UNIVERSITY
OF TECHNOLOGY
SYDNEY

Frauds in finance: Prevalence, determinants, and consequences

by Inji Allahverdiyeva

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy at University of Technology Sydney

under the supervision of Professor Talis Putnins, Professor Kathy Walsh, Professor Dave Michayluk, Professor Harry Scheule

University of Technology Sydney
Faculty of Business

June 17, 2024

Statement of Originality

I certify that this dissertation, titled "Frauds in finance: Prevalence, determinants, and consequences', has not previously been submitted for a degree nor has it been submitted as part of requirement for a degree except as fully acknowledged within the text.

I also certify that the dissertation has been written by me. Any help that I have received in my research and the preparation of the dissertation itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the dissertation.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: June 17, 2024

Working Papers

1. Aliyev, N., Allahverdiyeva, I. and Putnins, T. (2023a), 'Is public distrust of the finance sector warranted? Evidence from financial adviser misconduct', *Working paper* is based on Chapter 2 of this dissertation and has been presented in the following conferences/seminars:
   - 2023 Computing in Economics and Finance, Nice, France
   - 2023 Seminar series in Macquarie University, Sydney, Australia
   - 2023 UTS Business School doctoral presentation, Sydney, Australia
   - 2022 Financial Management Association Asia Pacific, Melbourne, Australia
2. Aliyev, N., Allahverdiyeva, I. and Putnins, T. (2023b), 'The real effects of market manipulation', *Working paper* is based on Chapter 3 of this dissertation and has been presented in the following conferences/seminars:
   - 2023 UTS Business School doctoral presentation, Sydney, Australia
3. Aliyev, N., Allahverdiyeva, I. and Putnins, T. (2023c), 'Scam alert: Can cryptocurrency scams be detected early?', *Working paper* is based on Chapter 4 of this dissertation and has been presented in the following conferences/seminars:
   - 2024 Seminar series in University of New South Wales, Sydney, Australia
   - 2023 Computing in Economics and Finance, Nice, France
   - 2023 UTS Business School doctoral presentation, Sydney, Australia
   - 2023 Financial Research Network (FIRN), Hobart, Australia
   - 2023 Market Microstructure Meeting, Sydney, Australia
4. Aliyev, N., Allahverdiyeva, I. and Putnins, T. (2023d), 'Once bitten twice shy: Learning about scams', *Working paper* is based on Chapter 5 of this dissertation and has been presented in the following conferences/seminars:
   - 2023 UTS Business School doctoral presentation, Sydney, Australia

To my family.

Acknowledgements

I am deeply grateful to my co-author Dr. Nihad Aliyev for his valuable contribution to my dissertation. His guidance and expertise have helped me to finalise this dissertation and influenced the way I approach academic challenges.

I am also deeply grateful to my academic supervisor, Prof. Talis Putnins. I have learnt a lot from his approach to research and his expertise in forensic finance. It has been a great experience collaborating with him, and I consider myself lucky to have had this opportunity.

I extend my sincere appreciation to my supervisory panel: Profs. Kathy Walsh, Harry Scheule, and Dave Michayluk. I also thank my friends and colleagues Profs. Tony He and Isa Hafalir, and Dr. Elif Hafalir for giving valuable advice during my PhD.

A/Prof Christina Nikitopoulos, Dr. Kristoffer Glover, and Andrea Greer tirelessly supported and facilitated my academic journey. The assistance provided by them in handling conference funding, managing forms, coordinating processes, and ensuring the smooth flow of administrative tasks has been very helpful to focus on my research. I also gratefully acknowledge the financial support of UTS Graduate Research School and Business School.

I extend my heartfelt appreciation to my fellow PhD colleagues and friends, Ali Vergili, Sasha Erakhtina, Xinyi Deng, Vlad Pyzhov, Windy Mai for their friendship and shared experiences during this doctoral journey. We have grown, struggled, and celebrated together.

Finally, I dedicate this dissertation to my family. I am full of gratitude to my mother Gulnara, my father Sehran, my sister Besti, my brother Toghrul, and my husband Nihad for their unconditional love. Your presence in my life is my greatest blessing.

Abstract

There is hardly any activity in finance untouched by fraud and misconduct. This dissertation investigates different aspects of financial fraud. It covers deceptive practices of financial advisers, market manipulation, cryptocurrency-related scams, and investigates whether investors falling victim to scams learn from their mistakes.

Chapter 2 analyzes the role of financial advisers in maintaining trust in the financial system. Using data from over one million financial advisers in the U.S., Chapter 2 estimates that up to 30% of advisers are involved in misconduct, but only one-third of these cases are reported by regulators. Advisers with a high propensity for misconduct oversee approximately $6.9 trillion assets under management (AUM). The rates of adviser misconduct and unreported cases increase during the GFC, paralleling the decline of trust in financial institutions. One misconduct costs a firm about 5 clients and $10 million AUM annually. This chapter also provides a list of characteristics for consumers, advisory firms, and regulators to help identify adviser misconduct.

Chapter 3 investigates the real economic consequences of market manipulation. Market manipulation distorts financial market prices, but does it have real economic effects on listed companies? Chapter 3 shows that it does. Increased manipulation makes stock price signals less useful for firm managers seeking to learn about potential investment opportunities, thereby decreasing the sensitivity of firms' investments to stock prices. This leads to a decline in the quality of firms' investment decisions, and consequently, firm operating performance also decreases. The findings in Chapter 3 suggest that the real economic consequences of market manipulation extend beyond the direct effects on secondary markets.

Chapter 4 explores decentralized exchanges (DEXs) and cryptocurrency scams in major DEXs. DEXs are gaining popularity, allowing investors to trade cryptocurrencies through liquidity pools. Chapter 4 analyzes how public blockchains have given rise to a new type of scam known as a "rug pull". In this chapter, we find these scams are pervasive: 44% of tokens in major decentralized exchanges (DEXs) are scams, causing losses of $1.5 billion to investors. This chapter shows that scams differ from legitimate tokens in key characteristics, including the token creators launch multiple liquidity pools, do not lock

their liquidity provider tokens, imitate other tokens, and create the liquidity pool shortly after the token's release. Using these characteristics, Chapter 4 develops a scam index that can predict cryptocurrency scams, with practical applications in cryptocurrency surveillance.

Based on methodologies developed in Chapter 4, Chapter 5 analyzes investors' trading behavior and investment approaches after falling victim to scams. Do investors learn from their mistakes after falling victim to a scam? Experiencing a scam in the recent past (resp. a 1% increase in the scam investment ratio) decreases the probability of investing in a scam again by approximately 4% (resp. 0.30%). However, experiencing a scam also decreases subsequent non-scam investment returns by around 14%. Victims increase their portfolio values, but the cash allocation and the standard deviation of daily portfolio returns suggest that victims become more (resp. less) risk-tolerant in the short (resp. long) term after being scammed. This chapter also explores characteristics associated with varying learning rates among investors and finds that less experienced investors tend to learn less and invest in multiple scams.

Overall, this dissertation quantifies and characterizes different forms of frauds in finance. Understanding how these fraudulent activities work, how common they are, what leads to them, and their consequences are important for designing effective surveillance tools, protecting consumers, designing better markets, maintaining trust in the broader financial system, and more generally, improving overall welfare in society.

Contents

Statement of Originality **i**

Acknowledgements **iv**

List of Figures **xi**

List of Tables **xii**

1 Introduction **1**

1.1 Literature review 8

1.1.1 Trust in finance 8

1.1.2 Financial adviser misconduct 9

1.1.3 Market manipulation 9

1.1.4 Investment-to-price sensitivity 10

1.1.5 Blockchain technology, cryptocurrencies, and scams 11

1.1.6 Investor behaviour after scam 11

1.2 Thesis outline 12

2 Is public distrust of the finance sector warranted? Evidence from financial adviser misconduct **14**

2.1 Introduction 14

2.2 Data and descriptive statistics 19

2.2.1 Data, filters and transformations 19

2.2.2 Sample of reported misconduct 25

2.2.3 Descriptive statistics 29

2.3 Empirical model of financial adviser misconduct and detection 31

2.3.1 Incomplete detection problem 31

2.3.2 Detection-controlled estimation 31

2.3.3 Model specifications 34

2.4 Results . . . 38
2.4.1 The extent of adviser misconduct . . . 38
2.4.2 The determinants of adviser misconduct and detection . . . 41
2.4.3 How does adviser misconduct vary through time? . . . 44
2.4.4 Adviser misconduct and trust . . . 49
2.5 Extensions and robustness tests . . . 52
2.5.1 Investment advisers versus brokers . . . 52
2.5.2 Alternative definition of adviser misconduct . . . 54
2.5.3 Relaxing exclusion restrictions . . . 54
2.6 Conclusion . . . 55
Appendix 2.1. . . . 56

3 The real effects of market manipulation 65
3.1 Introduction . . . 65
3.2 Related literature . . . 67
3.3 The model . . . 69
3.4 Data, measures and descriptive statistics . . . 72
3.4.1 Data . . . 72
3.4.2 Measure of stock price manipulation . . . 73
3.4.3 Measure of investment-to-price sensitivity . . . 74
3.4.4 Descriptive statistics . . . 77
3.5 Empirical analysis . . . 79
3.5.1 Market manipulation and investment-Q sensitivity . . . 79
3.5.1.1 Ordinary least squares (OLS) . . . 79
3.5.1.2 Two-stage least squares (2SLS) . . . 80
3.5.2 Cross-sectional analysis . . . 87
3.5.3 Market manipulation and operating performance . . . 89
3.6 Robustness and Extensions . . . 91
3.6.1 Monotonic effects of manipulation on investment-Q sensitivity . . . 91
3.6.2 Controlling for managerial information . . . 93
3.6.3 Controlling for analyst coverage . . . 94
3.6.4 Controlling for capital constraints . . . 96
3.6.5 Fama-MacBeth regressions . . . 96
3.6.6 Accounting for intangible capital . . . 99
3.7 Conclusion . . . 99
Appendix 3.1. . . . 100

4 Scam alert: Can cryptocurrency scams be detected early? 110
4.1 Introduction . . . 110
4.2 Related literature . . . 114
4.3 Institutional background and case studies . . . 116
4.3.1 The structure of Ethereum blockchain and ERC-20 tokens . . . 116

4.3.2 Trading through decentralized exchanges 116
4.3.3 Illustration of rug pulls 117
4.4 Data and methods 123
4.4.1 Data 123
4.4.2 Methods 124
4.5 Identifying scams 126
4.5.1 Hard rug pulls 126
4.5.2 Soft rug pulls 129
4.6 Analysis of scams 132
4.6.1 Characteristics of soft and hard rug pulls 132
4.6.2 Which pools are more likely to be a scam? 136
4.6.3 Measuring the probability of a scam 141
4.7 Validation and robustness tests 144
4.7.1 Validation 144
4.7.2 Robustness 146
4.7.2.1 Characteristics analysis with all scam pools 146
4.7.2.2 Collaborator addresses 146
4.7.2.3 Magnitude of the price run-up and collapse 146
4.7.2.4 Activity thresholds 147
4.8 Conclusion 147
Appendix 4.1. 149

5 Once bitten twice shy: Learning about scams **156**
5.1 Introduction 156
5.2 Related literature 159
5.3 Institutional background 160
5.3.1 ERC-20 tokens 161
5.3.2 Trading through decentralized exchanges 161
5.4 Data and methods 162
5.4.1 Data sources 162
5.4.2 Data filters and transformations 163
5.4.3 Identifying scams and victims 163
5.4.4 Investor characteristics 166
5.5 Results 167
5.5.1 Characteristics of victims 168
5.5.1.1 Victims versus non-victims 168
5.5.1.2 Before versus after scams 170
5.5.2 Impact of scams on further scam investments 171
5.5.3 Impact of scams on trading performance 174
5.5.4 Impact of scams on investor portfolio 176
5.5.5 Learning rate of investors 178

5.6 Conclusion 181
Appendix 5.1. 182

6 Conclusion **197**

Bibliography **203**

List of Figures

2.1. The time series of reported misconduct and non-misconduct-related disclosures . . . 28
2.2. Detection-controlled estimation model . . . 32
2.3. Expected utility of misconduct . . . 41
2.4. Estimated and reported adviser misconduct over time . . . 45
2.5. Estimated and reported misconduct by a new adviser over time . . . 47
2.6. Asset under management of active advisers with misconduct over time . . . 48
2.7. Financial trust index and adviser misconduct over time . . . 50
A2.1. Estimated and reported adviser misconduct over time . . . 61
A2.2. Estimated and reported misconduct by a new adviser over time . . . 63
A2.3. Asset under management of active advisers with misconduct over time . 64

3.1. The impact of market manipulation on investment-Q sensitivity . . . 85
3.2. The monotonic effects of market manipulation on investment-Q sensitivity 93
3.3. The yearly investment-Q sensitivity with and without market manipulation 98

4.1. The evolution of vshiba/weth liquidity pool . . . 118
4.2. Vshiba smart contract . . . 119
4.3. The evolution of ukraine/weth liquidity pool . . . 120
4.4. The number of weth withdrawn by the creator of the ukraine token . . . 122
4.5. The number of hard rug-pulls over time . . . 128
4.6. The number of soft rug-pulls over time . . . 131
4.7. Receiver Operating Characteristics (ROC) curves . . . 140
4.8. The distribution of scam index . . . 143
A4.1. Receiver Operating Characteristics (ROC) curves . . . 155

5.1. Number of investors and victims over time . . . 165

List of Tables

2.1. Definitions of variables . . . 22
2.2. Definitions of misconduct categories . . . 25
2.3. The number and percentage of advisers with each misconduct and disclosure category . . . 26
2.4. Differences in characteristics of financial advisers with and without reported misconduct . . . 30
2.5. Estimated misconduct frequency . . . 39
2.6. Model estimates . . . 43
2.7. Impact of adviser misconduct on number of clients and total AUM . . . 52
2.8. Estimated misconduct frequency: Investment advisers vs brokers . . . 53
2.9. Estimated misconduct frequency: Relaxing exclusion restrictions . . . 54
A2.1. Definition of non-misconduct-related disclosure categories . . . 57
A2.2. Coefficient estimates for the models reported in Table 2.6 . . . 59
A2.3. Misconduct frequency estimation based on individual misconduct categories . . . 60

3.1. Variable definitions . . . 76
3.2. Descriptive statistics . . . 78
3.3. Market manipulation and investment-Q sensitivity: OLS . . . 79
3.4. Market manipulation and investment-Q sensitivity: 2SLS . . . 83
3.5. Different levels of market manipulation and investment-Q sensitivity . . . 84
3.6. The impact of market manipulation on investment-Q sensitivity by stock characteristics . . . 88
3.7. Market manipulation and operating performance . . . 90
3.8. Monotonic effects of market manipulation on investment-Q sensitivity . . . 92
3.9. Market manipulation and investment-Q sensitivity: Controlling for additional factors . . . 95
3.10. Market manipulation and investment-Q sensitivity: Fama-Macbeth regression . . . 97
A3.1. Market manipulation measure . . . 101
A3.2. Market manipulation and investment-Q sensitivity: 2SLS IV first-stage results results for Table 3.4 . . . 102

A3.3. Market manipulation and operating performance: OLS version of Table 3.7 103
A3.4. Monotonic effects of market manipulation on investment-Q sensitivity: OLS version of Table 3.8 104
A3.5. The impact of market manipulation on investment-Q sensitivity by stock characteristics: CAPEX and CHGASSET versions of Table 3.6 105
A3.6. Market manipulation and investment-Q sensitivity - Controlling for additional factors: OLS version of Table 3.9 107
A3.7. Market manipulation and investment-Q sensitivity - OLS: Peters-Taylor Q version of Table 3.3 108

4.1. Definitions of variables 125
4.2. Hard rug pulls 127
4.3. Soft rug pulls 131
4.4. The number of soft and hard rug pulls by automated market makers . . . 133
4.5. Characteristics of soft and hard rug pulls 134
4.6. Characteristics of scams and non-scam liquidity pools 137
4.7. The ex ante determinants of scam pools: logit regression with all scams . 142
4.8. Validation test 145
4.9. Robustness tests 147
A4.1. Well-known locker addresses 150
A4.2. Characteristics of soft and hard rug pulls – All scams 151
A4.3. Characteristics of rug pulls and non-rug pulls – All scams 152
A4.4. The correlation matrix of variables 153
A4.5. The ex ante determinants of scam pools: logit regression with economically meaningful scams 154

5.1. Definitions of variables 167
5.2. Characteristics of victims and non-victims 169
5.3. Characteristics of victims before and after scam 170
5.4. Impact of scams on further scam investment 173
5.5. Impact of scams on subsequent investment returns 175
5.6. Impact of scams on portfolio characteristics (trades only) 177
5.7. Impact of scams on portfolio characteristics (trades and transfers) 179
5.8. Impact of investor characteristics on learning rate 180
A5.1. Descriptive statistics of portfolio characteristics (trades and transfers) . 183
A5.2. Impact of scams on portfolio value (trades only) 184
A5.3. Impact of scams on the number of tokens in the portfolio (trades only) . 185
A5.4. Impact of scams on daily portfolio returns (trades only) 186
A5.5. Impact of scams on daily portfolio standard deviation (trades only) . . . 187
A5.6. Impact of scams on daily portfolio Sharpe ratio (trades only) 188
A5.7. Impact of scams on daily portfolio cash allocation (trades only) 189

A5.8. Impact of scams on portfolio value (trades and transfers) 190
A5.9. Impact of scams on the number of tokens in the portfolio (trades and transfers) 191
A5.10. Impact of scams on daily portfolio returns (trades and transfers) 192
A5.11. Impact of scams on daily portfolio standard deviation (trades and transfers) 193
A5.12. Impact of scams on daily portfolio Sharpe ratio (trades and transfers) . 194
A5.13. Impact of scams on daily portfolio cash allocation (trades and transfers) 195
A5.14. Different definitions of scam date and scam amount 196

6.1. Summary of the thesis 198

Chapter 1

Introduction

There is hardly any activity in finance untouched by fraud and misconduct, including market manipulation, insider trading, corporate frauds, misconduct by financial advisers, financial intermediaries, and more recently cryptocurrencies. Understanding the mechanics of different forms of frauds, their prevalence, determinants, and consequences are important to protect consumers and maintain trust in the broader financial system. This dissertation explores different aspects of fraud in finance.

The key question that Chapter 2 explores is whether the public's increasing distrust of the finance sector warranted, drawing evidence from the main interface between the general public and the finance sector — financial advisers. Specifically, in Chapter 2, we examine whether the underlying true rates of misconduct (accounting for a detection bias) by financial advisers warrant the growing skepticism of the public, which has led to protests, disruption, parliamentary inquiries, and movements towards disintermediation and decentralization of financial services.

This dissertation also explores another long-standing malpractice in financial markets — market manipulation. In particular, Chapter 3 explores market manipulation and its real economic consequences. Market manipulation is prohibited in most financial markets around the world. The key question that Chapter 3 addresses is whether market manipulation should be prohibited. The answer depends on how severe are the harms caused by market manipulation, and do they outweigh the costs of regulation and compliance. The literature to date has focused on the direct consequences of market manipulation, such as distorting prices in secondary markets and decreasing market liquidity (e.g., Aggarwal and Wu 2006; Comerton-Forde and Putnins 2011a). But do the impacts of market manipulation extend beyond these direct effects? Does manipulation affect listed companies and real investment decisions? Does it affect where and how capital is deployed? These are the issues that we address in Chapter 3.

Chapter 4 investigates scams using cryptocurrencies. Cryptocurrencies offer many benefits. However, they also provide many avenues for illegal activities, such as scams, money laundering, terrorism financing, ransomware, and many more. Scams are often considered one of the largest forms of cryptocurrency-based crime. For example, more than $7.7 billion worth of crypto was stolen from victims globally in 2021, representing an 81% increase compared to 2020 (e.g., Chainalysis 2022). Chapter 4 investigates the prevalence of cryptocurrency scams in decentralized exchanges by developing a scam detection algorithm and examines whether they can be detected early.

Detecting crypto scams also enables us to investigate important questions in household finance. Households make serious investment mistakes, such as nonparticipation in financial markets or underdiversification (e.g., Campbell 2006, Gomes, Haliassos, and Ramadorai 2021). Among these investment mistakes, falling victim to investment scams is perhaps the most dramatic because they involve significant financial damage to households. For example, households lost more than $20 billion in known investment scams in 2022 (e.g., Knüpfer et al. 2023). Chapter 5 explores the impact of investment scams on households' trading behavior and investment approach and particularly, whether investors learn from their mistakes after falling victim to a scam.

Chapter 2: Is public distrust of the finance sector warranted? Evidence from financial adviser misconduct

Most financial investments and transactions require trust between counterparties. Individuals that are less trusting are less likely to participate in markets (e.g., Guiso, Sapienza, and Zingales 2008). Countries that have more social capital have higher generalized trust, leading to greater financial development (e.g., Guiso, Sapienza, and Zingales 2004). Trust is also important in investment management, which is a major component of most wealth/savings management (e.g., Gennaioli, Shleifer, and Vishny 2015; Garleanu and

Pedersen 2018). Put simply, finance does not work without trust and trust is a cornerstone that enables financial systems. It is therefore puzzling why the public do not trust the finance sector in general.

To investigate whether public's distrust of the finance sector warranted, Chapter 2 focuses on financial advisers because they are the main interface between the public and the finance sector. Over 750,000 financial advisers in the U.S. actively manage over $39 trillion as of 2020.[1] The advisory industry is one of the largest segments of the financial services sector, representing approximately 10% of total employment in the finance and insurance sector.[2]

Chapter 2 documents that approximately 30% of financial advisers in the U.S. with current AUM of $6.9 trillion are involved in misconduct, yet only about one-third of those (9% of advisers) are detected and reprimanded by regulators. In addition, adviser misconduct is pervasive throughout 2003-2020, but spikes during the GFC along with the percentage of unreported (undetected by regulators) misconduct. Chapter 2 also provides a comprehensive list of characteristics that impact advisers' propensity for misconduct. These characteristics can help consumers navigate the sector safely, firms design effective internal controls, and regulators improve regulatory efficiency.

We link adviser misconduct and trust in two ways. First, we correlate the estimated, reported, and unreported share of financial adviser misconduct with the Financial Trust Index (FTI) of Sapienza and Zingales.[3] We show that the estimated share of adviser misconduct and unreported misconduct is at its highest when the FTI is at its lowest during the GFC. Since the GFC, there has been a decreasing trend in the estimated, unreported, and reported share of adviser misconduct and an increasing trend in the FTI, inducing a negative correlation between the FTI and the (estimated, unreported, and reported) share of adviser misconduct. Second, we explore the relationship between the number of clients and AUM of advisory firms and their (estimated, unreported, and reported) number of adviser misconduct. We find that the estimated and unreported

[1] These estimates are based on the SEC Investment Adviser Public Disclosure database.

[2] See, for example, the occupational employment and wage statistics from the U.S. Bureau of Labor Statistics, https://www.bls.gov/oes/home.htm.

[3] The FTI measures the overall level of trust in the US financial institutions since 2009 based on interviews with 1,000 American households, randomly chosen and surveyed via telephone. This initiative is sponsored jointly by the Kellogg School of Management at Northwestern University and the University of Chicago Booth School of Business and administered by Social Science Research Solutions. See http://www.financialtrustindex.org/.

(but not reported) numbers of adviser misconduct have a negative impact on both the number of clients and AUM of advisory firms. Specifically, one adviser with estimated misconduct, on average, costs the firm around 5 clients and decreases the total AUM by around $10 million. Similarly, one adviser with unreported misconduct, on average, costs the firm around 7 clients and decreases the total AUM by around $13 million. These findings further suggest that adviser misconduct undermines trust in financial advisory firms and leads to a decrease in the use of financial advisory services.

This study is the first to quantify the true extent of adviser misconduct and link that to trust in the financial system. To do this, in Chapter 2, we address the problem of incomplete detection by jointly modeling the misconduct and detection process. Quantifying the true extent of adviser misconduct is important to assess the scale of the problem and identify regulatory resources to cope with the problem. In addition, in Chapter 2, we document the economic determinants of financial adviser misconduct and regulatory detection. This can help consumers to consistently use services from advisers with characteristics that do not signal potential misconduct. Furthermore, the economic determinants of adviser misconduct can help regulators and firms efficiently steer surveillance activities.

Chapter 3: The real effects of market manipulation

Market manipulation challenges the functioning and integrity of financial markets by distorting asset prices and creating uncertainty for investors. The literature on market manipulation generally defines it as actions that artificially affect asset prices for personal gain and provides insights into various techniques employed by market participants to manipulate stock prices, such as pump-and-dump schemes, wash trades, spoofing and layering. Market manipulation not only distorts market information and increases risk for investors, but also leads to the misallocation of resources resulting in deadweight economic losses to society.

Despite extensive research on market manipulation, there is limited empirical research on the detrimental effects market manipulation on resource allocation. This limitation stems partly from the unobservable nature of the true extent of market manipulation, with regulators detecting and prosecuting only a small fraction of manipulative activities in markets (e.g., Comerton-Forde and Putnins 2014). Consequently, the real economic consequences of market manipulation remain relatively unknown.

However, market manipulation could have real economic consequences by distorting corporate investment decisions. For example, Goldstein and Guembel (2008) show that the a feedback effect from the financial market to the real firm value creates an incentive for traders to manipulate stock prices, resulting in a weaker allocational role of financial markets. But in practice, is manipulation sufficiently frequent and distortionary to significantly impact firms? And beyond isolated manipulative attacks like the ones anticipated by Goldstein and Guembel (2008), does the ex-ante possibility that prices might be distorted by manipulation decrease the use of price signals in a way that harms real investment decisions?

Chapter 3 addresses these questions. Specifically, Chapter 3 develops a conceptual framework that shows how the possibility of manipulative price distortions can decrease the use of price signals by corporate managers. As a result, stock price manipulation has a negative impact on firm performance since managers relying less on stock prices tend to make less-informed investment decisions. Chapter 3 also takes these predictions to the data.

We examine the impact of stock price manipulation on investment-to-price sensitivity and firm performance using a large sample of stocks in US during 1997-2019. Consistent with theory, this chapter documents that stock price manipulation has a significant negative impact on firms' investment-to-price sensitivity. Then it provides evidence that stock price manipulation significantly harms the future operating performance of firms. In summary, Chapter 3 documents that market manipulation suppresses the production of new information to managers, thereby reducing the investment-to-price sensitivity and corporate profitability.

Chapter 4: Scam alert: Can cryptocurrency scams be detected early

Chapter 4 investigates the prevalence of cryptocurrency scams and their determinants. In particular, we focus on "hard" and "soft" rug pulls that trade in decentralized exchanges. Hard rug pulls are scams with malicious code in their token smart contracts designed to exploit investors. The malicious code often disables investors' ability to sell tokens or includes other hidden functionalities that are detrimental to investors. Soft rug pulls are harder to identify because they do not have malicious code in smart contracts. They often occur after trust has been built with investors. However, they are similar to hard rug pulls in that token creators steal liquidity and leave investors with worthless tokens in the end.

