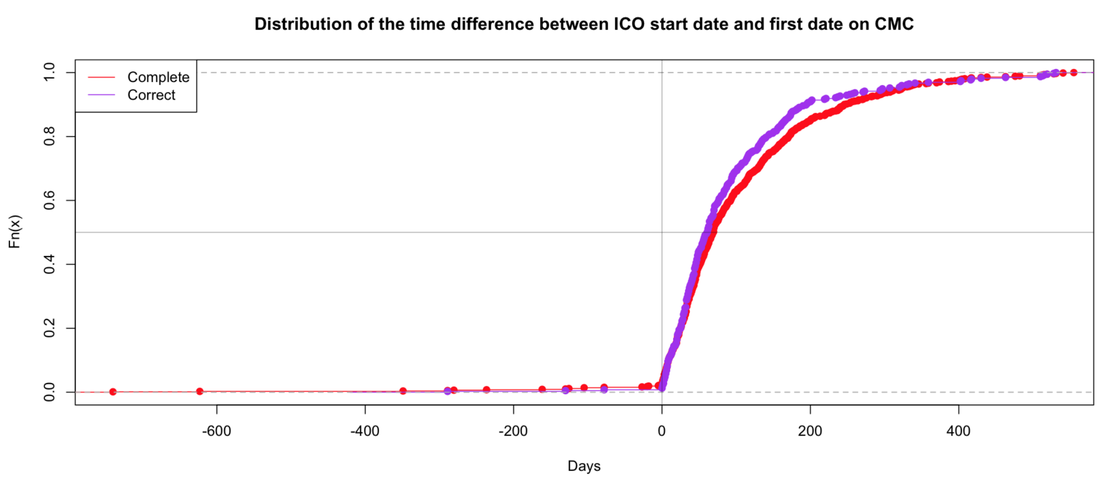


Figure 7.6: Time between ICO start date and first date on CMC.

There 18 tokens in the complete data and 3 tokens in the correct data with negative time differences (the token was found on `coinmarketcap.com` before the ICO start date). The negative time differences could be a result of human error, or it could be because the token was created and actively trading before the ICO was launched. Tokens do not need to be new and tradeless to be used for an ICO sale.

7.2 Measures of success

Success can be a difficult metric to define and reliably measure. In this section, token success is defined through the use of two different methods. In the first method the hardcap set by ICO organizers is compared to the funds raised during the Initial Coin Offering (ICO). If the hardcap value is met or exceeded then the ICO is flagged as successful and if the hardcap value is not met then the ICO is flagged unsuccessful. The second method explores success through simulated investments in tokens available at the first of each year with returns calculated at the first day of each subsequent year. Successful "investments" yield positive returns.

The analysis in this section relies on the complete dataset for its ICO values.

7.2.1 Hardcap and Token Returns

The "hardcap" value associated with an ICO is meant to be the maximum amount of money the organizers are looking to raise. Many organizers either increase the hardcap value in the middle of the ICO or perhaps ignore it altogether as some of the amounts raised are in excess of 100% of the hardcap value. The distribution of the percentages of success can be seen in Figure 7.7. An ICO is successful if it raises at least 100% of its hardcap goal and unsuccessful otherwise.

Although success is treated as a boolean, out of the 1,446 ICOs reporting the necessary attributes for a success calculation, only 155 have a success percentage >100 and the remaining 1,291 failed to meet their hardcap goal. Furthermore, out of the failed ICOs the mean and median success rates were 26% and 15% respectively.

Table 7.7 presents calculated returns at four points of interest: the first day observed on `coinmarketcap.com`, 90 days after the ICO start date, 6 months after the ICO start date, and finally, the last day observed on `coinmarketcap.com`. These return calculations use the

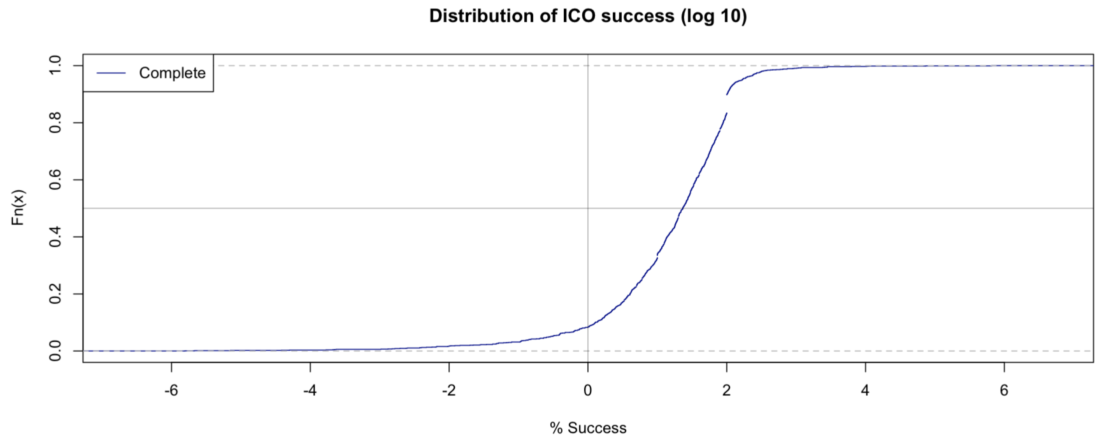


Figure 7.7: Plot of the distribution of the % ICO success (Log10).

	Unsuccessful ICO			Successful ICO			All ICO		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
First Day CMC	14.116	-56.639	2,111.51	68.17	-14.52	1,956.94	355.23	-22.67	88,700.00
ICO Start + 90 Days	101.83	-58.27	1,986.17	30.23	-41.28	1,340.99	123.32	-24.23	7,416.67
ICO Start + 6 months	58.27	-81.78	10,582.73	26.71	-75.17	2,357.70	170.29	-60.48	11,049.50
Last Day CMC	-71.43	-95.27	830.00	-74.63	-94.08	459.24	126.20	-93.36	81,600.00

Table 7.7: Summary statistics on token returns split by ICO success (binary).

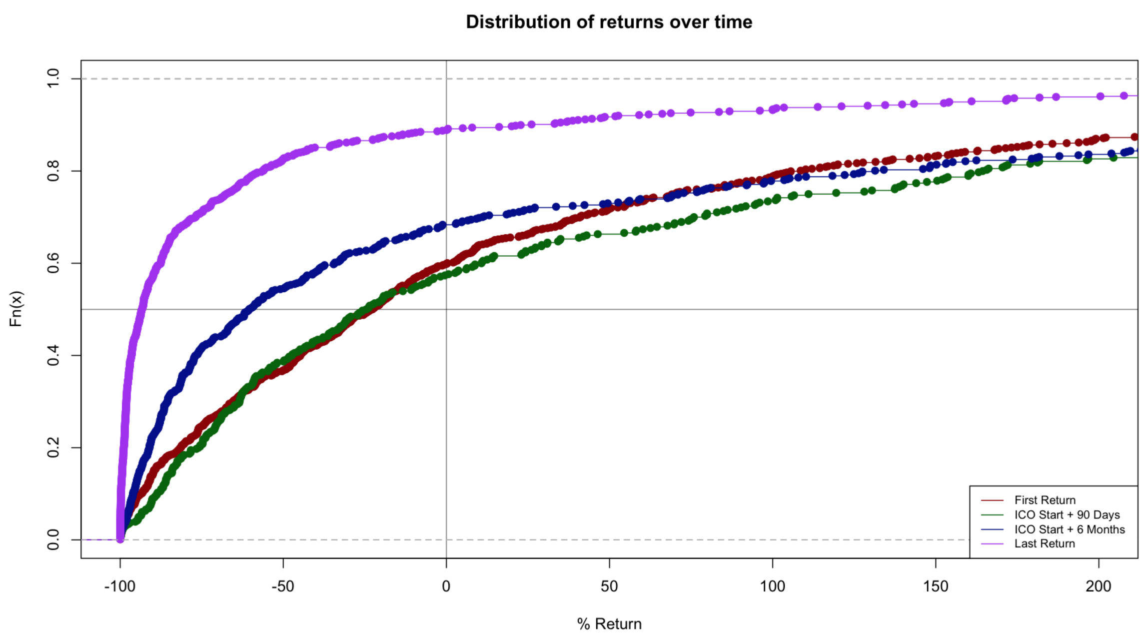


Figure 7.8: Distribution of returns over time split by periods of interest.

reported token price during the ICO as the starting price.

The median values for all returns regardless of success is negative. The average returns spike at the 90 day level, but are also negative at the end of the dataset. This means that there are a small number of tokens that make above average returns, propping up the return values for the remaining tokens in the data. Figure 7.8 shows the distribution of the returns for all ICOs.

In the plot the first return and 90 day return are similar and that is expected. Because of the gap in time between the ICO start date and the first date on `coinmarketcap.com` these two dates can be very close in time. As returns are examined farther out from the ICO start date the probability of a negative return value increases. Just over 70% of tokens have a negative return by the 6-month mark, and that number increases to around 90% by the end of the data.

7.2.2 Pseudo-Trading

A second measure of success was evaluated by creating a separate trading portfolio for each year between 2015 and 2019. On January 1st of each year a hypothetical $10,000 investment was traded in the group of the tokens reporting a market capitalization through CMC. The market capitalization reported for each token was converted into a percentage of the market capitalization for the entire group. Then a proportional share of each token was acquired to construct an "index" of tokens based on value on January 1 of each year. Table 7.8 shows a subset of the tokens purchased for the 2015 portfolio.

Token	Market Capitalization	Close (USD)	% MC	USD Bought	Shares Bought
bitcny	170,366	0.162797	0.43	43.11	264.82
bitusd	856,957	1.010000	2.17	216.86	214.71
maidsafecoin	24,447,786	0.054022	61.87	6,186.67	114,521.27
nushares	4,135,008	0.006819	10.46	1,046.39	153,452.17

Table 7.8: Subset of tokens purchased 2015 for portfolio

In total sixteen tokens were "purchased" for investment in 2015 and this table is fairly representative of the group. Market capitalizations vary from token to token, and subsequently so do the dollar amounts of each token purchased. Some of the tokens provide more opportunity for substantial portfolio growth simply because the low token value allows for more “shares” to be purchased. Although the number of tokens available for purchase at the beginning of each calendar year increases year to year the methodology remains constant. Table 7.9 shows summary values for the yearly portfolio returns at yearly intervals.

	COUNT	ALLTIME	1YR	2YR	3YR	4YR
2015	16	45.26	-68.27 (15)	35.59 (13)	2,222.72 (11)	79.50 (9)
2016	31	228.67	315.47 (28)	6,178.65 (27)	340.00 (23)	
2017	62	-6.34	1,791.38 (55)	52.62 (43)		
2018	491	-90.12	-87.17 (357)			
2019	1,189	-36.57				

Table 7.9: Summary statistics for yearly portfolio percent returns. Token count at return calculation in parenthesis.

For each of the investment portfolios returns were calculated on January 1st for each of the subsequent years with available data. Again, these calculations only included tokens that reported market capitalization values. Over time the number of tokens available through CMC varies; some of these tokens are de-listed from the service or they die completely. Regardless of the reason if they do not report a market capitalization value the “investment” is considered to be a loss and is removed from the portfolio calculation.

The data presented here shows that early token investors have the most potential to make money. Although token prices are influenced by the movements of more popular cryptocurrencies returns are diminishing over time. The influence of other entities can be seen on the diagonal from 3YR (2015) to 1YR (2017). These large returns are due to the last big bitcoin price bubble around the end of 2017.

The daily portfolio values over time can be seen below in Figure 7.9. The horizontal line in the plot represents the beginning USD value of each of the portfolios. As mentioned before, bitcoin is the rising tide that lifts all boats around 2018. However, post price bubble cryptocurrency values fall just as swiftly as they rose.

The 2015 through 2017 portfolio values follow roughly the same trends. However, the trend is mostly negative with the 2018 and 2019 portfolio values. Both see a slight rise in value after the start of the year but quickly drop below their initial investment value.

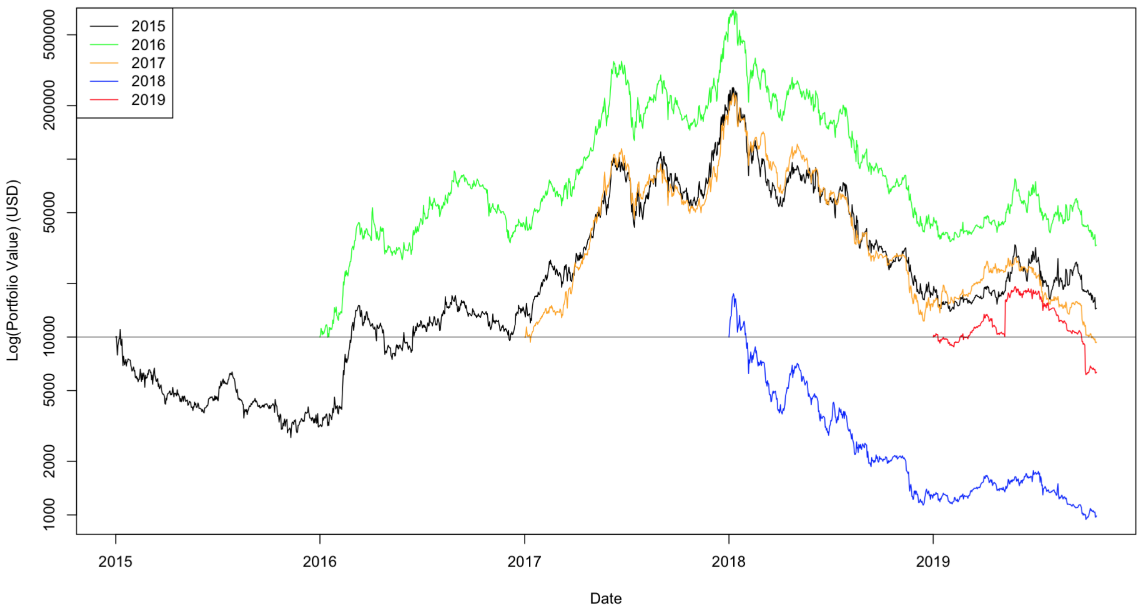


Figure 7.9: Daily values for yearly portfolios. Here the y-axis is log transformed.

Neither displays ideal portfolio performance.

7.3 Token Rise and Fall

Without modifying the established definition's requirements for abandonment and resurrection, all previous analysis from Chapter 6 was re-run on the token dataset for the exact same time window as the original research (ending on 2018-02-06). Applying the algorithm to the token data reveals 1,299 price peaks involving 701 tokens, and 1,426 volume peaks covering 706 tokens. Almost half of the 1,564 tokens in the dataset saw at least one price and volume peak. Although token prices are lower and less volatile than their coin equivalent volume was again utilized for a majority of the analysis because the price data continues to be more volatile than the volume data.

The token with the largest market capitalization by far is Tether with a value of 2,239,511,764 USD at the end of the dataset. A majority of the tokens available for use fail to command the same interest so to better study the characteristics of these smaller tokens the data was split into groups based on aggregate lifetime trading volume.

Table 7.10 shows summary statistics for peaks, abandonments, and resurrections for

	overall	<$1M	$1–10M	$10–100M	$100M-1B	>$1B
# tokens	725	126	191	176	161	71
# price peaks (total)	1,304	246	331	294	295	138
# price peaks (median)	2	2	2	1	1	2
% price increase 1st peak (median)	279	233	281	201	295	608
# volume peaks (total)	1,423	284	390	312	307	130
# volume peaks (median)	2	2	2	1	1	2
% volume increase 1st peak (median)	1,427	405	663	689	2,964	8,442
# tokens abandoned	53	22	21	6	3	1
% tokens abandoned	7	17	11	3	2	1
# abandonments	57	25	22	6	3	1
days abandoned (median)	328	341	300	288	390	182
# tokens resurrected	22	7	10	3	1	1
% tokens resurrected	42	5	5	2	1	1
# resurrections	23	8	10	3	1	1
months to resurrected (median)	3	3	3	2	3	2
# tokens permanently abandoned	34	17	12	3	2	0
% tokens permanently abandoned	5	13	6	2	1	0

Table 7.10: Summary statistics on token peaks, abandonment and resurrection, broken down by total trading volume per coin.

all of the tokens in the data used as an overall measure as well as split into the different volume groups. By examining grouped tokens, it can be seen that this data mainly consists of tokens with mid-level trading volume. This differs from the coin results as most of the coins were focused in the low trading volume group (< 1M USD). A total of 70 tokens have an aggregate trading volume over 1 billion USD, which is just north of the 57 coins in the same group.

Its not surprising that the group with the most tokens also reported the largest number of price and volume peaks; this mirrors the coin results. Again, the number of peaks per token is relatively consistent across all volume groups in Table 7.10, with the maximum deviation from the median being a single additional peak.

One important measure is the length of time to the first peak, as it provides an

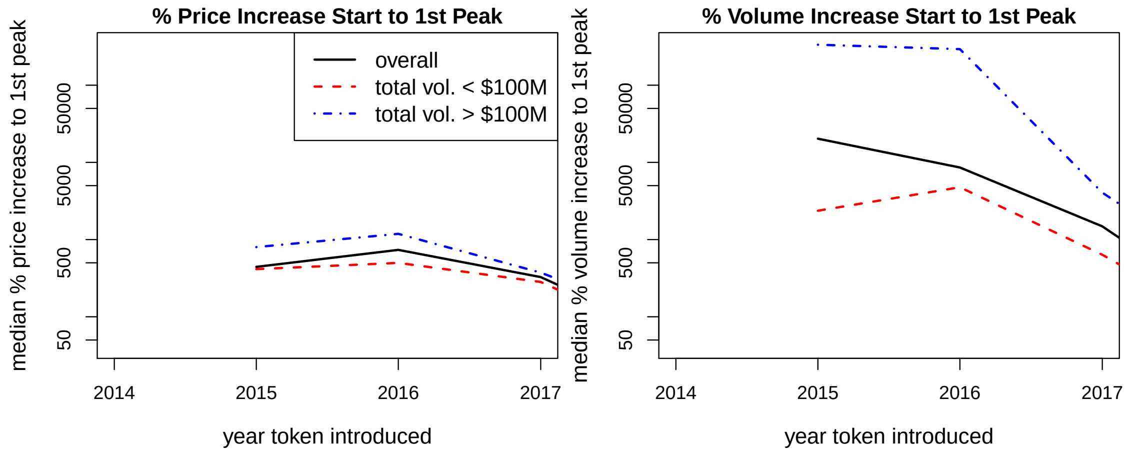


Figure 7.10: Percentage price and volume increase from a token's launch to first peak, based on the year in which the token was launched and its size. The vertical axis is logarithmic.

indication of how much profit the early token backers potentially made. The mean number of days between the beginning of trading activity on CMC and the first volume peak is 59 days and the first peak is an impressive 1,399% increase from the first day of trading on CMC. Because price does not increase at the same rate that volume does the first peak is lower, but a still impressive 279% increase over the first day of trading on CMC. This means that at least half of the early backers saw at least a two-fold rise in the price by the time the first peak was reached.

Figure 7.10 examines the relationship between when token is launched and the magnitude of the initial peak seen by CMC after launch. With the left graph plotting the median percent rise based on the token's launch year. Overall, tokens launched in 2015 saw a median initial price increase of 443%. This rose to 735% in 2016, and then fell to 327% in 2017. This suggests that even though these tokens saw healthy growth initially, the long term price movement is relatively flat.