Despite some evidence of well-known high-profile rug pulls, several important questions remain unanswered regarding their extent and characteristics. How are rug pulls typically formed? What is the scale of rug pulls, and how does it evolve over time? What are the distinguishing characteristics of scam tokens? Are there characteristics that can predict whether a given token will turn out to be a scam or rug pull on an ex ante basis? These questions are important for understanding cryptocurrency scams but are also challenging to address due to the difficulty in detecting such scams. Chapter 4 answers these questions by identifying hard rug pulls based on smart contract audits and soft rug pulls with a simple detection algorithm.

Based on external smart contract audits, Chapter 4 documents that approximately 3% of liquidity pools on major DEXs are hard rug pulls, resulting in an estimated theft of around $292 million between May 2020 and April 2022.

We also develop a simple soft rug pull detection algorithm based on three heuristic patterns: price, activity, and profit. Our algorithm identifies soft rug pulls, detecting around 41% of liquidity pools as soft rug pulls, with an estimated theft of about $1.28 billion from investors during the same period. By quantifying the extent and evolution of scams within major DEXs, Chapter 4 helps reduce the uncertainty surrounding the scale of scamming activities in DEXs and enables more informed decisions by policymakers assessing the costs and benefits of DEXs. Furthermore, this chapter develops a scam index based on ex-ante characteristics. The scam index can serve as a proactive tool for identifying potential scam pools, with potential applications in regulation, investor decision-making, academic research, and blockchain development.

Chapter 5: Once bitten twice shy: Learning about scams

Little evidence exists on the impact of scams on victims. Do victims learn from their mistakes or continue reinvesting in scams? How does their investment performance change after falling victim to a scam? Are there changes in their risk tolerance? Do they adjust their portfolios post-scam? What characterizes different learning rates among victims? These are important questions in household finance, but obtaining reliable answers is extremely challenging because they require high-quality data on scams and victims.

Identifying scams and victims in Chapter 4 enables us to address these questions. Chapter 5 addresses these challenges by leveraging blockchain's unique capabilities to detect scams and victims. Detecting cryptocurrency scams through smart contract audits and systematic analysis of transaction activity in DEXs in Chapter 4, not only helps identify scams but also enables us to track the trading behavior of investors who have fallen victim to these scams.

In Chapter 5, we construct trading and portfolio characteristics of investors. We then document that investors tend to become more active in the market after falling victim a scam. However, they tend to experience lower returns on their investments. On average, the post-scam portfolio of victims demonstrates a lower average daily return, along with a lower standard deviation of returns, and higher cash holdings (ratio of stablecoins in the portfolio) compared to their pre-scam portfolio.

Building upon these findings, we construct multivariate models to analyze the impact of scams on investor behavior. Using a probit model on trading characteristics, we find that investors are less likely to re-invest in scams after experiencing one. This is particularly true for investors who experienced significant losses in scams. Furthermore, victims of scams experience a negative effect on subsequent non-scam investment returns, which decrease by approximately 14%. Extending the analysis to portfolio characteristics, we find that after experiencing a scam, investors become more risk-tolerant in the short term (if the investor has been scammed in the past 30 and 60 days), but less risk-tolerant in the long term (if the investor has been scammed in the past 90 and more days), as measured by the 7-day moving average standard deviation of daily portfolio returns.

We also explore variations in learning outcomes among investors. Less experienced investors with shorter trading histories tend to learn less and are more likely to invest in multiple scam pools. Investors with higher trading volumes, lower returns on their non-scam investments, and lower cash allocation tend to learn less and invest in multiple scam pools. In addition, higher portfolio values and a greater number of tokens within portfolios increase the likelihood of investing in multiple scams.

1.1 Literature review

In this section, we review the general literature on economics of fraud, focusing on misconduct by financial advisers, market manipulation, cryptocurrency scams, and behaviour of victims. The objective is to provide an overview of recent trends and research on fraud in financial markets and show the relevance of the dissertation in contributing to these areas. We compare each chapter with related papers in the corresponding chapter. This dissertation is related to six main strands of literature.

1.1.1 Trust in finance

Chapter 2 of this dissertation aligns closely with existing literature that highlights the crucial role of trust in finance. The effective operation of financial investments and transactions requires the presence of trust between counterparties. Guiso, Sapienza, and Zingales (2004) highlight the role of trust in the financial development of countries. Guiso, Sapienza, and Zingales (2008) show that less trusting individuals are less likely to participate in financial markets. Additionally, Gennaioli, Shleifer, and Vishny (2015) and Garleanu and Pedersen (2018) demonstrate the importance of trust in delegated investment management, which is a major component of most wealth/savings management.

Chapter 2 is also related to the role of trust in choosing financial service providers. The literature shows that investors prioritize trust when selecting financial service providers (Hung et al., 2008). Furthermore, Burke and Hung (2021) show that trust influences investors' decisions to follow financial advice. Kostovetsky (2016) argues that a decline in trust due to changes in ownership leads investors to withdraw investments from funds. More broadly, Guiso (2010) and Sapienza and Zingales (2012) study the decrease of trust in financial institutions following the GFC. Zingales (2015) also shows that financial advisers consistently rank among the least trustworthy professionals, generally due to anti-finance sentiment and well-known misconduct scandals.

1.1.2 Financial adviser misconduct

This dissertation contributes to the literature on the economics of fraud and misconduct, dating back to seminal work of Becker (1968). Specifically, Chapter 2 focuses on fraud and misconduct in financial advice industry. Egan et al. (2019) show that 7% of advisers in the U.S. have at least one past misconduct record.

A growing literature in finance finds that many advisers exploit their clients by selling unsuitable investment products. Inderst and Ottaviani (2009, 2012*a*) show financial advisers harm investment performance by steering client funds in high fee products. Furthermore Hackethal, Haliassos and Jappelli (2012) demonstrate that financial advisers may affect portfolio outcomes by recommending unsuitable investment products.

More recently, there has been growing interest in the factors that impact adviser misconduct. For example, Qureshi and Sokobin (2015) show that brokers' disciplinary records, financial disclosures, and employment history are significant predictors of investor harm. Dimmock, Gerken, and Graham (2018) find that coworker misconduct influences adviser's propensity for misconduct, making adviser misconduct contagious. Kowaleski, Sutherland and Vetter (2020) demonstrate that investment advisers passing the qualification exam with fewer rules and ethics coverage are more likely to commit misconduct. Furthermore, using advisers' home addresses, Dimmock, Gerken and Van Alfen (2021) find that advisers who suffer more from household-level financial shocks are more likely to engage in misconduct. Egan et al. (2022) explore the gender punishment gap of adviser misconduct. Chapter 2 contributes to this literature by identifying a comprehensive list of characteristics associated with adviser misconduct and detection.

1.1.3 Market manipulation

Chapter 3 is closely related to the studies on market manipulation. Cumming, Li and Yu (2023), Pirrong (1995), and Brunnermeier and Oehmke (2013) show that market manipulation not only distorts market information and increases the risk for investors but is also associated with the misallocation of resources, resulting in economic losses for society.

The literature on market manipulation generally defines it as actions that artificially affect asset prices for personal gain and provides insights into various techniques employed by market participants to manipulate stock prices, such as pump-and-dump schemes, wash trades, spoofing and layering (see Allen and Gale 1992, Aggarwal and Wu 2006, and Putnins 2012). The other empirical studies on market manipulation include Merrick, Naik and Yadav (2005) on squeezes, Allen, Litov and Mei (2006) on corners, and Cumming, Li and Yu (2023) on the relationship between market manipulation and corporate venture capital.

Furthermore, Comerton-Forde and Putnins (2014) investigate the determinants and frequency of closing price manipulation, and Comerton-Forde and Putnins (2011b) show that closing price manipulation is associated with large increases in day-end returns, return reversals, trading activity, and bid-ask spreads. Using the closing price manipulation measure developed by Comerton-Forde and Putnins (2011b), Chapter 3 contributes to this literature by focusing on the real economic consequences of the closing price manipulation.

1.1.4 Investment-to-price sensitivity

Chapter 3 is also related to the literature that analyzes how stock prices affect firms' investment decisions (e.g., Morck, Shleifer, and Vishny 1990, Barro 1990) and how firm managers learn from stock prices (see Bond, Edmans, and Goldstein 2012 for a survey). Specifically, we examine managerial learning in the presence of market manipulation to measure the real effects of manipulative price distortions.

Theoretically, Dow and Gorton (1997) and Subrahmanyam and Titman (1999) show that firm managers learn from information in stock prices about the prospects of their firms, suggesting that stock prices play a more active role in driving the real economy. The theory suggests that stock prices aggregate information from many different market participants. These market participants do not generally have communication channels with the firm outside the trading process. Thus, stock prices may complement the inside information of managers and guide them in making corporate investment decisions, referred to as the managerial learning channel from stock prices in the literature.

Chapter 3 is closely related to the empirical studies of managerial learning channels, such as the role of private information (e.g., Chen, Goldstein, and Jiang 2007), cross-listing (e.g., Foucault and Frésard 2012), peers' stock prices (e.g., Foucault and Frésard 2014, Dessaint et al. 2019), and the importance of political connections among countries (e.g., Durnev 2010) in driving the investment-to-price sensitivity of firms. The managerial learning from stock prices is further investigated by Edmans, Jayaraman, Schneemeier (2017), Baker, Stein, and Wurgler (2003). Chapter 3 contributes to this literature by documenting that market manipulation suppresses the production of new information to firm managers, and thus reduces firms' investment-to-price sensitivity and operating performance.

1.1.5 Blockchain technology, cryptocurrencies, and scams

A large literature now studies the economics of blockchain and cryptocurrencies (see, for example, Easley et al. 2019, Yermack 2017, Huberman et al. 2021, Saleh 2021, Budish 2022, Rosu and Saleh 2021, Biais et al. 2023, Chiu and Koeppl 2017, 2019). Chapter 4 is related to a branch of literature that focuses on various aspects of blockchain technology and cryptocurrencies, with a specific emphasis on illegal activities. Some of these studies focus on the involvement of bitcoin in illegal activities (e.g., Meiklejohn et al. 2013; Foley, Karlsen and Putnins 2019), pump and dump manipulation in cryptocurrency markets (e.g., Dhawan and Putnins 2021; Li, Shin and Wang 2021), initial coin offerings (ICO) scams (e.g., Hu, Parlour and Rajan 2019; Howell, Niessner and Yermack 2020; Benedetti and Kostovetsky 2021; Phua et al. 2022), ransomware attacks (e.g., Cong et al. 2023), crypto laundering (e.g., Dupuis and Gleason 2020), and wash trading that artificially boosts trading volumes on cryptocurrency exchanges (e.g., Aloosh and Li 2019; Cong et al. 2021). Chapter 4 contributes to this literature by quantifying the prevalence of cryptocurrency scams in DEXs and developing a scam index that can detect cryptocurrency scams before they occur.

1.1.6 Investor behaviour after scam

Chapter 5 focuses on investors falling victim to a scam. Few studies that analyses the effects of investment scams on victims focus on a single isolated scam. For example, Knüpfer et al. (2023) study the long-run income loss experienced by victims who fell to a "Wincapita" Ponzi scheme in Finland during 2003-2008 (also see Rantala 2019). Furthermore, Giannetti and Wang (2016) show that federal securities enforcement actions against fraudulent firms reduce household market participation in the fraudulent firm's state (also see Georgarakos and Inderst 2014). Gurun, Stoffman and Yonker (2018) document that communities exposed to the Madoff Ponzi scheme withdraw assets from investment advisers and increase deposits at banks and relate their findings to eroding trust in the financial industry.

Unlike existing literature that focus on a single isolated scam, Chapter 5 contributes to this learning literature by focusing on learning from investments in cryptocurrency scams identified in Chapter 4. Specifically, we investigate the impact of scams on investors' subsequent trading behavior. There is some evidence suggesting the potential for learning through trading actions. For example, Nicolosi, Peng and Zhu (2009) and Seru, Shumway and Stoffman (2010) find that individuals' trading performance improves with trading experience. Learning through trading has been explored in various other settings. For example, Strahilevitz, Odean and Barber (2011) examine whether previous experiences with a stock influence investors' willingness to repurchase stocks, and List (2003) investigate whether experience eliminates investors' endowment effect (tendency of investors to retain an asset they own). Other papers that document investors' learning include Feng and Seasholes (2005), Linnainmaa (2010), Pástor, Taylor and Veronesi (2009), and Choi et al. (2009). Chapter 5 contributes to this literature by focusing on learning from scams.

1.2 Thesis outline

The remaining sections of this thesis are divided into four main chapters:

- Chapter 2: Is public distrust of the finance sector warranted? Evidence from financial adviser misconduct
- Chapter 3: The real effects of market manipulation
- Chapter 4: Scam alert: Can cryptocurrency scams be detected early?

- Chapter 5: Once bitten twice shy: Learning about scams

Chapter 6 summarizes the key findings and discusses their implications.

Chapter 2

Is public distrust of the finance sector warranted? Evidence from financial adviser misconduct

2.1 Introduction

Finance is built on trust. Yet, the public increasingly does not trust the finance sector. Is this distrust warranted? That is the key question that we explore in this chapter, drawing evidence from the main interface between the general public and the finance sector — financial advisers. Specifically, we examine whether the underlying true rates of misconduct (accounting for a detection bias) by financial advisers warrant the growing skepticism of the public, which has led to protests, disruption, parliamentary inquiries, and movements towards disintermediation and decentralization of financial services.

Our analysis reveals a solid basis for the distrust. We estimate that as much as one-third of financial advisers in the U.S. are involved in misconduct. Such levels of misconduct vastly exceed those in other professional services (e.g., medical professionals, lawyers, engineers), with the exception of real estate agents and used car salesmen, which arguably are not professionals and perhaps even embarrassingly part of a broader definition of the "financial system."

Finance does not work without trust. Most financial investments and transactions require trust between counterparties. Less trusting individuals are less likely to participate in financial markets (e.g., Guiso, Sapienza, and Zingales 2008). Countries that have more social capital have higher generalized trust, leading to greater financial development (e.g., Guiso, Sapienza, and Zingales 2004). Trust is also crucial in delegated investment management, which is a major component of most wealth/savings management (e.g., Gennaioli, Shleifer, and Vishny 2015; Garleanu and Pedersen 2018). Put simply, trust is a cornerstone that enables financial systems.

Financial advisers (including investment advisers and brokers) are the main interface between the public and the finance sector.[4] Over 750,000 financial advisers in the U.S. actively manage around $39 trillion as of 2020.[5] Examples include Blackrock Fund Advisors with over $2.4 trillion assets under management (AUM), J.P. Morgan ($2 trillion AUM), and Morgan Stanley ($1.18 trillion AUM) or smaller advisory firms such as Nichols Investment Advisory Services with one investment adviser representative managing $23 million AUM for 100 clients.[6] The advisory industry is one of the largest segments of the financial services sector, representing approximately 10% of total employment in the finance and insurance sector.[7]

Public trust of finance deteriorated during the GFC and never fully recovered. According to various survey data, financial service providers consistently rank among the least trustworthy professionals (e.g., Edelman Trust Barometer 2022; Zingales 2015).[8] This erosion of trust was exemplified by the Occupy Wall Street protests in September 2011, which brought attention to issues of greed and corruption within the financial services sector. In response to corporate greed and a growing lack of faith in traditional financial institutions, digital currencies like Bitcoin emerged as an alternative decentralized form

[4] Brokers are registered with the Financial Industry Regulatory Authority (FINRA) and the U.S. Securities and Exchange Commission (SEC) and are defined in the Securities and Exchange Act 1934 as "any person engaged in the business of effecting transactions in securities for the account of other". An investment adviser provides financial advice rather than transaction services. Both are often considered financial advisers and we refer to both as financial advisers throughout the chapter.

[5] These estimates are based on the SEC Investment Adviser Public Disclosure database.

[6] The AUMs are sourced from 2021 Form ADVs filed with the SEC Investment Adviser Public Disclosure database.

[7] See, for example, the occupational employment and wage statistics from the U.S. Bureau of Labor Statistics, https://www.bls.gov/oes/home.htm.

[8] See also "Brokers Are Trusted Less than Uber Drivers, Survey Finds" by Anna Prior, Wall Street Journal, July 28, 2015.

of money. These events collectively illustrate the enduring impact of the GFC on public perception of finance sector.

But perceptions do not always equal reality. In fact, finance academics and practitioners usually disregard the public perception of the finance sector and view the "anti-finance" or "anti-Wall Street" sentiment as the manifestation of ignorant populism (e.g., Sapienza and Zingales 2013; Zingales 2015). To examine whether the public distrust of finance is warranted and supported by the data, we extract comprehensive data on all U.S. financial advisers and use econometric techniques to estimate what fraction is involved in misconduct. We do not rely on reported cases of misconduct as there is a well-known under-reporting and under-detection bias. Instead, we use a sample of reported misconducts to jointly estimate a structural model of misconduct and detection processes to get at the true underlying rate of misconduct.

Our results are striking. We find that approximately 30% of financial advisers in the U.S. with current AUM of $6.9 trillion are involved in misconduct, yet only about one-third of those (9% of advisers) are detected and reprimanded by regulators. In addition, adviser misconduct is pervasive throughout 2003-2020, but spikes during the GFC along with the percentage of unreported (undetected by regulators) misconduct. We provide a comprehensive list of characteristics that impact advisers' propensity for misconduct. These characteristics can help consumers navigate the sector safely, firms design effective internal controls, and regulators improve regulatory efficiency.

We construct a database of all advisers in the U.S. with adviser-level characteristics for 1.12 million advisers and firm-level characteristics for 0.75 million advisers. We observe a complete set of disclosures such as customer disputes, disciplinary events, and other financial matters for each adviser grouped into 23 categories by the Financial Industry Regulatory Authority (FINRA). The disciplinary events include civil, criminal and regulatory events, and other disclosed investigations. Since disclosures are not always indicative of wrongdoing, following Egan, Matvos and Seru (2019), we conservatively define misconduct as the regulatory and criminal offenses, and customer events resolved against the adviser (six most severe disclosure categories).[9] Around 9% of advisers have reported misconduct in our sample.

[9] The misconduct-related disclosure categories are Customer Dispute—Settled, Customer Dispute—Award/Judgment, Regulatory—Final, Employment Separation After Allegations, Criminal—Final Disposition, and Civil—Final.

Starting from the sample of advisers with reported misconduct, we estimate two structural models. First, we estimate with adviser-level characteristics that more than 330,000 advisers (about 30% of all advisers) in the U.S. have been involved in misconduct. Complementing the adviser-level characteristics with firm characteristics of advisers in an augmented model uncovers a similar misconduct rate. The number of advisers estimated to be engaged in misconduct is around three times more than those with reported misconduct by regulators. These results indicate that misconduct in this industry accounts for a sizeable proportion of the sector, suggesting a rationale for why the public distrusts the finance sector.

We use several approaches to test the robustness of our findings. First, we estimate the model separately for investment advisers and brokers. We find that around 32% of brokers and 16% of investment advisers are likely to be involved in misconduct. Second, we adopt a stricter definition of misconduct (the most severe three disclosure categories) and find that about 18% of advisers are involved in misconduct, but only 8% of advisers are detected by regulators.[10] Third, we examine the impact of excluding specific instrumental variables in the model to show that our findings are not dependent on one particular instrumental variable but on the collection of variables that characterize adviser misconduct and detection.

Advisers involved in misconduct tend to have several interesting distinguishing characteristics compared to those with a clean record. They have a shorter tenure, work for more firms, and experience a shorter average employment duration. Additionally, advisers with a higher propensity for misconduct tend to be male and dual-registered with the SEC and FINRA. They have more disclosures, work in a more toxic environment (with a high office misconduct rate), spend a higher percentage of their tenure during economic recessions, and pass fewer professional industry exams. Regarding their firm characteristics, the firms they work for tend to be larger, and are more likely to charge fees per hour or based on a portion of AUM. Additionally, their firms have lower discretionary AUM and less sales interest in client transactions.

Our analysis of the timeseries of misconduct aligns with the timeseries of public distrust in the finance sector. The share of estimated and unreported misconduct is at its highest, and trust (as measured by the Sapienza and Zingales financial trust index) is at its lowest

[10] The most severe disclosure categories are Criminal—Final Disposition, Regulatory—Final, and Customer Dispute—Settled.

during the GFC. These trends indicate that the risk of adviser misconduct increases during financial stress (e.g., Dimmock, Gerken, and Van Alfen 2021). Furthermore, it raises concerns about regulators' response time to the increase in adviser misconduct, justifying the increased distrust in financial institutions during the GFC. Since the GFC, trust has partly recovered, and the estimated and unreported share of adviser misconduct has slightly decreased.

We further link misconduct and trust by showing that the estimated and unreported number of adviser misconduct decrease the number of clients and assets under management (AUM) of advisory firms. Specifically, one adviser with estimated (resp. unreported) misconduct, on average, costs the firm around 5 (resp. 7) clients and decreases the total AUM by around $10 million (resp. $13 million) per annum. These findings suggest that adviser misconduct can undermine trust in financial markets, leading to decreased market participation and higher costs of capital for firms.

Financial transactions are vital in market-based economies, connecting household savings with users of financial capital. For example, Giannetti and Wang (2016) show that federal securities enforcement actions against fraudulent firms reduce household market participation in the fraudulent firm's state (also see Georgarakos and Inderst 2014). Furthermore, Gurun et al. (2018) show that communities exposed to the Madoff Ponzi scheme withdraw assets from investment advisers and increase bank deposits. The ability to benefit from financial market returns is crucial for household welfare, and distrust in financial advisers is likely to lead to significant welfare losses.

This chapter relates to the literature on the role of trust in the financial system. The literature shows that investors prioritize trust when choosing financial service providers (Hung et al., 2008), and trust influences investors' likelihood of following advice (Burke and Hung 2021). Kostovetsky (2016) argues that a decline in trust due to changes in ownership leads investors to withdraw investments from funds. More broadly, Guiso (2010) and Sapienza and Zingales (2012) study the trust decline in financial institutions following the GFC. We contribute to this literature by showing that the trust decline is justified and supported by the data.

Chapter 2 also relates to other analyses of financial advisers. A growing literature in finance finds that many advisers exploit their clients by selling unsuitable investments.[11] Recently, there has been growing interest in the factors that impact adviser misconduct.[12] For example, Qureshi and Sokobin (2015) show that brokers' disciplinary records, financial disclosures and employment history are significant predictors of investor harm. Egan et al. (2019) show that 7% of advisers in the U.S. have at least one past misconduct record. Unlike these papers that focus on only reported adviser misconduct cases, we quantify and characterize the true extent of adviser misconduct and its evolution through time. Quantifying the true extent of adviser misconduct is important to assess the scale of the problem and identify resources to cope with the problem. Thus, Chapter 2 is also related to the literature that evaluates various policy responses in regulating consumer financial products.[13]

2.2 Data and descriptive statistics

2.2.1 Data, filters and transformations

We construct a unique database of all financial advisers in the U.S. from multiple data sources. We extract the complete record of investment advisers from the Investment Adviser Public Disclosure (IAPD) database jointly operated by the U.S. SEC and North American Securities Administrators Association (NASAA) and brokers from the FINRA BrokerCheck database.[14] Our initial dataset contains 1,246,001 financial advisers, including 361,890 currently registered active investment advisers, 166,099 previously registered inactive investment advisers, 623,629 active brokers, and 545,310 inactive brokers. Due to the dual registration with the SEC/NASAA and FINRA, 88.2% of investment advisers are also brokers.

[11] See Inderst and Ottaviani (2009, 2012*a*, 2012*b*), Bergstresser, Chalmers and Tufano (2008), Hackethal, Haliassos and Jappelli (2012), Hackethal, Inderst and Meyer (2012), Mullainathan, Noeth and Schoar (2012)), and Chalmers and Reuter (2020).

[12] See Dimmock, Gerken and Graham (2018), Dimmock, Gerken and Van Alfen (2021), Kowaleski, Sutherland and Vetter (2020), Egan, Matvos and Seru (2022), and Charoenwong, Kwan and Umar (2019).

[13] See Campbell (2006), Campbell et al. (2011), Agarwal et al. (2009, 2015), Inderst and Ottaviani (2010).

[14] See https://adviserinfo.sec.gov/ for IAPD platform and https://brokercheck.finra.org/ for BrokerCheck platform.

We observe the full name, unique Central Registration Depository (CRD) number, adviser flags (four dummies for an active and inactive investment adviser or broker) and data on the complete employment history of advisers, including the firm name, firm CRD, the status of the adviser in the firm (investment adviser or broker), office location and employment dates. Along with employment details, we observe all related industry qualification exams passed by advisers.[15] We also observe all disclosures of advisers, such as customer disputes, disciplinary events and other potential wrongdoings reported under 23 disclosure categories and the corresponding date of each disclosure.

The IAPD and BrokerCheck platforms do not provide the gender of advisers. We determine the genders based on advisers' first names using the GenderChecker and Genderise platforms. We create a gender dummy variable equal to one if the adviser is male.

We complement our dataset with the yearly number of employees (full-time equivalents) in the Office of Compliance Inspections and Examinations (OCIE) department of the SEC from 2003 to 2020 from the annual reports of the SEC.[16] OCIE conducts examinations and inspects financial advisers and other market participants to improve compliance and prevent fraud in financial markets. We scale the number of full-time OCIE employees with the number of advisers each year to calculate the annual number of inspectors per adviser, which serves as a proxy for the regulatory budget. We calculate the average regulatory budget during the employment period of each adviser.

We refer to the data extracted from the IAPD and BrokerCheck platforms and complemented with the gender and regulatory budget data as the adviser-level data. We apply multiple filters and transformations to the adviser-level data to construct adviser-level variables.

1. If the adviser's active investment adviser flag is zero, but the adviser's status in the firm they are working for is an investment adviser, we set that employment information to missing. We do this consistently for inactive investment advisers and active and inactive brokers.

2. We remove advisers with no active and inactive investment adviser or broker flags.

[15] The industry exam categories include (i) State Securities Law exams, (ii) General Industry/Products exams and (iii) Principal/Supervisory exams with different series of exams within each category.

[16] The number of employees in the OCIE department of the SEC is not reported in the annual reports before 2003.

3. We remove employment with missing firm CRD.

4. We remove employment with fewer than 90 days to account for a standard probation period of three months.

5. We remove disclosures with missing disclosure categories or dates.

6. For some advisers, the office location is missing. We first replace the missing office location with the location of the same firm that the adviser has worked with a different status or in a different year. If the location is still missing, we replace the office location with the primary address of the firm (the one with the highest number of advisers). Lastly, if the location is still missing, we replace the location with the firm id (i.e., all advisers in that firm work in the same office).

7. We remove advisers with unavailable gender data.

8. The number of OCIE employees is available from 2003 onwards. Thus, we remove advisers who became inactive before 2003.[17]

After applying these filters, we calculate a collection of adviser-level variables for each of the 1,120,771 advisers in our sample: office misconduct rate, other disclosure count, economic recession, firm count, employment duration, gender, reported misconduct rate, regulatory budget, Dodd-Frank switch, tenure, exam count, employee count, active financial adviser, and dual registration. Table 2.1 reports the detailed definitions of these variables.

We complement the adviser-level data with the firm-level data by going through multiple Form ADVs submitted by investment adviser firms. Investment adviser firms use Form ADV to register with the SEC and/or states. The form is updated annually by investment adviser firms.[18]

---

[17] Note that we keep advisers who have started to work before 2003 but have not become inactive before 2003. For example, if an adviser worked during 2000-2005, he/she is assigned the average regulatory budget from 2003 to 2005.