The volume for these tokens mirrors that of the coins of the same time period. Median percentage jumps were consistently higher for the first volume peak but stayed relatively level for tokens launched in 2015 through 2016. However, mirroring the coin volume movement the initial volume rise fell sharply in 2017. Examined together these plots show that volume increases continue to be high, while price increases have become less impressive over time.

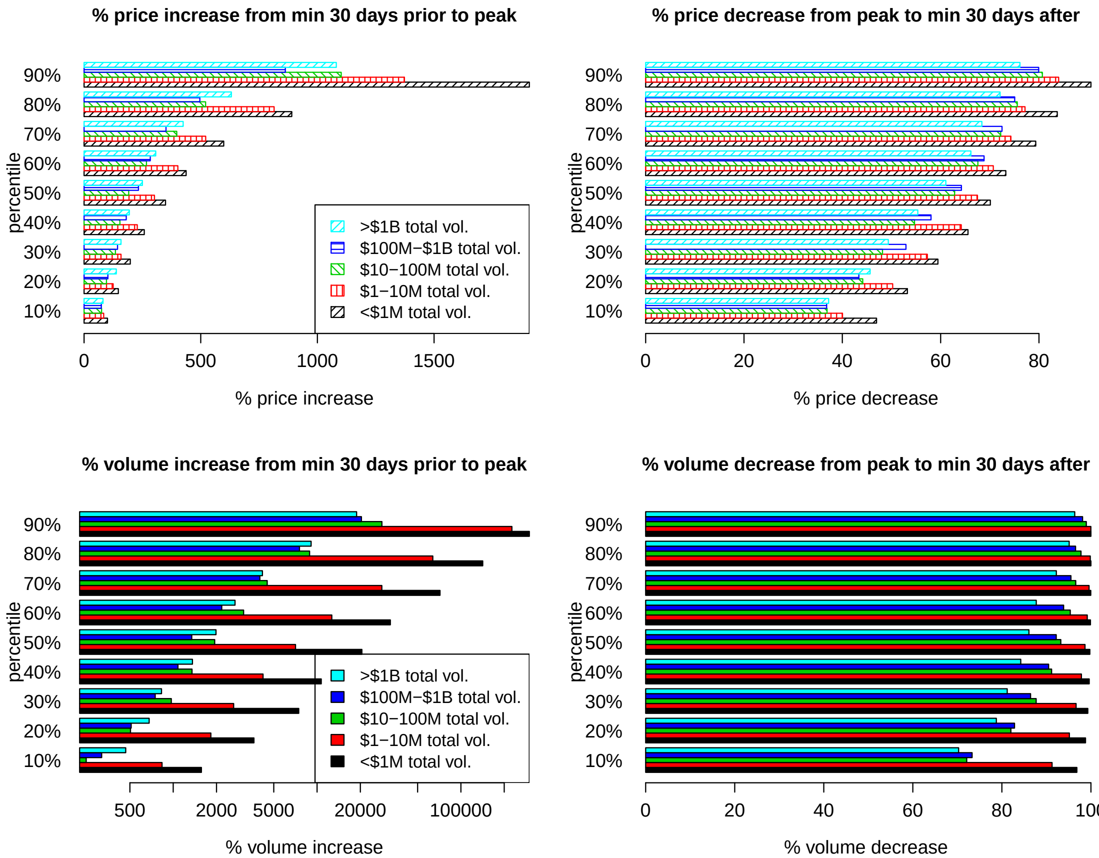


Figure 7.11: Deciles of percent price and volume rises from the smallest value in the month prior to a peak (left graphs) and deciles of percent price and volume falls from the peak to the smallest value in the month following the peak (right graphs).

Figure 7.11 examines the rise and subsequent fall surrounding all of the peaks discovered. Recall that for a point to be a peak the value must be, at a minimum, 50% greater than the lowest value in the previous 30 days. The top two plots show the distribution of the rise and fall associated with the peak's trading price.

Beginning with the top left plot, the 10th to 90th percentiles of the peak's percent increase relative to the smallest value in the previous 30 days is shown[5]. To better understand movements of tokens with different popularity, data was further divided by the same coin size used in Table 7.10. The top 10% of price increases range from 812% for tokens with a total trading volume between 100 million USD and 1 billion USD to 1904% for tokens with

[5]For this analysis we exclude any price or volume rises from peaks occurring in the first week of a coin's operation, as well as any falls within the last week of its operation. This is to deal with edge effects from the 7-day rolling average used to compute peaks

a total trading volume less than 1 million USD. Within the coin data the second smallest group always grew the fastest within each subset, and the second largest group grew the slowest; the same cannot be said for tokens. For a majority of the data the smallest group realizes the largest increases, and the middle group sees the slowest growth.

The top right plot in Figure 7.11 shows the fall that occurs following a peak. It can be seen that the smallest tokens always fall farther than the other groups. Nine out of 10 tokens lost more than 40% of their value in the 30 days following a peak. Furthermore half of all coins lost 60-70% of their value shortly after a peak.

The bottom row of plots in Figure 7.11 shows the movements around volume peaks instead of price. The plot on the left shows that the group with the smallest total trading volume almost always experiences the largest volume increase leading up to a peak. Median increases range from 248% for the middle subset of total trading volume to over 1,000 times that for the tokens with the second highest aggregate trading volume. Only 2% of the tokens in the dataset saw their volume drop to zero following a volume peak.

Finally, similar to the percentage price decreases following a peak, volume decreases post peak are extreme. For almost every coin group, trading volume falls more than 80% in the 30 days following a peak.

7.3.1 Abandonment, and Resurrection

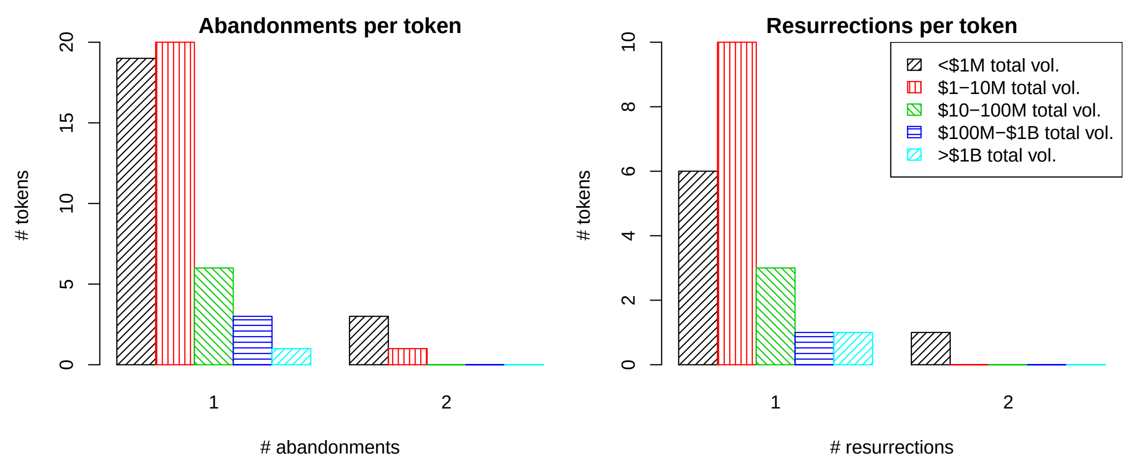


Figure 7.12: Abandonments (left) and resurrections (right) per token, split by total trading volume.

Even though some tokens saw impressive falls following peaks in price and volume only 7% of tokens experienced an abandonment. Out of the 54 tokens abandoned 41% (22) were resurrected. Recall from the definition that the volume must rise to at least 10% of the previous peak value following abandonment to become "resurrected." Figure 7.12 shows the number, per token, of abandonments in the plot on the left and the number of resurrections in the plot on the right. The low number of abandonments and resurrections per coin suggests that large token price movements occur over a much longer period of time ($>$ 30 days) instead of rapidly as seen within the coin dataset.

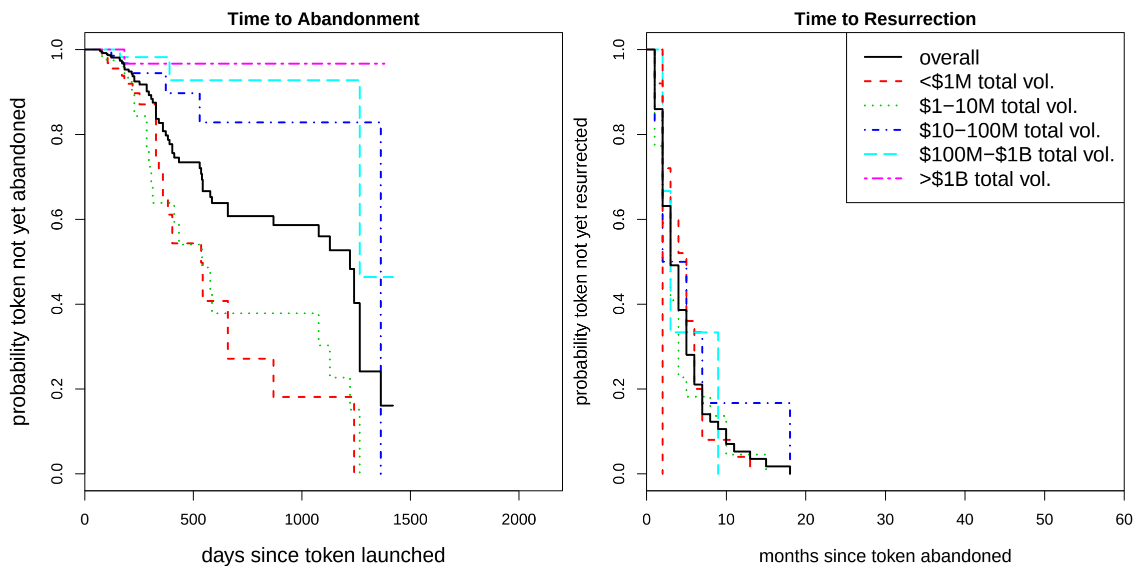


Figure 7.13: Survival probabilities for the time to abandonment (left) and time to resurrection (right).

On average, tokens are abandoned within 5 months of reaching their peak value with a 2 month median. This shows that when tokens do fail it can happen quickly. Token resurrection can also occur at a fast pace with a 10 month average and a 6.5 month median. For a more comprehensive look at the time to abandonment and resurrection survival probabilities were computed, as shown in Figure 7.13, using Kaplan Meier estimators. This allows for empirical estimation of the time from launch to abandonment using the duration of all tokens, even those that had not been abandoned. Overall, the median time to abandonment for tokens is 184.1 days. Not all volume groups are abandoned at the same rate. For tokens with a volume between 1 million USD and 10 million USD the median time from launch to abandonment is 154.26 days. Tokens with an aggregate trading volume between 100 million USD and 1 billion USD have a median time to abandonment of 196.56 days.

The top right plot in Figure 7.13 shows the estimated time to resurrection. The time to resurrection is shorter than the time to abandonment. Overall the median time from abandonment to resurrection is 3 months. While there is variation within the size groups it is only 1 month on either side of the median.

7.3.2 Relationships Between Key Variables

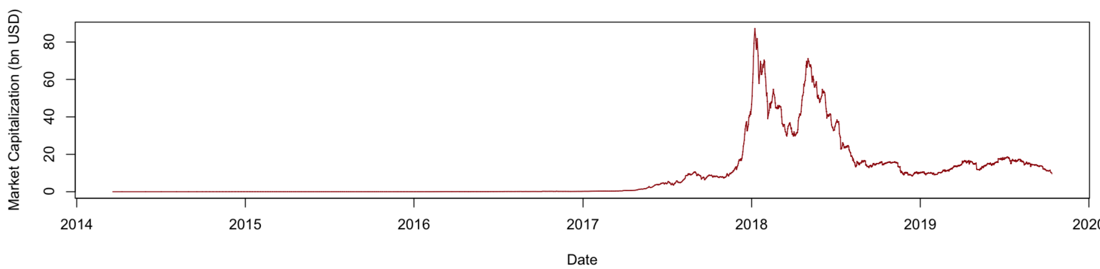


Figure 7.14: Market capitalization of all tokens traded (in billions of dollars).

As shown in Figure 7.14, unprecedented growth was seen in all tokens at the end of 2017. This was in part due to an impressive increase in the price of all cryptocurrencies at the time, which was driven by the popularity of Bitcoin. With this growth came new and diverse offerings trying to "ride the wave" of success.

Similar to the growth of coins during the same time period, the trend of token creation follows the fluctuations in the price of Bitcoin. When the price is high more tokens are released and when the price falls new token issuance slows. This movement is explained in greater detail later in this section with Figure 7.16.

The two plots in Figure 7.15 show a significant correlation between the number of price and volume peaks each month. Unlike the coin peak activity which saw an almost constant number of peaks between 2014 and 2016, the bottom graph shows tokens experience price and volume peaks much more infrequently. However, the longterm number and percentage growth mirrors that of coins for the same time period. The number of peaks for both price and volume are initially low but increase rapidly beginning in mid-2017 ending the dataset

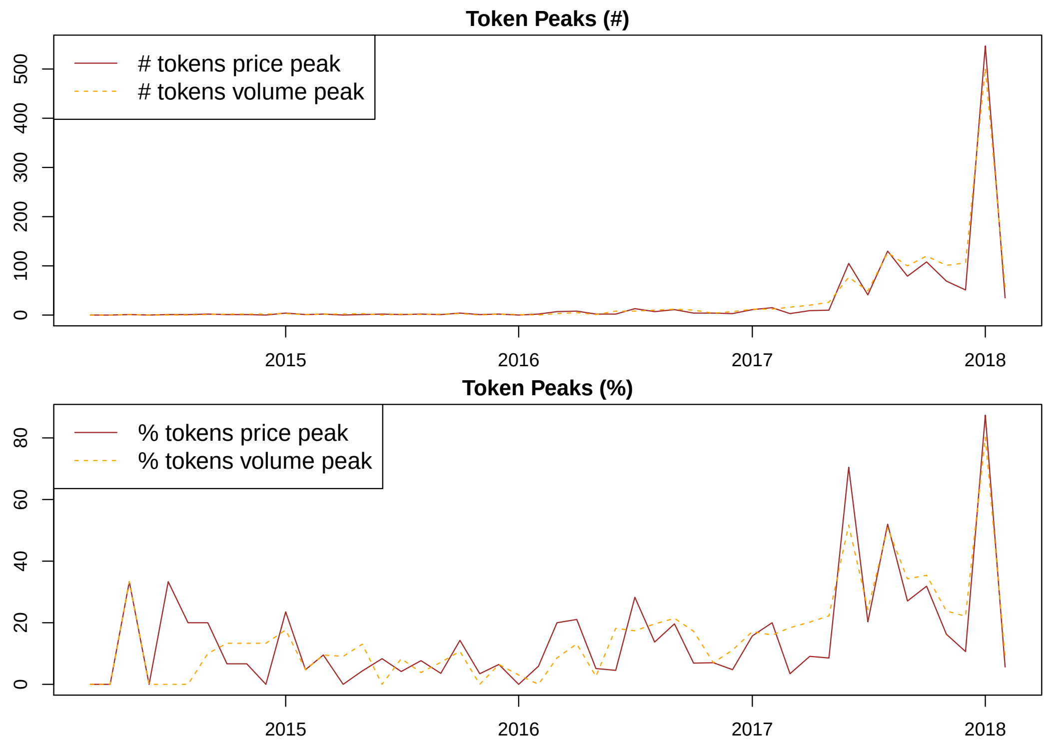


Figure 7.15: Number (top plot) and percentage of active tokens (bottom plot) experiencing price and volume peaks over time.

with around 50% of tokens experiencing both price and volume peaks.

The relationship between Bitcoin price, coin creation, abandonment, and resurrection are visible in the plots found in Figure 7.16.

Similar to coins, tokens benefit from cryptocurrency's popularity and typically enter the market with substantial trading volume. However, unlike coins, token trading volume can initially be pushed even higher through the use of an initial coin offering (ICO). Following this spike when the popularity inevitably declines the markets correct and already low token prices fall further. This activity occasionally results in the token becoming abandoned. Token abandonment is expected to lag behind the price trend. Certain tokens seem to endure periods of low trading activity as very few tokens in the largest (> 1 billion USD) aggregate trading volume group have been abandoned.

The top plot in Figure 7.16 shows the differences over time within the number of newly created and abandoned tokens per month. Abandonment is relatively flat throughout

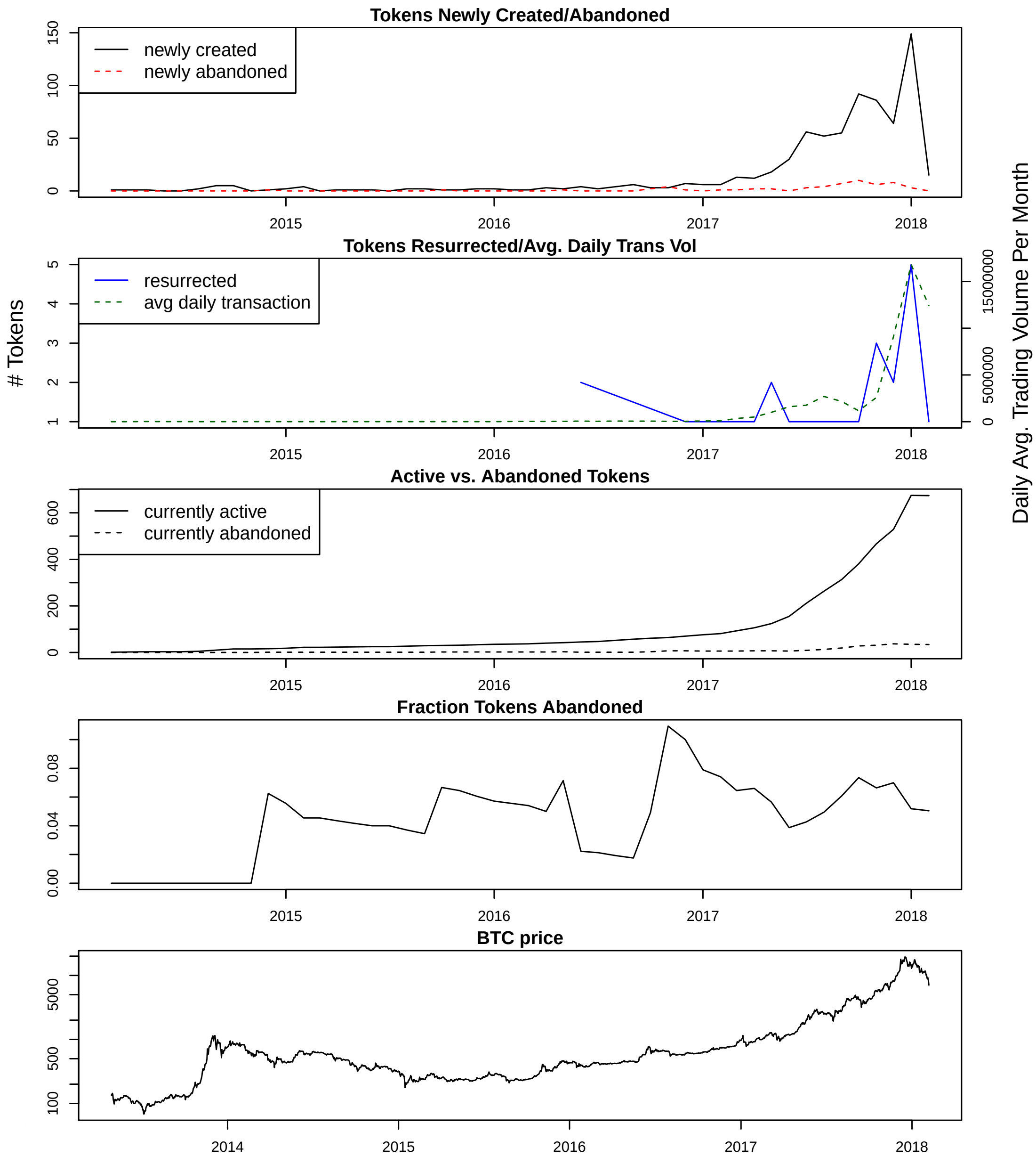


Figure 7.16: Token summary statistics including creation, resurrection, abandonment, and daily average trading volume.

the entire time window only slightly increasing when the number of coins created increases rapidly in 2017. Token creation was also fairly flat until early 2017 when the price of Bitcoin began to rise.