[18] Part 1 of the Form ADV is accessible from https://www.sec.gov/about/forms/formadv-part1a.pdf. Some advisers who do not have to register with the SEC or the states (i.e., exempt reporting advisers) also complete some of the questions in Form ADV for purposes of reporting to the SEC and/or states. We exclude exempt reporting advisers from the analysis for two reasons. First, we do not observe all necessary firm-level data fields of exempt reporting advisers since they only fill in certain parts of the Form ADV. Second, they are marginally small advisory firms with small AUM, making them unimportant in scale.

Table 2.1. Definitions of variables

This table reports the variables that we compute for each adviser. Panel A defines the adviser-level variables computed from the Investment Adviser Public Disclosure (IAPD) and the Financial Industry Regulatory Authority (FINRA) BrokerCheck databases complemented with GenderChecker and Genderise platforms, and the SEC annual reports. Panel B defines the firm-level variables computed from Form ADVs submitted by investment adviser firms. The last two columns specify whether a variable is used in the misconduct equation (M) and the detection equation (D) or both equations of the DCE model. The symbols + and - indicate that a variable is included in the model and reports the sign of our prior belief of its impact on the probability of misconduct (M) or detection (D). The symbol ? also indicates that a variable is included in the model, but we do not have a strong prior about its impact on the probability of misconduct or detection.

Variable	**Definition**	**M**	**D**
Panel A: Adviser-level variables			
Office misconduct rate	The average yearly reported misconduct rate in each office location of an adviser during the employment period. Office misconduct rate is calculated as an average of the yearly ratio of misconduct count to adviser count in the adviser's office during the employment period of an adviser.	+	
Other disclosure count	The total number of non-misconduct related disclosures of an adviser that are not misconduct.	+	
Economic recession	Percentage of experience in years worked during the following economic recession years: Great Depression (1929-1933), Roosevelt recession (1937-1938), Union Recession (1945), Post-War Recession (1948-1949), Post-Korean War Recession (1953-1954), Eisenhower Recession (1957-1958), "Rolling Adjustment" Recession (1960-1961), Nixon Recession (1969-1970), Oil Shock Recession (1973-1975), Energy Crisis Recession (1980-1982), Gulf War Recession (1990-1991), Dot Com Recession (2001), Global Financial Crisis (2007-2009), COVID-19 Recession (2020).	+	
Firm count	The total number of advisory firms a financial adviser has worked for.	+	
Employment duration	The average employment duration of an adviser in financial advisory firms.	-	
Gender	Dummy variable that is equal to one if an adviser is male and zero if female.	+	
Reported misconduct rate	The average yearly misconduct rate during the employment of an adviser. Misconduct rate is a ratio of the number of advisers with reported misconduct each year to the total number of advisers during that year.		+
Regulatory budget	The average yearly regulatory budget during the employment period of an adviser. The regulatory budget is calculated as the yearly number of employees (full-time equivalents) in the SEC Office of Compliance Inspections and Examinations (OCIE) scaled with the yearly number of advisers.		+
Dodd-Frank switch	Dummy variable that is equal to one if an adviser switches from the SEC to state regulation in 2012 with the implementation of the Dodd-Frank Act, which required "midsize" advisory firms ($25 million to $100 million in AUM) to transition from the SEC to state regulation.		+
Tenure	The total employment period of an adviser (in years) excluding the employment gaps.	-	?
Exam count	The total number of relevant industry (broker and investment adviser) exams a financial adviser has passed throughout his/her career. The relevant exam list includes a series of exams related to State Securities Law exams, General Industry/Products exams, and Principal/Supervisory exams.	-	+
Employee count	The average total number of employees working in the adviser's firm during the employment period of an adviser.	+	-

Table 2.1. Continued

Variable	Definition	M	D
Active financial adviser	Dummy variable that is equal to one if an adviser currently practices and zero if not.	-	-
Dual registration	Dummy variable that is equal to one if an adviser is or has been registered as both broker with FINRA and investment adviser with SEC/NAASA during his/her career.	+	+
Panel B: Firm-level variables			
Portion of AUM fees	Experience-weighted average of the dummy variable that is equal to one if an adviser has worked for a firm that is primarily compensated based on a portion of AUM.	?	
Per hour fees	Experience-weighted average of the dummy variable that is equal to one if an adviser has worked for a firm that is primarily compensated based on per-hour fees.	?	
Fixed fees	Experience-weighted average of the dummy variable that is equal to one if an adviser is compensated based on the fixed fees.	?	
Total AUM	Experience-weighted average of the latest available total assets under management (AUM) of the firm that an adviser has worked for. The total AUM of each adviser is computed as Total AUM $= \sum_{i=1}^{n} \frac{Tenure_i}{Total\ tenure} \cdot AUM_i$, where $Tenure_i$ is the experience of an adviser (in years) in firm i and AUM_i is the AUM of firm i. The computation of the firm-level variables is similar.	+	-
Discretionary AUM	Experience-weighted average of the latest available discretionary AUM of the firm that an adviser has worked. The asset is considered under discretionary authority if a firm decides which securities to purchase and sell for the client or which investment advisers to retain on behalf of the client.	?	+
Custody AUM	Experience-weighted average of the latest available AUM under direct custody of the firm that an adviser has worked for. The asset is considered under direct custody if an adviser firm directly holds client funds (cash or access to bank accounts), securities or has any authority to obtain possession of them.	?	+
Number of clients	Experience-weighted average of the latest available client account count of the firm that an adviser has worked for.	+	-
Sales interest	Experience-weighted average of the dummy variable that is equal to one if an adviser has sales interest in client transactions. Sales interest includes (i) cross-agency transactions, (ii) recommend purchase of securities to advisory clients for which the adviser or any related person serves as underwriter, general or managing partner, or purchaser representative, and (iii) recommend purchase or sale of securities to advisory clients for which the adviser or any related person has any other sales interest.	?	+

To access the ADV filings of advisory firms, we extract the monthly SEC-registered investment adviser reports from November 2009 to June 2020.[19] We also access the recent Form ADV filings of the current state-registered investment adviser firms from the IAPD database.[20] For consistency, we only keep the latest non-missing values of variables

[19] The monthly SEC-registered investment adviser reports are from the SEC's webpage and are accessible through https://www.sec.gov/help/foiadocsinvafoiahtm.html. The data fields in these monthly reports refer to specific questions within Form ADV (e.g., 9A(1) refers to Form ADV Item 9A(1)). There are some changes to the format of questions in Form ADV over time (e.g., the percentage of clients are asked as a check-the-box format in ranges until 2017 and after that are asked as a fill-in-the-blank format), and the item numbers may change from one period to another. We manually review these changes to ensure consistent data fields for firm-level variables.

[20] Unlike the SEC-registered firms, we only observe Form ADV filings of the current state-registered investment adviser firms from https://adviserinfo.sec.gov/compilation.

for the SEC-registered firms.

The firm-level data includes the total AUM, number of clients, and AUM in different account types: (i) client accounts under custody and (ii) discretionary accounts.[21] We also observe whether a firm has a sales interest in clients' transactions.[22] Lastly, we observe the fee structure of firms' advisory services. From the combined SEC and state registered firm-level data, we construct 8 firm variables based on their latest available values: total AUM, discretionary AUM (AUM in discretionary accounts), custody AUM (AUM in custody accounts), number of clients, sales interest (one if an firm has sales interest in client transactions), and three dummy variables for the fee structure of firms (portion of AUM, per-hour fees, and fixed fees). Table 2.1 also reports the detailed definitions of firm-level variables.

We merge the firm-level data with the adviser-level data and refer to this sample as the firm-level sample. To merge the firm-level data with the adviser-level data, we keep the adviser-level variables intact where the unit of observation is an adviser and transform the firm-level variables by taking an experience-weighted average of the firm data based on the adviser's experience in that firm. During the merge, we only keep advisers with more than 50% of experience with available firm data.[23]

Complementing the adviser-level data with firm-level data reduces the number of advisers from 1,120,771 to 753,833, but increases the number of model variables. Therefore, we separately conduct the empirical analysis based on the adviser-level sample with an entire universe of advisers and the firm-level sample with both adviser- and firm-level characteristics.[24]

[21] The account is considered under custody if a firm has access to client funds (cash or access to bank accounts) or securities. The account is discretionary if a firm decides which securities to purchase and sell for the client or which investment advisers to retain on behalf of the client.

[22] Sales interests include cross-agency transactions (e.g., a broker-dealer executes trades for brokerage customers in which advisory client securities are sold to or bought from the brokerage customer) or recommending the purchase of securities to advisory clients for which the adviser has sales interest.

[23] The main results are not very sensitive when we increase this threshold, but the number of advisers reduces substantially as we increase the threshold.

[24] Note that adviser- and firm-level variables refer to the data source they are constructed from. Adviser-level variables are mainly constructed from IAPD and BrokerCheck platforms, and firm-level variables are constructed from Form ADV filings of advisory firms. Adviser-level variables also include the characteristics of advisers' firms, such as employee count, though it is available for more advisers, given that it is constructed from IAPD and BrokerCheck.

2.2.2 Sample of reported misconduct

We observe the full set of disclosures for each financial adviser categorized into 23 categories by the SEC/NAASA and FINRA. Following Egan et al. (2019), we conservatively define misconduct as regulatory, internal investigations and customer events that were resolved against the adviser. The disclosure types that we classify as adviser misconduct are described in Table 2.2. In our empirical models, we categorize the remaining 17 disclosure categories in one group as other disclosures and include the count of other disclosures as a predictor of misconduct.[25]

Table 2.2. Definitions of misconduct categories
This table defines 6 of the 23 disclosure categories that we define as misconduct. The definitions of the remaining 17 disclosure categories are reported in Table A2.1 of the Appendix.

Misconduct category	Definition
Customer Dispute—Settled	This type of disclosure event involves a consumer-initiated, investment-related complaint, arbitration proceeding, or civil suit containing allegations of sales practice violations against the adviser that resulted in a monetary settlement to the customer.
Customer Dispute—Award/Judgment	This type of disclosure event involves a final, consumer-initiated, investment-related arbitration or civil suit containing allegations of sales practice violations against the adviser that resulted in an arbitration award or civil judgment for the customer.
Regulatory—Final	This type of disclosure event involves (1) a final, formal proceeding initiated by a regulatory authority (e.g., a state securities agency, self-regulatory organization, federal regulatory agency such as the SEC, foreign financial regulatory body) for a violation of investment-related rules or regulations or (2) a revocation or suspension of an adviser's authority to act as an attorney, accountant, or federal contractor.
Employment Separation After Allegations	This type of disclosure event involves a situation in which the adviser voluntarily resigned, was discharged, or was permitted to resign after being accused of (1) violating investment-related statutes, regulations, rules, or industry standards of conduct; (2) fraud or the wrongful taking of property; or (3) failure to supervise in connection with investment-related statutes, regulations, rules, or industry standards of conduct.
Criminal—Final Disposition	This type of disclosure event involves a criminal charge against the adviser that has resulted in a conviction, acquittal, dismissal, or plea. The criminal matter may pertain to any felony or certain misdemeanor offenses, including bribery, perjury, forgery, counterfeiting, extortion, fraud, and wrongful taking of property.
Civil—Final	This type of disclosure event involves (1) an injunction issued by a court in connection with investment-related activity, (2) a finding by a court of a violation of any investment-related statute or regulation, or (3) an action brought by a state or foreign financial regulatory authority that is dismissed by a court pursuant to a settlement agreement.

[25] The disclosure categories that are not misconduct are Civil—Pending, Civil—on Appeal, Civil Bond, Customer Dispute—Closed—No Action, Customer Dispute—Withdrawn, Customer Dispute—Dismissed Customer Dispute—Denied, Customer Dispute—Final, Customer Dispute—Pending, Judgment/Lien, Financial—Final, Financial Pending, Investigation, Criminal—on Appeal, Criminal—Pending Charge, Regulatory—on Appeal, Regulatory—Pending. Table A1 in the Online Appendix contains the description of each category obtained from https://brokercheck.finra.org/.

Table 2.3. The number and percentage of advisers with each misconduct and disclosure category
This table presents the number and percentage of advisers with at least one disclosure in each category as reported by FINRA, SEC, and NASAA in our final sample of 1,120,771 financial advisers. The first column shows the category's name, the second column shows the number of advisers with at least one instance of misconduct/disclosure, and the third column shows the percentage of advisers with misconduct/disclosure in each category. Panel A reports misconduct categories, and Panel B reports the remaining non-misconduct-related disclosure categories. The last row of each Panel represents the total number of advisers with at least one instance of misconduct or other disclosure category. The sum of advisers in the 6 misconduct categories and 17 other disclosure categories does not equate to the total, as one adviser may belong to multiple categories.

Panel A. Misconduct		
	Number of financial advisers	Percentage of financial advisers
Customer Dispute—Settled	38,275	3.415
Customer Dispute—Award/Judgment	7,107	0.634
Regulatory—Final	29,626	2.643
Employment Separation After Allegations	20,683	1.845
Criminal—Final Disposition	27,919	2.491
Civil—Final	1,375	0.123
Total	102,157	9.115
Panel B. Other disclosure		
	Number of financial advisers	Percentage of financial advisers
Civil—Pending	306	0.027
Civil—on Appeal	13	0.001
Civil Bond	239	0.021
Customer Dispute—Closed—No Action	9,619	0.858
Customer Dispute—Withdrawn	2,057	0.184
Customer Dispute—Dismissed	192	0.017
Customer Dispute—Denied	34,341	3.064
Customer Dispute—Final	263	0.024
Customer Dispute—Pending	4,400	0.393
Judgment/Lien	20,872	1.862
Financial—Final	33,539	2.993
Financial—Pending	3,396	0.303
Investigation	421	0.038
Criminal—on Appeal	32	0.003
Criminal—Pending Charge	809	0.072
Regulatory—on Appeal	70	0.006
Regulatory—Pending	312	0.028
Total	96,584	8.618

Table 2.3 reports the number and percentage of financial advisers with at least one misconduct record. Overall, there are 102,157 advisers with at least one misconduct record, representing 9.12% of all advisers in our sample. Within the six categories of misconduct, the three largest groups are advisers with Customer Dispute—Settled (38,275 advisers representing 3.42% of all advisers), followed by Regulatory-Final (29,626 advisers representing 2.64% of all advisers), and Criminal—Final Disposition (27,919 advisers representing 2.49% of all advisers). Table 2.3 also shows that 96,584 advisers have at

least one of the remaining 17 disclosure categories, representing 8.62% of all advisers, with the largest three categories being Customer Dispute—Denied (3.06%), followed by Financial—Final (2.99%), and Judgment/Lien (1.86%).

Table 2.3 shows that the sample of advisers with reported misconduct is a substantial proportion of the industry, as documented by Egan et al. (2019).[26] This is important for our analysis because a rich characterization of reported misconduct allows us to empirically estimate the full extent of adviser misconduct and the characteristics of advisers with a higher propensity for misconduct. In addition, the fact that the sample of advisers with misconduct is drawn from multiple categories of misconduct with diverse characteristics is likely to help the models to converge to the full extent of adviser misconduct and better characterize the misconduct and detection processes.

Figure 2.1 illustrates the number of misconduct-related disclosures, other disclosures (remaining 17 disclosure categories), and total disclosures (all 23 disclosure categories) over time and contrasts their dynamics with the number of active advisers each year during 1990-2020. We also calculate the annual misconduct and other disclosure rates during the same period.[27] The misconduct rate is the ratio of the number of misconduct to the number of active advisers each year. Similarly, the other disclosure rate is the ratio of the number of remaining 17 disclosures to the number of active advisers each year. The total disclosure rate is the sum of the two.

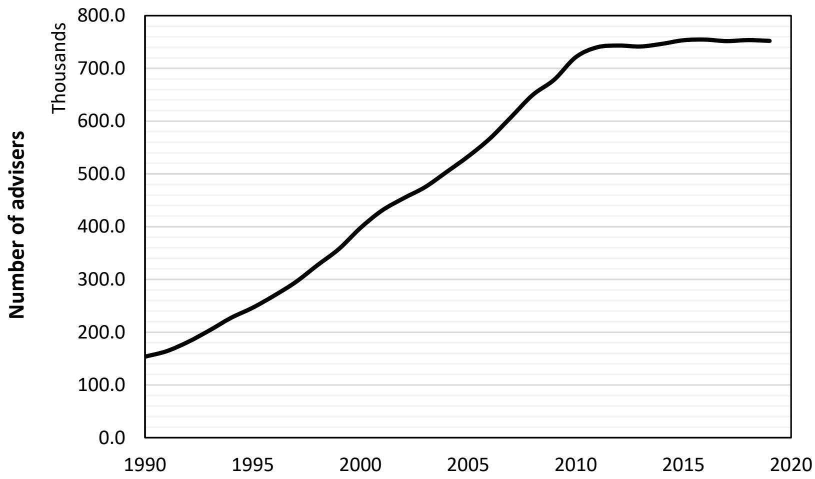
Panel A of Figure 2.1 shows that the number of advisers has been increasing since 1990, reaching more than 750,000 in 2020. The growth in the number of advisers is mainly due to the massive growth in the U.S. financial system during this period. Panel B shows that the number of adviser misconduct is pervasive during 1990-2020 and non-misconduct related disclosures has significantly increased during this period. The number of disclosures (both misconduct and non-misconduct-related) have two major spikes, one during the dot-com bubble (1998 to 2001) and the other during the GFC (2007 to 2010), suggesting that adviser misconduct increases during stressful market conditions.

[26] The percentage of reported misconduct is slightly higher in our sample compared to Egan et al. (2019) as our sample includes all advisers with available adviser-level characteristics until 2020. In contrast, Egan et al. (2019) analyze advisers during 2005-2015.

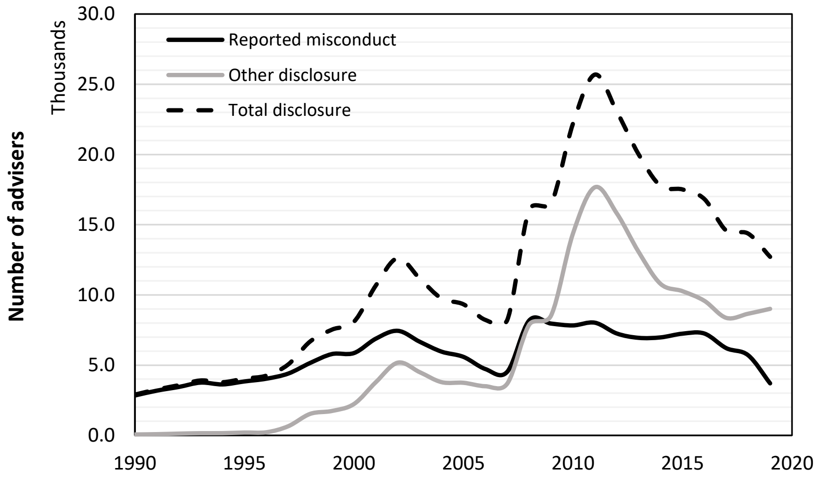
[27] We only examine the dynamics of the number of active advisers and misconducts and other disclosures from 1990 since the industry is not mature before that. For example, there were less than 150,000 advisers before 1990 and less than 50 advisers before 1950. Our sample goes back to 1900 when there were only 2 advisers. With fewer advisers, a record of misconduct increases the misconduct rate substantially (e.g., 2 of the 4 active advisers in 1908 have misconduct, making the misconduct rate 50%).

Figure 2.1. The time series of reported misconduct and non-misconduct-related disclosures
This figure illustrates the time series of the number of financial advisers (Panel A) and number of advisers with misconduct-related disclosures, non-misconduct-related disclosures, and total disclosures (Panel B), and that of Panel B in percentage terms (Panel C). The number of advisers in Panel A represents active advisers each year during 1990-2020. Panel B plots the number of advisers with misconduct (6 misconduct-related disclosure categories), other disclosure (17 non-misconduct-related disclosure categories), and total disclosure (all 23 disclosure categories) during 1990-2020. Panel C plots the misconduct rate (the ratio of the number of advisers with misconduct to the total number of active advisers each year), the other disclosure rate (the ratio of the number of advisers with other disclosures to the total number of active advisers each year), and the total disclosure rate (the sum of the misconduct and other disclosure rates).

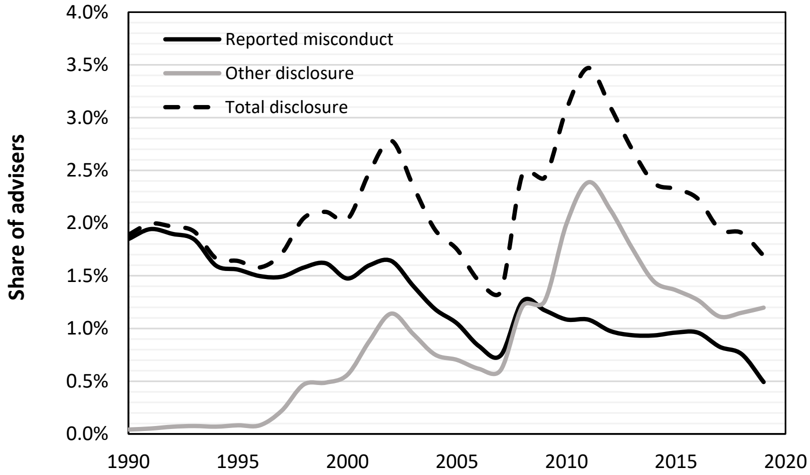
Panel A. The number of advisers over time



Panel B. The number of advisers with reported misconduct, other disclosure, and total disclosure over time



Panel C. The reported misconduct rate, other disclosure rate, and total disclosure rate over time



Panel C shows that the misconduct rate starts around 2% early in the sample and declines to 0.5% due to the increase in the number of advisers, and the total disclosure rate ranges between 1.5% and 3.5%. These results indicate that even reported adviser misconduct is pervasive across years and accounts for a sizeable proportion of the industry.

2.2.3 Descriptive statistics

Our adviser-level sample has 1,120,771 financial advisers with 14 adviser-level variables, and the firm-level sample has 753,833 advisers with additional 8 firm-level variables. Note that the unit of observation is an adviser in both samples. To see which adviser- and firm-level characteristics influence the advisers' propensity for misconduct, Table 2.4 divides the adviser-level and firm-level samples into advisers with and without reported misconduct and reports the descriptive statistics of the variables in both subsamples.[28]

We perform a univariate analysis and compare the means of variables in two subsamples. We evaluate the significance of the mean difference between the groups using paired t-tests. There is significant differences in the characteristics of the two groups. For example, advisers with misconduct tend to work in a more toxic environment (with a high office misconduct rate) and have more non-misconduct-related disclosures. Table 2.4 compares

[28] The full sample descriptive statistics of the model variables are close to that of the advisers without misconduct.

Table 2.4. Differences in characteristics of financial advisers with and without reported misconduct

This table reports the descriptive statistics of the model variables and differences in mean characteristics of advisers with and without reported misconduct. Panel A reports the descriptive statistics for the adviser-level variables, and Panel B reports for the firm-level variables. The adviser-level variables are available for 1,120,771 advisers, and firm-level variables are available for 753,833 of the 1,120,771 advisers. The last two columns in the table report the difference in mean characteristics of advisers with and without detected or reported misconduct. The significance of the difference in means is computed with t-statistics. The standard errors used to compute the t-statistics are double clustered by the years of experience and the state of an adviser. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. For the definitions and calculations of variables, refer to Table 2.1. The unit of observations is an adviser, and the units of variables are in parentheses following the variable name.

	Financial advisers with misconduct (1)			Financial advisers without misconduct (2)				
	Mean	Median	Std. dev.	Mean	Median	Std. dev.	Difference (1)-(2)	t-stat
Panel A. Adviser-level variables								
Office misconduct rate (%)	6.65	4.08	8.21	3.44	1.23	6.46	3.21***	3.21
Other disclosure count	0.50	0.00	1.39	0.11	0.00	0.53	0.39***	10.11
Economic recession	0.21	0.21	0.12	0.21	0.21	0.15	0.005	0.96
Firm count	4.21	4.00	3.11	2.61	2.00	2.20	1.60***	10.81
Employment duration (years)	6.27	4.65	5.67	5.78	4.11	5.37	0.48	1.43
Gender (dummy)	0.86	1.00	0.34	0.71	1.00	0.45	0.15***	39.00
Reported misconduct rate (%)	1.22	1.19	0.35	0.96	0.97	0.29	0.26***	9.36
Regulatory budget (%)	0.13	0.13	0.01	0.13	0.13	0.01	0.002***	5.17
Dodd-Frank switch (dummy)	0.02	0.00	0.13	0.01	0.00	0.10	0.01***	9.21
Tenure (years)	20.98	21.00	12.79	13.55	11.00	10.53	7.43***	7.80
Exam count	3.93	4.00	1.68	3.39	3.00	1.46	0.54***	8.51
Employee count	6,919	4,033	7,610	8,011	4,452	8,948	-1,092***	-7.07
Active financial adviser (dummy)	0.46	0.00	0.50	0.58	1.00	0.49	-0.12***	-6.38
Dual registration (dummy)	0.50	0.00	0.50	0.34	0.00	0.47	0.16***	11.55
N obs.	102,157			1,018,614				
Panel B. Firm-level variables								
Portion of AUM fees (dummy)	0.93	1.00	0.21	0.89	1.00	0.28	0.04***	5.33
Per hour fees (dummy)	0.50	0.52	0.41	0.39	0.11	0.44	0.11***	4.61
Fixed fees (dummy)	0.74	1.00	0.37	0.65	1.00	0.43	0.09***	7.96
Total AUM ($ million)	205,658	70,132	264,924	178,878	32,924	261,809	26,781***	4.57
Discretionary AUM ($ million)	101,874	45,196	120,097	94,154	21,117	125,789	7,719***	3.38
Custody AUM ($ million)	152,686	22,581	240,332	127,534	8,761	229,625	25,152***	5.40
Number of clients	619,878	228,071	792,749	570,591	81,271	857,229	49,281**	2.05
Sales interest (dummy)	0.75	1.00	0.36	0.73	1.00	0.40	0.02	1.03
N obs.	64,107			689,726				

other characteristics of advisers with and without misconduct. Note that the univariate tests are based on the reported misconduct and do not consider the share of unreported misconduct.

2.3 Empirical model of financial adviser misconduct and detection

2.3.1 Incomplete detection problem

The analysis of reported adviser misconduct without accounting for the non-random regulatory detection process can lead to substantial biases in inferences about the frequency and the determinants of adviser misconduct. This problem is overlooked or inadequately addressed in much of the recently growing financial adviser empirical literature. The foundation of our empirical approach is based on the premise that many instances of misconduct remain undetected (i.e., incomplete detection problem). The incomplete detection problem jeopardizes any inferences drawn from reported misconduct, since the data on reported misconduct is not representative of all misconduct.

The econometric problems caused by incomplete detection are documented by Poirier (1980) and Feinstein (1987, 1990). Feinstein (1990) develops a detection-controlled estimation (DCE) method that allows inference about undetected misconduct. Various versions of the DCE method are applied to quantify the extent of misconduct in different contexts, including income tax evasion (Feinstein 1991), corporate fraud (Wang, Winton, and Yu 2010), market manipulation (Comerton-Forde and Putnins 2014), and illegal bitcoin transactions (Foley, Karlsen and Putnins 2019).

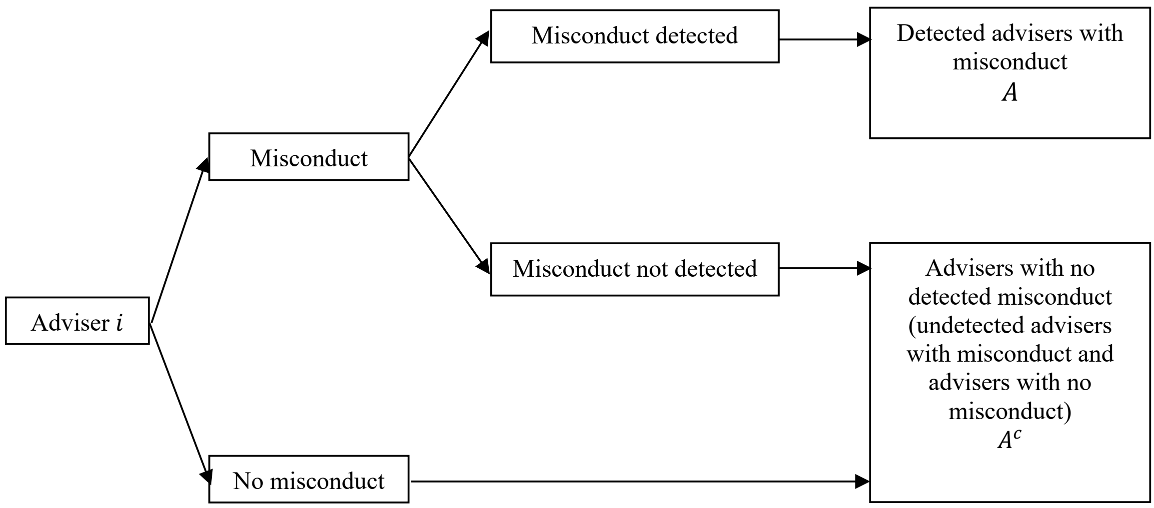
2.3.2 Detection-controlled estimation

We follow Feinstein (1990) and the subsequent applications of the DCE to infer the unreported/undetected share of adviser misconduct. The DCE model jointly estimates the adviser misconduct and detection processes. Intuitively, the model exploits the differences in the characteristics of advisers with and without reported misconduct to probabilistically identify the true population of advisers who are involved in misconduct.

If we had a random sample of advisers with and without misconduct and a set of characteristics that differ between advisers with and without misconduct, this task would be relatively simple and could be achieved with standard techniques such as a probit or logit model. However, a complication arises because detection by regulators is not random, and this nonrandomness must be accounted for to obtain unbiased estimators.

The DCE model is more similar to Heckman-style selection bias correction models because both explicitly model the process that causes the sample to be a non-random subset of the population. However, in Heckman-style models, the outcome of the selection process is observed (e.g., nonrespondents in a survey). In contrast, in the incomplete detection problem, the outcome that some misconduct is undetected is not observed, further complicating the problem. Therefore, the classic Heckman models to deal with sample selection bias cannot be applied, and the DCE approach is a more suitable method to estimate the probability of misconduct and detection.

Figure 2.2. Detection-controlled estimation model
The figure illustrates the structure of the two-stage detection-controlled estimation (DCE) model. Stage 1 models how advisers with and without misconduct differ in characteristics. Stage 2 models the determinants of the probability that an adviser with misconduct was detected (reported). Both stages are simultaneously estimated using maximum likelihood to select parameter values that maximize the likelihood of the observable adviser classifications, A and A^c.



More formally, the model is based on two simultaneously estimated equations: one modeling misconduct and the other modeling detection conditional on misconduct has occurred. Figure 2.2 illustrates our baseline DCE model.

In the figure, a full sample of advisers falls into two observed disjoint sets, A and A^c. Set A consists of advisers who have engaged in misconduct and subsequently detected by regulators. Set A^c consists of two groups of advisers: (i) advisers who have either not engaged in misconduct and (ii) advisers who have engaged in misconduct but have not been detected or reported by regulators.

We model the propensity of adviser i to engage in misconduct as a continuous latent variable, Y_{1i}^*, that is a function of adviser-specific attributes (e.g., tenure, employment duration, firm count, other disclosure counts), X_{1i}, i.e.,

$$Y_{1i}^* = X_{1i} \cdot \beta_1 + \varepsilon_{1i}, \tag{2.1}$$

$$Y_{1i} = \begin{cases} 1 \text{ (misconduct)} & \text{if} \quad Y_{1i}^* > 0, \\ 0 \text{ (no misconduct)} & \text{if} \quad Y_{1i}^* \leq 0. \end{cases} \tag{2.2}$$

Y_{1i} is the binary variable that takes the value one if the adviser i has engaged in misconduct and zero otherwise. Y_{1i} is not directly observable. Instead, we observe advisers who have engaged in misconduct and are detected by authorities. Similarly, the detection process is modeled as a continuous latent variable, Y_{2i}^*, that is also a function of adviser-specific attributes (e.g., reported misconduct rate, regulatory budget), X_{2i}, i.e.,

$$Y_{2i}^* = X_{2i} \cdot \beta_2 + \varepsilon_{2i}, \tag{2.3}$$

$$Y_{2i} = \begin{cases} 1 \text{ (detected)} & \text{if} \quad Y_{2i}^* > 0, \\ 0 \text{ (not detected)} & \text{if} \quad Y_{2i}^* \leq 0. \end{cases} \tag{2.4}$$

We define monotonic link functions $M(\cdot)$ and $D(\cdot)$ that map $X_{1i}\beta_1$ and $X_{2i}\beta_2$ to probability of misconduct and detection, respectively.[29] The log-likelihood of observations is the sum of log likelihood of advisers with reported misconduct (A) and advisers with unreported misconduct (A^c), i.e.,

$$L = \sum_{i \in A} \log(M \cdot D) + \sum_{i \in A^c} \log(1 - M \cdot D). \tag{2.5}$$

[29] We use logistic functions for link functions $M(X_{1i}\beta_1) = \frac{1}{1+\exp(X_{1i}\beta_1)}$ and $D(X_{2i}\beta_2) = \frac{1}{1+\exp(X_{2i}\beta_2)}$.

Note that the second term in Eq. (2.5) is the log of the sum (over advisers in A^c) of the joint probability that an adviser is clean $(1-M)$ and an adviser with misconduct is not reported $(M \cdot (1-D))$. We select parameter vectors β_1 and β_2 such that Eq. (2.5) is maximized. From the estimates, we compute each adviser's probability of being involved in misconduct and construct a binary classification of advisers with and without misconduct.

2.3.3 Model specifications

We apply DCE to the adviser-level sample (Model 1) and firm-level sample (Model 2) by jointly modeling adviser misconduct and misconduct detection processes as in Eqs. (2.1)-(2.4). With two joint adviser misconduct and regulatory detection processes, we rely on instrumental variables that affect one process but not the other for identification. The instrumental variables with strong exclusion restrictions that go into either misconduct or detection (not both) equations are critical in robustly identifying the model.[30] Table 2.1 defines variables used in both models and the equation they appear in the model with our ex-ante belief about the sign of the variable's impact on the probability of misconduct (M) and detection (D).

Model 1 is based on the adviser-level sample with 1,120,771 advisers, where 11 variables appear in the misconduct equation (Eq. 2.1) and 8 variables in the detection equation (Eq. 2.3). The variables that only impact adviser misconduct but not detection are office misconduct rate, other disclosure count, economic recession, firm count, employment duration, and gender.

1. We include the office misconduct rate in the misconduct equation following Dimmock et al. (2018), who show that misconduct is contagious — coworkers of financial advisers influence their propensity to commit misconduct. Conceptually, this can occur because advisers can learn about the profitability and techniques of misconduct from their peers (e.g., Sah 1991) or alternatively, advisers may want to

[30] Identification of the model without instruments is still possible, relying on functional forms and distributional assumptions. We take a more robust instrumental variables approach with strong exclusion restrictions. In robustness tests, we examine the sensitivity of the results to a given instrument by relaxing the assumed exclusion restrictions one at a time. The results are not particularly sensitive to any individual instrumental variable's exclusion restriction.

mimic their coworkers' behavior due to concerns about their relative position (e.g., DeMarzo, Kaniel, and Kremer 2004).

2. We include other disclosure count in the misconduct equation following Qureshi and Sokobin (2015), who find that an adviser's disclosure is a strong predictor of investor harm. Thus, we hypothesize that non-misconduct-related disclosure count is positively associated with the probability of misconduct.

3. We include economic recession in the misconduct equation because a negative shock to advisers' wealth can increase their propensity for misconduct (e.g., Dimmock et al. 2021). Alternatively, misconduct can function as a form of risky borrowing (e.g., Andreoni 1992), with advisers possibly comparing conventional borrowing costs with misconduct penalties, akin to the classical Becker (1968) model. The increase in the costs of conventional borrowing during economic recessions can increase the relative attractiveness of borrowing via misconduct, resulting in higher adviser misconduct.

4. The intuition of including firm count in the misconduct equation is that advisers with misconduct are more likely to switch between firms, and repeat offenders are more likely to switch between multiple firms. Therefore, an adviser is more likely to engage in misconduct if the firm count is higher. We do not include the firm count in the detection equation because it is not apparent why regulators would target advisers with more or less firm count. We could not find any information from the SEC or FINRA that negates this intuition.

5. The intuition of including employment duration in the misconduct equation is that employers are more likely to end contracts after misconduct, leading to shorter employment. Thus, we expect a negative association between an adviser's employment duration and the probability of misconduct.

6. We include gender in the misconduct equation following Egan et al. (2022), who find male advisers are more likely to engage in misconduct. Intuitively, the regulatory detection of misconduct should not depend on the gender of an adviser because it is not so obvious why regulators would do such discrimination.

The instrumental variables that primarily impact detection but not misconduct are reported misconduct rate, regulatory budget, and Dodd-Frank switch.

1. The reported misconduct rate measures the effectiveness of the regulators in detecting adviser misconduct during the employment period of an adviser. Therefore, we expect it to be positively related to the probability of detection.

2. The regulatory budget (number of inspectors per adviser) measures the resources available at the SEC to conduct potential misconduct investigations during the employment period of an adviser. We expect larger regulatory budgets are likely associated with a greater capacity to prosecute advisers and a higher probability of detection.

3. We include Dodd-Frank switch in the detection equation following Charoenwong et al. (2019) who show that the transition of "midsize" advisory firms ($25 million to $100 million in AUM) from the SEC to state regulation with the implementation of the Dodd-Frank Act in 2012 increased client complaints of those switching to state regulation.[31] Therefore, we expect the switch to be positively associated with the probability of detection.

The control variables in both misconduct and detection equations of the model are tenure, exam count, employee count, active financial adviser, and dual registration.

1. We expect more experienced advisers will respect their duties and legal requirements, leading to a negative association between tenure and the probability of misconduct. The impact of tenure on the probability of detection is unclear. On the one hand, more time in the industry gives regulators more time to detect if any misconduct occurs, increasing the probability of detection. On the other hand, experienced advisers may be better at concealing their tracks, reducing the probability of detection. Since it is difficult to predict ex ante which effect dominates, we leave this prediction to the data.

2. The employee count proxying the firm size impacts the probability of misconduct because advisers working for large firms have a large client base and client funds

[31] The purpose of the switch was free up regulatory resources at the SEC to regulate "private funds" such as hedge funds, which were newly enrolled under SEC oversight by the Act. See, for example, SEC Investor Bulletion on the "Transition of Mid-Sized Investment Advisers from Federal to State Registration".

(sometimes with direct access to those funds), potentially leading to more misconduct. The firm size also impacts the probability of detection because, in general, different regulatory procedures apply for different firm sizes.

3. We include the exam count in the misconduct equation following Kowaleski et al. (2020), who show the critical role of qualification exams in predicting adviser misconduct. We also include the exam count in the detection equation because regulators conduct these exams to screen advisers. We expect an adviser's exam count to be negatively associated with the probability of misconduct and positively associated with the probability of detection as reducing misconduct and increasing detection is one of the main purposes of examination.

4. Advisers dual registered as an investment adviser with the SEC and a broker with FINRA offer brokerage and advisory services to the same client with sometimes conflicting interests and cross-agency transactions, potentially increasing the probability of misconduct. Dual registration is also likely to increase detection because of more scrutiny from the SEC and FINRA.

5. Active advisers are less likely to do misconduct because the current regulatory framework is more transparent and stringent. We expect inactive advisers to be detected more because misconduct detection generally leads an adviser to cease operations (either loss of license for a certain period or bar from the industry).

In Model 2, we add eight firm-level variables to Model 1. We include three dummy variables for firms' fee structure (portion of AUM, per-hour fees, and fixed fees) only in the misconduct equation because some fee types can create more perverse incentives than others. We do not have strong priors about the impact of fee types on the probability of misconduct.

We add the remaining five firm-level variables in both misconduct and detection equations. We include total AUM and number of clients representing the firm size in both equations with the same reasoning as the employee count in Model 1. We include discretionary AUM, custody AUM and sales interest in the probability of detection because direct access to client funds with a potential conflict of interest indicates high risk, and regulators prioritize inspections of high-risk companies.[32] We also include them in the misconduct equation as control variables. On the one hand, they may increase the probability of misconduct because advisers with discretionary funds or funds under custody can steer the funds, unlike advisers who do not have direct access to funds. On the other hand, they may decrease the probability of misconduct because advisers who have direct access to client funds or have sales interest in clients' transactions are either clean or are aware that they are more likely to be inspected by regulators. All variables before entering the model are standardized to have mean zero and standard deviation of one and winsorized at +/- 3 standard deviations to reduce the influence of extreme values.

2.4 Results

2.4.1 The extent of adviser misconduct

The models arrive at probabilistic classifications of advisers into those primarily engaged in misconduct and those with a clean sheet. Once we group advisers with and without misconduct, we use the size of the two groups to measure the extent of adviser misconduct. Table 2.5 presents the results.

Panel A presents the estimated size of the groups in terms of the number of advisers. We present the number of advisers with reported misconduct, and the estimated number of advisers with misconduct based on the DCE model, $\Pr(Y_1 = 1)$. We also present the number of advisers with unreported misconduct, $\Pr(Y_1 = 1|Y_2 = 0)$, and the number of advisers with reported misconduct but estimated not to be involved in misconduct based

[32] The Regulation of Investment Advisers by the U.S. SEC document notes that: "The SEC staff utilizes a risk-based process, identifying higher-risk investment advisers for examination consideration and focusing examination resources on certain higher-risk activities at selected investment advisers, including conflicts of interest, portfolio management, valuation, performance, advertising and asset verification" (retrieved from https://www.sec.gov). For example, the OCIE department of the SEC began to conduct enhanced asset verification procedures for advisers with custody of client assets starting from 2009.

on the DCE, $\Pr(Y_1 = 0|Y_2 = 1)$.[33] Panel B re-expresses these values as percentages of the total number of advisers.

Table 2.5. Estimated misconduct frequency
This table presents the frequency of adviser misconduct as estimated by the detection-controlled estimation (DCE) model, applied to both the adviser-level sample with 1,120,771 advisers (Model 1), and the firm-level sample with 753,833 advisers (Model 2). Panel A presents the number of advisers with reported misconduct, the number of advisers with misconduct based on two DCE models (estimated misconduct), the number of advisers with unreported misconduct that are estimated to be involved in misconduct based on the DCE models (estimated misconduct when unreported misconduct), and the number of advisers with reported misconduct that are classified as advisers with no misconduct based on the DCE models (no estimated misconduct when reported misconduct). Panel B reports that of Panel A in terms of percentages of total advisers.

	Model 1	Model 2
Panel A. Number of advisers		
Reported misconduct	102,157	64,107
Estimated misconduct, $\Pr(Y_1 = 1)$	334,660	237,392
Estimated misconduct when unreported misconduct, $\Pr(Y_1 = 1\|Y_2 = 0)$	279,456	205,540
No estimated misconduct when reported misconduct, $\Pr(Y_1 = 0\|Y_2 = 1)$	46,953	32,255
Panel B. Percentage of advisers		
Reported misconduct	9.115	8.504
Estimated misconduct, $\Pr(Y_1 = 1)$	29.860	31.491
Estimated misconduct when unreported misconduct, $\Pr(Y_1 = 1\|Y_2 = 0)$	24.934	27.266
No estimated misconduct when reported misconduct, $\Pr(Y_1 = 0\|Y_2 = 1)$	4.189	4.279

The number of advisers estimated to be involved in misconduct (334,660 in Model 1 and 237,392 in Model 2) is around three times more than the number of advisers with reported misconduct records (102,157 in the adviser-level sample and 64,107 in the firm-level sample). This estimate is similar across models (3.3 times in Model 1 and 3.7 times in Model 2). In percentage terms, the models estimate that roughly 30% (29.9% in Model 1 and 31.5% in Model 2) of advisers in the U.S. are involved in misconduct. Table 2.5 also reports that the share of unreported $\Pr(Y_1 = 1|Y_2 = 0)$ and incorrectly reported misconduct $\Pr(Y_1 = 0|Y_2 = 1)$ are around 25% and 4% of advisers, respectively.

[33] The classification of advisers into misconduct and non-misconduct groups can result in advisers with reported misconduct being reclassified in the non-misconduct group. For example, suppose an adviser was involved in misconduct but is mainly a good adviser. Such an adviser will have characteristics similar to those in the non-misconduct group, leading to a classification by the DCE model into the non-misconduct group. In contrast, an adviser who is predominantly involved in misconduct, even if not detected by authorities, is likely to have characteristics similar to those in the misconduct category. Therefore, after controlling for the differences in characteristics due to nonrandom detection, such an adviser is likely to be classified in the misconduct category by the model.

Out of the 334,660 advisers flagged for potential misconduct in Model 1, 220,256 are investment advisers, and 327,851 are brokers (213,447 are dual). The higher likelihood of misconduct among brokers compared to investment advisers may stem from the fact that brokers are subject to a less stringent suitability standard by FINRA, whereas investment advisers adhere to a fiduciary standard mandated by the SEC.[34] Among the 220,256 investment advisers with estimated misconduct, 138,288 are presently active, collectively managing around \$6.9 trillion in AUM as calculated by the firm AUM divided by the number of investment advisers in the firm in 2020. In contrast, 34,211 current investment advisers with reported misconduct collectively manage \$1.5 trillion. Among the 327,851 brokers with estimated misconduct, 165,976 are also currently active, executing transactions in securities for the account of others.

From these findings, a natural question is why does a significant amount of adviser misconduct go unreported? One explanation is that regulators have resource constraints, and therefore target inspecting firms that are more likely to do misconduct and prioritize penalizing high-profile cases. It's not necessary for the estimated misconduct cases to be as large-scale as schemes like Bernard Madoff's or Jordan Belfort's. For example, FINRA requires that brokers have a reasonable basis to believe a recommended transaction or investment strategy is suitable for the customer (suitability standard). It seems quite plausible that many brokers sacrifice basic suitability standards for customers in pursuit of additional gains.

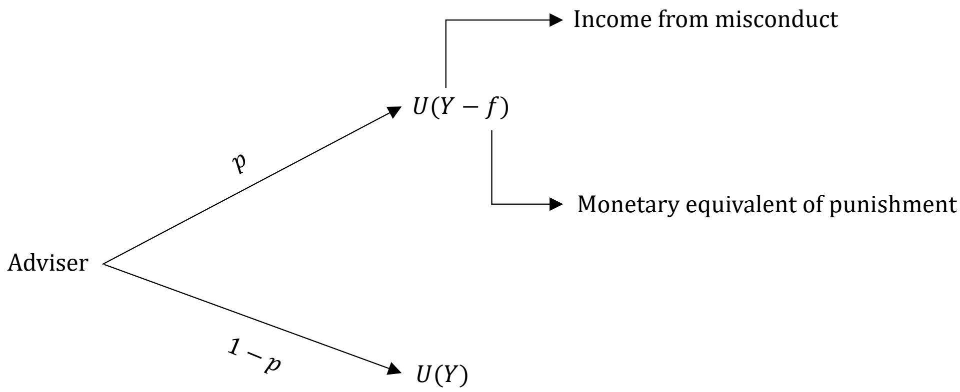
Another related explanation is that there is a societal cost associated with combating misconduct. From an adviser's perspective, the expected utility of engaging in misconduct is

$$E[U] = p \cdot U(Y - f) + (1 - p) \cdot U(Y), \tag{2.6}$$

where Y represents the income from misconduct, p is the probability of being caught, and f is the monetary equivalent of the punishment, as illustrated in Figure 2.3. Drawing on the classic models of Becker (1968) and Ehrlich (1973), if the primary goal of the regulator is deterrence, the probability of being caught (p) could be increased to nearly 1, and punishments (f) could be increased such that the number of misconduct cases reduces to 0. However, an increase in p or f also increases the social cost of misconduct through their effects on the cost of combating misconduct and the cost of punishments.

[34] In robustness tests, we estimate the model for brokers and investment advisers separately.

Figure 2.3. Expected utility of misconduct
The figure illustrates the expected utility of misconduct from an adviser's perspective. p represents the probability of being caught, Y is the income from misconduct, f is the monetary equivalent of the punishment, and U is the utility function of the adviser.



Therefore, in equilibrium, there exists an optimal level of p and f where the marginal cost of detecting/reporting misconduct equals the marginal benefit. This equilibrium maintains a balance between deterring misconduct and minimizing the overall societal cost associated with combating and penalizing misconduct, and thus always features unreported misconduct.

Two important conclusions follow from our findings. First, without empirical methods, the size of misconduct in the financial advisory industry that can be inferred from reported misconduct vastly underestimates the full extent of misconduct, highlighting the importance of extrapolating the full scale of adviser misconduct using statistical methods. Second, misconduct accounts for a sizeable proportion of the financial advisory industry, suggesting a rationale for why the public distrusts Wall Street and the financial services industry. Unfortunately, finance professionals often overlook this erosion of trust and lack of faith in traditional financial institutions. Our results strongly support the validity of such distrust.

2.4.2 The determinants of adviser misconduct and detection

We now assess the economic determinants of adviser misconduct and detection by exploiting the estimated coefficients of the models. Table 2.6 reports the marginal effects and statistical significance (in parentheses) of determinants of adviser misconduct and detection.

Note that all determinants are standardized with mean zero and standard deviation of one for easier comparison. Therefore, marginal effects measure the percentage change in the probability of misconduct or detection for a one standard deviation change in the value of a given variable, indicating that variable's relative importance.[35]

The signs of coefficients across models are consistent with each other. Advisers who are involved in misconduct tend to have a high office misconduct rate, a high non-misconduct-related disclosure count, work more during economic recessions, work for more firms, have a lower average employment duration in each firm, and are more likely to be male. The signs of these instrumental variables that impact the probability of misconduct are consistent with the hypothesized signs discussed in Section 2.3.3 and reported in Table 2.1. Additionally, they tend to have a shorter tenure and pass fewer industry exams. Advisers working for large firms (in terms of employee count), those who are inactive, and those who are dual-registered with the SEC and FINRA are also more likely to be involved in misconduct.

Table 2.6 provides the magnitude (relative importance) of the effects. The magnitudes generally indicate that most of the determinants are economically significant. In particular, other disclosure count and office misconduct rate, followed by gender, appear to be the most critical variables in determining the probability of misconduct. A one standard deviation increase in other disclosure count and office misconduct rate increases the probability of misconduct by a factor of 0.557 (55.7% of what the probability of misconduct would otherwise be) and 0.488, respectively.

Model 2 adds to these findings by showing that advisers who are involved in misconduct

[35] The marginal effects follow from the partial derivatives of the probability of misconduct (M) and detection (D) with respect to each of the characteristics, scaled by the estimated corresponding probability, i.e., $\frac{\partial M/\partial x}{M} = \frac{\beta_{1i} \cdot e^{-Y_1^*}}{(1+e^{-Y_1^*})^2 \cdot M}$ and $\frac{\partial D/\partial x}{D} = \frac{\beta_{2i} \cdot e^{-Y_2^*}}{(1+e^{-Y_2^*})^2 \cdot D}$. The former equation is the marginal effect of variable x on the probability of misconduct, and the latter is the marginal effect of variable x on the probability of detection. We calculate marginal effects for each observation and then average over the entire sample. As an example of how to interpret the marginal effect of 0.488 for office misconduct rate, if an adviser's probability of misconduct is, say, 0.3, the effect of a 1-standard-deviation increase in the office misconduct rate, holding all else constant, is to increase the probability of misconduct to $0.3 \cdot 1.488 = 0.446$, an increase of 48.8% of what the probability would otherwise be. We report the coefficient estimates of the models in the Appendix.

Table 2.6. Model estimates

This table reports the results of two detection-controlled estimation (DCE) models. Model 1 (resp. Model 2) is a two-equation DCE model applied to the adviser-level sample with 1,120,771 advisers (resp. firm-level sample with 753,833 advisers). The numbers not in parentheses are the marginal effects, and the numbers in parentheses are the corresponding t-statistics. Marginal effects are computed as the partial derivatives of the probability of misconduct (M) and detection (D) with respect to each variable, scaled by the estimated corresponding probability. Marginal effects are computed for each observation and averaged over the entire sample. Pseudo R^2 is McFadden's likelihood ratio index (one minus the ratio of the log-likelihood with all predictors and the log-likelihood with intercepts only). Significance at the 10%, 5%, and 1% levels is indicated by *, **, and ***, respectively.

	Model 1		Model 2	
	M	D	M	D
Panel A. Adviser-level variables				
Office misconduct rate	0.488***		0.323***	
	(64.235)		(34.997)	
Other disclosure count	0.557***		0.507***	
	(57.772)		(47.054)	
Economic recession	0.033***		0.036***	
	(7.344)		(6.538)	
Firm count	0.078***		0.068***	
	(16.932)		(10.505)	
Employment duration	-0.041***		-0.092***	
	(-8.907)		(-15.331)	
Gender	0.278***		0.275***	
	(72.633)		(58.847)	
Reported misconduct rate		0.878***		0.861***
		(100.027)		(72.824)
Regulatory budget		0.328***		0.286***
		(87.925)		(58.042)
Dodd-Frank switch		0.039***		0.027***
		(13.463)		(8.116)
Tenure	-0.205***	0.009	-0.197***	0.140***
	(-30.165)	(0.946)	(-20.401)	(10.516)
Exam count	-0.042***	0.040***	-0.026***	0.028**
	(-6.800)	(5.557)	(-2.869)	(2.591)
Employee count	0.088***	-0.013	0.070***	0.092***
	(9.074)	(-1.618)	(4.725)	(6.484)
Active financial adviser	-0.072***	-0.011	-0.025***	-0.075***
	(-10.650)	(-1.303)	(-2.956)	(-6.774)
Dual registration	0.136***	0.043***	0.111***	0.075***
	(21.285)	(6.500)	(12.304)	(7.658)
Panel B. Firm-level variables				
Portion of AUM fees			0.050***	
			(10.400)	
Per hour fees			0.128***	
			(21.617)	
Fixed fees			-0.046***	
			(-7.111)	
Total AUM			0.215***	-0.184***
			(4.077)	(-4.009)
Discretionary AUM			-0.254***	0.146***
			(-8.621)	(5.496)
Custody AUM			0.009	0.006
			(0.348)	(0.258)
Number of clients			0.149***	-0.121***
			(6.589)	(-6.824)
Sales interest			-0.065***	0.014
			(-6.224)	(1.222)
Pseudo R^2	17.51%		15.20%	
N obs.	1,120,771		753,833	

tend to work for firms that charge fees based on a portion of AUM or per hour, rather than fixed fees. They tend to work for larger firms (in terms of AUM and number of clients), but with lower discretionary AUM and sales interest in client transactions. Firms with higher discretionary AUM and sales interest in client transactions either employ good advisers or are aware that they are likely to be inspected by regulators, decreasing the probability of misconduct. In contrast, custody AUM of firms does not significantly impact advisers' probability of misconduct.