The second plot from the top examines the rate of coin resurrection per month and the daily transaction volume over time. There is a substantial correlation between the two: token resurrection remained flat through 2016 and rose at the end of 2017. This implies that as more money enters the cryptocurrency ecosystem investors seek out opportunities with significant upward potential.

The next two plots in Figure 7.16 show how the rate of abandonment affect the overall number of token offerings over time. The middle plot shows the number of available tokens increasing at a steady pace, and although it appears to be flat the number of abandoned tokens is also rising just at a much slower pace. Because the total number of tokens abandoned at one time never exceeds 10 its difficult to see any real decrease in the number of active tokens.

The final plot shows the USD price of Bitcoin over time on a logarithmic scale. Increases on this plot regularly coincide with points of interest on other plots in this figure.

	# Tokens Abandoned	# Tokens Resurrected	# Tokens Created	Trade Volume	log_{10}(Average BTC Price)	# Price Peaks	# Volume Peaks
# Abandoned	1						
# Resurrected	0.47	1					
# Created	0.75	0.78	1				
Trade Volume	0.35	0.72	0.71	1			
log_{10}(Average BTC Price)	0.65	0.75	0.81	0.90	1		
# Price Peaks	0.37	0.75	0.85	0.78	0.67	1	
# Volume Peaks	0.48	0.80	0.91	0.83	0.77	0.99	1

Table 7.11: Monthly correlations between key variables in the ecosystem.

To further explore the relationships between these variables the correlations between were calculated and are displayed in Table 7.11. Through close examination of the coefficients in this table a few key trends emerge.

- There exists a high positive correlation between token abandonment and tokens created (0.75), suggesting that new tokens are created to fill the void left by abandoned tokens.

- Resurrection is highly correlated with price and volume peaks, 0.75 and 0.80 respectively which suggests that many of the tokens are "riding the wave" created by the huge increases within the cryptocurrency market. Additionally, token creation is also highly correlated with resurrection (0.78). This relationship, coupled with the high creation/abandonment correlation suggests that heavy competition exists within these tokens.

- Token creation is highly correlated with trade volume (0.71), the log of the Bitcoin price (0.81), and the number of price (0.85) and volume peaks (0.95).

Many of these trends can be seen in Figure 7.16. The Bitcoin price was again included as the bottom plot of this figure because it continues to be the market leader and set the trend for other cryptocurrencies. These Bitcoin backed trends are apparent through the examination of high correlations found involving the log transformed Bitcoin price in Table 7.11.

7.4 Conclusion

This chapter utilized previously established methods for identifying peaks in trading volume and price for over 1,000 tokens. Using those peaks periods of abandonment and resurrection were identified for each of the tokens. Similar to the results of the coin focused analysis, lower volume tokens were at greater risk of abandonment but were also more likely to resurrect as the prices of other cryptocurrencies rose. As expected, tokens also ride "the wave" created by exceptional increases in price found throughout the cryptocurrency market.

Additionally, it was found that although token prices may increase over the short term, most are unprofitable over a longer time frame.

This analysis, like other sections of this dissertation rely heavily on what is reported by `coinmarketcap.com`. It was found that only 11% of the tokens found within the four ICO data sources used made it to CMC by the end of the dataset. Future work should focus on locating reliable sources for pricing and volume data to fill the gap left by CMC.

CHAPTER 8

MEASURING THE IMPACT OF CRYPTOCURRENCY "PUMP-AND-DUMP" SCHEMES

With the growing popularity of cryptocurrencies, it's important to know in which ways these platforms are susceptible to manipulation especially with the recent investments from mainstream finance, and the willingness of countries to implement payment systems that accept cryptocurrencies. Cryptocurrencies are no longer a niche market. The total market capitalization for all cryptocurrencies exceeded 800 billion USD at the end of 2017 when bitcoin experienced a phenomenal increase in its price. Even with the fall in valuation that followed the peak of 2017, the market capitalization of cryptocurrencies was still around 186 billion USD at the time of writing.

In this chapter, a single type of price manipulation, "pump-and-dump" schemes, found within the cryptocurrency ecosystem is examined. Through a coordinated effort, these schemes temporarily inflate the price of a cryptocurrency allowing early backers to realize a healthy profit. At the beginning of an organized "pump-and-dump," a signal indicating the currency to buy is transmitted to insiders via a group messaging platform. Ideally, from the standpoint of the "pumpers," the coordinated buying increases the trading activity and begins to drive up the price. When outside buyers are attracted and begin making purchases, the price rises further. Finally, the "pumpers" "dump" the positions they acquired previously at a profit.

The growing cryptocurrency audience, in addition to the widespread availability of free, or cheap, messaging platforms, has made it extremely easy to conduct "pump-and-dump" schemes.

The goal herein is to describe how the pumps work in the cryptocurrency realm, to quantify the extent of the phenomenon, and to examine what factors (e.g., coin popularity - the number of exchanges on which it is traded) affect the "success" of a pump.

8.1 Methodology

In this section, the methodology used to collect relevant pump signals from social media as well as public and private messaging sources is discussed. The methods used to collect cryptocurrency pricing data and how "pump-and-dump" success is measured are also detailed.

8.1.1 Pump Signals Data From Discord and Telegram

A pump signal is an announcement used to motivate traders to buy into a cryptocurrency with the intention of driving up the price through a surge of trading activity. The two main platforms where these pump messages are posted and shared are Telegram and Discord. Telegram is a cloud-based messaging and VoIP service, and, similarly, Discord is a VoIP service that specializes in text chat. Before collecting any data, it was necessary to become familiar with the platforms and how users post and format pump signals.

A public list of URLs for large Discord pump groups found on the `bitcointalk.org` forum was the starting point. Then, data was collected from from all pump groups with over 4,000 members that were listed on `https://padl.mine.nu`, an Android app that tracks "pump-and-dump" group popularity. Having this data allowed keyword based filters for each of the channels used, based on the organizer's posting patterns, to be developed. This required significant effort as language and communication style were not uniform across these channels. Then the filtered posts were manually inspected and verified as to whether a pump was actually being discussed. Any channel discovered within another channel from which data was being collected was subsequently added to the dataset. Data collection methods employed allow for a high level of confidence in the coverage of relevant channels during the period examined: January to June 2018.

One of the challenges faced during data collection was invitation only pump groups.

Invitation only and paid pump groups typically publish pump signals first and the information slowly makes its way to the free groups. Another challenge is to make sure that an announcement is actually a pump signal. To weed out junk signals, the posts attempting to predict the future price of a coin were removed as were signals to "hodl" coins. "Hodl" is a cryptocurrency meme for holding coins for a long period of time based on a misspelling of "hold." Based on conversations with "pump-and-dump" channel members, the decision was made to not include pump channels with low member counts because the users rarely participated in any of the pumps posted to the groups.

Data collection on Discord went beyond simply gathering information from pump signals. Users on Discord join servers and those servers can host multiple channels which these users can join and as a member of a server they have access to all of the server's hosted channels. These individual channels can be used to narrow the focus of the group discussion. Through the Discord API it was possible to record a specific server's member count. Unfortunately, this data is not channel or pump specific, but it does allows for the measurement of the potential market for participating in a pump scheme promoted within the channels on a server.

Telegram is a cloud-based service where individual channels are set up by individual operators and hosted on Telegram's infrastructure. Hence, there is no analogous variable to number of members that belong to a specific server in the Telegram data.

Through the data collection efforts made it was possible to find repeated patterns within the channels used. The discovery of how pumps work was based on those patterns and through them it was possible to characterize the channels into three broad categories:

- Transparent Pumps: These channels used the words "pump" and "dump" everywhere, including in the name of their channels.

- Obscure Pumps: These channels usually avoided the words "pump" and "dump." The main concern that was reflected in their "chatrooms" was that members were not sure if pump and dump was legal, so they avoided using the terminology.

- Copied Pumps: This category contains signals copied from other sources.

After painstakingly going through the pumps, it was discovered that there were significant differences between the way the "transparent" and "obscured" pumps operated. The following section describes the key characteristics of the three categories of pump signals.

Transparent Pumps:

This category was the most straightforward to identify as these channels included a liberal use of the words "pump" and "dump," occasionally even in the title of the channel. They essentially used a "countdown" strategy for releasing pump information to the group. They usually posted the first announcement 24 to 48 hours before the pump. Then, they would post many other announcements about timing and the cryptocurrency exchange where the pump would occur. At the time of the pump the group organizer would post the name of the cryptocurrency. They would typically post pump results a few hours afterward, along with the date of the next pump.

These channels usually had a premium membership offering. To gain access to a premium channel members could either recruit individuals for the free channel, or buy a premium membership plan. Based on the type of plans, premium members would receive the pump signals a certain amount of time before others giving them a slight advantage.

Finally, these channels did not typically pump the same coins over and over.

Obscured Pumps:

This category was not as brazen as the first because it typically avoided the words "pump" and "dump." The main concern of these groups was the questionable legality of "pumping" cryptocurrencies, so they avoided using the phrase "pump-and-dump." Further, since broadcasting a countdown clearly demonstrates a coordinated pumping, the obscured pumps designed their signals differently. Instead of a countdown, they typically gave target prices along with the coins, exhorting channel members not to sell below the target price. Moving forward this is termed a "price target" strategy.

Whether caused by a lack of sophistication or a desire to avoid detection, the obscured pumps lacked many of the hallmarks of coordination. These channels typically did not have a premium membership option. Unlike the first group, they did not make multiple announcements about a particular pump. They typically posted the name of the coin and its current price, without any previous announcement.

Importantly, unlike the transparent pumps, they often pumped the same coins many times.

Differentiation between “countdown” vs. “price signal”:

To be sure that the two types of pumps typically had differentiated strategies (“countdown” vs. “price signal”) 125 randomly selected Telegram signals were inspected in detail. 105 of the 125 pump signals included either a countdown or a price target. Of the 53 transparent pumps, 32 included a countdown but no target, 13 had both a countdown and a target, and 8 had a target but no countdown. 50 of the 52 inspected obscured pumps only had a target, with the remaining 2 including a target and countdown. Thus it can be concluded that transparent pumps mostly use countdowns while obscured pumps almost exclusively set price targets.

Copied pumps:

The third, and final, category consisted completely of pump announcements copied from other sources. Although these groups posted the pump signal hours after the scheduled time, they copied the timestamp from the original announcement and the source of the signal. These sources were not used for pump signals, but were used to gain a more complete data coverage, i.e., find the pump source and follow it. However, the data collected was included when access to the source group could not be obtained.

Copied pumps were actually an important source; it accounted for 4 Discord channels with 514 pump signals not found elsewhere. There were no Telegram groups in this category because of the complete overlap between the groups found here and others found in Category 1 and Category 2.

Although a majority of the copied pumps originated in other Discord channels, approximately five percent of the Discord data overlaps with the data from Telegram. The copied pumps are included for completeness. However, the results are qualitatively unchanged if the small number of these copied pumps from the Discord analysis are removed.

Summary of Pump Signals:

In the case of Telegram, 88 percent of the signals were obscured and 12 percent were transparent.

In the case of Discord, 42 percent of the signals were obscured, 40 percent of the signals were copied, and 11 percent transparent.

8.1.2 Pricing Data on Cryptocurrencies

Pricing data was collected on nearly 2,000 coins and tokens (henceforth referred to as "coins") across 220 exchanges, as reported on `coinmarketcap.com`, the leading website of aggregated data on cryptocurrency trading. All price data for each of the coins listed on `coinmarketcap.com` from mid-January 2018 through early July 2018 was gathered and provided a total of 316,244,976 collective volume and price data points across all of the coins listed. The data collected are at the finest granularity of a 5-minute interval presented by `coinmarketcap.com` at the time of collection.

There are limitations to this method of data collection. For instance, `coinmarketcap.com` does not list every coin or token available for purchase or trade. Further, these data are slightly different from what would be available for collection from an exchange API. Since the website is collecting data from so many sources, it reports a volume weighted average of all of the prices reported at each exchange to calculate the price it reports. An advantage is that this approach is more comprehensive in the number of exchanges and coins covered.

Every internet service experiences outages, planned or otherwise, and the cryptocurrency services are not exceptions to that rule. During the initial data exploration phase, gaps in the data were discovered. To make sure those gaps were recorded in the data and not a result of collection efforts, the data were programmatically checked for proper intervals. If a gap existed in the data that spanned a time period equal to or greater than 7.5 minutes, that data point was reported as missing. Seven and a half minute intervals were chose because of the 5 minute average interval in the data collected. After reviewing the timeline of each of the coins, an hour long window surrounding the missing data points was created and `coinmarketcap.com` was queried for that data. If the gap persisted after the additional data collection, it was believed to have been caused by an outage due to the exchange or `coinmarketcap.com`. Approximately 3,806,474 volume and price records across all of the coins, about 1% of the data, were found to be missing.

Matching Discord/Telegram Information with Trading Data:

For the purpose of this study, it was essential to ensure a consistent mapping between what was announced in the pump signal to what was associated with the trading data. In particular, pump signals were not consistent in regards to the coin names used in the messages. Some users referred only to the coin ticker name such as DOGE, the ticker name for Dogecoin. This is not a good practice because there is no equivalent to NYSE or NASDAQ to enforce the uniqueness of ticker symbols and several cryptocurrencies employ identical tickers. Others use the full coin or token name but that can also be problematic because many coins have similar names. For instance, the cryptocurrency IOTA has the ticker MIOTA and the ticker for IoTex is IOTX. Still others use some combination of the ticker name and full or partial coin or token name. For example, "Bitcoin (BCD)" refers to Bitcoin Diamond, not Bitcoin. The ticker for Bitcoin is BTC.

The name used by `coinmarketcap.com` was used to normalize reports. To do this, a name map that contained several variations of the actual cryptocurrency name based on observations in the data was created. Then, special characters were removed from the names reported in Discord a case insensitive comparison to the map that had been created was performed. If a match was found the pump name was replaced with a clean version that matched the name elsewhere in the data. Some of the names required manual replacement because cryptocurrencies have the ability to rebrand. In this way 1,034 of the Discord pump signals as well as 3,767 of the Telegram pump signals were mapped to more than 300 cryptocurrencies.[1]

Identifying pump Timing and Success:

Throughout the processes of aggregating, combining, and cleaning the data, it became increasingly apparent that using the time of a pump signal to mark the beginning of a period of anomalous trading activity was not reliable.[2]

So, instead of taking the pump signal time as given, it was treated as the starting point to identify associated spikes in trading activity. Forty-eight hours before and after the time of the reported signal was inspected to find the maximum percentage jump between the consecutive price data points which were typically spaced 5 minutes apart.

In the data analyses described in the next section, the maximum 5-minute percentage increase in the 96 hour period in the coin's price relative to BTC was used as the measure of pump success.

8.1.3 Data Summary

The Discord and Telegram data span the six month period from January to June 2018. A small number of observations were duplicates in the sense that they involved the same on on the same day and roughly at the same time (within an hour) on the same exchanges. The duplicates were eliminated, but the results are quantitatively unchanged if they are included. Once the duplicate observations were eliminated along with a few observations for which the data was incomplete, the dataset was reduced to 952 observations with complete data on Discord and 2,469 observations with complete data on Telegram. This gives a sense of the scope of the pump and dump phenomenon on these platforms.[3]

[1]There are more total pumps than that, but approximately 5% do not have complete data and cannot be used in the analysis.

[2]This may be because "insiders," i.e., those running the pump, strategically purchase before the agreed upon time. This is consistent with the other work in this area. [29] noticed that pumps sometimes occurred exactly when a signal was put out and other times occurred afterwards. [31] collected more pump signal information than [29] and observed the same effect. [57] collected *hourly* market data and found that the markets moved as much as 72 hours before an announced pump.

[3]It is possible that there are a small number of pumps that occur both on Telegram and Discord, primarily in May 2018. This is not a problem since the Discord and Telegram data are analyzed separately. Analysis

It is found that ten percent of the pumps on Telegram (Discord) increase the price by 16.3 percent (15.6 percent) in just five minutes. Recall that January-June 2018 time period was a period in which cryptocurrency prices were falling significantly; hence "moderate" percentage increases were an achievement for the pump.

8.2 Informal Analysis of the Ecosystem

8.2.1 A Bit of Theory...

Economic theory suggests that the cryptocurrency "pump and dump" ecosystem would not succeed over time for the following reasons:

- Pump and dump schemes need outside investors to succeed. The idea is that the initial surge in volume attracts additional traders. Such (honest) traders in cryptocurrencies would learn how to recognize pump and dumps and adjust their strategies accordingly, so as not to fall prey to the schemes. It has been documented in the literature that it is fairly straightforward to adjust investment strategies to account for cryptocurrency pump and dumps.

- Many insider members of the pump and dump schemes actually lose money. This is because, as has been documented, administrators/insiders of the schemes typically make purchases before the "beginning" time of the pump. This would make it less attractive to participate over time.

- Regulators might begin to react if the phenomenon becomes prolific. A few pump and dump schemes will not have an effect on regulatory policy, but hundreds or thousands of pumps might eventually lead regulators to act.

8.2.2 Dynamics over time in the ecosystem

This dynamic dataset allows the examination of whether profitability has in fact declined over time. From the data, the median profitability of the pumps go down over time

is split because some of the variables are not available for both of the platforms. The results are robust to eliminating these potential duplicates.

for the six months this data covers (January to June 2018.) This is true both for Telegram and for Discord.[4] See Table 8.1. The decline is steep on both platforms: On Telegram profitability was essentially 50% lower on average in June than in January. In the case of Discord, profitability was essentially 60% lower on average in June than in January.

Table 8.1: Pump (Median) Success/Profitability by Month in Percentage Terms

Month	Discord	Telegram
Jan 2018	5.4	6.4
Feb 2018	4.1	4.9
Mar 2018	3.9	5.2
Apr 2018	3.2	4.2
May 2018	2.9	2.8
Jun 2018	2.2	3.2

While the interesting question of whether this decline in profitability reduced pumps over time cannot be definitively answered,[5] it is interesting to note the following: "Google Trends" data suggest that interest in pump and dump schemes took off during the increase in bitcoin's massive increase in price in late 2017 and declined sharply after June 2018. See Figure 8.1.[6] While the intention is not to push this point, it does provide some support for a decline in the cryptocurrency pump and dump phenomenon, which is consistent with declining profits over time.