Table 2.6 also presents the determinants of misconduct detection. The reported misconduct rate and regulatory budget are critical instrumental variables impacting the probability of detection. A one standard deviation increase in the reported misconduct rate and regulatory budget increases the detection probability by around 87.8% and 32.8%, respectively. The Dodd-Frank switch also positively impacts the probability of detection. Moreover, advisers with a longer tenure and more professional industry exams are more likely to be detected. That means advisers engaged in misconduct eventually get caught when they stay in the industry longer and participate in more exams. Detection probability is also higher for inactive and dual-registered advisers. Again, these effects are consistent with the hypothesized signs since misconduct detection generally leads an adviser to cease operations and dual registration means more scrutiny from the SEC and FINRA.

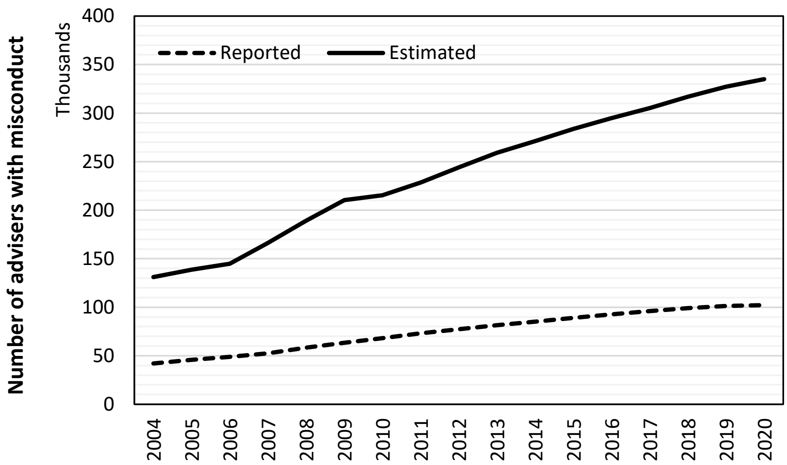
In terms of firm characteristics, advisers working for smaller firms (lower total AUM and number of clients) and higher discretionary AUM and sales interest in client transactions are more likely to be detected. These results indicate regulators tilt their focus toward inspecting smaller firms with potential red flags, such as discretionary AUM and sales interest in client transactions.

2.4.3 How does adviser misconduct vary through time?

To estimate how adviser misconduct varies through time, we re-construct the adviser-level variables (except gender and exam count) for 13,868,156 adviser-year from 2004 to 2020 for 1,120,771 adviser, starting from their industry entry year. Gender and exam count are static and thus filled forward after an adviser joins the industry. We estimate the number of advisers involved in misconduct for each year in the sample using the baseline model coefficients (reported in the Appendix Table A2.2) and yearly adviser characteristics. There is an interesting time-series variation in the number/share and AUM of advisers with misconduct. Figures 2.4 to 2.6 compare the estimated and reported number/share and AUM of advisers with misconduct from 2004 to 2020.[36]

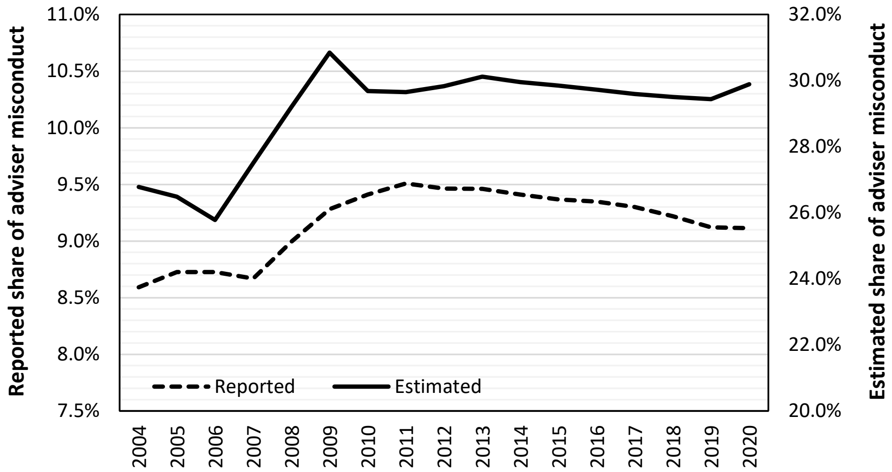
Figure 2.4. Estimated and reported adviser misconduct over time
This figure illustrates the yearly number and share of advisers with estimated misconduct during 2004-2020 in Panel A and B, respectively and compare them to advisers with reported misconduct during the same period. In Panel B, the reported share of advisers with misconduct is plotted with the dashed line using the left-hand-side axis, and the estimated share of advisers with misconduct is plotted with the solid line using the right-hand-side axis. Panel C plots the number of advisers with unreported/undetected misconduct (the difference between estimated and reported misconduct). To estimate the number of advisers involved in misconduct over time, we re-construct the adviser-level variables for each adviser-year during 2004-2020 and use the baseline model coefficients to calculate the probability of misconduct for each adviser-year. The sample includes 13,868,156 adviser-year observations for 1,120,771 advisers. Advisers with a probability of misconduct greater than 0.5 are identified as engaging in misconduct.

Panel A. Yearly number of advisers with estimated and reported misconduct

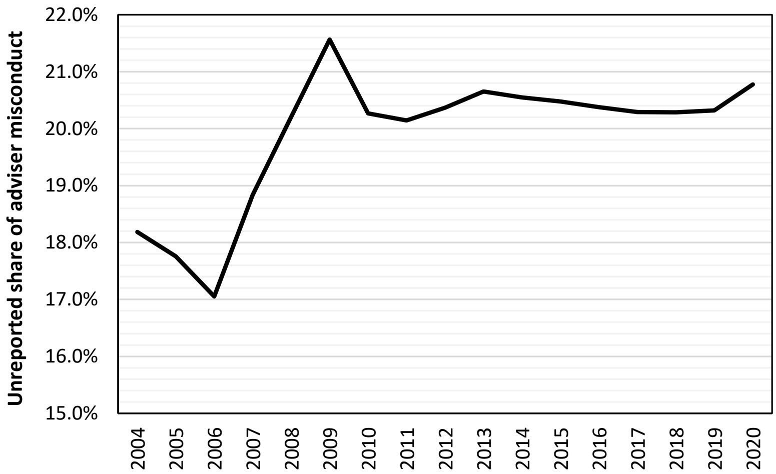


[36] Alternatively, we can track 334,660 advisers with estimated misconduct from the baseline model over time, starting from their industry entry date. While this approach does not reveal the year in which an adviser is estimated to have committed misconduct, it generates similar patterns. Appendix Figures A2.1-A2.3 present the plots using this approach.

Panel B. Yearly share of advisers with estimated and reported misconduct



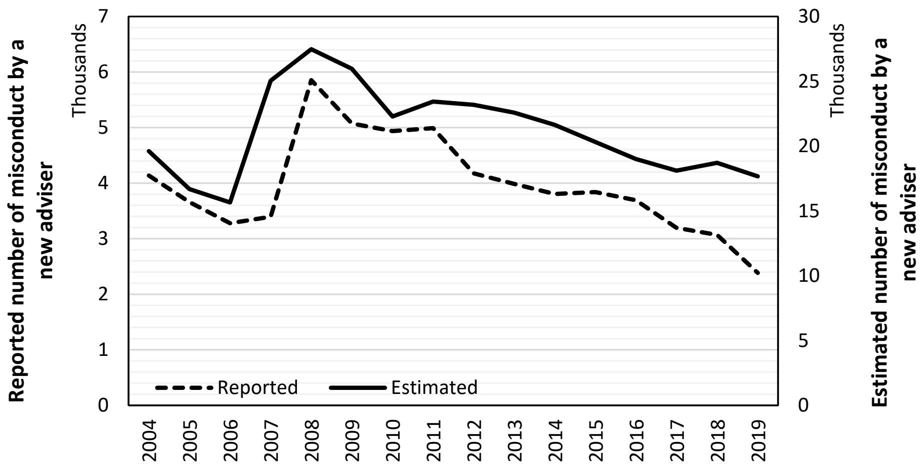
Panel C. Yearly share of advisers with unreported misconduct



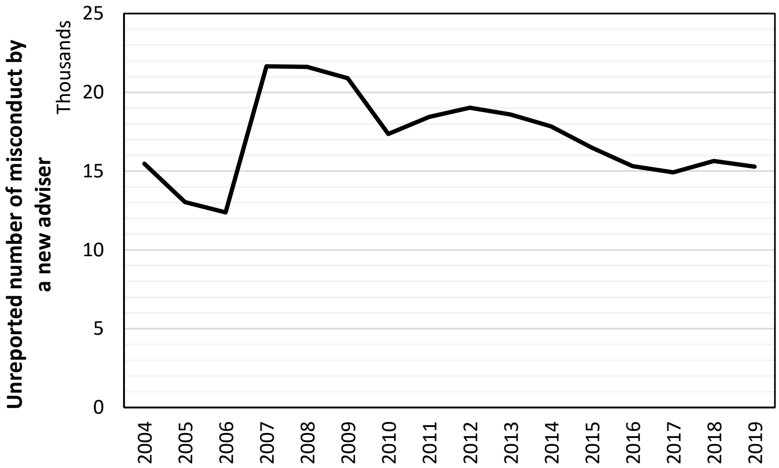
A pattern observed in Figure 2.4 is that the number of advisers estimated to be involved in misconduct (as reported misconduct) has increased over time, reaching 334,660 advisers in 2020 (Panel A). The percentage of advisers with (reported and estimated) misconduct experienced a sharp spike during the GFC and subsequently remained around the same level (Panel B). Particularly important is the larger increase in estimated misconduct compared to reported misconduct, implying that the share of unreported misconduct also increased during the GFC (Panel C).

Figure 2.5. Estimated and reported misconduct by a new adviser over time
This figure illustrates the estimated and reported number of misconduct done by a new adviser (with no previous misconduct record) during 2004-2019 in Panel A. In Panel A, the reported number of misconduct by a new adviser is plotted with the dashed line using the left-hand-side axis, and the estimated number of misconduct by a new adviser is plotted with the solid line using the right-hand-side axis. Panel B plots the difference between the estimated and reported number of misconduct by a new adviser. To estimate the number of advisers involved in misconduct over time, we re-construct the adviser-level variables for each adviser-year during 2004-2020 and use the baseline model coefficients to calculate the probability of misconduct for each adviser and year. Advisers with a probability of misconduct greater than 0.5 are identified as engaging in misconduct. We count the number of advisers with first misconduct each year during 2004-2019.

Panel A. Yearly number of advisers with estimated and reported first misconduct

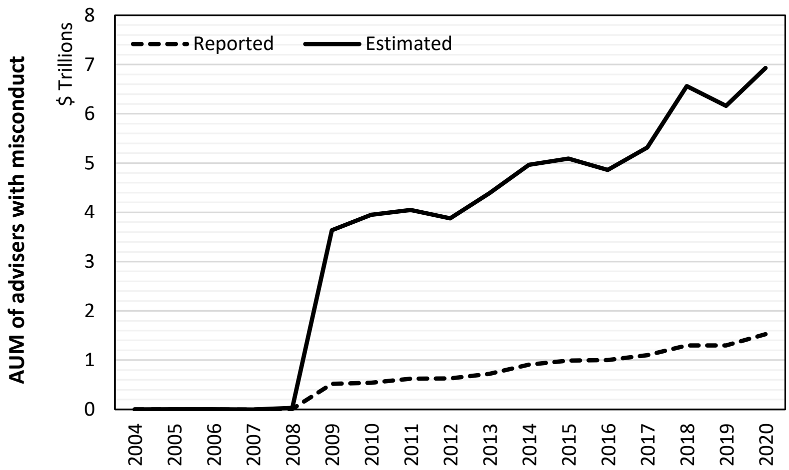


Panel B. Yearly share of unreported misconduct by a new adviser



Similar patterns emerge when we plot the number of misconduct by a new adviser each year in Figure 2.5. The number of advisers with first misconduct spikes during the GFC (Panel A), and the share of unreported misconduct by a new adviser also spikes during the GFC (Panel B).[37]

Figure 2.6. AUM of active advisers with misconduct over time
This figure plots the annual assets under management (AUM) in trillions of dollars for active advisers with estimated and reported misconduct during the period 2004-2020. We calculate the AUM for each adviser in a given year by dividing the firm's annual AUM by the number of investment advisers within that firm for that year. We then aggregate the AUM of active investment advisers with misconduct on an annual basis. To estimate the number of advisers involved in misconduct over time, we re-construct the adviser-level variables for each adviser-year during 2004-2020 and use the baseline model coefficients to calculate the probability of misconduct for each adviser-year. Advisers with a probability of misconduct greater than 0.5 are identified as engaging in misconduct.



To get a sense of the dollar value of assets (mis)managed by advisers with misconduct, we compute an adviser's AUM for a specific year by scaling the firm's yearly AUM (reported in yearly Form ADVs) according to the number of investment advisers within that firm for that year. We then track the AUM of active investment advisers with both reported and estimated misconduct annually. Figure 2.6 shows that the divergence in the AUM of reported and estimated misconduct starts in 2008 and increases subsequently. As of 2020, 138,288 active investment advisers with estimated misconduct collectively manage around $6.9 trillion AUM, while 34,211 active investment advisers with reported misconduct have a combined $1.5 trillion AUM.

[37] We omit 2020 in Figure 2.5 because it plots changes across years rather than cumulative values and our data is incomplete for 2020, covering until April 2020.

These findings have several other implications. First, the increase in the share of unreported misconduct suggests that regulators are slow in reacting to the rise in adviser misconduct during critical periods. The share of unreported misconduct is important for evaluating the effectiveness of the regulatory effort. Second, with new reforms regulating financial institutions after the GFC (e.g., the Dodd-Frank Act and Consumer Protection Act), the increase in adviser misconduct has halted, and the number of new advisers engaging in misconduct each year has slightly decreased. However, note that adviser misconduct is not specific to the GFC and is indeed pervasive in other years as well, suggesting that misconduct has become a feature of this industry (Zingales 2015).

2.4.4 Adviser misconduct and trust

Our findings so far provide strong validation for the public's perception of the finance sector, offering a rationale for why public trust in finance deteriorated during the GFC and never fully recovered. The increase in the share of unreported misconduct during the GFC parallels the public distrust of the finance sector. Conceptually, one can interpret the effect of the GFC on adviser misconduct in two different ways (as the economic recession variable in the model). The first is that a negative shock to advisers' wealth during the GFC can increase their willingness to commit misconduct (e.g., Dimmock et al. 2021). The second is that misconduct can function as risky borrowing (e.g., Andreoni 1992). Advisers may compare the costs of conventional borrowing against the expected misconduct penalties. The increase in conventional borrowing costs during the GFC can enhance the relative attractiveness of borrowing via misconduct, resulting in higher adviser misconduct.

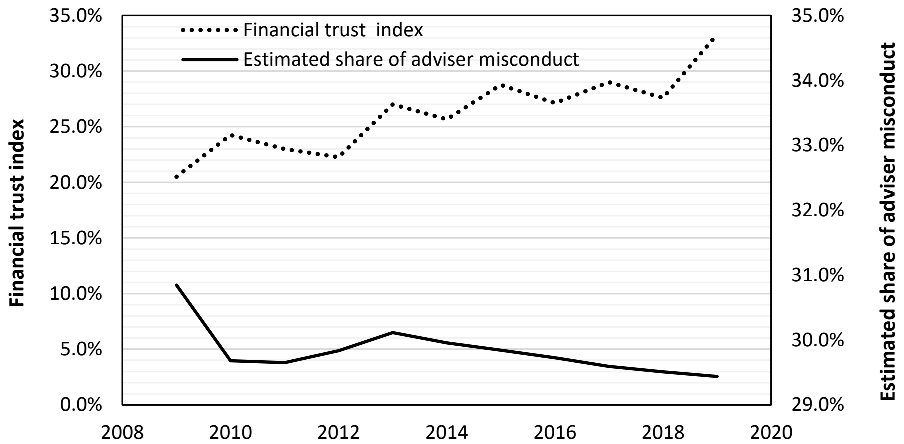
We investigate the relation between adviser misconduct and trust in two ways. First, we correlate the estimated, reported, and unreported share of financial adviser misconduct with the financial trust index (FTI) of Sapienza and Zingales. The FTI measures the overall level of trust in US financial institutions since 2009 based on interviews with 1,000 American households, randomly chosen and surveyed via telephone.[38]

[38] This initiative is sponsored jointly by the Kellogg School of Management at Northwestern University and the University of Chicago Booth School of Business and administered by Social Science Research Solutions. See http://www.financialtrustindex.org/.

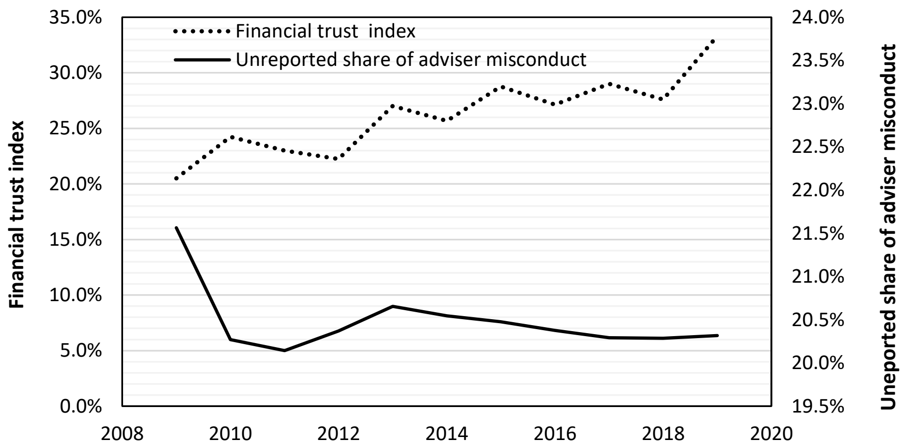
Figure 2.7 shows an upward trend in the Financial Trust Index (FTI) since the GFC, starting at around 20% in 2009 and increasing to 30% in 2019. In Panel A, the estimated share of adviser misconduct declines from approximately 31% in 2009 to 29% in 2019.

Figure 2.7. Financial trust index and adviser misconduct over time
This figure compares the yearly financial trust index (FTI) of Sapienza and Zingales with the yearly shares of estimated, unreported, and reported adviser misconduct during 2009-2019 in Panels A, B, and C, respectively. The FTI measures the overall level of trust in US financial institutions based on telephone interviews with 1,000 randomly chosen American households. The FTI is plotted with the dashed line on the left-hand-side axis and the estimated, unreported, and reported share of adviser misconduct is plotted with the solid line in the right-hand-side axis.

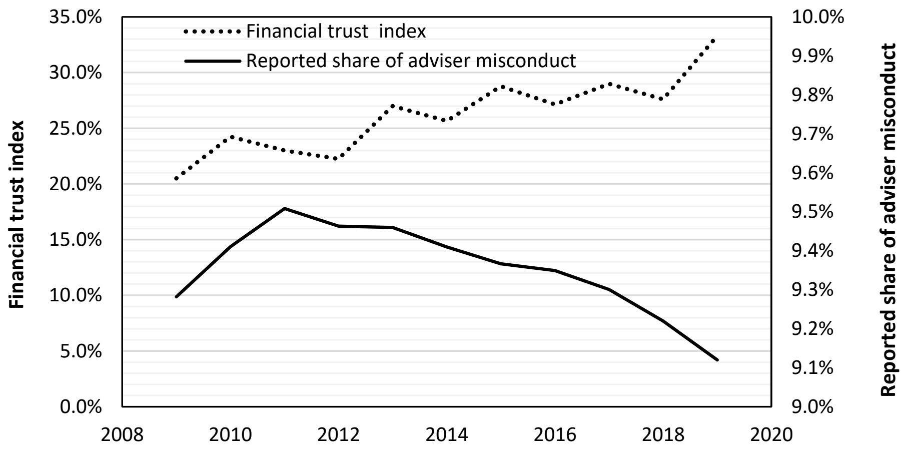
Panel A. Yearly estimated share of adviser misconduct and financial trust index



Panel B. Yearly unreported share of adviser misconduct and financial trust index



Panel C. Yearly reported share of adviser misconduct and financial trust index



Panel B similarly shows a decreasing trend in the unreported share of adviser misconduct, starting at 22% in 2009 and dropping to 20% over the sample period. The reported share of adviser misconduct starts at 9.3%, slightly increases to 9.5% two years after the GFC, but subsequently decreases to around 9%. These plots indicate a negative relation between the FTI and the estimated, unreported, and reported shares of adviser misconduct. The correlation coefficients between the FTI and the estimated, unreported, and reported shares of adviser misconduct are -0.62, -0.44, and -0.61, respectively.[39]

Next, we explore the relationship between the number of clients and AUM of advisory firms and their (estimated, unreported, and reported) shares of adviser misconduct. To do this, we run panel regressions for 9,341 financial advisory firms with 68,846 firm-year observations, where the dependent variables are the yearly number of clients and AUM of advisory firms and independent variables are lagged values of estimated, unreported, and reported shares of adviser misconduct in different versions of the model. We only include firms with a minimum of three years of observations and control for the total number of advisers in firms and include firm and year-fixed effects. The standard errors used to compute t-statistics are double-clustered by firm and year.

Table 2.7 reports that the estimated and unreported numbers of adviser misconduct have a negative impact on both the number of clients and AUM of advisory firms. Specifically,

[39] We also correlate the FTI with the estimated, unreported, and reported shares of misconduct by new advisers (see Figure 2.5), and the correlation coefficients are around -0.89.

one adviser with estimated misconduct, on average, costs the firm around 5 clients and decreases the total AUM by around \$10 million. Similarly, one adviser with unreported misconduct, on average, costs the firm around 7 clients and decreases the total AUM by around \$13 million. We do not find a significant relation between reported misconduct and the number of clients or AUM of firms. These findings further suggest that adviser misconduct undermines trust in financial advisory firms and leads to a decrease in the use of financial advisory services.

Table 2.7. Impact of adviser misconduct on number of clients and total AUM
This table reports the results of the panel regression:

$$Y_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot X_{i,t} + \beta_2 \cdot Total\,advisers_{i,t} + \varepsilon_{i,t},$$

where $Y_{i,t+1}$ is the firm i's number of clients and total AUM (million \$) in year $t+1$, $X_{i,t}$ is the number of advisers in firm i in year t with estimated misconduct (columns 1 and 4), unreported misconduct (columns 2 and 5), and reported misconduct (columns 3 and 6). All versions of the model include firm and year-fixed effects (α_i and γ_t, respectively). The standard errors used to compute the t-statistics (in brackets) are double clustered by firm and year. *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	Number of clients			Total AUM ($\cdot 10^6$ USD)		
	(1)	(2)	(3)	(4)	(5)	(6)
Estimated misconduct	-4.92***			-10.17***		
	(-2.99)			(-4.16)		
Unreported misconduct		-7.15***			-13.19***	
		(-3.54)			(-3.77)	
Reported misconduct			3.56			-4.42
			(0.46)			(-0.42)
Total advisers	3.41***	4.00***	0.82***	7.55***	8.10***	2.98***
	(3.69)	(4.22)	(2.59)	(5.37)	(5.15)	(6.69)
R^2	1.5%	1.8%	0.1%	4.6%	4.9%	3.3%
N obs.	68,846					

2.5 Extensions and robustness tests

2.5.1 Investment advisers versus brokers

An adviser can register as a broker with FINRA, an investment adviser with the SEC/-NAASA, or both. Our baseline analysis pools investment advisers and brokers together. However, separately estimating the extent of investment adviser and broker misconduct is important because they differ in registration, duties, legal requirements, and clientele.

Therefore, we redo the baseline analysis for investment advisers and brokers separately.[40] In our sample, there are 498,056 investment advisers and 1,061,933 brokers.

Table 2.8 presents the reported and estimated broker and investment adviser misconduct separately. We estimate that brokers are more likely to be involved in misconduct than investment advisers. 31.5% of brokers and 16.4% of investment advisers are estimated to be involved in misconduct, while the reported misconduct rate for brokers (9.4%) is lower than that for investment advisers (11.8%).[41] This is not surprising given that brokers are subject to a less stringent suitability standard by FINRA, whereas investment advisers have fiduciary responsibilities to their clients regulated by the SEC.

Table 2.8. Estimated misconduct frequency: Investment advisers vs brokers
This table reports the frequency of investment adviser and broker misconduct estimated by the detection-controlled estimation using the adviser-level sample with 498,056 investment advisers and 1,061,933 brokers and contrasts estimated frequencies with the frequency of reported misconduct. Panel A presents the number of advisers with reported misconduct, the number of advisers with misconduct estimated based on the DCE model (estimated misconduct), the number of advisers with unreported misconduct that are classified as advisers with misconduct based on the DCE model (estimated misconduct when unreported misconduct), and the number of advisers with reported misconduct that are classified as advisers with no misconduct based on the DCE model (no estimated misconduct when reported misconduct). Panel B reports that of Panel A in terms of percentages of total investment advisers and brokers.

	Investment advisers	Brokers
Panel A. Number of advisers		
Reported misconduct	58,582	99,795
Estimated misconduct, $\Pr(Y_1 = 1)$	81,692	334,754
Estimated misconduct when unreported misconduct, $\Pr(Y_1 = 1 \mid Y_2 = 0)$	58,567	279,358
No estimated misconduct when reported misconduct, $\Pr(Y_1 = 0 \mid Y_2 = 1)$	35,457	44,399
Panel B. Percentage of advisers		
Reported misconduct	11.762	9.398
Estimated misconduct, $\Pr(Y_1 = 1)$	16.402	31.523
Estimated misconduct when unreported misconduct, $\Pr(Y_1 = 1 \mid Y_2 = 0)$	11.759	26.307
No estimated misconduct when reported misconduct, $\Pr(Y_1 = 0 \mid Y_2 = 1)$	7.119	4.181

[40] The results are similar with both adviser-level and firm-level variables (Model 2). In the interest of conciseness, we only report the results of Model 1 with the full sample of advisers.

[41] We do not observe whether a given misconduct relates to the investment advisory or brokerage services for dual-registered advisers. This should not be a problem because it is hard to imagine an adviser going with the client-first approach in investment advisory services when he/she engages in misconduct in brokerage, and vice versa.

2.5.2 Alternative definition of adviser misconduct

We now assess the robustness of our findings using a more stringent definition of misconduct. We define misconduct as the most severe three disclosure categories: Criminal—Final Disposition, Regulatory—Final, and Customer Dispute—Settled.[42] With a stricter definition of misconduct, we estimate that the true adviser misconduct is around 2.27 times more than the reported misconduct — 17.60% of advisers are estimated to be involved in misconduct, but only 7.75% of advisers are reported.

2.5.3 Relaxing exclusion restrictions

To ensure that the estimated scale of adviser misconduct is not dependent on one instrumental variable, we relax the exclusion restriction for each instrumental variable one at a time. We run 12 different variations of the DCE model and report the results in Table 2.9. The estimated scale of adviser misconduct varies between 20% and 33%.

Table 2.9 Estimated misconduct frequency - Relaxing exclusion restrictions
This table reports the frequency of adviser misconduct estimated by the detection-controlled estimation when one instrumental variable is excluded from the model. We use Model 1 with 1,120,771 advisers. The results of this model are reported in Table 2.5 with an estimated misconduct frequency of 29.86% when reported misconduct frequency is 9.12%. The first column shows which instrumental variable is excluded from that model, and the second column shows the estimated misconduct frequency.

The following instrumental variables are excluded from the model	Estimated misconduct
Office misconduct rate	26.685
Other disclosure count	27.777
Economic recession	31.689
Firm count	32.501
Employment duration	32.172
Gender	33.405
Reported misconduct rate	20.256
Regulatory budget	23.408
Dodd-Frank switch	31.280
Portion of AUM fees	31.326
Per hour fees	29.855
Fixed fees	31.468

[42] Table A2.3 in the Appendix reports the estimates for the individual misconduct categories separately.

2.6 Conclusion

Financial advisers, as the main interface between the public and the finance sector, serve as primary representatives of the industry. We investigate the public's distrust of the finance sector in the context of financial adviser misconduct. Using comprehensive data from over one million advisers, we estimate that almost one in three advisers in the U.S. is involved in misconduct, yet only about one-third of those cases are detected by regulators. As of 2020, we estimate that $6.9 trillion in AUM is managed by 138,288 currently active investment advisers who we flag for potential misconduct based on their probability of misconduct, and 165,976 brokers with potential misconduct actively execute transactions on behalf of households.

Furthermore, we find that the frequency of adviser misconduct and the proportion of unreported misconduct increase significantly during the GFC, aligning with the erosion of public trust in the finance sector, and decrease thereafter, corresponding with the partial recovery of public trust. The estimated and unreported misconduct impact the number of clients and total AUM of financial advisory firms, highlighting the link between misconduct and trust in financial institutions. We provide an extensive list of characteristics associated with advisers' propensity for misconduct. These characteristics can help consumers make informed choices, assist advisory firms in designing effective internal controls, and enable regulators to use their resources more efficiently.

Our analysis indicates a significant level of misconduct among financial advisers, but translating this into concrete welfare effects for clients and the broader market is not straightforward. While we quantify the impact of adviser misconduct on the number of clients and AUM of advisory firms, direct financial losses, diminished trust in financial institutions, and potential long-term impacts on market participation are difficult to quantify precisely. Moreover, the welfare effects may vary significantly among different demographic groups, regions, and types of investors.

Future research should comprehensively evaluate the welfare implications of adviser misconduct, including its influence on investor behavior and market participation. Understanding the deterrent effect of adviser misconduct on market participation is crucial, as reduced market participation can lead to higher capital costs for firms and fewer investment opportunities for households. In addition, future studies should explore the broader socio-economic impacts of adviser misconduct, such as effects on investor confidence and household financial stability.

Appendix 2.1.

Table A2.1
Definition of non-misconduct-related disclosure categories
This table defines 17 non-misconduct-related disclosure categories that we define as other disclosures: (1) Civil—Pending, (2) Civil—on Appeal, (3) Civil Bond, (4) Customer Dispute—Closed—No Action, (5) Customer Dispute—Withdrawn, (6) Customer Dispute—Dismissed, (7) Customer Dispute—Denied, (8) Customer Dispute—Final, (9) Customer Dispute—Pending, (10) Judgment/Lien, (11) Financial—Final, (12) Financial—Pending, (13) Investigation, (14) Criminal—on Appeal, (15) Criminal—Pending Charge, (16) Regulatory—on Appeal, (17) Regulatory—Pending. All definitions are from https://brokercheck.finra.org/.

Disclosure category	**Definition**
(1) Civil—Pending	This type of disclosure event involves a pending civil court action that seeks an injunction in connection with any investment-related activity or alleges a violation of any investment-related statute or regulation.
(2) Civil—on Appeal	This type of disclosure event involves an injunction issued by a court in connection with an investment-related activity or a finding by a court of a violation of any investment-related statute or regulation that is currently on appeal.
(3) Civil Bond	This type of disclosure event involves a civil bond for the adviser that has been denied, paid, or revoked by a bonding company.
(4-8) Customer Dispute—Closed —No Action / Withdrawn / Dismissed / Denied / Final	This type of disclosure event involves (1) a consumer-initiated, investment-related arbitration or civil suit containing allegations of sales practice violations against the individual adviser that was dismissed, withdrawn, or denied or (2) a consumer-initiated, investment-related written complaint containing allegations that the adviser engaged in sales practice violations resulting in compensatory damages of at least $5,000, forgery, theft, or misappropriation, or conversion of funds or securities, which was closed without action, withdrawn, or denied.
(9) Customer Dispute—Pending	This type of disclosure event involves (1) a pending consumer-initiated, investment-related arbitration or civil suit that contains allegations of sales practice violations against the adviser or (2) a pending, consumer-initiated, investment-related written complaint containing allegations that the adviser engaged in sales practice violations resulting in compensatory damages of at least $5,000, forgery, theft, or misappropriation, or conversion of funds or securities.
(10) Judgment/Lien	This type of disclosure event involves any unsatisfied and outstanding judgments or liens against the adviser.
(11) Financial—Final	This type of disclosure event involves a bankruptcy, compromise with one or more creditors, or Securities Investor Protection Corporation liquidation involving the adviser or an organization the adviser controlled that occurred within the last 10 years.
(12) Financial—Pending	This type of disclosure event involves a pending bankruptcy, compromise with one or more creditors, or Securities Investor Protection Corporation liquidation involving the adviser or an organization the adviser controlled that occurred within the last 10 years.
(13) Investigation	This type of disclosure event involves any ongoing formal investigation by an entity such as a grand jury, state or federal agency, self-regulatory organization, or foreign regulatory authority. Subpoenas, preliminary or routine regulatory inquiries, and general requests by a regulatory entity for information are not considered investigations and therefore are not included in a Broker-Check report.

(continued)

Table A2.1
Continued

Disclosure category	Definition
(14) Criminal—on Appeal	This type of disclosure event involves a conviction for any felony or certain misdemeanor offenses, including bribery, perjury, forgery, counterfeiting, extortion, fraud, and wrongful taking of property that is currently on appeal.
(15) Criminal—Pending Charge	This type of disclosure event involves a formal charge for a crime involving a felony or certain misdemeanor offenses, including bribery, perjury, forgery, counterfeiting, extortion, fraud, and wrongful taking of property that is currently pending.
(16) Regulatory—on Appeal	This type of disclosure event involves (1) a formal proceeding initiated by a regulatory authority (e.g., a state securities agency, self-regulatory organization, federal regulatory agency such as the SEC, foreign financial regulatory body) for a violation of investment-related rules or regulations that is currently on appeal or (2) a revocation or suspension of an adviser's authority to act as an attorney, accountant, or federal contractor that is currently on appeal.
(17) Regulatory—Pending	This type of disclosure event involves a pending formal proceeding initiated by a regulatory authority (e.g., a state securities agency, self-regulatory organization, federal regulatory agency such as the SEC, foreign financial regulatory body) for alleged violations of investment-related rules or regulations.

Table A2.2
Coefficient estimates for the models reported in Table 2.6

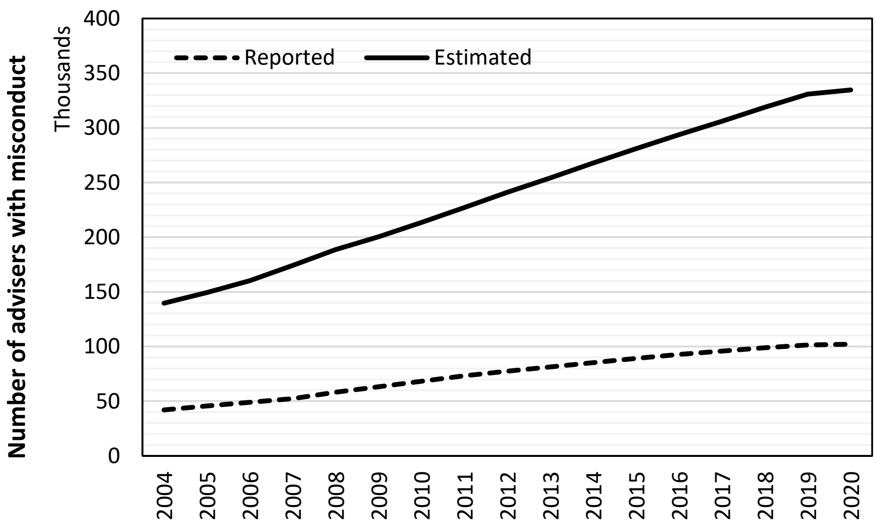
	Model 1		Model 2	
	M	D	M	D
Panel A. Adviser-level variables				
Office misconduct rate	0.849***		0.559***	
	(64.235)		(34.997)	
Other disclosure count	0.968***		0.876***	
	(57.772)		(47.054)	
Economic recession	0.057***		0.063***	
	(7.344)		(6.538)	
Firm count	0.135***		0.117***	
	(16.932)		(10.505)	
Employment duration	-0.071***		-0.159***	
	(-8.907)		(-15.331)	
Gender	0.484***		0.475***	
	(72.633)		(58.847)	
Reported misconduct rate		1.113***		1.090***
		(100.027)		(72.824)
Regulatory budget		0.415***		0.362***
		(87.925)		(58.042)
Dodd-Frank switch		0.049***		0.034***
		(13.463)		(8.116)
Tenure	-0.357***	0.011	-0.341***	0.177***
	(-30.165)	(0.946)	(-20.401)	(10.516)
Exam count	-0.074***	0.051***	-0.045***	0.035**
	(-6.800)	(5.557)	(-2.869)	(2.591)
Employee count	0.153***	-0.016	0.121***	0.116***
	(9.074)	(-1.618)	(4.725)	(6.484)
Active financial adviser	-0.125***	-0.014	-0.044***	-0.095***
	(-10.650)	(-1.303)	(-2.956)	(-6.774)
Dual registration	0.236***	0.055***	0.191***	0.095***
	(21.285)	(6.500)	(12.304)	(7.658)
Panel B. Firm-level variables				
Portion of AUM fees			0.087***	
			(10.400)	
Per hour fees			0.221***	
			(21.617)	
Fixed fees			-0.079***	
			(-7.111)	
Total AUM			0.371***	-0.234***
			(4.077)	(-4.009)
Discretionary AUM			-0.439***	0.185***
			(-8.621)	(5.496)
Custody AUM			0.015	0.008
			(0.348)	(0.258)
Number of clients			0.258***	-0.153***
			(6.589)	(-6.824)
Sales interest			-0.112***	0.017
			(-6.224)	(1.222)
Pseudo R^2	17.51%		15.20%	
N obs.	1,120,771		753,833	

Table A2.3
Misconduct frequency estimation based on individual misconduct categories
This table presents the frequency of adviser misconduct as estimated by the detection-controlled estimation (DCE) model, applied to the adviser-level sample with 1,120,771 advisers (Model 1) by changing the definition of reported misconduct. In Panel A, misconduct is defined as three of the most severe misconduct categories (Criminal—Final Disposition, Regulatory—Final, and Customer Dispute—Settled). Panel B, C, D, E, F and G report that of Panel A when misconduct is defined as Criminal—Final Disposition, Regulatory—Final, Customer Dispute—Settled, Customer Dispute—Award/Judgment, Employment Separation After Allegations, and Civil—Final, respectively.

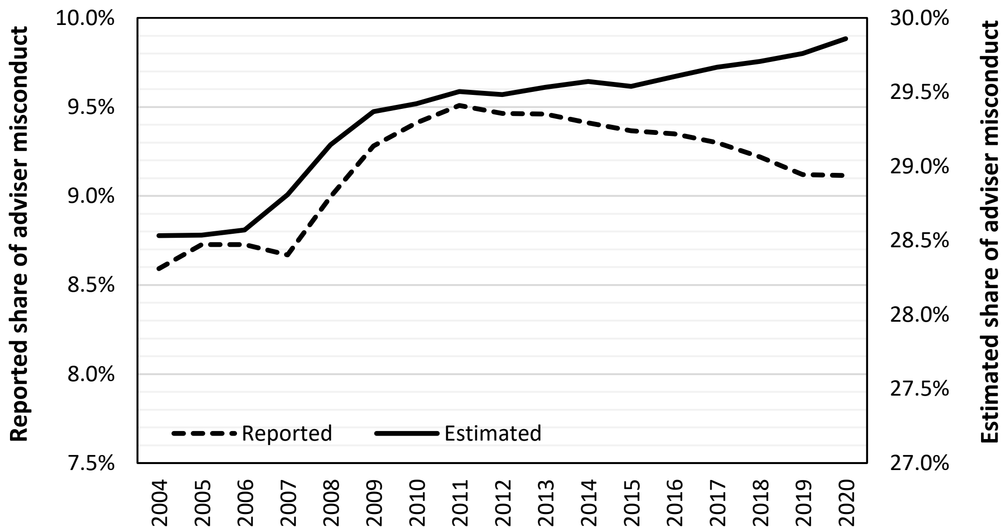
	Percentage
Panel A. Criminal—Final Disposition, Regulatory—Final, Customer Dispute—Settled	
Reported misconduct, $\Pr(Y_2 = 1)$	7.754
DCE misconduct, $\Pr(Y_1 = 1)$	17.596
DCE misconduct when undetected misconduct, $\Pr(Y_1 = 1 \mid Y_2 = 0)$	14.153
DCE no misconduct when detected misconduct, $\Pr(Y_1 = 0 \mid Y_2 = 1)$	4.311
Panel B. Criminal—Final Disposition	
Reported misconduct, $\Pr(Y_2 = 1)$	2.491
DCE misconduct, $\Pr(Y_1 = 1)$	3.651
DCE misconduct when undetected misconduct, $\Pr(Y_1 = 1 \mid Y_2 = 0)$	3.451
DCE no misconduct when detected misconduct, $\Pr(Y_1 = 0 \mid Y_2 = 1)$	2.291
Panel C. Regulatory—Final	
Reported misconduct, $\Pr(Y_2 = 1)$	2.643
DCE misconduct, $\Pr(Y_1 = 1)$	8.284
DCE misconduct when undetected misconduct, $\Pr(Y_1 = 1 \mid Y_2 = 0)$	6.995
DCE no misconduct when detected misconduct, $\Pr(Y_1 = 0 \mid Y_2 = 1)$	1.355
Panel D. Customer Dispute—Settled	
Reported misconduct, $\Pr(Y_2 = 1)$	3.415
DCE misconduct, $\Pr(Y_1 = 1)$	4.905
DCE misconduct when undetected misconduct, $\Pr(Y_1 = 1 \mid Y_2 = 0)$	3.675
DCE no misconduct when detected misconduct, $\Pr(Y_1 = 0 \mid Y_2 = 1)$	2.186
Panel E. Customer Dispute—Award/Judgment	
Reported misconduct, $\Pr(Y_2 = 1)$	0.634
DCE misconduct, $\Pr(Y_1 = 1)$	5.954
DCE misconduct when undetected misconduct, $\Pr(Y_1 = 1 \mid Y_2 = 0)$	5.612
DCE no misconduct when detected misconduct, $\Pr(Y_1 = 0 \mid Y_2 = 1)$	0.292
Panel F. Employment Separation After Allegations	
Reported misconduct, $\Pr(Y_2 = 1)$	1.845
DCE misconduct, $\Pr(Y_1 = 1)$	5.159
DCE misconduct when undetected misconduct, $\Pr(Y_1 = 1 \mid Y_2 = 0)$	4.548
DCE no misconduct when detected misconduct, $\Pr(Y_1 = 0 \mid Y_2 = 1)$	1.234
Panel G. Civil—Final	
Reported misconduct, $\Pr(Y_2 = 1)$	0.123
DCE misconduct, $\Pr(Y_1 = 1)$	2.319
DCE misconduct when undetected misconduct, $\Pr(Y_1 = 1 \mid Y_2 = 0)$	2.265
DCE no misconduct when detected misconduct, $\Pr(Y_1 = 0 \mid Y_2 = 1)$	0.068

Figure A2.1
Estimated and reported adviser misconduct over time
This figure illustrates the yearly number and share of advisers with estimated misconduct during 2004-2020 in Panel A and B, respectively and compare them to advisers with reported misconduct during the same period. In Panel B, the reported share of advisers with misconduct is plotted with the dashed line using the left-hand-side axis, and the estimated share of advisers with misconduct is plotted with the solid line using the right-hand-side axis. Panel C plots the number of advisers with unreported misconduct (the difference between estimated and reported misconduct). To estimate the number of advisers involved in misconduct over time, we track 334,660 advisers with estimated misconduct from the baseline model (see Table 2.5) over time, starting from their industry entry date.

Panel A. Yearly number of advisers with estimated and reported misconduct



Panel B. Yearly share of advisers with estimated and reported misconduct



Panel C. Yearly share of advisers with undetected misconduct

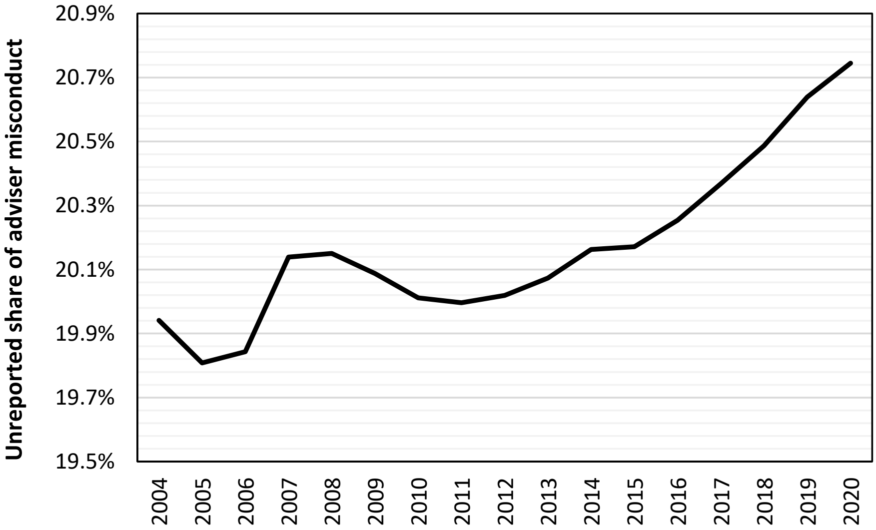
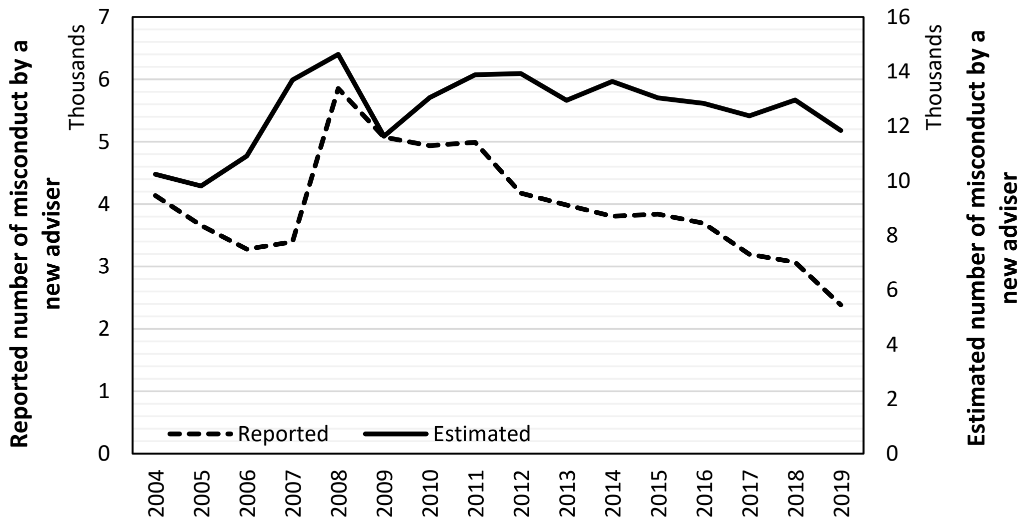


Figure A2.2
Estimated and reported misconduct by a new adviser over time
This figure illustrates the estimated and reported number of misconduct done by a new adviser (with no previous misconduct record) during 2004-2019 in Panel A. In Panel A, the reported number of misconduct by a new adviser is plotted with the dashed line using the left-hand-side axis, and the estimated number of misconduct by a new adviser is plotted with the solid line using the right-hand-side axis. Panel B plots the number of undetected misconduct by a new adviser (the difference between the estimated and reported number of misconduct by a new adviser). To estimate the number of advisers involved in misconduct over time, we track 334,660 advisers with estimated misconduct (see Table 2.5) from the baseline model over time, starting from their industry entry date. We count the number of advisers with first misconduct each year during 2004-2019.

Panel A. Yearly number of advisers with estimated and reported first misconduct



Panel B. Yearly share of undetected misconduct by a new adviser

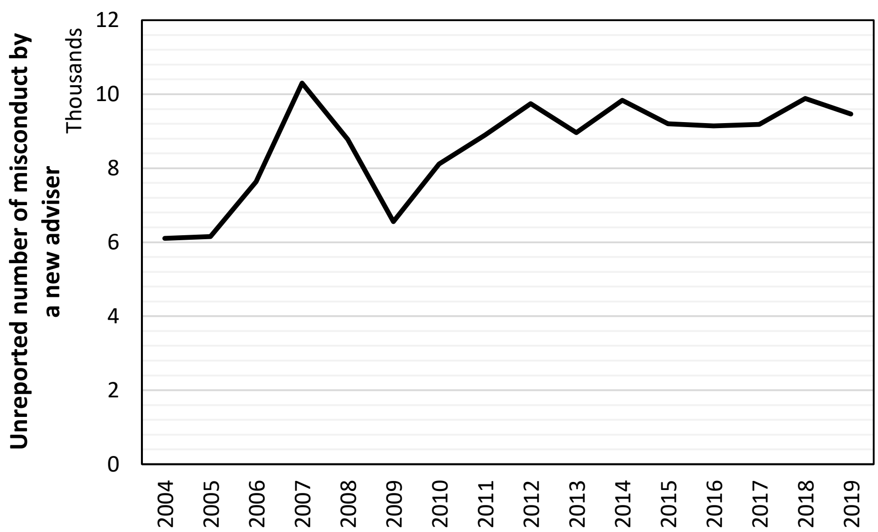
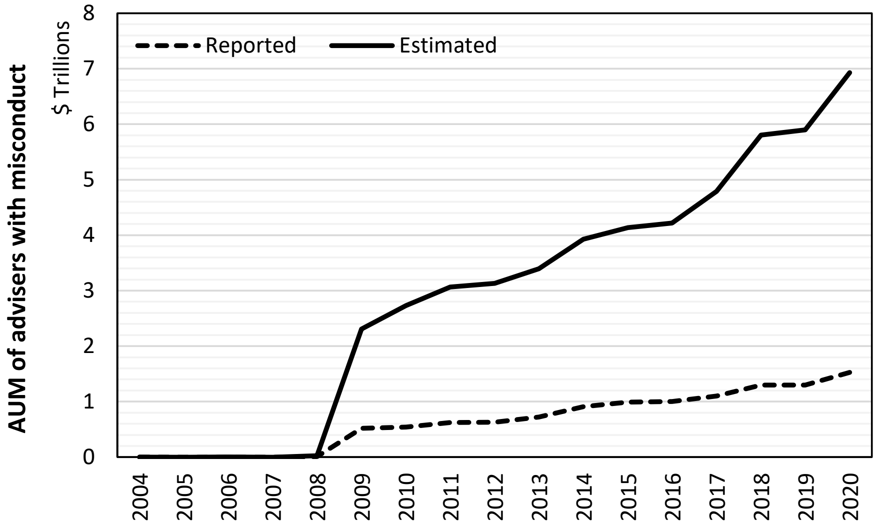


Figure A2.3

Asset under management of active advisers with misconduct over time

This figure plots the annual assets under management (AUM) in trillions of dollars for active advisers with estimated and reported misconduct during the period 2004-2020. We calculate the AUM for each adviser in a given year by dividing the firm's annual AUM by the number of investment advisers within that firm for that year. We then aggregate the AUM of active investment advisers with misconduct on an annual basis. To estimate the number of advisers involved in misconduct over time, we track 334,660 advisers with estimated misconduct from the baseline model (see Table 2.5) over time, starting from their industry entry date.



Chapter 3

The real effects of market manipulation

3.1 Introduction

Market manipulation is prohibited in most financial markets around the world. But should it be? The answer depends on how severe are the harms caused by market manipulation, and do they outweigh the costs of regulation and compliance. The literature to date has focused on the direct consequences of market manipulation, such as distorting prices in secondary markets and decreasing market liquidity (e.g., Aggarwal and Wu 2006; Comerton-Forde and Putnins 2011a). But do the impacts of market manipulation extend beyond these direct effects? Does manipulation affect listed companies and real investment decisions? Does it affect where and how capital is deployed? These are the issues that we address in this chapter.

Theoretically, it is clear that market manipulation could have real economic consequences by distorting corporate investment decisions. For example, Goldstein and Guembel (2008) show that the presence of a feedback effect from the financial market to the real value of a firm creates an incentive for traders to manipulate stock prices, resulting in a weaker allocational role of financial markets. But in practice, is manipulation sufficiently frequent and distortionary to significantly impact firms? And beyond isolated manipulative attacks like the ones anticipated by Goldstein and Guembel (2008), does the possibility that prices might be distorted by manipulation decrease the use of price signals in a way that harms real investment decisions?

To address these questions we develop a conceptual framework that shows how the possibility of manipulative price distortions can decrease the use of price signals by corporate managers. We show that when managers use stock prices and internal information for optimal investment decisions, they rely less on stock prices as the intensity of price manipulation increases. The noise in stock prices due to manipulation diminishes managers' incentive to use stock prices as an additional source of information, resulting in a decrease in the sensitivity of investment to stock prices (known as investment-to-price or investment-Q sensitivity). As a result, stock price manipulation has a negative impact on firm performance since managers relying less on stock prices tend to make less-informed investment decisions.

We then take these predictions to the data. To measure market manipulation, we employ the closing stock price manipulation measure of Comerton-Forde and Putnins (2011b). This measure uses standardized trading and stock price metrics such as abnormality of return, reversal, trading frequency, and bid-ask spread to distinguish manipulated stock prices from non-manipulated prices. To measure firms' investment, we consider capital expenditures (with and without research and development expenses), as well as changes in total assets. To measure firm performance, we use return on assets and asset turnover. Subsequently, we examine the impact of stock price manipulation on investment-to-price sensitivity and operating performance of a large sample of firms in US during 1997-2019.