8.2.3 Concentration in the Ecosystem

The data also enable the examination of other key questions about the ecosystem: (I) Was the ecosystem dominated by a few channels (running a lot of pumps) or were there many active channels? (II) Did pumps occur on many exchanges or just a few? (III) Were coins pumped repeatedly? High levels of concentration were (perhaps surprisingly) found in both exchanges employed for pumps and channels involved in running the pumps.

[4]In the formal analysis, regressions are run. The regression results show that even after controlling for other factors that affect pump and dump success, success falls over time, with a steep drop-off near the end of the period for which data exists.

[5]It was not possible to collect detailed information following the end of this data, since pump channels removed the relied upon accounts from the groups.

[6]It is well known that interest in bitcoin in Google trends data is very highly correlated with the price of bitcoin.

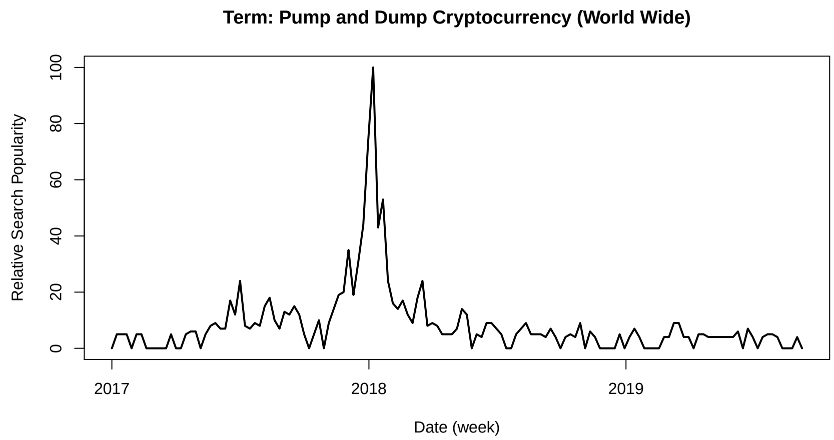


Figure 8.1: Google Trends results from searching for pump and dump cryptocurrency.

- Cryptocurrency Exchanges: From the data, Binance and Bittrex were by far the most popular exchanges for pump and dump schemes. Binance and Bitfinex together accounted for 86% (87%) of the pumps for pumps that listed/recommended an exchange on Telegram (Discord). During this period and afterwards, Binance was the largest cryptocurrency exchange, by trading volume, and Bittrex had large trading volume as well. Both exchanges offer trading in hundreds of cryptocurrencies, which likely made them attractive to organizers of the pump and dumps.

- The perceived wisdom was that there were many channels running pumps. It turns out that, like exchanges, this aspect of the ecosystem was highly concentrated as well. In the case of Telegram for example, six channels accounted for more than 70 percent of the pumps.

- Additionally, twenty-three coins were pumped 18 or more times (thus on average, these coins were pumped at least three times a month during the six month period.). These twenty-three coins accounted for more than 20% of all pumps during that period on Telegram. Similarly on Discord, the top 20 coins accounted for 28 percent of the pumps. Again, this suggests a concentrated industry. This information should be helpful to regulators.

8.3 Analysis

Here declining profitability over time is examined to see if it holds when controlling for various other factors such as the rank of the coin. An additional and important part of the analysis here is to examine what factors explain the success of the pump and dump scheme, where success means that the pump increased the price significantly.

This is first examined by including all pump and dump types together in the analysis. Then, revisiting the analysis the same issue is examined separately for the two types of pump groups.

8.3.1 Both Types of Pump Groups Examined Together

The maximum % price increase (as described above) is located within the 48 hours preceding and following the pump as the dependent variable. This variable will be referred to as % Price Increase. Most of the cryptocurrencies cannot be directly traded with USD, but they can be traded with bitcoin. Hence, coin prices are in bitcoin.[7]

8.3.2 Independent Variables

For the regressions, the following independent variables were employed:

- Exchanges: the number of exchanges on which the coin can be traded. This variable is measured twice: once at the end of 2017 and once in September 2018. The correlations are above 0.99 and the results are unchanged regardless which date was choosen. The 2018 variable has more observations, so that one is used.

- Rank: the rank of the coin in terms of market capitalization. Bitcoin is #1. Coins with higher rank have lower market capitalization.

- Pair Count: the number of other coins that the coin can be traded with.[8]

[7]Because of this, the very small number of pumps using bitcoin itself cannot be included.

[8]Similar to exchanges, this variable is measured twice, once at the end of 2017 and once in September 2018. The correlations are above 0.99 and the results are unchanged regardless which date was choosen. The 2018 variable has more observations, so that one is used.

- Server-Member-Count (Discord Only): the number of members that belong to a server (which is not specific to a particular pump). This variable essentially measures the potential market for participating on pump schemes promoted on that server.

- Views: (Telegram only) Number of views per pump.[9]

- Dummy variables for February, March, April, May and June 2018.

- Dummy variables for Binance-only, Bittrex-only, and Binance-Bittrex. A non-trivial potion of the pumps were on both exchanges. In that case, Binance-Bittrex takes on the value one.

- other-exchange takes on the variable one when the pump lists an exchange other than Binance or Bittrex.

- no-exchange is a dummy variable that on the value one if no exchange was listed in the pump.

Table 8.2: Descriptive Statistics: Discord, N=1,034

Variable	Obs	Mean	Std. Dev.	Min	Max
Max % Price inc.	1,034	6.96	16.78	0.64	221.90
Exchanges	1,034	21.48	26.72	1	182
Pair Count	1,034	24.07	89.90	1	759
Rank	1,034	256.98	308.17	2	1,863
Server Member Count	1,034	5,616.05	9,741.03	141	84,823

Table 8.3: Descriptive Statistics: Telegram, N=3,767

Variable	Obs	Mean	Std. Dev.	Min	Max
Max % Price inc.	3,767	9.96	21.29	0.42	341.99
Exchanges	3,767	18.51	24.45	1	182
Pair Count	3,767	19.11	72.20	1	759
Rank	3,767	394	433	2	2,036

Descriptive statistics for all variables used in the analyses appear in Table 8.2 and Table 8.3.

[9]With the possible exception of views, all of these variables are clearly exogenous to the pump. Views is essentially exogenous as well. Results are unchanged if views is not included in the analysis.

These table shows that in the case of Telegram (Discord), 45 percent (50 percent) of the pumps occurred on either Binance, Bittrex, or both and 48 percent (46 percent) occurred without an exchange listed.

Table 8.4: Median Price Increases by Coin Rankings.

	Discord				Telegram			
	Coins		Signals	Price	Coins		Signals	Price
Rank	#	%	#	Inc %	#	%	#	Inc %
$\leq$ 75	52	69.33	342	3.51	56	74.67	1,000	4.81
76-200	58	46.40	257	5.22	62	49.60	736	6.46
201-500	75	25.00	285	5.32	84	28.00	948	8.10
$>$ 500	73	5.33	150	23.23	176	11.46	1,083	18.74

Table 8.4 groups coins by rank (in terms of market capitalization.) In Table 8.4 shows that while many of the pumps involve coins with light trading and low market capitalization (similar to penny stocks), pumps are not limited to obscure coins. Coins with greater market caps experience smaller spikes in prices: the median price increase for the top 75 coins (in rank) is 2.4% for Discord and 2.6% for Telegram. The median return for coins ranked between 500 and 1000 was 5.8% for Discord and 7.1% for Telegram. See Table 8.4 for the full breakdown.

The pumping of more "mainstream" coins may be because it is not always easy to pump obscure coins that are traded on a small number of exchanges. Additionally, there is less volatility in mainstream coins, and some "investors" (pumpers) may have preferred a relatively lower risk level.

Overall, in the case of Discord data, the median (mean) percentage price increase was 3.5% (7.4%), while the 75th percentile of the distribution was 6.3%. In the case of Telegram data, the median (mean) percentage price increase was 5.1% (9.8%), while the 75th percentile of the distribution was 9.2%. Recall that the January–June 2018 period was a period in which cryptocurrency prices and trading volume were falling significantly; hence "moderate" percentage increases were an achievement for the pump.

From the above discussion, it is not surprising that the coin rank is the independent

Table 8.5: Correlations Among Variables: Discord, N=1,034

Variable	% Price inc.	Exchanges	Pair Count	Rank	Server Member Count
% Price inc.	1				
Exchanges	−0.14	1			
Pair Count	−0.056	0.71	1		
Rank	0.45	−0.43	−0.18	1	
Server Member Count	−0.011	−0.0090	0.017	0.0015	1

Table 8.6: Correlations Among Variables: Telegram, N=3,767

Variable	% Price inc.	Exchanges	Pair Count	Rank
% Price inc.	1			
Exchanges	−0.16	1		
Pair Count	−0.067	0.69	1	
Rank	0.35	−0.45	−0.19	1

variable that is most highly correlated with the percent price increase of the pump, both on Discord (0.48) and Telegram (0.35.) The correlations among the variables are shown in Table 8.5 and Table 8.6. As Table 8.5 and Table 8.6 show, the correlations are similar across the Discord and Telegram platforms.

8.3.3 Analyzing the Two Types of Pump Groups Separately: Telegram Data Only

The analysis by pump-and-dump category is only conducted for Telegram. This is because Discord only has 71 transparent pumps, while Telegram has 271.

Descriptive statistics for the same variables used in the "unified" analysis are shown separately for the transparent and obscured pump types separately in Table 8.7 and Table 8.8. The following key differences were found:

- The transparent type of pump and dumps did not pump the same coin over and over. In particular, 167 different coins were used in the 271 transparent pumps. In the case of obscured pumps, just 276 different coins were used in 2,198 pumps! Thus although the second category had roughly eight times as many pumps as the first category, they employed less than twice the number of coins.

- Transparent pumps were more likely to pick a particular trading platform (exchange)

Table 8.7: Descriptive Statistics: (Telegram) Transparent Pumps, N=271

Variable	Obs	Mean	Std. Dev.	Min	Max
Max % Price inc.	271	29.44	53.4	0.49	341.99
Exchanges	271	13.5	20.3	1	163
Pair Count	271	14.56	71.25	1	759
Rank	271	699.53	613.43	3	2,036
Views	271	3,241.33	3,021.19	0	14,498
January 2018	271	0.14	0.35	0	1
February 2018	271	0.13	0.34	0	1
March 2018	271	0.24	0.43	0	1
April 2018	271	0.23	0.42	0	1
May 2018	271	0.16	0.37	0	1
June 2018	271	0.1	0.3	0	1
Binance-only	271	0.18	0.38	0	1
Bittrex-only	271	0.11	0.32	0	1
Binance-Bittrex	271	0.06	0.24	0	1
other exchange	271	0.42	0.5	0	1
No exchange	271	0.23	0.42	0	1

Table 8.8: Descriptive Statistics: (Telegram) Obscured Pumps, N=2,198

Variable	Obs	Mean	Std. Dev.	Min	Max
Max % Price inc.	2,198	7.12	13.64	0.42	309.09
Exchanges	2,198	18.24	22.72	1	182
Pair Count	2,198	17.18	63.25	1	759
Rank	2,198	334.73	366.64	2	1,935
Views	2,198	10,438.91	10,070.62	0	77,266
January 2018	2,198	0.16	0.37	0	1
February 2018	2,198	0.11	0.32	0	1
March 2018	2,198	0.12	0.32	0	1
April 2018	2,198	0.28	0.45	0	1
May 2018	2,198	0.2	0.4	0	1
June 2018	2,198	0.13	0.34	0	1
Binance-only	2,198	0.22	0.42	0	1
Bittrex-only	2,198	0.19	0.4	0	1
Binance-Bittrex	2,198	0.05	0.22	0	1
other exchange	2,198	0.02	0.15	0	1
No exchange	2,198	0.51	0.5	0	1

for the schemes, while the obscured type often did not specify and exchange. Transparent pumps were also more likely to stay away from the dominant exchanges.

- The transparent pump and dump schemes achieved a 7.7% median rate of return, while the obscured pump and dump schemes achieved a 4.1% median return.

Additionally, there is high concentration for both channels in terms of pump origin. In the case of the transparent traders, three channels accounted for more than 65 percent of the pumps. In the case of the obscured traders, six channels accounted for more than 75 percent of the pumps.

8.4 Formal Regression Results

8.4.1 All Pump and Dump Schemes Included Together

In the regressions in Table 8.9, the percentage price increase was used as the dependent variable. Because the variables in the analysis are skewed, a log/log OLS regression was run using the natural logarithm of the variables, both the dependent variable and the independent variables.[10] Clustered standard errors at the level of the coin were employed since many of the coins appear more than once in the data set.

The regression results when all pump and dump schemes are included together (see Table 8.9 are as follows:

- In the case of Telegram, the log/log regression has an adjusted R-squared of 0.32 versus 0.30 for Discord.

- The ranking of the coin is positively associated with success for both Discord and Telegram. This effect is highly significant in both cases.[11] Coins with lower market capitalization typically have lower average volume. Lower average volume gives the pump scheme a greater likelihood of success.

- The number of exchanges on which the coin can be traded is negatively associated with success and the effect is statistically significant for both Discord and Telegram. This makes intuitive sense, because with fewer exchanges, pump schemes have better control over the total volume of the coin.

[10]Not surprisingly, the log/log regression has much higher explanatory power (in the sense that it has a much higher adjusted R-squared) than either a log/linear or linear/linear specification. This is true both for Discord and Telegram.

[11]Recall that higher rank means more obscure.

Table 8.9: Examining What Affects Success of Pump and Dump Schemes:

Independent Variables	Telegram Dept. Var. % Price Increase log/log	Discord Dept. Var. % Price Increase log/log
Exchanges	-0.29***	-0.23***
	(0.057)	(0.067)
Pair Count	0.034	0.15**
	(0.05)	(0.066)
Rank	0.16***	0.24***
	(0.043)	(0.050)
Server Members		-0.007
		(0.020)
Views	-0.061***	
	(0.013)	
February 2018	-0.036	-0.24**
	(0.067)	(0.087)
March 2018	-0.046	-0.13
	(0.069)	(0.091)
April 2018	-0.20***	-0.40***
	(0.052)	(0.11)
May 2018	-0.43***	-0.49***
	(0.070)	(0.079)
June 2018	-0.26***	-0.66***
	(0.075)	(0.11)
Binance Only	-0.31***	-0.24***
	(0.055)	(0.062)
Bittrex Only	-0.17***	-0.23***
	(0.050)	(0.075)
Binance-Bittrex	-0.41***	-0.38***
	(0.086)	(0.080)
Observations	2,649	952
Adjusted R^2	0.32	0.30

Standard errors in parentheses: They are clustered at the level of the coin.
* significant at the 90% level
** significant at the 95% level
*** significant at the 99% level

- The number of other coins that the coin can be traded with is positive and statistically associated with success in the case of Discord. In the case of Telegram, the estimated coefficient is positive but is insignificant. One possibility is that more trading pairs allow greater flexibility for those involved in the pumps.

- In the case of Discord, the estimated coefficient on the variable "Server Member Count" is negative, but not significant. In the case of Telegram, the variable "Views" is negatively associated with success and the effect is statistically significant. One possible interpretation is that it is hard to coordinate if there are too many people potentially involved in the pump.[12]

- Pumps on Binance and Bittrex do worse than pumps not on those exchanges. It might be that, since these are dominant exchanges, more people are involved in the pumps - and coordination is more difficult.

- Perhaps most importantly, the declining "success" rate over time, as shown by the negative coefficients on the monthly dummy variables holds, even after controlling for the other factors. In both Telegram and Discord, the estimated coefficients associated with April, May and June are statistically significant, suggesting a deep decline in profitability over time.

8.4.2 Transparent and Obscured Analyzed Separately: Telegram Only

The regression results when the pump and dump schemes are analyzed separately by category are shown in Table 8.10. The key results are as follows:

- In the case of transparent traders, the log/log regression has an adjusted R-squared of 0.60 versus 0.26 for obscured traders.

- The ranking of the coin is positively associated with success for both types. This effect is highly significant for transparent traders and significant for obscured traders.

- The number of exchanges on which the coin can be traded is negatively associated with success and the effect is statistically significant for both groups.

- Again, pumps on Binance and Bittrex do worse than pumps not on those exchanges.

[12] It might also be because "Views" could be endogenous. All of the other results are robust to excluding "Views" from the analysis.

Table 8.10: Examining What Affects Success of Pump and Dump Schemes:

Independent Variables	Transparent Max % Price inc.	Obscured Max % Price inc.
Exchanges	-0.38**	-0.21***
	(0.14)	(0.055)
Pair Count	0.11	-0.04
	(0.12)	(0.05)
Rank	0.39***	0.09*
	(0.11)	(0.039)
Views	0.11*	-0.04**
	(0.049)	(0.012)
February 2018	0.92***	-0.21**
	(0.23)	(0.061)
March 2018	0.75***	-0.17*
	(0.21)	(0.069)
April 2018	0.57*	-0.27***
	(0.25)	(0.054)
May 2018	0.77*	-0.55***
	(0.34)	(0.065)
June 2018	0.87*	-0.42***
	(0.33)	(0.072)
Binance Only	-0.88***	-0.2***
	(0.21)	(0.054)
Bittrex Only	-0.72***	-0.03
	(0.19)	(0.052)
Binance-Bittrex	-1.09**	-0.31***
	(0.32)	(0.084)
Constant	-0.67	2.33***
	(0.89)	(0.32)
Observations	271	2,198
Adjusted R^2	0.60	0.26

Standard errors in parentheses: They are clustered at the level of the coin.
* significant at the 90% level
** significant at the 95% level
*** significant at the 99% level

- Interestingly, the variable “Views” is positively associated with success and the effect is statistically significant for transparent traders. However, this variable is negatively associated with success and the effect is statistically significant for obscured traders . This may be because, in general, there are fewer “viewers” for the transparent pumps.

Recall that the transparent pumpers typically restricted membership.

- For the obscured traders, the declining "success" rate over time, as shown by the negative coefficients on the monthly dummy variables holds, even after controlling for the other factors.

- Strikingly for the transparent traders, the "success" rate is virtually constant over time as shown by the very similar coefficients for the dummy variables on the months February through June 2018.[13]

8.4.3 What happens after the pump is over

An interesting question is what happens after the pump is over. To address this issue, two additional variables were calculated.

- **Starting price**: this is the starting price associated with the maximum five minute percentage increase in price. It can be interpreted as the "pre-pump" price.

- **End price**: This is the minimum price in the 48 hours after pump.

- The following variable was then calculated: $\frac{\text{End price}-\text{Starting price}}{\text{Starting price}}$. This is the percentage change in price from the pre-pump period to the post-pump period.

The following was found: The median percentage change in price from the pre-pump period to the post pump period is -41% for Discord data and -38% for Telegram data. Overall, more than 60% of the coins have a lower "post-pump" price than the "pre-pump" price. Even though prices were generally falling during this period, a 40% fall in prices in 48 hours is large.[14]

[13]It appears that the success rate jumped from January to February, and stayed at that level over time.

[14]Regressions were run using the percentage change in price from the pre-pump period to the post-pump period as the dependent variable, and the right-hand-side variables as the independent variables. In these regressions, the adjusted R-squared was virtually zero.

8.4.4 Trading Volume Data

Corresponding volume data does not exist for the data utilized since volume data on `coinmarketcap.com` is reported continuously over the preceding 24 hour period and it is not clear how often volume information is updated.[15]

But the following volume variable was calculated: "Per-change volume after," which equals the maximum (24 hour) volume in the 24 hours following a pump signal less the minimum (24 hour) volume in the 24 hours following a pump signal divided by the minimum (24 hour) volume in the 24 hours following a pump signal. The following was found:

- On both Discord and Telegram, there is approximately a 30 percent correlation between (i) the maximum five-minute percentage change in price and (ii) the "Per-change volume after."

- Since the price signal occurs before the changes in volume, a regression could be run with volume change as the dependent variable and put the maximum five-minute percentage change in price as a right-hand-variable along with the other independent variables used in the price regressions. In such a case, only that variable is significant and the adjusted R-squared is relatively large. Even though the timing cannot be confidently determined it does suggest the following: The pump organizers buy first, increasing the price. Then the "herd" jumps in, where the herd is comprised of other people who received the pump signal and outsiders (some of whom may be using trading algorithms.) During this period, the original "pumpers" are likely selling there shares as well.

- The two points above suggest that the maximum five-minute percentage change in price is a good proxy for success.

[15] Since the data utilized does not include delineated trading volume, profitability from the pumps cannot be quantified. Even if did have trading volume by time, it still would be impossible to measure profitability. This is the "pumpers" act as individuals and others can trade as well. The only way to measure profitability would be to have access to trading activity over time at the individual level; labeled trading data is not available.

8.4.5 The number of exchanges mentioned by the pump

Finally, the data regarding the number of exchanges mentioned by the pump is briefly summarized. Exchanges are observed for 546 pumps on Discord, around half of the total. This data was scraped from the pump signal, counting any exchanges directly mentioned in the signal message.

Table 8.11: Correlation Table: Pump Exchange and Independent Variables, N=546

Variable	Pump Exchanges	Exchanges	Pair Count	Rank75
Pump Exchanges	1			
Exchanges	0.29	1		
Pair Count	0.13	0.73	1	
Rank	−0.25	−0.40	−0.18	1

While most pumps mention a single exchange, more than 18 percent of the pumps mention more than one exchange. Correlations among the number of pump exchanges and the independent variables are shown in Table 8.11. Not surprisingly, the number of exchanges used in the pump is negatively correlated with the rank of the coin (-0.25) and positively correlated with the number of exchanges the coins are traded on (0.29). These numbers give additional confidence that actual pumps are being chosen.

8.5 Conclusion

In this chapter the phenomenon of "pump-and-dump" schemes for cryptocurrencies was examined. The proliferation of cryptocurrencies and changes in technology have made it relatively easy for individuals to coordinate these activities.

In terms of scope, this "pump-and-dump" phenomenon was found to be widespread on both Discord and Telegram. The most important variable in explaining success of the pump was found to be the ranking of the coin. While there were attempts to pump coins spanning a wide range of popularity, "pumping" obscure coins gave the pump scheme the potential for greater success at the expense of increased risk, i.e., volatility. In some sense, the choice between using lower or higher ranking coins is similar to using conservative and risky investment strategies. The benefit of investing in assets with low expected returns is that the volatility is low. The key difference, of course, is that deliberately "pumping" cryptocurrencies for profit is unethical.

The research carried out in this chapter is an important foundation for further exploring malicious trading activities that utilize the sometimes extreme price variations brought about by "pump-and-dump" schemes. For instance, short sellers can take advantage of the expected price drop following a pump to make money when the price increases and when the price decreases. Short selling is a form of speculative trading in which an individual bets the price of a cryptocurrency will go down. If the price drops as predicted the trader receives the difference between the price when the order was placed and the current price. In theory, in the middle of a pump and dump scheme, when the price is high, a trader can submit a short sell order for the price drop that follows. In practice, this may be difficult because of the required transaction confirmations as well as the quick timing of some of the "pump-and-dump" schemes. pump schemes that utilize targets instead of relying on scheme timing would be better suited for this type of malicious activity.

CHAPTER 9

MARKET MANIPULATION THROUGH ORGANIZED, TARGET-BASED TRADING

Despite the fact that pump and dump signals can be quite heterogeneous, analysis in Chapter 8 examined all pump signals. Chapter 8 outlined two main approaches: countdown pumps and target pumps. Recall that in the countdown category pump information is released to the group incrementally, with the cryptocurrency name being posted at the pump time. The target group, on the other hand, releases all information at once and includes trade target values, which countdown pumps typically do not do. Chapter 8 found that pumps in the group that mostly used countdowns were more successful. This could be explained by the differences in the pump signal, or possibly how success is measured. In this chapter, target pumps are more closely examined.

9.1 Methodology

In this section the data sources are examined, and the data normalization procedure is explained.

9.1.1 Extracting Targets from Pump Signals

Using the data found in Chapter 8 as a starting point, pump signal collection continued through January 2019 with the same collection methodology. Using the chat application accounts that were still active, pump signals were programmatically collected from the associated channels. This extended data contains 12,252 target and countdown pump signals between July 2017 and January 2019. The format of the targets within each pump signal needs to be known before non-target signals could be reliably removed. Therefore random signals were selected and post formats were examined for consistency.

Figure 9.1: Example pump signal from Whale Pump Group.

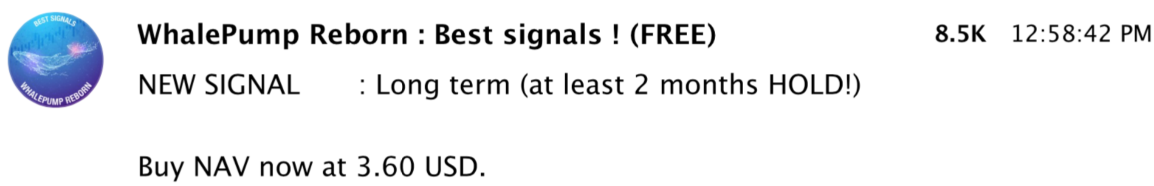


Figure 9.2: Example pump signal from Whale Pump Group.

Figure 9.3: Example target pump signal from Mega Pump Group.

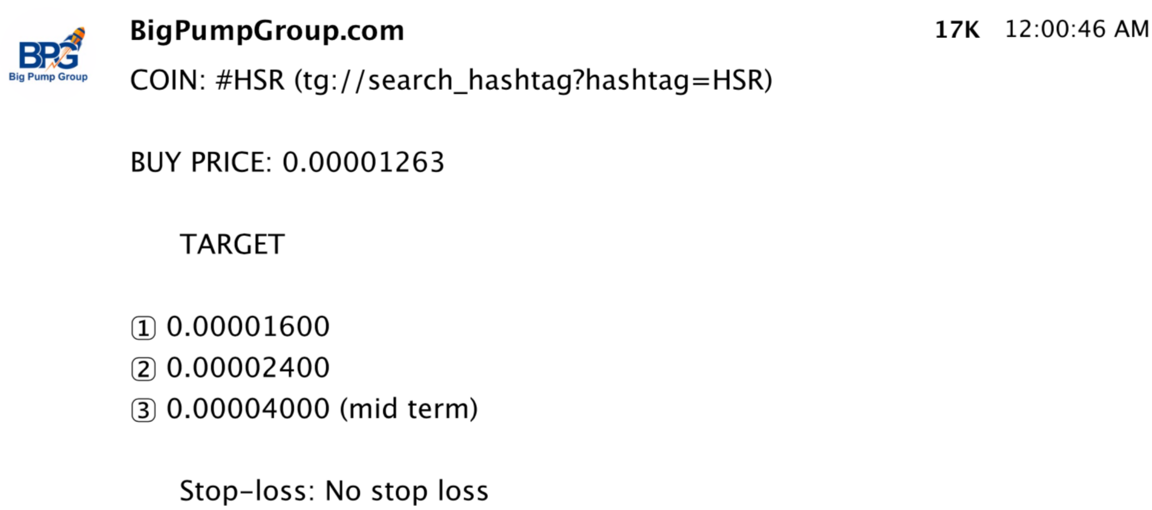


Figure 9.4: Example pump signal from Big Pump Group.

Figures 9.1-9.4 show examples of the various formatting irregularities found within pump signals. Figure 9.1 and Figure 9.2 only give buy targets which are more than likely the cryptocurrency's current price. Figure 9.3 is an example of an incomplete target signal as it gives a buy and a sell target but nothing else. Complete target signals have a minimum of one value for each of the targets: stop-loss, buy, and sell. Traders can extrapolate missing values based on pump signal patterns. However, this analysis only uses values from explicitly stated targets. The last signal in Figure 9.4 gives more information than the previous pump signals but would be considered an incomplete signal because of the lack of an explicit stop-loss target.

Reducing the data to only include target signals required considerable effort. As the example posts show signal text formatting is rarely consistent between pump groups and is occasionally inconsistent within a single pump group. The process began with the development of robust regular expressions that would accurately locate target data within each of the pump signals. Valuable target data is not always posted in a vertical list as displayed in Figure 9.4. Occasionally this data is a comma or hyphen delimited list, sometimes colorful emojis are used to show which target is which type, sometimes the target type is spelled out. In other words, numerous cases need to be accounted for. Once automated data scraping had completed, each of the pump signals was manually compared to the scraped data for correctness. Missing or incorrect data was corrected during the manual inspection process.

During data discovery it was found that not all monetary values were consistent within a post. Pump group organizers would switch between USD, BTC, EUR, percent and satoshis[1] for target values. Percentage values were almost always based on the buy price of the cryptocurrency. In the event that no buy target existed the remaining target values could not be calculated. Pump signal organizers would also occasionally switch between numerical abbreviations and their whole number equivalents. To normalize the number formatting throughout a signal an algorithm was developed to convert abbreviations to whole numbers. For example, 18k-22.3-24k-26.7k would be converted to 18000-22300-24000-26700; these values could then easily be converted to satoshi. Many of the cryptocurrencies targeted by these pump groups cannot be traded with USD, they can only be traded with bitcoin.

[1]Satoshis are a fraction of a whole bitcoin. A satoshi is 1/100,000,000 of a bitcoin.

For this reason all monetary values not already represented in satoshi were converted.

With the data now in an easily digestible format, record pairing could begin. Pump signals with no target values are not technically target pumps; 3,375 pump signals with no monetary target values were removed. The collected pump signals contain a varying number of each of the targets: 0-5 stop-loss values, 0-3 buy values, and 0-12 sell values. It is worth noting that the 10th through 12th sell targets are never acquired for any of the pump signals for which they are explicitly given. The 9th target, which is more of a long-term target anyway, was only reached for one signal in the data and it was crossed roughly six months after the pump signal was posted. A further 1,409 signals were removed since the cryptocurrencies were not tracked by `coinmarketcap.com`, so the corresponding pricing data could not be gathered. An additional 1,471 signals were removed as they were complete duplicates.

Clearing the dataset of unusable records left a total of 3,683 meaningful pump signals between July 2017 and January 2019 with at least one target value.

9.1.2 Pricing Data

This analysis in following sections relies heavily on `coinmarketcap.com` for cryptocurrency price and volume data. `Coinmarketcap.com` is the leading website for aggregate cryptocurrency trading data. Minute level data has yet to be made available through `coinmarketcap.com`, however, the service does present data at roughly 5-minute intervals. This 5-minute data provides prices in bitcoin as well as a 24-hour aggregate volume. This fine grained data was chosen over daily values because of the inherent volatility that comes with cryptocurrencies, as well as the fast pace at which pump and dump activity can motivate price movements. These quick increases and equally quick decreases can be seen in Figure 9.5. This is the activity that resulted from a countdown pump signal that originated within the Big Pump Signal Telegram group on January 13th, 2018. This pump reached its peak value approximately 9 minutes after the coin name was posted to the channel.

Pump and dump price movements could be extrapolated from daily max and min price values, but calculations yield more meaningful results with more finely grained data.

Figure 9.5: Pump and dump using Genesis Vision as observed on Coinmarketcap.

Scraping all available cryptocurrency price and volume data from `coinmarketcap.com` gives access to over 2,000 coins and tokens as well as over 220 cryptocurrency exchanges. This data was reduced to only cover the 293 cryptocurrencies observed in the cleaned pump signal data. The date range was also modified to roughly cover the the same time window as the pump signal data. The final set contains roughly 145,000,000 data points covering January 2017 to November 2019.

There are limitations to collecting data from `coinmarketcap.com`. This data aggregation service only collects coin and token data from a relatively short list of exchanges. Cryptocurrencies only appear on `coinmarketcap.com` after they sustain trading activity exceeding a threshold, and they can be removed for a variety of reasons including inactivity.

9.1.3 Matching Pump Information with Trade Data

With no central naming organization, cryptocurrency names can change and identifiers can be used for multiple entities without consequence. `Coinmarketcap.com` provides a reasonable solution to any naming issues that may be encountered. A URL is issued to each cryptocurrency tracked so that site visitors can view coin or token information specific to a single entity. Unlike cryptocurrency names, URLs must be unique.

Beginning with the URL values reported on `coinmarketcap.com` a name map was created. This map was extended to include the coin name and ticker value, as well as several combinations of the two to ensure complete coverage. Once completed this map was used to associate price and volume values with the coins or tokens present in each of the pump signals. This strategy allowed mapping between 293 distinct cryptocurrencies and 3,683 unique pump signals.

9.2 Analysis

The goal in the following section is to describe the dynamics of target based pump and dump signals through the data made available by `coinmarketcap.com`. Success here is defined by movements between buy targets and stop-loss targets; if a sell target is reached within the range then the pump was a success.

9.2.1 Identifying Target Cycles

Identification of successful pump and dump signals began with the development of a method that would flag points of interest within the cryptocurrency price data. These points of interest include the pump signal time as well as every occurrence of the cryptocurrency's price crossing a target value. Figure 9.6 gives an example of the method output.

In the event that multiple targets exist for a pump signal they grouped by type and numbered with the lowest value having the lowest associated label number. Figure 9.6 shows one stop-loss, two buy targets, and three sell targets. Moving through these targets from the bottom of the figure to the top they are numbered as follows: stoploss1, buy1, buy2, sell1, sell2, and sell3.

An important detail about the flagging method is only the first target crossing is

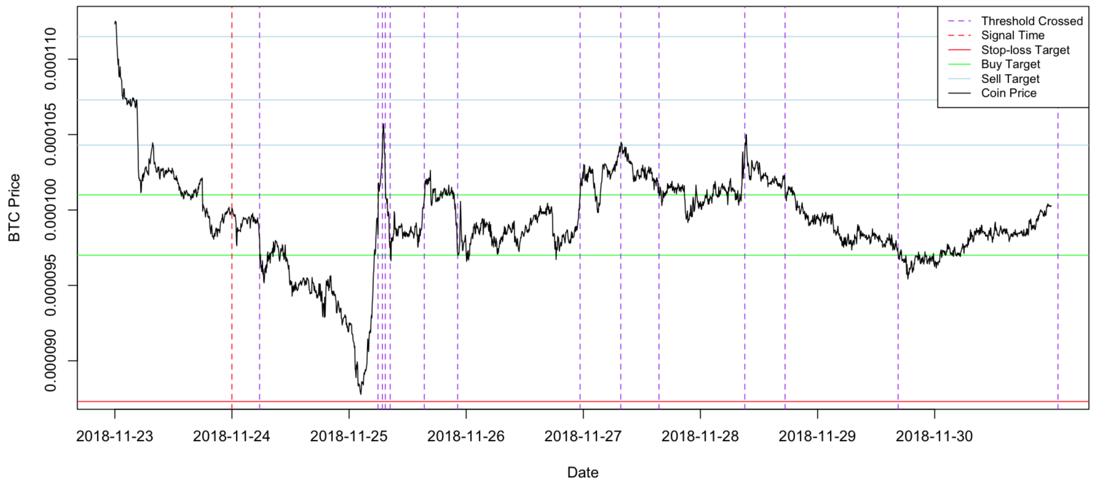


Figure 9.6: Ark pump activity from November 2018.

flagged. The price can fluctuate between the next highest and next lowest target and as long as the price never crosses either target it will stay flagged at the current target. Therefore, if the buy1 target is crossed, as it is in Figure 9.6 on 11-24, it stays there until it crosses another target, which it does on 11-25 after almost touching the stoploss1 target earlier that same day.

# Signals	No Data	No Cross	Stoploss1	Buy1	Buy2	Buy3	Sell1	Sell2	Sell3
3,683	43 1.2%	293 8%	39 1.1%	1,442 39%	1,243 34%	2 0.05%	398 10.8%	95 2.6%	84 2.3%
% Above	0%	0%	38.5%	50.7%	45.8%	50%	80.2%	54.7%	19%
# Signals	Sell4	Sell5	Sell6	Sell7	Sell8	Sell9	Sell10	Sell11	
3,683	28 0.8%	4 0.1%	6 0.16%	1 0.03%	2 0.05%	1 0.03%	1 0.03%	1 0.03%	
% Above	3.6%	0%	0%	0%	0%	0%	0%	0%	

Table 9.1: First observed target for each pump signal.

Figure 9.6 shows the expected price movement through targets following the posting of a pump signal. However, not all pump signal organizers post correct price values or time pumps properly and the result of this minor misstep can be seen in Table 9.1.

Table 9.1 shows that a majority (73%) of the pump signals are immediately followed by the price crossing one of the buy targets. This is a good sign as it most likely means that these pump signals are not copied pumps. Conversely, pump signals that are directly followed by the price achieving a sell target are suspected to be copied pumps. 336 of the pump signals cross no thresholds for one of two reasons: the first being a gap in the data for that cryptocurrency, and the second being the prices simply never cross a defined target.

The following analysis treats pump signals as if they are actively being used to place buy and sell orders. Instead of locating the pump by searching a 48-hour window on either side of the pump signal, price-based orders are placed at the pump signal time and the movements are observed. This is done because it reflects the most likely response to target pump signals by members of the pump group. The targets provide the information needed to place orders for buying, selling, and setting a stop-loss. This also makes our analysis easier in two ways. First, the analysis does not have to rely so heavily on the pump signal time to locate the start of a pump, and second trading cycles can be defined completely independent of the signal time. Figure 9.6 displays the consequences of this decision. For this pump, the cycle does not start at the pump signal time, instead it begins shortly after when the price crosses the buy1 target.

Once posted by the pump group organizers, these signal values never change. This allows for long term pump activity observation assuming the prices stay within the target range. To programmatically simulate trading based on target pump signals an algorithm was developed to identify cycles within the crossed targets. Following cryptocurrency trading logic, a pump cycle adheres to the following rules:

- Starts at one of the defined buy targets
- The end of the cycle is identified either at the stop-loss following a buy signal or at the end of the data for that cryptocurrency (whichever comes first)

The trading entry point is whichever buy target is crossed first following a pump signal, and the exit point is either the highest sell target achieved achieved within this cycle or the first stop-loss crossed. In this regard, a pump is successful if it reaches a sell target within this cycle and it is unsuccessful if the price only travels from buy to stop-loss without crossing a sell target. Because this method relies so heavily on the target values it can only use complete pump signals. Recall from Section 9.1.1 that pump signals contain varying levels of information and many are incomplete. To proceed, 2,259 incomplete pump signals were removed from the pump signal data. An additional 172 rows were removed because no signals were crossed (126), the pump signal target values were not in order (24), or no cycles were detected (22)[2]. This leaves 1,252 records with complete pump signal data.

Pump signal targets technically could be inferred where they are missing. By examining the differences between targets for complete pump signals it was found that targets are, on average, roughly 10% apart. In this regard, these methods could also be extended to pump signals without target values by using the current price as a buy target. However, these assumptions would potentially take this research back to examining all pump signals and not just target signals. For this reason, extending incomplete signals through implied targets has been earmarked for future work.