In a simple ordinary least squares (OLS) setting, on average, stock price manipulation decreases the effect of a one standard deviation shock to the stock price on the next year's capital expenditure ($CAPEX$) by approximately 20%, capital expenditures including research and development ($CAPEXRND$) by 17%, and change in total assets ($CHGASSET$) by 12%. The effect of market manipulation on investment-to-price sensitivity becomes more severe once we establish causality using different instrumental variables, such as average manipulation by stock's size group, lagged market manipulation, and the SEC regulatory budget interacted with stock's liquidity in a two-stage least squares (2SLS) setting. Furthermore, the average stock price manipulation decreases the next year's return on assets (ROA) by about 5.5% and the next year's asset turnover (ATO) by 15%.

We also investigate which stock characteristics influence the magnitude of the impact of stock price manipulation on investment-to-price sensitivity. We find that the negative impact of manipulation on investment-to-price sensitivity is stronger in illiquid stocks (measured by the relative bid-ask spread) and in more volatile stocks (measured by the standard deviations of returns), as well as in those with more private information (measured by the probability of informed trading). This means the negative impact of manipulation on investment-to-price sensitivity is amplified when manipulators have the ability to move stock prices (it is hard to move the prices of liquid and less volatile stocks) and when stock prices convey more private information to firm managers. These findings support the idea that firm managers use signals from stock prices to make investment decisions, known as the "managerial learning hypothesis" in the literature (e.g., Chen, Goldstein, and Jiang 2007; Foucault and Frésard 2012).

We consider several extensions and robustness issues. First, we explore the impact of different levels of stock price manipulation on investment-to-price sensitivity. We split the stocks into quintiles based on the probability of stock price manipulation. Then we show that the detrimental effect of manipulation on investment-to-price sensitivity increases monotonically from the low-manipulated stocks (bottom quintile) to the high-manipulated stocks (top quintile). Second, we control for several additional factors. We find that the association between stock price manipulation and investment-to-price sensitivity remains negative and significant after controlling for managers' existing information (intensity of insider trading), analyst coverage, and firms' capital constraints. Lastly, we run yearly Fama-MacBeth regressions and document that firms with highly manipulated stocks have lower investment-to-price sensitivity in the cross-section. The Appendix contains additional robustness tests.

3.2 Related literature

Market manipulation challenges the functioning and integrity of financial markets by distorting asset prices and creating uncertainty for investors. The literature on market manipulation generally defines it as actions that artificially affect asset prices for personal gain and provides insights into various techniques employed by market participants to manipulate prices, such as pump-and-dump schemes, wash trades, spoofing and layering (e.g., Allen and Gale 1992*a*, Putnins 2012, and Aggarwal and Wu 2006). Manipulation not only distorts market information and increases risk for investors, but also leads to the misallocation of resources resulting in deadweight economic losses to society (e.g., Pirrong 1995, and Brunnermeier and Oehmke 2013).

Despite extensive research on market manipulation, there is limited research on the detrimental effects market manipulation on resource allocation. This limitation stems partly from the unobservable nature of the true extent of market manipulation, with regulators detecting and prosecuting only a small fraction of manipulative activities in markets (e.g., Comerton-Forde and Putnins 2014). Consequently, the real economic consequences of market manipulation remain relatively unknown. We address this question by analyzing the impact of stock price manipulation on firms' investment-to-price sensitivity and operating performance.

Chapter 3 makes contributions to two distinct research streams. First, it contributes to the empirical studies on market manipulation, such as Merrick, Naik and Yadav (2005) on squeezes, Aggarwal and Wu (2006) on pump-and-dump manipulation, Allen, Litov and Mei (2006) on corners, and Cumming, Li and Yu (2023) on the relationship between market manipulation and corporate venture capital. We contribute to this literature by focusing on the real economic consequences of market manipulation. To our knowledge, we are the first to show that the prevalence of stock price manipulation has a significant negative effect on the sensitivity of firms' investment to stock prices, consequently harming the operating performance of firms.

Second, we contribute to the literature that analyzes how firm managers learn from stock prices – managerial learning (e.g., Chen, Goldstein, and Jiang 2007; Foucault and Frésard (2012, 2014); Edmans, Jayaraman, Schneemeier 2017; Baker, Stein, and Wurgler 2003) and how stock prices affect firms' investment decisions (e.g., Barro 1990; Morck, Shleifer, and Vishny 1990). Stock prices pools information from different types of traders who do not generally have open communication channels with the firm other than the trading process. Therefore, firm managers learn from information in stock prices about the prospects of their firms (e.g., Dow and Gorton 1997; Subrahmanyam and Titman 1999). Our results indicate that stock price manipulation impedes the generation of new information for firm managers, leading to a decrease in firm's investment-to-price sensitivity and operating performance.

3.3 The model

Consider a publicly traded firm with assets and growth opportunity. The asset payoff is given by

$$\tilde{V} = \bar{V} + \tilde{\theta}, \tag{3.1}$$

where $\bar{V}$ is a positive constant and $\tilde{\theta} \sim N(0, \sigma_\theta^2)$. The growth opportunity is given by

$$\tilde{G}(K) = K \cdot \tilde{\theta}^2 - 0.5 \cdot K^2, \tag{3.2}$$

where the firm manager invests K units in a growth opportunity at a cost $0.5 \cdot K^2$.[43] Following Subrahmanyam and Titman (1999) and Foucault and Gehrig (2008), this structure assumes that the profitability of the growth opportunity is perfectly correlated with the asset payoff (both driven by $\tilde{\theta}$).

The firm manager faces uncertainty regarding the asset payoff and growth opportunity profitability, $\tilde{\theta}$. To make an optimal investment decision K^*, the manager relies on two sources of information: stock price P and a managerial signal $\tilde{s} = \tilde{\theta} + \tilde{m}$, where $\tilde{m} \sim N(0, \sigma_m^2)$. Given these two sources of information, the manager chooses the investment in the growth opportunity K^* that maximizes the expected firm value (the assets payoff plus the profit from the growth opportunity).

We consider securities that are a claim only to assets in place, rather than the sum of assets and growth opportunities. Therefore, we let the price function that the manager learns from be given by $P = \bar{V} + \tilde{\delta}$, where $\tilde{\delta} = \sqrt{\rho} \cdot \tilde{\theta} + \sqrt{1-\rho} \cdot \tilde{\varepsilon}$ and ρ is the correlation between the stock price and the firm value (e.g., Foucault and Gehrig 2008; Edmans, Jayaraman and Schneemeier 2017).[44] We further assume that $\tilde{\theta}$ and $\tilde{\varepsilon}$ are identically distributed with the same variance $\tilde{\varepsilon} \sim N(0, \sigma_\theta^2)$. This assumption is critical because the identical distribution of the noise $\tilde{\varepsilon}$ and the fundamental value $\tilde{\theta}$ with the same variance hinders the firm manager from perfectly learning the value of $\tilde{\theta}$ from the stock price.

[43] The variables with $\sim$ on top are uncertain variables.

[44] We do not explicitly model the demand functions of market participants that determines the stock price P. Rather we focus on how the price manipulation disrupts the manager's signal extraction problem. The assumption that the price function is independent of the optimal investment substantially simplifies the model because it means that the investment decision is influenced by the stock price, but the stock price does not depend on the investment decision. If the security were also a claim to the new investment, the price would no longer be normally distributed, and the manager's signal extraction problem becomes intractable.

We focus on (i) optimal investment level K^*, (ii) investment-to-price sensitivity $\partial K^*/\partial P$, and (iii) expected growth opportunity conditional on available information $E[\tilde{G}(K)|\tilde{s}, P]$, and how (i)-(iii) are impacted by the stock price manipulation. Formally, Proposition 1 derives (i)-(iii) and allows us to investigate (iv).

Proposition 3.1. *Let* $w = \frac{1}{\left(1+\tau_m \cdot \sigma_\theta^2 \cdot (1-\rho)\right)}$ *where* $\tau_m = \frac{1}{\sigma_m^2}$ *is the precision of the managerial signal.*

(i) The optimal investment is given by $K^* = w \cdot \tilde{\delta} + (1-w) \cdot \tilde{s}$.

(ii) The investment-to-price sensitivity is given by $\frac{\partial K^*}{\partial P} = w$.

(iii) The expected growth opportunity conditional on available information is given by $E[\tilde{G}(K)|\tilde{s}, P] = \left(\frac{1-w\cdot(1-\rho)}{2}\right) \cdot \sigma_\theta^2$.

All three remarks of Proposition 1 are intuitive. First, the optimal investment K^* is a weighted average of the price signal $\tilde{\delta}$ and the managerial signal $\widetilde{s}$, with the weight assigned to the price (resp. managerial) signal denoted as w (resp. $1-w$). These weights depend on three factors: (i) the correlation between the stock price and the firm value (ρ), (ii) the volatility of the stock price (σ_θ^2), and (iii) the precision of the managerial signal (τ_m). As the correlation between the stock price and the firm value increases, the weight of the price signal increases and the weight of the managerial signal decreases ($\frac{\partial w}{\partial \rho} > 0$). As the stock price volatility increases, the weight of the price signal decreases and the weight of the managerial signal increases ($\frac{\partial w}{\partial \sigma_\theta^2} < 0$). As the managerial signal becomes more precise, the weight of the price signal decreases and the weight of the managerial signal increases ($\frac{\partial w}{\partial \tau_m} < 0$).

Second, the sensitivity of investment to stock prices is the weight w given by the manager to the price signal. It increases with the correlation between the stock price and the firm value and decreases with the volatility of the stock price and the precision of the managerial signal.

Third, the expected growth opportunity conditional on available information is the scaled volatility of the growth opportunity σ_θ^2, where the scale factor is $S = \frac{1-w\cdot(1-\rho)}{2}$. The scale factor increases with the correlation between the stock price and the firm value ($\frac{\partial S}{\partial \rho} > 0$) and the precision of the managerial signal ($\frac{\partial S}{\partial \tau_m} > 0$). That means the expected growth opportunity increases with the quality of both price and managerial signals.

Let μ represent the intensity of stock price manipulation and $I_P = \rho \cdot \sigma_\theta^2$ represent stock price informativeness about firm value, reflecting the reduction in the unconditional variance of firm value σ_θ^2 after observing the price,

$$Var(\widetilde{\theta}|P) = \sigma_\theta^2 \cdot (1-\rho) = \sigma_\theta^2 - I_P. \tag{3.3}$$

If the intensity of price manipulation μ influences the stock price informativeness I_P, it will impact investment, investment-to-price sensitivity, and expected growth opportunity.

To link the intensity of price manipulation (μ) and stock price informativeness (I_P), consider the classic market microstructure models of Glosten and Milgrom (1985) and Kyle (1985). These models involve informed traders, uninformed traders, and competitive market makers. Upgrading uninformed traders (e.g., Allen and Gorton 1992*b*) or informed traders (e.g., Chakraborty and Yılmaz 2004) in these models to employ manipulative strategies reduces stock price informativeness by introducing additional noise. Therefore, stock price informativeness decreases with the intensity of price manipulation ($\partial I_P/\partial \mu < 0$), leading to the following results.

1. The optimal investment is positively (negatively) related to the intensity of manipulation when manipulators decrease (increase) the stock price beyond the manager's expectation based on the managerial information; $\frac{\partial K^*}{\partial \mu} > 0$ if and only if $P < \bar{V} + \tilde{s}$.
2. The sensitivity of investment to prices decreases with the intensity of manipulation; $\frac{\partial}{\partial \mu}\left(\frac{\partial K^*}{\partial P}\right) < 0$.
3. The expected growth opportunity conditional on available information decreases with the intensity of manipulation; $\frac{\partial E[\tilde{G}(K)|\tilde{s},P]}{\partial \mu} \leq 0$.

The effect of price manipulation on optimal investment (Result 1) differs from its effect on investment-to-price sensitivity (Result 2). Result 1 indicates that the amount of investment can either increase or decrease depending on whether manipulated stock prices convey positive or negative information to the firm manager. In contrast, Result 2 shows that the investment-to-price sensitivity always decreases with the intensity of price manipulation. Furthermore, Result 3 shows that stock price manipulation negatively impacts the expected profitability of growth opportunities, suggesting that managers with more manipulative activities in their stocks tend to make less-informed decisions that could harm future firm performance. In the empirical analysis, we focus on testing Results 2 and 3 because they show an unambiguous effect of price manipulation on investment-to-price sensitivity and firm performance.

3.4 Data, measures and descriptive statistics

3.4.1 Data

We obtain data from multiple sources. First, we obtain firms' investment and other financial data from Compustat. For each firm, we obtain data for firms' financial policy and profitability such as total assets, market value of equity, capital expenditures, research and development (R&D) expenses, cash flows, and sales for the period 1997–2019.

As in Foucault and Fresard (2012), we exclude international affairs and non-operating establishments (between 9000 and 9999 SIC codes) and financial firms (between 6000 and 6999 SIC codes) because of the differences in the accounting practices of these industries. We exclude firm-year observations with less than ten million book value of equity and with missing information on the market value of equity, total assets, capital expenditures, and sales. We also exclude firms that have less than 3 years of observations. We winsorize all variables from Compustat at 1% from each tail.

Second, we obtain firms' daily stock price, closing bid and ask quotes, trading volume, and shares outstanding from Center for Research in Security Prices (CRSP). We then calculate the yearly averages of daily relative bid-ask spreads and dollar volume to proxy for stock's liquidity, standard deviations of daily returns to measure stock's volatility, and month-end market capitalizations to proxy for size.

Third, we obtain 5-minute intraday bid and ask quotes, last trade price, and number of trades in each interval from Refinitiv Tick History (RTH) for stocks of all Compustat firms. We then use 5-minute intraday data from RTH to calculate the probability of daily closing price manipulation measure of Comerton-Forde and Putnins (2011b) and the probability of informed trading (PIN) measure of Easley, Hvidkjaer and O'Hara (2002) to capture the private information environment of firms. Our sample contains all stocks listed on NYSE, AMEX, and NASDAQ during 1997 - 2019. We apply standard filters (remove crossed and outlier bid-ask spreads as per Chordia, Roll and Subrahmanyam (2001), erroneous entries, and stock-days that do not contain at least one trade and one quote). We also exclude stocks that traded less than 30 days each year.

We use Compustat variables to capture firms' investment and operating performance, as well as other control variables. We use CRSP variables to identify stocks characteristics that impact the effects of market manipulation on investment-to-price sensitivity. We use RTH variables to compute the stock price manipulation. For models that incorporate Compustat and RTH data, our final sample comprises 52,865 firm-year observations and 4,671 firms. For models that use data from all three sources (CRSP, Compustat, and RTH), our final sample consists of 46,028 firm-year observations and 4,000 firms.

We augment our primary datasets by including the yearly averages of the number of analysts following the stock obtained from I/B/E/S on WRDS at a monthly frequency, the yearly regulatory budget of the SEC retrieved from the SEC website, insider trading activities from Thomson Financial's TFN database, and Peters-Taylor Q obtained from Wharton Research Data Services (WRDS) to run additional tests.

3.4.2 Measure of stock price manipulation

Following Comerton-Forde and Putnins (2011b), we identify the closing price manipulation for a given stock-day using stock returns, bid-ask spreads, trading frequencies and return reversals. We do this in three steps. First, we estimate day-end metrics using 5-minute intraday data from RTH as follows:

1. Return is the natural logarithm of the ratio of the closing price (price at 4 pm) to the bid-ask midpoint at x minutes before the close. We choose the value of x as the smallest interval among $\{15, 20, 30, 60, 90\}$ minutes prior to the market close, where at least one trade occurs within the interval.

2. Return reversal is the natural logarithm of the ratio of the closing price to the bid-ask midpoint the following morning at 11 am.

3. Trade frequency is the ratio of the number trades in x minutes before close to x minutes.

4. Bid-ask spread is the difference between the bid and ask prices at the close relative to the bid-ask midpoint.

Second, we standardize the difference between a specific stock-day metric and that stock's previous trading using sign statistics from non-parametric sign tests. To do this, we compare the metric's value on a given day to the metric's values over the preceding 42 trading days, which ends 1 year prior to the analyzed day. For each stock-day, we calculate the number of positive changes n_+ and the number of negative changes n_-. From these values, we derive the sign statistic as $S_i = (n_+ - n_-)/2$. To mitigate the influence of market trends, we compute the difference between the sign statistics for a given stock on each day s and the median sign statistics of all other stocks' S_i on the same exchange for that day, denoted as $\Delta_i^{sign} = S_i - \text{med}_s(S_i)$.

Third, we calculate the probability of manipulation M for a given stock-day as

$$M = \left(1 + \exp\left(-(a_0 + a_1 \cdot \Delta^{sign}_{return} + a_2 \cdot \Delta^{sign}_{reversal} + a_3 \cdot \Delta^{sign}_{frequency} + a_4 \cdot \Delta^{sign}_{spread})\right)\right)^{-1}, \quad (3.4)$$

where $a_0 = -7.5$, $a_1 = 4.2$, $a_2 = 3.6$, $a_3 = 8, 5$, and $a_4 = 1.8$ as in Comerton-Forde and Putnins (2011b). The manipulation measure in Eq. (3.4) captures the idea that abnormal daily return, reversal, trading frequency, and bid-ask spread can differentiate manipulated closing prices from non-manipulated ones. Table A3.1 in the Appendix shows the detailed description and calculation of the market manipulation measure.

3.4.3 Measure of investment-to-price sensitivity

To measure investment-to-price sensitivity, we use the baseline equation used in Foucault and Frésard (2012, 2014) and Chen, Goldstein, and Jiang (2007),

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot M_{i,t} + \beta_3 \cdot M_{i,t} \cdot Q_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1}, \quad (3.5)$$

where $I_{i,t+1}$ represents firm i's investment in year $t+1$ and α_i and γ_t are firm and year fixed effects, respectively. $Q_{i,t}$ is the normalized stock price for firm i in year t computed as the market value of equity plus the book value of assets minus the book value of equity, scaled by the book value of asset. $M_{i,t}$ is the average probability of price manipulation for stock i in year t.[45] We include cash flow ($CF_{i,t}$) and $\ln(Asset_{i,t})$ to control for the effects of cash flow and total assets on firms' investments.

We use three proxies for investment $I_{i,t+1}$ in Eq. (3.5): (i) capital expenditure in year $t+1$ scaled by total assets in year t ($CAPEX_{i,t+1}$), (ii) the sum of capital expenditure and R&D expenses in year $t+1$ scaled by total assets in year t ($CAPEXRND_{i,t+1}$), and (iii) the percentage change from year t to $t+1$ in the assets' book value ($CHGASSET_{i,t+1}$). These measures show different aspects of firms' investment activities. $CAPEX_{i,t+1}$ and $CAPEXRND_{i,t+1}$ is the direct measure of firms' ongoing investment and R&D activities and $CHGASSET_{i,t+1}$ shows firms' acquisition and divestiture activities. In all estimations of the model, we double cluster standard errors by firm and year following Petersen (2009).

It follows from Eq. (3.5) that the marginal effect of price manipulation ($M_{i,t}$) on the amount of investment is given by $\frac{\partial I_{i,t+1}}{\partial M_{i,t}} = \beta_2 + \beta_3 \cdot Q_{i,t}$ (Result 1 in the model). However, our main focus in Eq. (3.5) is the impact of stock price manipulation on the sensitivity of investment to stock price, captured by the coefficient $\beta_3 = \frac{\partial}{\partial M}(\frac{\partial I_{i,t+1}}{\partial Q_{i,t}})$. The coefficient β_3 measures the extent to which stock price manipulation impacts firms' investment-to-price sensitivity. We expect β_3 to be negative and significant (Result 2 in the model). Table 3.1 reports the complete list of variables, their descriptions, and sources (main model variables in Panel A and supplementary variables in Panel B).

[45] We calculate the book value of equity as the sum of the stockholder's equity (Compustat item SEQ) and balance sheet deferred taxes and investment tax credit (Compustat item $TXDITC$) minus preferred stocks. The preferred stocks value is the first non-missing value of redemption (Compustat item $PSTKRV$), liquidation ($PSTKL$), or par value ($PSTK$) in that order following Fama and French (1993).

Table 3.1. Variable definitions

This table reports the definitions (notation in the first column, description in the second column, and data source in the third column) of variables. Panel A presents the main variables used in the model, and Panel B presents the supplementary variables.

Variable	Descripion	Data source
Panel A: Main model variables		
$CAPEX_{i,t+1}$ (%)	Capital expenditure for firm i in year $t+1$ scaled by total assets in year t.	Compustat
$CAPEXRND_{i,t+1}$ (%)	Capital expenditure plus R&D for firm i in year $t+1$ scaled by total assets in year t.	Compustat
$CHGASSET_{i,t+1}$(%)	Change in total book value of assets for firm i from year t to year $t+1$.	Compustat
$ROA_{i,t+1}$ (%)	Return on asset for firm i in year $t+1$ is earnings before interest, tax depreciation, and amortization ($EBITDA_{i,t+1}$) as a percentage of total assets.	Compustat
$ATO_{i,t+1}$ (%)	Asset turnover for firm i in year $t+1$ is total revenue as a percentage of total assets.	Compustat
$Q_{i,t}$	The normalized stock price for firm i in year t is the market value of equity plus the book value of assets minus the book value of equity, scaled by the book value of assets.	Compustat
$M_{i,t}$	Yearly average probability of daily closing price manipulation for firm i during the year t. The variable is computed based on the probability of stock price manipulation using stock returns, spreads, trading frequencies and return reversals (see Appendix).	RTH
$CF_{i,t}$ (%)	Cash flow of firm i in year t scaled by book value of assets.	Compustat
$\ln(Asset_{i,t})$	Natural logarithm of the total assets for firm i in year t.	Compustat
Panel B: Supplementary variables		
$AvgM_{i,t}$	Average yearly probability of stock price manipulation of stock i's size group in year t. We split stocks into quartiles based on their market capitalizations. $AvgM_{i,t}$ is calculated as the average $M_{i,t}$ for each size group in year t.	RTH
RB_t	Annual SEC regulatory budget in year t.	SEC
$RSpread_{i,t}$ (in bps)	Yearly average of the daily relative spreads for stock i in year t. The daily relative spread is computed as the difference between closing ask and bid prices for each day divided by the midpoint of closing ask and bid prices.	CRSP
$Volume_{i,t}$ ($\div 10^9$ USD)	Yearly average of the daily stock price multiplied by the daily volume for stock i in year t.	CRSP
$MktCap_{i,t}$ ($\div 10^6$ USD)	Yearly average the month-end market capitalization for stock i in year t. The month-end market capitalization for stock i is calculated as the number of outstanding shares multiplied by the share price for stock i at the end of each month.	CRSP
$Volatility_{i,t}$ (%)	Standard deviation of the daily percentage returns for stock i in year t.	CRSP
PIN	The probability of informed trading for stock i in year t is computed using the maximum likelihood estimation (MLE) following Easley, Hvidkjaer and O'Hara (2002). The number of buy and sell trades in each 5-minute interval is classified using the bulk volume classification (BVC) following Easley, López de Prado and O'Hara (2012).	RTH
$ACoverage_{i,t}$	Yearly average of the monthly number of analysts issuing EPS forecasts for the next year of firm i in year t.	I/B/E/S
$Insider_{i,t}$ (%)	The ratio of insider transactions of firm i in year t to the total number of transactions.	RTH / CRSP
$KZ_{i,t}$	The weighted sum of cash flow, cash dividends, cash balances, and leverage ratio following Baker, Stein, and Wurgler (2003).	Compustat

3.4.4 Descriptive statistics

Table 3.2 presents the descriptive statistics for the main and supplementary variables. Specifically, firms have approximately 5.80% of their total assets as capital expenditure ($CAPEX_{i,t+1}$) in the upcoming year, with a standard deviation of 6.49%. The capital expenditure, combined with research and development ($CAPEXRND_{i,t+1}$), amounts to 11.17%, with a standard deviation of 11.70%. On average, firms experience an annual increase of 13.12% in the value of their total assets during the sample period, with a standard deviation of 35.87%. The average annual cash flow of firms accounts for approximately 13% of total assets, with a standard deviation of 15%. On average, firms achieve a 9% return on assets ($ROA_{i,t+1}$) and a 99% asset turnover ($ATO_{i,t+1}$) annually in terms of operating performance.

We split firms into two subsamples (high and low) based on whether their manipulation measure ($M_{i,t}$) is above or below the median value to analyze the factors affecting the likelihood of becoming a target for manipulators. We assess the significance of the differences in each characteristic between the high and low manipulation subsamples using a paired t-test.

As reported in Panel A of Table 3.2, there are significant differences in the investment ($CAPEX_{i,t+1}$, $CAPEXRND_{i,t+1}$, $CHGASSET_{i,t+1}$) and operating performance ($ATO_{i,t+1}$) of firms with high and low stock price manipulation. Firms with high stock price manipulation tend to over-invest in the upcoming year in terms of all three investment measures and perform poorly in terms of $ATO_{i,t+1}$. Specifically, the average $CAPEX_{i,t+1}$ for firms with high stock price manipulation (6.26%) is approximately 17% greater than that of firms with low stock price manipulation (5.33%). Similarly, the average $CAPEXRND_{i,t+1}$ and $CHGASSET_{i,t+1}$ for firms with high-manipulated stocks are 17% and 59% higher than those of firms with low-manipulated stocks, respectively. The difference in $ATO_{i,t+1}$ indicates that firms with high-manipulated stocks perform poorly in the upcoming year. Furthermore, firms with high-manipulated stocks tend to have higher stock prices and cash flows.

Panel B in Table 3.2 compares additional characteristics of firms with high and low-manipulated stocks, including liquidity as measured by relative spread and dollar volume, market capitalization, volatility, private and public information environment of stocks as measured by PIN and analyst coverage, respectively. The results indicate that firms with high stock price manipulation tend to have more liquid (in terms of bid-ask spread and trading volume) and more volatile stocks, and lower level of insider trading activities than firms with low stock price manipulation. There are no significant differences in the market capitalization, PIN, and analyst coverage of firms with high and low stock price manipulation.

Table 3.2. Descriptive statistics
This table reports the summary statistics across all stocks and the differences in mean characteristics of high (M_{high}) versus low manipulated stocks (M_{low}). Panel A reports the summary statistics of the main variables used in the model. Panel B reports the summary statistics of the supplementary variables. Stocks are split into high and low groups depending on whether their average yearly probability of stock price manipulation is above (high) or below (low) the median value each year. The significance of the difference in means is computed with t-statistics in the last column. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors following Petersen (2009). ***, **, and * indicate statistical significance levels of 1%, 5%, and 10%, respectively. The unit of observations is firm-year.