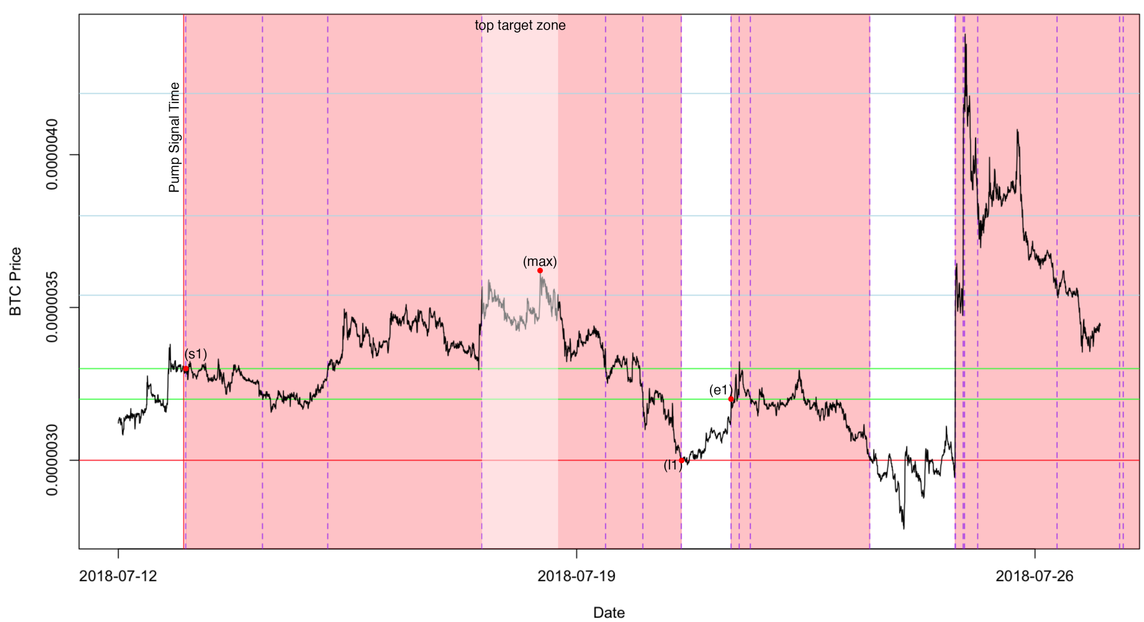


Figure 9.7: Cycle identification within cryptocurrency time series data (iostoken).

[2]These pump signals began trading within the sell or stop-loss target zone and stayed within that zone. They never crossed a buy target.

Within each cycle certain points of interest are identified. The top target crossed is flagged and the max price within that top zone is recorded. This top zone starts the first time the price crosses the top target and ends the last time the price crosses the target when it begins moving toward the subsequent stop-loss target. Within this zone the price can and occasionally does cross lower sell targets but for simplicity it will be treated as a "max zone." Finally, the lowest value is identified in the area between the max zone and the end of the current cycle. Figure 9.7 displays an example cycle and related points of interest.

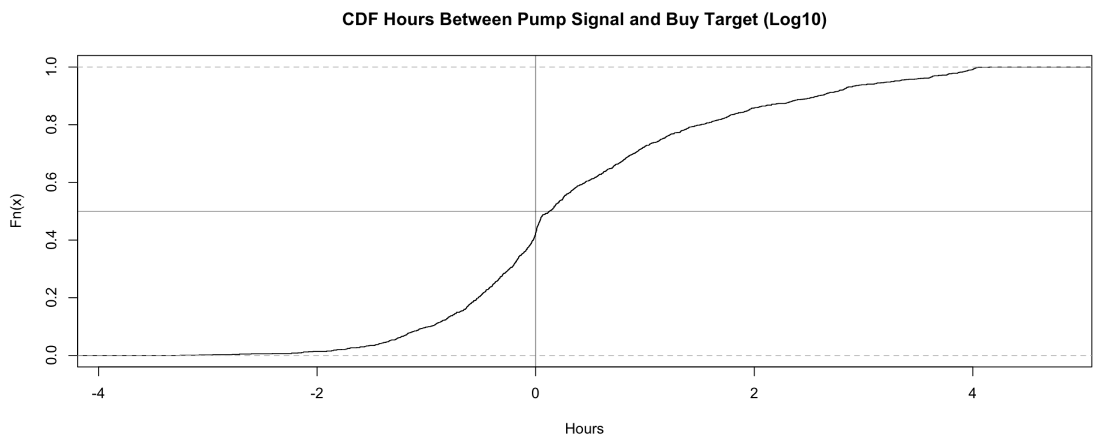


Figure 9.8: Hours between pump signal and the price crossing the first buy target.

Figure 9.7 shows a very nice example of a pump signal where the first buy target is met very close to the pump signal time. Recall from Table 9.1 that the first target crossed is not always a buy target. Figure 9.8 shows how long it takes to get to the first buy target following a pump signal. 75% of the complete pump signals reach a buy target within 14 hours. 13% of pump signals do not reach a buy target until at least 7 days after a pump signal and the max distance between a pump signal and a first buy target is nearly 592 days. In these cases, the pumps could be more accurately considered to have failed, since the success is achieved so late compared to the signal time.

With a median difference of 1.3 hours from pump signal to buy target, some price variability is expected. Figure 9.9 plots a CDF of the percent difference between the signal price and the price at the first buy target crossed. 708 of the pump signals saw a price increase to the first buy target and 543 saw a drop in price between the signal and the first

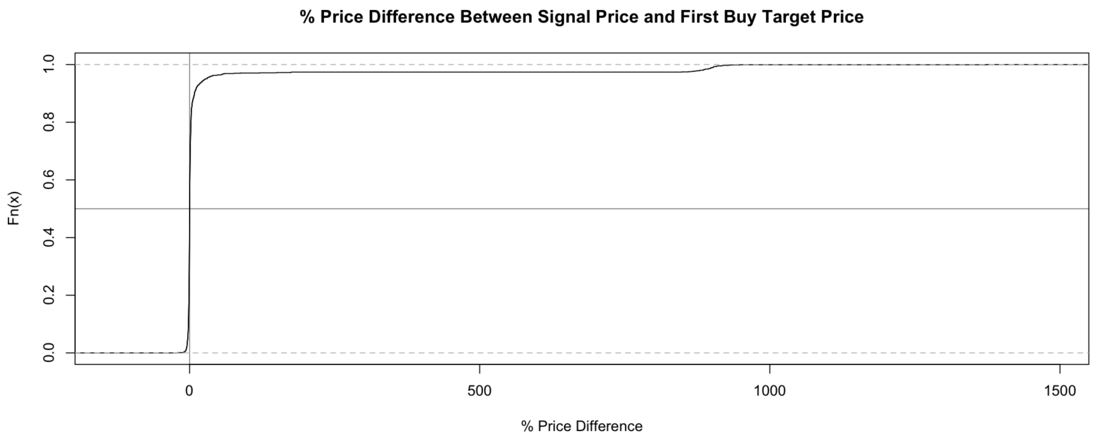


Figure 9.9: Difference between cryptocurrency price at the pump signal and the price at the first buy target.

buy target. Only 37 pump signals experienced a 100% price increase or more between the signal time and the first buy target. This sounds like an incredible price increase. However, the average signal price for those 37 cryptocurrencies is only 0.000540170 bitcoin[3].

9.2.2 Success in Waves

Cyclic price movements through target values grant the ability to look at success farther out than the time window immediately following a pump signal. These movements also allow the ability to calculate success through one or more of these cycles. Ultimately, success was defined by observing the first cycle for each pump signal, however, Table 9.2 includes summary values for two additional pump signal cutoff values.

Measure	Successful	Unsuccessful
First cycle only	647	605
Overall (cutoff at next signal)	716	536
Overall	1,002	250

Table 9.2: Success based on inclusion or exclusion of trading cycles.

By focusing only on the first cycle following a pump signal a measure of immediate success can be calculated. Hitting a sales target within the first cycle is marginally more

[3]At the time of writing 0.000540170 bitcoin is only worth 4.93 USD.

probable than getting heads on a random coin flip. First cycles have a 51.7% chance of being successful.

If the time frame for success is extended beyond the first cycle many more pump signals are considered successful. The first measure of overall success ends the search window at the next pump signal for the coin being observed; if no future signals exist for a cryptocurrency then the cutoff date of the price data is used. There is an 11% increase in success between the first cycle and this first measure of overall success.

Furthermore, if all subsequent pump signals are ignored and all trading cycles for a pump signal are grouped together then a 55% increase in success can be observed when comparing to the first cycle group. This last measure of success is somewhat misleading because a handful of these pump signals have up to two years to experience success within a cycle.

The two overall measures are not necessarily realistic success measures when it comes to relating this analysis back to real-world trading activities. If traders lose out on the first cycle of the pump signal it is unlikely that they will submit the same trades and risk further losses.

Days since start	All	% Successful Pump & Hit Stop-loss	% Unsuccessful Pump & Hit Stop-loss
1	16.5%	2.5%	31.6%
2	23.6%	4.5%	44.0%
3	28.2%	6.2%	51.7%
4	32.9%	8.3%	59.2%
5	37.1%	11.3%	64.8%
6	40.3%	13.0%	69.4%
7	43.1%	13.9%	74.4%
# Signals	1,252	647	605

Table 9.3: Stoploss crossed within X days of cycle start (first cycle only).

Now that cycles and success have been defined, these two measures can be used to peer into the first few days of activity following a pump signal to see when cycles end. Table 9.3 shows the percentage of pump signals that cross a stop-loss target after 1-7 days. After the first day only 16.5% of all pump signals have completed a cycle and crossed a stop-loss target. By the middle of the seven day window the percentage of signals that cross a stop-loss value doubles to 33%, and by the end only 43% of pump signals have crossed a stop-loss value.

If this group is split by success, two different trends emerge. Successful pumps are slow to reach stop-loss targets with only 14% hitting a stop-loss target after 7 days. Unsuccessful pumps, on the other hand, reach stop-loss targets much faster. 32% cross a stop-loss target within the first 24 hours and just under 75% end their trading cycle within seven days of crossing the buy target. These differences in speed to the bottom could be the result of trader behavior after realizing the outcome of the pump scheme. Unsuccessful pump attempts could be abandoned early, driving the price down at an accelerated pace. Conversely, successful pumps could see an influx of new money propping up the price for an extended period of time.

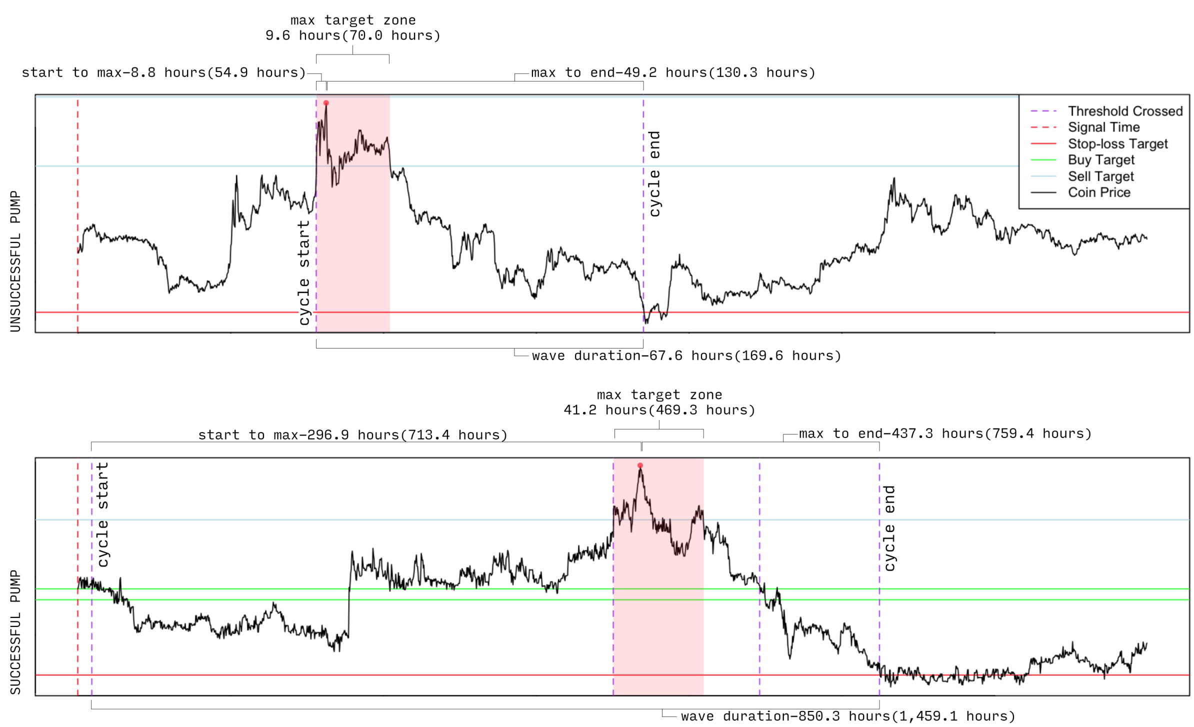


Figure 9.10: Cycle timing summary statistics - median(mean) hours.

Figure 9.10 displays summary statistics for the time elapsed between points of interest within a trading cycle. These summary values are grouped by pump success with an unsuccessful pump displayed on the top plot and a successful pump on the bottom plot. These hourly values are median values with associated mean in parenthesis. The red rectangle in the plot simply displays the area in which the pump is in its "max zone."

Further explaining the values found in Table 9.3, unsuccessful pumps outpace their successful counterparts through every section of the trading cycle. Successful pumps take 34 times as long to reach the max value, spend 4.3 times as long in the max zone, and take 8.9 times as long for pries to fall back to the stop-loss target when compared to the unsuccessful pumps (comparisons use median values).

9.2.3 Price Movements within Cycles

To quantify the magnitude of the price movements within a trading cycle summary statistics were calculated for the points of interest outlined in Figure 9.7. These values have been made available through Table 9.4.

Range	Min	Max	Median	Mean	Success
Start to max value	0.7	963.8	29.6	54.2	TRUE
Max value to stop-loss	-1.7	-2,510,825.0	-51.9	-4,363.5	TRUE
Start to max value	0.0	395.2	2.7	5.3	FALSE
Max value to stop-loss	-0.6	-1,154.2	-13.6	-20.5	FALSE

Table 9.4: Percentage price movements between points of interest.

On average successful pumps see an 30% median increase between the start target and the max value in the cycle. Looking back at Figure 9.10 successful pumps spend a median time of 41.2 hours in this zone. Afterwards it takes a median of 437.3 hours to drop down 52% (median) to the stop-loss value. Both successful and unsuccessful pumps see a price increase between the start and the max value, however, the drop in price that follows the max value is consistently greater than the rise. The max value seen in the max to stop-loss for successful pumps is somewhat misleading but it makes sense when the low trading price of this coin and the 6-months between the max value and the stop-loss are taken into account.

9.3 Conclusion

In this chapter a subset of pump and dump schemes targeting cryptocurrencies was examined. Trading cycles were defined utilizing these target signals and these cycles were then used to define success. It was found that within the first cycle success is almost as likely as heads turning up after the toss of a coin. However, differences arose when the timelines of trade cycle events was examined. Successful pumps have trading cycles that are on average 8.6 times longer than trading cycles for unsuccessful pumps. Additionally, it was found that on a long enough timeline every pumped coin or token drops back down to the stop-loss value defined in the signal.

In Chapter 8, it was found that the median (mean) percentage price increase 3.5% (7.4%) and 5.1% (9.8%) for Discord and Telegram respectively. The analysis that produced those measures relied on the fastest increase in price within a window surrounding the pump signal. The analysis in the current chapter looked more at slower movements between target values and found that the median (mean) price increase following a pump signal was 9.7% (31.7%). Although the median values are close, it is difficult to directly compare the two methods because of the vastly different success definitions.

CHAPTER 10

CONCLUDING REMARKS

In a relatively short period of time, cryptocurrencies have grown from a niche market to an ecosystem with a current market capitalization of roughly 188 billion USD. To put this in perspective, the fourth and fifth largest banks in the United States have respective market capitalizations of 205 billion USD, and 100 billion USD. Cryptocurrencies' impressive gains come with equally impressive losses. Autonomous Research says that, as of July 2018, hacking-related losses from cryptocurrency exchanges total 1.63 billion USD [48]. Furthermore, CipherTrace claims first quarter 2019 cryptocurrency losses from thefts, scams, and other losses across the ecosystem may total 1.2 billion USD [49].

Due to the difficulty of navigating the safeguards built into many cryptocurrencies, malicious actors tend to focus their efforts on exchanges, wallets, and third-party cryptocurrency services. In this dissertation attention has been focused on cryptocurrency exchanges and the ways in which individuals take advantage of these services through strategic distributed denial-of-service (DDoS) attacks, unauthorized trades, and "pump-and-dump" schemes.

The work done herein falls within the broad category of security economics. Instead of viewing security issues through a purely technical lens, these threats are examined empirically. Because of the public nature of cryptocurrency transactions, it has been possible to to study financial manipulations in ways that are very infrequently available outside of this ecosystem.

In Chapter 5, the Bitcoin ecosystem's response to the shocks and exchange based manipulation outlined in Chapters 3 and 4 was quantified. Utilizing the leaked Mt. Gox transaction database, which consists of approximately 14 million matching buy and sell records recorded between 2013-04-14 to 2013-11-30, it was possible to identify what happens after a shock to the ecosystem. Two regressions that used skewness and kurtosis of the daily transaction volume as the dependent variables were built. Both skewness and kurtosis were found to decrease following a shock showing there are fewer trades with very large volumes occurring after these events. In Chapter 5, two distinct actors that consist of 50 unique accounts with distinct trading patterns were examined. Those accounts were able to fraudulently acquire around 600,000 bitcoin, worth approximately 112 million USD during the period studied. To quantify the impact of that trading activity regressions that used the daily difference of the bitcoin exchange rates and the percentage change in the daily rate as the dependent variables were developed. It was found that the trading activity of those two actors, consisting of 49 separate accounts, was associated with a 4% daily rise in the exchange rate. The combined efforts of those accounts was highly correlated with the impressive BTC/USD price increase from 150 USD to 1,000 USD in late 2013.

In Chapter 6, two algorithms for locating peaks and periods of abandonment within financial, in this case cryptocurrency, time series data were defined. Both algorithms utilized simple thresholds for discovery: a peak value must be at least 50% higher than the minimum value in the previous 30 days; a coin is abandoned if the monthly aggregate trading volume falls below 1% of the peak value; and, a coin is resurrected if the monthly aggregate trading volume moves above 10% of the same peak value used to define its abandonment. It was found that 44% of all coins at the time of writing had been abandoned at least once with 18% of those abandonments being permanent. Additionally, many of the new entrants in the cryptocurrency market, as well as coin resurrections, were found to be riding "the wave" created by the huge increase in the cryptocurrency market capitalization.

In Chapter 7, tokens and their fundraising counterparts were explored. Methods were developed for combining ICO tracker data into an easily digestible format with source and record level data reliabily measures. It was found that only 11% of the ICOs reporting the necessary values were successful. Through the examination of token returns at four points of interest it was found that initially many of these tokens realize impressive average returns (123.32%) but these averages are propped up by a small handful of successful tokens. Repeating the analysis found in Chapter 6, it was found that only 7% of tokens were abandoned, and of those 42% had been resurrected by the end of the data. The differences in coin and token abandonment could be a result of the prices they trade for and the volatility that comes with higher coin prices. Again, it was found that new entrants as well as existing tokens were riding "the wave" created by the huge increase in the cryptocurrency market capitalization.

In Chapter 8, an exhaustive search of two popular online messaging platforms, Discord and Telegram, was performed seeking cryptocurrency "pump-and-dump" groups. When the data collection efforts were complete it was possible to to categorize the posts into three distinct groups: "transparent pumps," "obscured pumps," and "copied pumps." When combined, the three groups accounted for 1,034 pump signals across 55 channels on Discord and 3,767 pump signals across 25 channels on Telegram. These pump signals were coupled with with roughly 5-minute increment price data from `coinmarketcap.com`. Data were collected on nearly 2,000 coins and tokens across 220 exchanges from mid-January 2018 to early July 2018, giving nearly 316 million data points across all coins. The most important indicator of pump success was found to be the rank of the cryptocurrency being pumped.

In Chapter 9, long term analysis was performed on pump and dump schemes by examining pump signals that include target values. By treating the analysis as if it were being used to place trade orders, trading cycles could be identified within the price data. By investigating the trade cycles for 1,252 complete pump signals it was found that pump scheme success was marginally better than the toss of a coin (51.7%). Within this trading cycle it was found that successful pumps on average see a 54.2% increase between the start and max value, where an unsuccessful pump only averages a 5.3% price increase. When compared to the pump and dump results from Chapter 8, the success rates are higher, however, a direct comparison is difficult because of how success is defined. One interesting takeaway from trading cycle analysis is that on a long enough timeline every cryptocurrency in the data saw its price fall back down to the pump signal stop-loss value. In other words, there is not enough sustained interest in small cryptocurrencies to support increasing prices.

10.0.1 Future Work

The research outlined in this dissertation lends itself to future work in both measuring crime in the cryptocurrency ecosystem, and the implementation of tools to thwart this type of behavior.

Anomalous Activity Detection: Section 5.2 shows that price manipulation is possible through strategic trading activity. This is partially due to how trading bots operate and partially due to how the cryptocurrency markets react to periods of elevated trading volume. Based on the work done in Section 5.2, a tool could be built to detect spurious trading activity and flag it for manual verification on an exchange. Using exchange order book data which consists wholly of incomplete trades in addition to publicly available anomaly detection tools, such as Prophet or AnomalyDetection, outliers could easily be detected. And, more reactive approaches than the anomaly detection method could be used to detect known bot trading patterns within the data.

Time Series Point-of-Interest Detection: The tools developed in Chapter 6 use simple thresholds for peak, abandonment, and resurrection detection. However, cryptocurrency price movement cannot always be bound by such simple definitions. Machine learning or other advanced algorithms may be better suited to detect the patterns observed throughout the life cycle of a cryptocurrency.

“Pump-and-Dump” Early Detection: Chapter 8 outlined the breadth and depth of the world of cryptocurrency “pump-and-dump” schemes. Because of simple post type classification and widespread use of success thresholds a tool could be built to help exchanges and law enforcement specialists detect, and potentially stop this type of trading activity. Such a tool could utilize public data from the chat groups as well as order book data directly from an exchange to paint a picture of why users are trading.

Pump Signal Analysis Through Inferred Targets: Through careful analysis of the cryptocurrency ecosystem it was found that there are essentially two forms of pump and dump signals: obvious, and target. A majority of the target pump signals do not present pump group members with a complete listing of target values, and obvious pumps rarely contain any target values. These target pumps could be for novice traders who are unsure where to set buy, sell, and stop-loss targets. Regardless, in Chapter 9 it was found that the average distance between target values was roughly 10%. By inferring targets across all pump signals longer term analysis could help identify which signals, or groups are more successful.

BIBLIOGRAPHY

[1] Free willy! – identifying the gox buy bot., January 2014. `https://www.reddit.com/r/Bitcoin/comments/20k4zc/free_willy_identifying_the_gox_buy_bot/`.

[2] Peter R's theory on the collapse of Mt. Gox, March 2014. `https://www.reddit.com/r/Bitcoin/comments/1zdnop/peter_rs_theory_on_the_collapse_of_mt_gox/`.

[3] The Willy Report, May 2014. `https://willyreport.wordpress.com/`.

[4] Pincoin ico crypto exit scam dupes 32,000 investors in $660 million heist, Oct 2018.

[5] Saman Adhami, Giancarlo Giudici, and Stefano Martinazzi. Why do businesses go crypto? An empirical analysis of Initial Coin Offerings. *Journal of Economics and Business*, 2018.

[6] Rajesh K. Aggarwal and Guojun Wu. Stock market manipulations*. *The Journal of Business*, 79(4):1915–1953, 2006.

[7] Ryan Amsden and Denis Schweizer. Are blockchain crowdsales the new gold rush? success determinants of initial coin offerings. *SSRN Electronic Journal*, 2018.

[8] David B Audretsch and José Mata. The post-entry performance of firms: Introduction. *International Journal of Industrial Organization*, 13(4):413–419, 1995.

[9] Hugo E Benedetti and Leonard Kostovetsky. Digital Tulips? Returns to Investors in Initial Coin Offerings. *SSRN Electronic Journal*, 2018.

[10] Shuqing Bian, Zhenpeng Deng, Fei Li, Will Monroe, Peng Shi, Zijun Sun, Wei Wu, Sikuang Wang, William Yang Wang, Arianna Yuan, et al. IcoRating: A deep-learning

system for scam ICO identification. Available at: https://arxiv.org/pdf/1803.03670, 2018.

[11] Rainer Böhme, Nicolas Christin, Benjamin Edelman, and Tyler Moore. Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives*, 29(2):213–38, 2015.

[12] Rainer Böhme and Thorsten Holz. The effect of stock spam on financial markets. In *Workshop on the Economics of Information Security*, 2006.

[13] Wilko Bolt and Maarten R.C. van Oordt. On the value of virtual currencies. http://www.bankofcanada.ca/wp-content/uploads/2016/08/swp2016-42.pdf.

[14] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A. Kroll, and Edward W. Felten. Research Perspectives and Challenges for Bitcoin and Cryptocurrencies. In *IEEE Symposium on Security and Privacy*, May 2015.

[15] David Chaum. Blind signatures for untraceable payments. In David Chaum, Ronald L. Rivest, and Alan T. Sherman, editors, *Advances in Cryptology*, pages 199–203, Boston, MA, 1983. Springer US.

[16] Shaen Corbet, Andrew Meegan, Charles James Larkin, Brian M. Lucey, and Larisa Yarovaya. Exploring the dynamic relationships between cryptocurrencies and other financial assets. *SSRN Electronic Journal*, 2017.

[17] Amir Feder, Neil Gandal, JT Hamrick, and Tyler Moore. The impact of DDoS and other security shocks on Bitcoin currency exchanges: Evidence from Mt. Gox. In *15th Workshop on the Economics of Information Security (WEIS)*, 2016.

[18] Laura Frieder and Jonathan Zittrain. Spam works: Evidence from stock touts and corresponding market activity. *Hastings Comm. & Ent. LJ*, 30:479, 2007.

[19] Neil Gandal and Hanna Halaburda. Can we predict the winner in a market with network effects? *Games*, 7(3), 2016. DOI: 10.3390/g7030016.

[20] Neil Gandal and Hanna Halaburda. Can we predict the winner in a market with network effects? competition in cryptocurrency market. *Games*, 7(3):16, 2016.

[21] Neil Gandal, Jt Hamrick, Tyler Moore, and Tali Oberman. Price manipulation in the bitcoin ecosystem. *Journal of Monetary Economics*, 95:86–96, 2018.

[22] Paul A Geroski. What do we know about entry? *International Journal of Industrial Organization*, 13(4):421–440, 1995.

[23] John M. Griffin and Amin Shams. Is bitcoin really un-tethered? *SSRN Electronic Journal*, 2018.

[24] Michael Hanke and Florian Hauser. On the effects of stock spam e-mails. *Journal of Financial markets*, 11(1):57–83, 2008.

[25] Adam S. Hayes. Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics*, May 2016.

[26] Danny Yuxing Huang, Hitesh Dharmdasani, Sarah Meiklejohn, Vacha Dave, Chris Grier, Damon McCoy, Stefan Savage, Nicholas Weaver, Alex C Snoeren, and Kirill Levchenko. Botcoin: Monetizing stolen cycles. In *Proceedings of the Network and Distributed System Security Symposium*, 2014.

[27] Danny Yuxing Huang, Kirill Levchenko, and Alex C Snoeren. Short paper: Estimating profitability of alternative cryptocurrencies. In *Financial Cryptography and Data Security*, 2018.

[28] Winifred Huang, Michele Meoli, and Silvio Vismara. The Geography of Initial Coin Offerings. SSRN Scholarly Paper ID 3206234, Social Science Research Network, Rochester, NY, July 2018.

[29] Josh Kamps and Bennett Kleinberg. To the moon: defining and detecting cryptocurrency pump-and-dumps. *Crime Science*, 7(1):18, 2018. Available at: `https://link.springer.com/article/10.1186/s40163-018-0093-5`.

[30] Peter M. Krafft, Nicolas Della Penna, and Alex Sandy Pentland. An experimental study of cryptocurrency market dynamics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 605:1–605:13, New York, NY, USA, 2018. ACM.

[31] Tao Li, Donghwa Shin, and Baolian Wang. Cryptocurrency pump-and-dump schemes. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3267041., 2018.

[32] Xin Li and Chong Alex Wang. The technology and economic determinants of cryptocurrency exchange rates: The case of bitcoin. *Decision Support Systems*, December 2016.

[33] Evgeny Lyandres, Berardino Palazzo, and Daniel Rabetti. Do Tokens Behave like Securities? An Anatomy of Initial Coin Offerings. SSRN Scholarly Paper ID 3287583, Social Science Research Network, Rochester, NY, April 2019.

[34] Ryan Mac. Bitcoin scammers are using this app to fleece people. *Buzzfeed News*, January 2018. Available at: https://www.buzzfeednews.com/article/ryanmac/cryptocurrency-scammers-are-running-wild-on-telegram.

[35] Nadia Massoud, Saif Ullah, and Barry Scholnick. Does it help firms to secretly pay for stock promoters? *Journal of Financial Stability*, 26:45–61, 2016.

[36] Micah Lerner Matthew Hougan, Hong Kim. Economic and non-economic trading in bitcoin: Exploring the real spot market for the world's first digital commodity, May 2019. https://www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5574233-185408.pdf.

[37] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. A fistful of Bitcoins: Characterizing payments among men with no names. In *Proceedings of the Internet Measurement Conference*, pages 127–140. ACM, 2013.

[38] Paul P. Momtaz. Initial Coin Offerings. SSRN Scholarly Paper ID 3166709, Social Science Research Network, Rochester, NY, July 2018.

[39] Tyler Moore and Nicolas Christin. Beware the middleman: Empirical analysis of Bitcoin-exchange risk. In *Financial Cryptography and Data Security*, volume 7859 of *Lecture Notes in Computer Science*, pages 25–33. Springer, April 2013.

[40] Tyler Moore, Nicolas Christin, and Janos Szurdi. Revisiting the risks of bitcoin currency exchange closure. *ACM Trans. Internet Technol.*, 18(4):50:1–50:18, September 2018.

[41] Malte Möser, Rainer Böhme, and Dominic Breuker. An inquiry into money laundering tools in the Bitcoin ecosystem. In *Proceedings of the Seventh APWG eCrime Researcher's Summit*, pages 1–14. IEEE, 2013.

[42] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2009. Available at: `http://www.bitcoin.org/bitcoin.pdf`.

[43] Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder. *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton University Press, Princeton, NJ, USA, 2016. Available at: `https://lopp.net/pdf/princeton_bitcoin_book.pdf`.

[44] Miroslava Rajcaniova Pavel Ciaian and d'Artis Kancs. The economics of bitcoin price formation. *Applied Economics*, 48:1799–1815, May 2016.

[45] Becky Peterson. Prodeum crypto scammers allegedly stole its teams' identities before vanishing and leaving only a website with the word 'penis', Jan 2018.

[46] Sylvain Ribes. Chasing fake volume: a crypto-plague – sylvain ribes – medium, Mar 2018.

[47] Dorit Ron and Adi Shamir. Quantitative analysis of the full Bitcoin transaction graph. In *Financial Cryptography and Data Security*, volume 7859 of *Lecture Notes in Computer Science*, pages 6–24. Springer, 2013.

[48] Steven Russolillo and Eun-Young Jeong. Cryptocurrency exchanges are getting hacked because it's easy, Jul 2018.

[49] Adam Schlabach. Q1 2019 cryptocurrency anti-money laundering report, Feb 2020.

[50] Shane Shifflett and Paul Vigna. Some traders are talking up cryptocurrencies, then dumping them, costing others millions. *The Wall Street Journal*, August 2018. Available at: `https://www.wsj.com/graphics/cryptocurrency-schemes-generate-big-coin/`.

[51] William Suberg. Mt. Gox trial update: Karpeles admits 'Willy Bot' existence. *Coin Telegraph*, jul 2017. `https://cointelegraph.com/news/mt-gox-trial-update-karpeles-admits-willy-bot-existence`.

[52] Marie Vasek, Joseph Bonneau, Ryan Castellucci, Cameron Keith, and Tyler Moore. The Bitcoin brain drain: a short paper on the use and abuse of bitcoin brain wallets. In *Financial Cryptography and Data Security*, Lecture Notes in Computer Science. Springer, 2016.

[53] Marie Vasek and Tyler Moore. There's no free lunch, even using Bitcoin: Tracking the popularity and profits of virtual currency scams. In Rainer Böhme and Tatsuaki Okamoto, editors, *Financial Cryptography and Data Security*, volume 8975 of *Lecture Notes in Computer Science*, pages 44–61. Springer, January 2015.

[54] Marie Vasek, Micah Thornton, and Tyler Moore. Empirical analysis of denial-of-service attacks in the Bitcoin ecosystem. In *1st Workshop on Bitcoin Research*, volume 8438 of *Lecture Notes in Computer Science*, pages 57–71. Springer, March 2014.

[55] JR Willett. The second bitcoin whitepaper. *white paper*, January 2012. Available at: `https://sites.google.com/site/2ndbtcwpaper/2ndBitcoinWhitepaper.pdf`.

[56] Peng Xie, Hailiang Chen, and Yu Jeffrey Hu. Network structure and predictive power of social media for the bitcoin market. *SSRN Electronic Journal*, 2017.

[57] Jiahua Xu and Benjamin Livshits. The anatomy of a cryptocurrency pump-and-dump scheme. Available at: `https://arxiv.org/abs/1811.10109`., 2018.

APPENDIX A

DESCRIPTIVE STATISTICS AND OTHER TABLES

Table A.1: Summary statistics of independent and dependent variables

	Mean	SD	Min	Max
"Markus"	0.09	0.29	0	1
"Willy"	0.14	0.34	0	1
DDOS	0.08	0.27	0	1
Day after DDOS	0.08	0.27	0	1
Other Attacks	0.02	0.13	0	1
Mt.Gox daily rate change ($)	3.24	22.39	-139.78	257.5
Bitstamp daily rate change ($)	3.06	19.53	-132.99	190.91
Bitfinex daily rate change ($)[1]	4.25	33.30	-295.97	294
Btce daily rate change ($)	2.86	19.28	-134.30	198.14
Mt.Gox daily % rate change	1.4%	6.6%	-28%	49%
Bitstamp daily % rate change	1.5%	6.9%	-49%	40%
Bitfinex daily % rate change[2]	1.4%	8.4%	-37%	59%
Btce % daily rate change	1.4%	6.7%	-50%	41%
N	365			

[1] N=244 for this variable.
[2] N=244 for this variable.

Table A.2: Correlation between daily rate changes and the independent variables

	Mt.Gox Rate Change	Bitstamp Rate Change	Bitfinex Rate Change	Btce Rate Change
"Markus"	0.001	0.01	-0.02	0.00009
"Willy"	0.33	0.35	0.23	0.34
DDoS	-0.06	-0.06	-0.05	-0.06
Day After DDoS	-0.07	-0.07	-0.05	-0.06
Other Attacks	0.02	0.02	0.013	0.02
N	365	365	244	365

Table A.3: Correlation between daily percent rate changes and the independent variables

	Mt.Gox % Rate Change	Bitstamp % Rate Change	Bitfinex % Rate Change	Btce % Rate Change
"Markus"	0.14	0.16	0.07	0.13
"Willy"	0.21	0.2	0.22	0.2
DDoS	-0.1	-0.05	-0.05	-0.06
Day After DDoS	-0.09	-0.06	-0.08	-0.06
Other Attacks	0.07	0.04	0.02	0.04
N	365	365	365	365

Table A.4: Correlation between independent variables

	"Markus"	"Willy"	DDoS	Day After DDoS	Other Attacks
"Markus"	1				
"Willy"	-0.1	1			
DDoS	0.05	-0.06	1		
Day After DDoS	0.05	-0.06	0.33	1	
Other Attacks	0.03	-0.05	-0.04	0.04	1
N	365				

Table A.5: Suspicious trading activity and price changes on Bitstamp

		Days with no STA		Days with STA	
		days	%	Days	%
“Markus”	Daily rate decrease	88	45	6	18
	Daily rate increase	105	55	27	82
“Willy”	Daily rate decrease	6	40	9	18
	Daily rate increase	9	60	41	82
Total	Daily rate decrease	94	45	15	18
	Daily rate increase	114	55	67	82

Table A.6: “Willy”: Volume activity (period 4)

	mean	median	N
Volume bought by “Willy” (Mt. Gox)	4,962	3,881	50
Total BTC volume on Mt. Gox (“Willy” active)	30,854	25,939	50
Total BTC volume on Mt. Gox (“Willy” inactive)	17,472	10,444	41
Total BTC volume on Bitstamp (“Willy” active)	26,084	23,684	50
Total BTC volume on Bitstamp (“Willy” inactive)	14,793	10,505	41
Total BTC volume on Bitfinex (“Willy” active)	12,981	11,756	50
Total BTC volume on Bitfinex (“Willy” inactive)	6,467	3,829	41
Total BTC volume on BTCE (“Willy” active)	20,691	18,661	50
Total BTC volume on BTCE (“Willy” inactive)	7,529	3,737	41
Total BTC volume (“Willy” active)	90,611	82,779	50
Total BTC volume (“Willy” inactive)	46,263	29,476	41

Table A.7: "Markus": Volume activity (period 3)

	mean	median	N
Volume bought by "Markus" (Mt. Gox)	10,056	8,901	17
Total BTC volume on Mt.Gox ("Markus" active)	39,619	42,022	17
Total BTC volume on Mt.Gox ("Markus" inactive)	27,672	17,421	75
Total BTC volume on Bitstamp ("Markus" active)	13,547	12,840	17
Total BTC volume on Bitstamp ("Markus" inactive)	10,299	8,850	75
Total BTC volume on Bitfinex ("Markus" active)	5,976	5,622	17
Total BTC volume on Bitfinex ("Markus" inactive)	4,331	3,197	75
Total BTC volume on BTCE ("Markus" active)	4,840	4,699	17
Total BTC volume on BTCE ("Markus" inactive)	4,660	3,589	75
Total BTC volume ("Markus" active)	63,984	67,691	17
Total BTC volume ("Markus" inactive)	46,962	31,173	75

Table A.8: Examining Skewness and Kurtosis Through Suspicious Trading Activity

Independent Variables / Dependent Variable	Skewness	Kurtosis
"Markus"	-0.03	-0.06
	(0.09)	(0.18)
"Willy"	-0.10	-0.11
	(0.07)	(0.15)
DDoS	-0.09	-0.09
	(0.10)	(0.20)
Day After DDoS	0.03	0.11
	(0.10)	(0.20)
Other Attacks	-0.27	-0.47
	(0.20)	(0.39)
Constant	2.92	6.28
	(0.03)	(0.06)
N	338	338
adj. R^2	-0.0024	-0.0082

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

APPENDIX B

SENSITIVITY ANALYSIS OF PEAK DETECTION

The algorithm used to discover peaks in the dataset utilizes several values that, when increased or decreased, affect the number of peaks returned. The three main variables whose values can be modified easily are: the window size on each side of the data point, the minimum value increase for peak, and threshold for minimum peak size. The values accepted for each are days (30 day default), percent (50% default), and percent (5% default) respectively.

	% of max peak	% price jump for peak 25	50	100	200
Volume (median)	10	2	2	2	2
	5	3	3	3	3
	0	7	7	7	6
Volume (total)	10	3 482	3 452	3 377	3 241
	5	4 185	4 148	4 054	3 867
	0	9 746	9 643	9 381	8 677
Price (median)	10	3	2	2	2
	5	3	3	2	2
	0	5	5	4	2
Price (total)	10	3 384	3 064	2 549	1 991
	5	4 078	3 650	2 963	2 227
	0	7 459	6 260	4 593	3 046

Table B.1: Sensitivity Analysis of Peak Definition Algorithm

Table B.1 displays the results from modifying the values. The cells shaded blue show the numbers obtained from the algorithm configuration used in the paper. Reducing the minimum required price jump has no effect on the median number of peaks found per currency, but it increases the total number of peaks discovered. Additionally, removing the restriction on the minimum peak value compared to the maximum peak found essentially doubles the number of peaks found for all currencies. This is concerning as most small peaks are within the domain of normal trading and do not lead us to believe they are result of anomalous trading activity.

The number shown in the tables associated with appendix B and C are higher than the numbers reported earlier in the study. This is due to the fact that an updated dataset was used for the sensitivity analysis which spans 2013-04-28 to 2018-05-15 instead of the original 2013-04-28 to 2018-02-07. However, these results are consistent with the numbers generated with the earlier dataset.

APPENDIX C

SENSITIVITY ANALYSIS OF ABANDONMENT AND RESURRECTION DETECTION

The abandonment and resurrection algorithm, like the peak algorithm, utilizes two threshold values to determine if a coin/token is abandoned and resurrected. The first variable, used to detect abandonment, uses a default value of 10%. If the price following a peak drops below 10% of the peak value the currency is considered to be abandoned. The second variable, used to detect resurrection following a period of abandonment, uses a default value of 1%. If the price following abandonment increases to or above 1% of the abandonment value then the currency is said to be resurrected.

To examine how modifications would alter the results we tested a multitude of different values for both abandonment and resurrection. These values can be seen in Tables C.1, C.2, and C.3.

The values chosen for our analysis find a reasonable balance between too many and not enough abandonments.

abandonment threshold (%)	resurrection threshold (%) 1.0	2.0	5.0	10.0	20.0	30.0
0.0	109	108	105	104	104	103
0.1	335	328	316	305	299	294
1.0	818	773	697	645	608	591
2.0	1 121	1 036	898	819	757	730
5.0	1 696	1 541	1 232	1 081	962	911
10.0	2 192	2 021	1 631	1 373	1 186	1 096

Table C.1: Total Number of Abandonments (Sensitivity Analysis)

abandonment threshold (%)	resurrection threshold (%) 1.0	2.0	5.0	10.0	20.0	30.0
0.0	273	336	366	425	456.5	488
0.1	151	183	228.5	276	365	366
1.0	62	92	153	184	215	243
2.0	31	62	122	153	184	212
5.0	31	31	92	123	153	184
10.0	31	31	61	92	151	153

Table C.2: Median Number of Abandonments (Sensitivity Analysis)

abandonment threshold (%)	resurrection threshold (%) 1.0	2.0	5.0	10.0	20.0	30.0
0.0	38 033	41 913	45 491	49 940	53 966	56 564
0.1	98 898	113 732	124 156	134 987	145 062	149 344
1.0	137 793	174 138	202 585	221 512	238 413	245 693
2.0	141 652	185 281	225 671	247 349	264 658	272 882
5.0	146 169	197 586	248 743	280 190	303 058	314 647
10.0	153 929	203 780	264 042	301 590	327 755	340 271

Table C.3: Total Duration of Abandonments (Sensitivity Analysis)

APPENDIX D

CRYPTOCURRENCY TRADING RISK AND REWARD

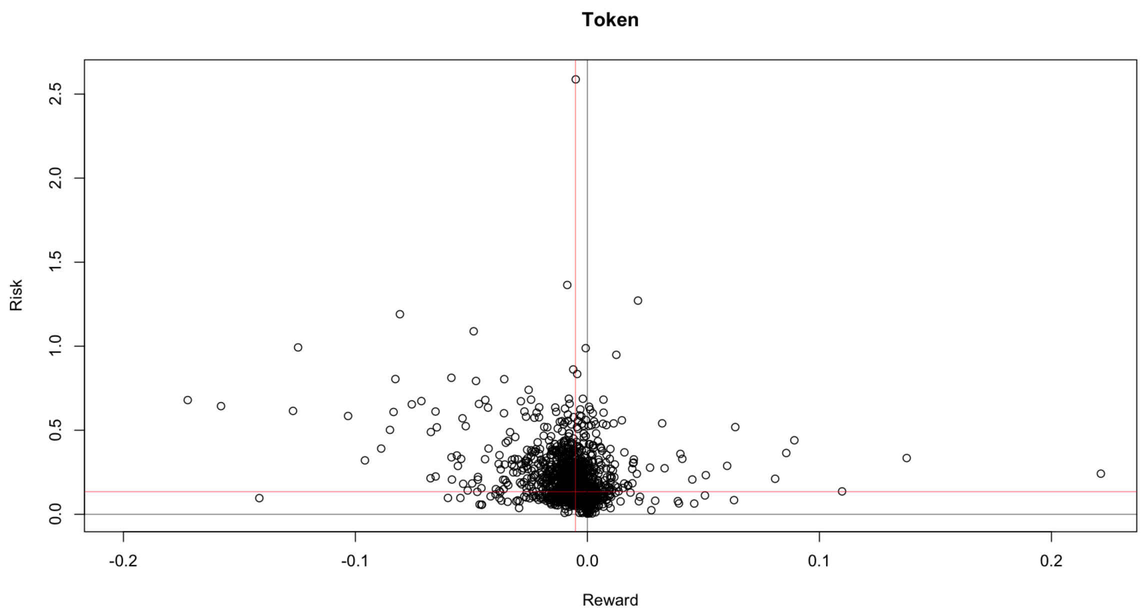


Figure D.1: Scatterplot of token risk and reward.

Within this analysis risk is calculated by taking the average of the daily log return for each of the cryptocurrencies in the dataset. Reward is calculated by taking the standard deviation of the daily log return for each of those cryptocurrencies. The values presented here were calculated using coin and token data from `coinmarketcap.com`. These date include daily values for open, high, low, close price, as well as the volume and market capitalization. The coin dataset spans April 2013 to October 2019 and the token dataset spans March 2014 to October 2019. In total the coin dataset consists of 1,236,090 daily values for 1,589 coins and the token dataset consists of 746,092 daily values for 1,905 tokens. Figure D.1 and Figure D.2 plot the risk and reward per token and coin respectively. Moreover, both plots include horizontal and vertical lines that represent the median values for risk and reward.

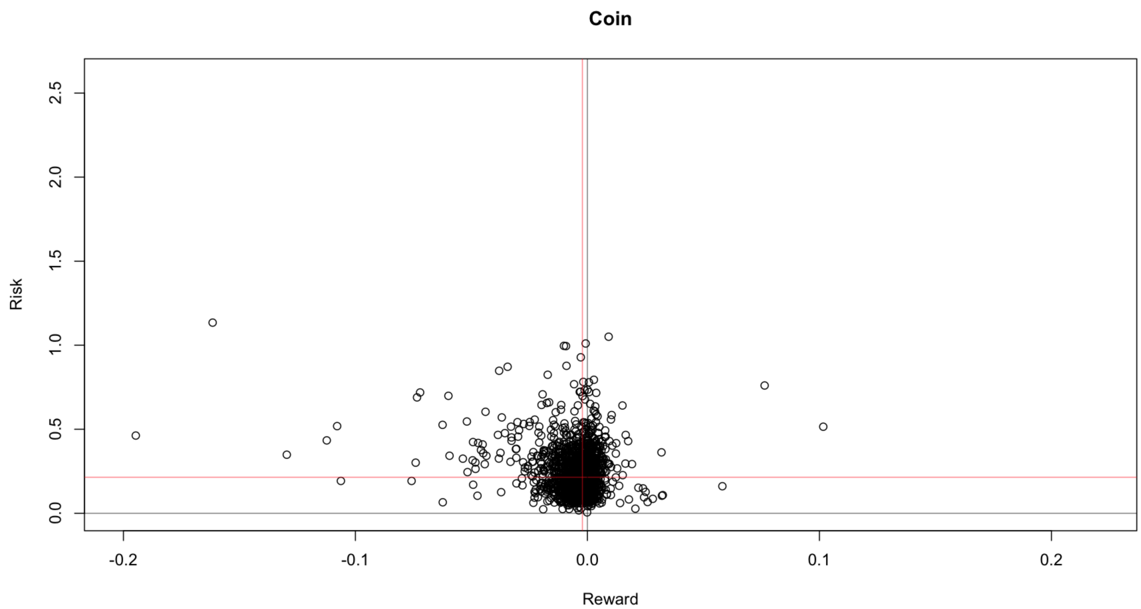


Figure D.2: Scatterplot of coin risk and reward.

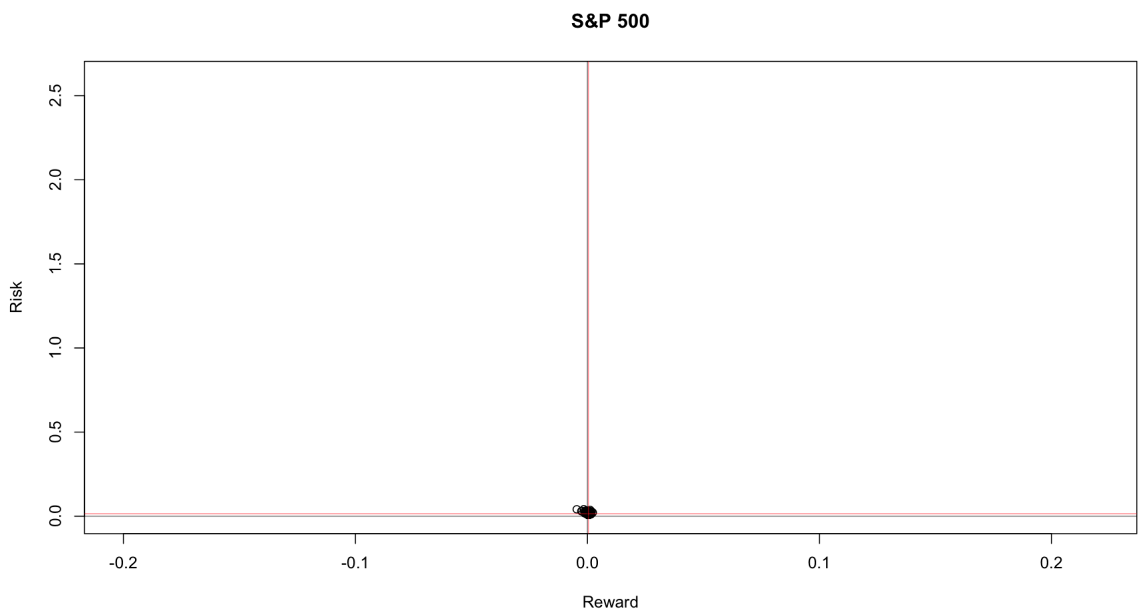


Figure D.3: Scatterplot of S&P 500 stock risk and reward.

Both tokens and coins have moderate median risk values and negative median reward values. The averages for risk and reward are slightly less appealing than the median values for both tokens and coins. The risk for tokens increases from 0.135 (median) to 0.182 (mean) and the reward decreases from -0.0051 (median) to -0.0061 (mean).

For the token subset, only 298 out of the 1,847 available have a positive mean log return value. Of these 298, 80% of the rewards are below 0.01. Coins perform similarly, with only 473 out of the 1,562 available having a positive mean log return value. Unlike tokens, more coins (95%) see reward values below 0.01. These risk-reward calculations are long-term, utilizing all of the available data for each of the tokens. The values are not necessarily representative of investments sold before the end of the dataset. However, long term very few tokens are worthwhile investments.

Because this risk-reward calculation was borrowed from traditional financial literature here a comparison is made to some traditional stocks. Figure D.3 plots the risk and reward relationship for the stocks included in the S&P 500. These values were calculated using 5 years of publicly available daily trade data between February 2013 and February 2018. All three figures displaying the risk and reward relationships cover the same x and y limits making differences easy to spot. The risk and reward values for these stocks are more tightly clustered around 0 than the coin and token measures. The S&P 500 stocks have a median reward value of 0.014445 and a median risk value of 0.0004249. In other words, unlike coins and tokens, traditional stocks have lower risk scores and average positive returns.

APPENDIX E

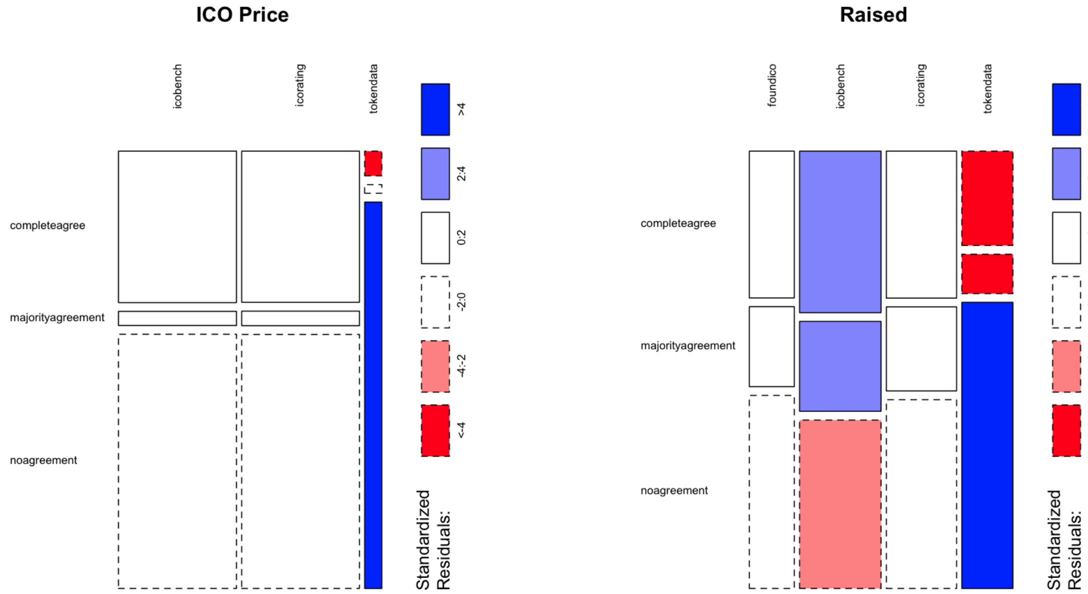
ADDITIONAL TOKEN STATISTICS

Cryptocurrency	Transition	Timestamp	Current Price
time-new-bank	STARTTOSELL1	2018-07-01 22:44:01	0.00000370784
time-new-bank	SELL1TOSELL2	2018-07-01 23:04:01	0.00000408808
time-new-bank	SELL2TOSELL3	2018-07-02 03:39:02	0.00000421311
time-new-bank	SELL3TOSELL4	2018-07-03 08:29:01	0.00000455368
time-new-bank	SELL4TOSELL3	2018-07-04 06:14:01	0.00000419579
time-new-bank	SELL3TOSELL2	2018-07-05 20:29:02	0.00000389236
time-new-bank	SELL2TOSELL1	2018-07-06 04:34:01	0.00000369518
time-new-bank	SELL1TOSELL2	2018-07-06 19:04:01	0.00000393351
time-new-bank	SELL2TOSELL1	2018-07-07 10:09:01	0.00000367417
time-new-bank	SELL1TOBUY1	2018-07-11 05:19:08	0.00000317042

Table E.1: Price movement through target values for Time New Bank.

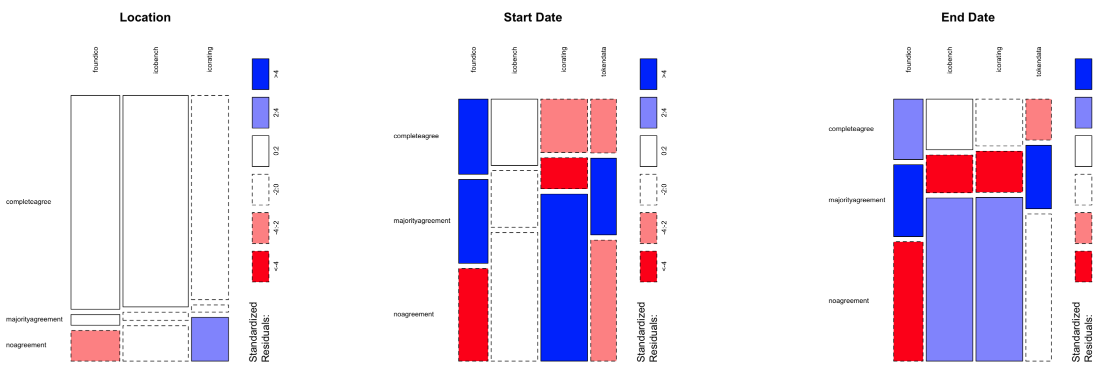
Measure	Min	Max	Median	Mean	Success
First cycle	0	131.581	0.818	3.555	TRUE
First cycle	0	443.854	1.012	7.644	FALSE

Table E.2: Hours between pump signal and first target crossed (not necessarily a buy target).



(a) (b)

Figure E.1: Numerical value source categorization.



(a) (b) (c)

Figure E.2: Categorical value source categorization.

	Min	Max	Median	Mean	No. Signals
Overall	1	749	18	34.6	3,640
Cutoff at next signal	0	397	3	8.6	3,640

Table E.3: Number of targets crossed per pump signal.

Table E.3 shows the results of two different methods used for counting the number of targets crossed per pump signal. The overall method counts all targets crossed between the pump signal time and the data cutoff in November 2019. The cutoff at next signal method does just that. Pump signal target crossings are counted until a subsequent pump on the same coin is encountered; if no more pumps occur for a specific cryptocurrency then the data cutoff is used as the stopping point. The max value is inflated because there are several pump signals with no pump signal afterward. With no following pump signal to reduce the pump's duration it could have years to accumulate crossings.

For example, the max number of crossings in the overall group was from a coin named XWC. The pump signal was posted in late November 2017 and the data cutoff is November 2019. This pump crossed target values for two years. A pump on the same coin is posted in January 2018, so if the cutoff is utilized the number of runaway counts is reduced.