Panel A: Main variables							
Variable	**Full sample**			$\mathbf{M_{high}}$	$\mathbf{M_{low}}$	**Mean difference**	
	Mean	Median	Std.dev.	Mean	Mean	M_{high}-M_{low}	t stat.
$CAPEX_{i,t+1}$ (%)	5.80	3.68	6.49	6.26	5.33	0.93***	7.00
$CAPEXRND_{i,t+1}$ (%)	11.17	7.50	11.70	12.03	10.30	1.73***	9.13
$CHGASET_{i,t+1}$ (%)	13.12	5.53	35.87	16.13	10.12	6.01***	8.12
$ROA_{i,t+1}$ (%)	8.99	11.11	15.02	9.11	8.86	0.25	0.52
$ATO_{i,t+1}$ (%)	99.16	82.73	72.89	97.15	101.17	-4.02**	-2.14
$Q_{i,t}$	1.99	1.50	1.46	2.19	1.78	0.40***	6.63
$M_{i,t}$	0.45	0.44	0.18	0.59	0.31	0.28***	51.66
$CF_{i,t}$ (%)	13.41	8.15	15.08	13.74	13.07	0.67**	2.29
$\ln(Asset_{i,t})$ (%)	6.55	6.43	1.99	6.61	6.50	0.11	1.24
N obs.	52,886			26,447	26,439		
Panel B: Supplementary variables							
$AvgM_{i,t}$	0.45	0.46	0.05	0.46	0.45	0.01***	5.45
$RB_t \cdot RSpread_{i,t}$ ($\div 10^6$)	91.19	34.11	139.96	84.34	98.08	-13.74*	1.76
$RSpread_{i,t}$(in bps)	121.81	44.52	173.21	108.31	135.41	-27.10***	-2.64
$Volume_{i,t}$ ($\div 10^9$ USD)	6.60	0.66	16.04	7.01	6.20	0.81***	7.04
$MktCap_{i,t}$ ($\div 10^6$ USD)	3.31	0.45	8.85	3.29	3.34	-0.05	-0.75
$Volatility_{i,t}$ (%)	3.61	2.79	3.99	3.64	3.57	0.07**	2.18
$PIN_{i,t}$	0.18	0.16	0.10	0.181	0.177	0.003	1.11
$ACoverage_{i,t}$	7.53	5.25	6.89	7.70	7.38	0.32	0.80
$Insider_{i,t}$ (%)	0.14	0.01	0.66	0.11	0.16	-0.05***	-6.29
$KZ_{i,t}$	-0.16	0.06	3.81	-0.20	-0.13	-0.06*	-1.86

3.5 Empirical analysis

3.5.1 Market manipulation and investment-Q sensitivity

3.5.1.1 Ordinary least squares (OLS)

Table 3.3 presents three versions of the model in Eq. (3.5), using three different investment measures ($CAPEX_{i,t+1}$, $CAPEXRND_{i,t+1}$, and $CHGASSET_{i,t+1}$), to explore the impact of stock price manipulation on investment-Q sensitivity. Consistent with prior literature (e.g., Chen, Goldstein, and Jiang 2007; Foucault and Frésard 2012), Table 3.3 shows a positive and significant relationship between firms' investment activity ($I_{i,t+1}$) and their stock price ($Q_{i,t}$).

Table 3.3. Market manipulation and investment-Q sensitivity: OLS
This table reports the impact of market manipulation on investment-Q sensitivity using the following OLS model:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot M_{i,t} + \beta_3 \cdot M_{i,t} \cdot Q_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1},$$

where $I_{i,t+1}$ is the firm i's investment in year $t+1$, defined as $CAPEX_{i,t+1}$ (capital expenditures divided by lagged total assets), $CAPEXRND_{i,t+1}$ (capital expenditures plus R&D expenses divided by lagged total assets), and $CHGASSET_{i,t+1}$ (annual change in total assets divided by lagged total assets). $Q_{i,t}$ is the normalized stock price, $M_{i,t}$ is the average probability of stock price manipulation, $CF_{i,t}$ is the cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of total assets. Refer to Table 3.1. for the definitions of variables. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. All versions of the model include firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	$\mathbf{CAPEX_{i,t+1}}$	$\mathbf{CAPEXRND_{i,t+1}}$	$\mathbf{CHGASSET_{i,t+1}}$
$Q_{i,t}$	1.14***	2.23***	10.94***
	(11.32)	(13.13)	(11.53)
$M_{i,t}$	1.99***	2.24***	9.68***
	(5.40)	(4.60)	(3.70)
$\mathbf{M_{i,t} \cdot Q_{i,t}}$	**-0.51*** **	**-0.85*** **	**-2.91** **
	(-4.28)	**(-3.55)**	**(-2.48)**
$CF_{i,t}$	-0.00	0.02***	0.01
	(-0.37)	(2.67)	(0.53)
$\ln(Asset_{i,t})$	-1.40***	-3.36***	-17.37***
	(-17.28)	(-20.40)	(-18.57)
N obs.	52,865	52,865	52,865
R^2	9%	17%	17%

The coefficients of manipulation ($M_{i,t}$) and the interaction term ($M_{i,t} \cdot Q_{i,t}$) combined enable us to calculate the overall effect of manipulative activities in a firm's stock on its investment. The overall effect of average stock price manipulation on $CAPEX_{i,t+1}$ is positive and equals to 0.98 ($\beta_2 + \beta_3 \cdot Average(Q) = 1.99 - 0.51 \cdot 1.99$). That means, all else equal, a 1% change in the probability of manipulation increases $CAPEX_{i,t+1}$ by around 0.98%. Alternatively, all else equal, a 1% change in the probability of manipulation increases $CAPEXRND_{i,t+1}$ and $CHGASSET_{i,t+1}$ by 0.55% and 3.89%, respectively.

Our main focus is whether stock price manipulation diminishes the ability of firm managers to learn from prices, as indicated by the interaction term coefficient ($M_{i,t} \cdot Q_{i,t}$). We find that in all versions of the model, the coefficient of the interaction term is negative and statistically significant. This indicates that investment-Q sensitivity is lower for firms whose stock prices are more manipulated. In other words, firm managers are less inclined to use stock prices as guidance for their investment decisions when manipulation is more prevalent in their stocks.

To calculate the economic impact, consider a one standard deviation shock to the stock price $Q_{i,t}$. In the absence of manipulation, this shock corresponds to an average increase of 1.66% ($\beta_1 \cdot StdDev(Q) = 1.14 \cdot 1.46$) in $CAPEX_{i,t+1}$. However, taking manipulation ($M_{i,t}$) into account, this increase is reduced to 1.33% ($\beta_1 \cdot StdDev(Q) + \beta_3 \cdot StdDev(Q) \cdot Average(M) = 1.14 \cdot 1.46 - 0.51 \cdot 1.46 \cdot 0.45$), representing a decrease of approximately 20% compared to the initial increase. Alternatively, when investment is measured by $CAPEXRND_{i,t+1}$ and $CHGASSET_{i,t+1}$, a one standard deviation shock to $Q_{i,t}$ results in an increase of 3.26% and 15.97% in $CAPEXRND_{i,t+1}$ and $CHGASSET_{i,t+1}$, respectively. However, accounting for stock price manipulation reduces the increase in $CAPEXRND_{i,t+1}$ to 2.70% (17% decrease), and in $CHGASSET_{i,t+1}$ to 14.06% (12% decrease).

3.5.1.2 Two-stage least squares (2SLS)

The findings in the previous subsection support our main hypothesis that market manipulation reduces firms' investment-to-price sensitivity. However, it is possible that the relationship between market manipulation and investment is not causal. An exogenous shock on investment can change the incentives for manipulators, resulting in a change in the level of manipulation. This mechanism can introduce a correlation between the market manipulation measure and the error term of Equation (3.5), rendering the OLS model inconsistent.

To establish the causal effect of stock price manipulation on investment-Q sensitivity, we re-estimate Eq. (3.5) using two stage least square (2SLS) regression with three instrumental variables for stock price manipulation: (i) average manipulation in all other stocks in the corresponding size quartile (based on market capitalization) $AvgM_{i,t}$ (Model 1), (ii) lagged manipulation $M_{i,t-1}$ (Model 2), and (iii) the interaction of the SEC regulatory budget with the relative spread $RB_t \cdot RSpread_{i,t}$ (Model 3). All three instruments are correlated with the manipulation measure ($M_{i,t}$) and uncorrelated with the error term ($\varepsilon_{i,t}$), meaning that they have no direct effects on investment ($I_{i,t+1}$).

The first instrument follows from Hasbrouck and Saar (2013) and instruments the manipulation in each stock-year using the average manipulation of all other stocks in the corresponding size quartile ($AvgM_{i,t}$). The manipulation level in other firms' stocks is unlikely to be driven by the investment activities of the given firm's stock. In addition, the manipulation in other stocks in the same size group is correlated with the manipulation level of a particular stock in that group (95% confidence interval for the pooled Pearson correlation coefficient is 0.249–0.265). Thus, $AvgM_{i,t}$ satisfies requirements of a strong instrument.

The second instrument, the lagged value of manipulation ($M_{i,t-1}$), is common practice in economics (see, for example Chapter 12 in Greene 2008). In estimations, $M_{i,t-1}$ is not part of the second stage of the estimation, but it is sufficiently correlated with the endogenous manipulation measure, making it an effective instrument. Furthermore, lagged manipulation ($M_{i,t-1}$) can hardly be driven by firms' two year ahead investment activities ($I_{i,t+1}$).

As the third instrument, we interact the SEC regulatory budget and the stock's liquidity ($RB_t \cdot RSpread_{i,t}$). The regulatory budget is negatively correlated with manipulation and has no association with firms' investment activities. By incorporating the interaction with stock liquidity, we account for regulators' tendency to focus their attention on stocks with lower liquidity in order to detect stock price manipulation (e.g., Comerton-Forde and Putnins 2014). To confirm the validity of all three instruments, we use an F-test in the first stage regression. Our results strongly reject the null hypothesis that the instruments are not the determinants of manipulation ($M_{i,t}$).

We estimate nine versions of 2SLS models using three instruments ($AvgM_{i,t}$, $M_{i,t-1}$, and $RB_t \cdot RSpread_{i,t}$) and three investment measures ($CAPEX_{i,t+1}$, $CAPEXRND_{i,t+1}$, and $CHGASSET_{i,t+1}$). The models are identical, except different instrumental variables are used in the first stage and different investment measures are used in the second stage. Table 3.4 reports the results of the second stage regression for all 2SLS models. The findings in Table 3.4 support the OLS results, indicating that stock price manipulation reduces firms' investment-to-price sensitivity.

Table 3.4. Market manipulation and investment-Q sensitivity: 2SLS

The table reports the impact of market manipulation on investment-Q sensitivity using the following model:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot \hat{M}_{i,t} + \beta_3 \cdot \hat{M}_{i,t} \cdot Q_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1},$$

where $I_{i,t+1}$ is the firm i's investment in year $t+1$, defined as $CAPEX_{i,t+1}$ (capital expenditures divided by lagged total assets), $CAPEXRND_{i,t+1}$ (capital expenditures plus R&D expenses divided by lagged total assets), and $CHGASSET_{i,t+1}$ (annual change in total assets divided by lagged total assets). $Q_{i,t}$ is the normalized stock price, $CF_{i,t}$ is the cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of total assets. $\hat{M}_{i,t}$ is the predicted value of the average yearly probability of stock price manipulation ($M_{i,t}$) based on three instrumental variables. The instrumental variable in Models 1 to 3 is the average manipulation by stock i's size quartile group ($AvgM_{i,t}$), the lagged manipulation ($M_{i,t-1}$), and the SEC regulatory budget (RB_t) interacted with the relative spread ($RSpread_{i,t}$), respectively. Refer to Table 3.1 for the definitions of all variables. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. All versions of models include firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	$\mathbf{CAPEX_{i,t+1}}$			$\mathbf{CAPEXRND_{i,t+1}}$			$\mathbf{CHGASSET_{i,t+1}}$		
	Model 1 $\mathbf{AvgM_{i,t}}$	**Model 2** $\mathbf{M_{i,t-1}}$	**Model 3** $\mathbf{RB_t \cdot RSpread_{i,t}}$	**Model 1** $\mathbf{AvgM_{i,t}}$	**Model 2** $\mathbf{M_{i,t-1}}$	**Model 3** $\mathbf{RB_t \cdot RSpread_{i,t}}$	**Model 1** $\mathbf{AvgM_{i,t}}$	**Model 2** $\mathbf{M_{i,t-1}}$	**Model 3** $\mathbf{RB_t \cdot RSpread_{i,t}}$
$Q_{i,t}$	3.46***	1.82***	1.36***	4.92***	3.50***	2.52***	26.00***	20.50***	16.34***
	(13.36)	(7.73)	(5.48)	(10.26)	(9.67)	(7.10)	(9.81)	(9.97)	(10.05)
$\hat{M}_{i,t}$	7.58***	24.65***	25.23***	9.67***	12.68	29.45***	65.12***	19.9	90.30***
	(6.36)	(4.00)	(5.16)	(4.72)	(1.35)	(4.10)	(7.60)	(0.49)	(3.16)
$\mathbf{\hat{M}_{i,t} \cdot Q_{i,t}}$	**-3.63*** **	**-4.05*** **	**-3.64*** **	**-4.46*** **	**-5.00*** **	**-4.56*** **	**-24.16*** **	**-27.72*** **	**-25.70*** **
	(-11.41)	**(-13.30)**	**(-10.90)**	**(-7.31)**	**(-8.39)**	**(-7.83)**	**(-7.76)**	**(-10.26)**	**-10.86)**
$CF_{i,t}$	-0.00	-0.01**	-0.02***	0.01**	0.01	-0.00	-0.01	0.03	-0.03
	(-0.86)	(-2.10)	(-3.55)	(2.10)	(1.22)	(-0.07)	(-0.31)	(0.82)	(-0.93)
$\ln(Asset_{i,t})$	-1.38***	-1.52***	-1.71***	-3.36***	-3.11***	-3.93***	-17.68***	-15.64***	-18.50***
	(-15.79)	(-9.91)	(-11.87)	(-22.31)	(-12.00)	(-14.39)	(-20.21)	(-12.12)	(-14.78)
N obs.	52,865	47,225	46,028	52,865	47,225	46,028	52,865	47,225	46,028
R^2	10%	9%	10%	17%	16%	18%	18%	16%	18%

In the 2SLS setting, the coefficient of the interaction term is higher than that of the OLS model. For example, the magnitude of the interaction coefficient in Model 1 is 7 times bigger for $CAPEX_{i,t+1}$ (-3.63/-0.51), 5 times bigger for $CAPEXRND_{i,t+1}$, and 8 times bigger for $CHGASSET_{i,t+1}$. To further examine this, we compare the economic impacts of market manipulation in all models (three OLS and nine 2SLS models). Table 3.5 presents the effect of a one standard deviation shock to $Q_{i,t}$ on firms' investments for different levels of manipulation.

Table 3.5. Different levels of market manipulation and investment-Q sensitivity
This table presents the effects of a one standard deviation shock to stock price ($Q_{i,t}$) on firms' investment at different level of stock price manipulation (M),

$$I = \beta_1 \cdot \sigma_Q + \beta_3 \cdot \sigma_Q \cdot M,$$

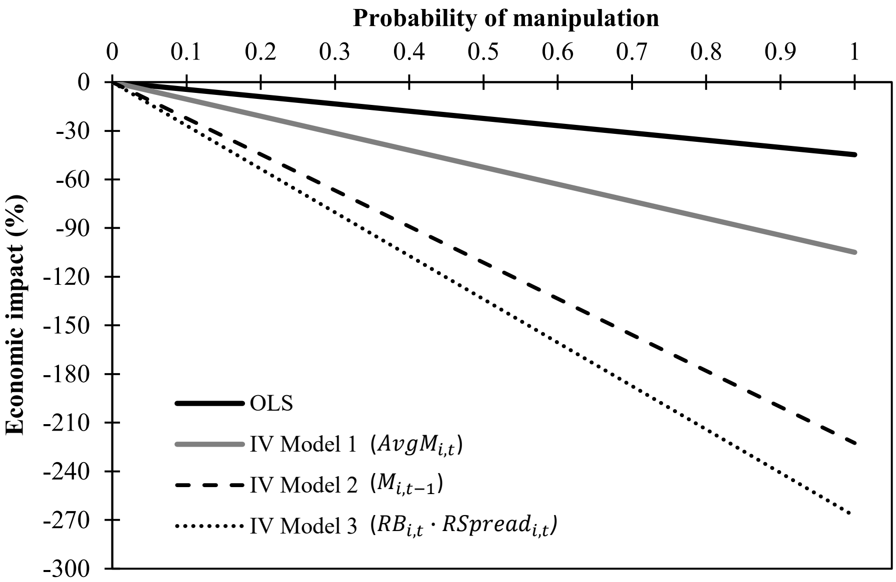
where I is measured by $CAPEX$, $CAPEXRND$, and $CHGASSET$ in Columns (i), (ii), and (iii) of each model, respectively. β_1 is the coefficient of Q, σ_Q is the standard deviation of Q, and β_3 is the interaction coefficient between M and Q. The first column shows different levels of market manipulation. The remaining of the table shows the effect of a one standard deviation shock to Q on investment I with different levels of market manipulation. The results of the OLS Model is calculated based on the coefficients from Table 3.3 and IV Models 1-3 are based on the coefficients from Table 3.4. In IV Models 1-3, the instrumental variables are the average manipulation by the size quartile of stock i ($AvgM_{i,t}$), the lagged manipulation ($M_{i,t-1}$), and the SEC regulatory budget (RB_t) interacted with the relative spread ($RSpread_{i,t}$), respectively. The last row shows the results for the average probability of manipulation ($Average(M)$) in different models. $Average(M)$ is 0.45 in OLS, 0.57 in IV Model 1, 0.23 in IV Model 2, and 0.20 in IV Model 4.

	OLS Model			**IV Model 1**			**IV Model 2**			**IV Model 3**		
M	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
0	1.66	3.26	15.97	5.05	7.18	37.96	2.51	4.83	28.29	2.03	3.75	24.35
0.1	1.59	3.13	15.55	4.52	6.53	34.43	1.95	4.14	24.46	1.48	3.08	20.52
0.2	1.52	3.01	15.12	3.99	5.88	30.91	1.39	3.45	20.64	0.94	2.40	16.69
0.3	1.44	2.88	14.70	3.46	5.23	27.38	0.83	2.76	16.81	0.40	1.72	12.86
0.4	1.37	2.76	14.27	2.93	4.58	23.85	0.28	2.07	12.99	-0.14	1.04	9.03
0.5	1.29	2.64	13.85	2.40	3.93	20.32	-0.28	1.38	9.16	-0.69	0.36	5.20
0.6	1.22	2.51	13.42	1.87	3.28	16.80	-0.84	0.69	5.34	-1.23	-0.32	1.37
0.7	1.14	2.39	13.00	1.34	2.63	13.27	-1.40	0.00	1.51	-1.77	-1.00	-2.46
0.8	1.07	2.26	12.57	0.81	1.97	9.74	-1.96	-0.69	-2.31	-2.31	-1.68	-6.29
0.9	0.99	2.14	12.15	0.28	1.32	6.21	-2.52	-1.38	-6.14	-2.85	-2.36	-10.12
1	0.92	2.01	11.72	-0.25	0.67	2.69	-3.08	-2.07	-9.96	-3.40	-3.04	-13.95
Average(M)	**1.33**	**2.70**	**14.06**	**2.03**	**3.47**	**17.85**	**1.23**	**3.24**	**19.49**	**0.94**	**2.40**	**16.69**

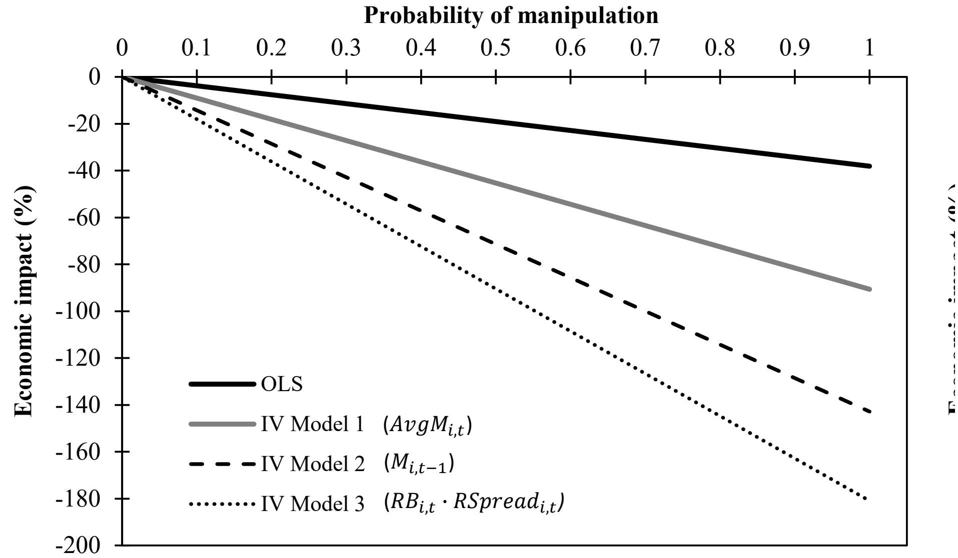
On average, in Model 1, a one standard deviation shock to stock price is associated with 5.05% ($\beta_1 \cdot StdDev(Q) = 3.46 \cdot 1.46$) increase in $CAPEX_{i,t+1}$ in the absence of manipulation. However, considering the average level of manipulation ($\hat{M}_{i,t} = 0.57$), this increase is reduced to 2.03% ($\beta_1 \cdot StdDev(Q) + \beta_3 \cdot StdDev(Q) \cdot Average(\hat{M}) = 3.46 \cdot 1.46 - 3.13 \cdot 1.46 \cdot 0.57$), resulting in a reduction of approximately 60% compared to the initial increase. The use of the lagged value of manipulation ($M_{i,t-1}$) and the interaction between the SEC regulatory budget and the stock's liquidity ($RB_t \cdot RSpread_{i,t}$) decrease the sensitivity of $CAPEX_{i,t+1}$ to price shocks by 51% and 54%, respectively. Similarly, on average, the sensitivity of $CAPEXRND_{i,t+1}$ (resp. $CHGASSET_{i,t+1}$) to a one standard deviation price shock in Models 1, 2, and 3 is reduced by the average predicted manipulation by 52%, 33%, and 36% (resp. 53%, 31%, and 31%), respectively.

Figure 3.1. The impact of market manipulation on investment-Q sensitivity
This figure graphically illustrates the economic magnitude of the impact of different levels of stock price manipulation on the effect of a one standard deviation shock to Q on firms' investment in four different models (OLS and IV Models 1-3). In IV Models 1-3, the instrument is the average manipulation by stock i's size quantile group ($AvgM_{i,t}$), lagged manipulation ($M_{i,t-1}$), and the SEC regulatory budget (RB_t) interacted with the relative spread ($RSpread_{i,t}$), respectively. Investment is measured by $CAPEX$, $CAPEXRND$, and $CHGASSET$ from Panel A to C, respectively.

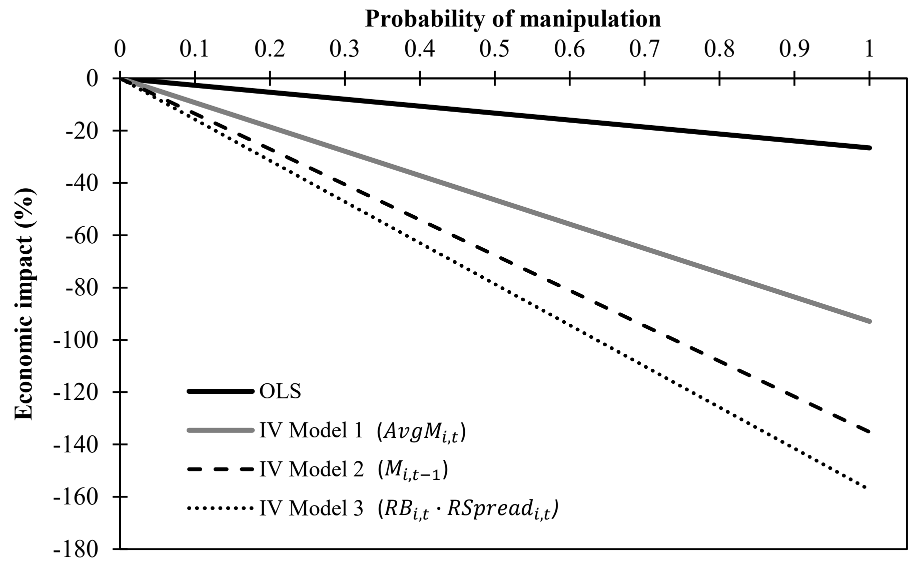
Panel A. The impact of different probabilities of market manipulation on the effect of a one standard deviation shock to price (Q) on $CAPEX$



Panel B. The impact of different probabilities of market manipulation on the effect of a one standard deviation shock to price (Q) on $CAPEXRND$



Panel C. The impact of different probabilities of market manipulation on the effect of a one standard deviation shock to price (Q) on $CHGASSET$



In addition to the average economic effects of market manipulation on investment-Q sensitivity, Figure 3.1 graphically illustrates the investment-Q sensitivity for $CAPEX_{i,t+1}$, $CAPEXRND_{i,t+1}$, and $CHGASSET_{i,t+1}$ for different stock price manipulation levels (betwen 0 and 1). The figure demonstrates a stronger negative impact of market manipulation on investment-Q sensitivity in the 2SLS models (Models 1-2-3 in ascending order).[46] Given the simplicity of OLS and conservative results and better performance of $AvgM_{i,t}$ as an instrument (from the first stage results), we focus on the OLS and Model 1 for subsequent analysis.[47]

3.5.2 Cross-sectional analysis

In this subsection, we analyze the influence of various stock characteristics on the relationship between market manipulation and investment-to-price sensitivity. We test this by allocating stocks into high (above-median) and low (below-median) groups each year based on stock characteristics. We create a dummy variable D_h that equals one if a firm's stock belongs to the high group according to each of the following measures: (i) relative spread ($RSpread$), (ii) dollar volume ($Volume$), (iii) market capitalization ($MktCap$), (iv) volatility ($Volatility$), (v) private information as measured by PIN, and (vi) public information ($ACoverage$). Subsequently, we re-estimate the OLS model in Eq. (3.5) by including D_h, $Q_{i,t} \cdot D_h$, $M_{i,t} \cdot D_h$, and $M_{i,t} \cdot Q_{i,t} \cdot D_h$ for each of the six stock characteristic dummies as

$$\begin{aligned} I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot M_{i,t} + \beta_3 \cdot M_{i,t} \cdot Q_{i,t} + \beta_4 \cdot D_h + \beta_5 \cdot Q_{i,t} \cdot D_h \\ + \beta_6 \cdot M_{i,t} \cdot D_h + \beta_7 \cdot M_{i,t} \cdot Q_{i,t} \cdot D_h + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1}. \end{aligned} \tag{3.6}$$

The coefficient of interest in Eq. (3.6) is β_7, which represents the difference in the impact of manipulation on investment-to-price sensitivity in the high group compared to the low group. A positive (resp. negative) β_7 indicates that the negative impact of manipulation on investment-to-price sensitivity is amplified in the low (resp. high) group. The results presented in Table 3.6 show significant differences in the $CAPEXRND$-to-price sensitivity across stock characteristics.[48] The negative impact of manipulation on $CAPEXRND$-to-price sensitivity is amplified when stocks are illiquid (high relative spread) and more volatile, suggesting that market manipulation have more tangible effects on managerial learning in illiquid and volatile stocks. This is intuitive because a key element to a successful manipulation is to move stock prices and it is hard for a manipulator to move the prices of highly liquid stocks by any significant amount.

[46] Besides the average negative effects of market manipulation on investment-to-price sensitivity, the figure also shows hypothetical extreme scenarios where a manager, becoming certain of price manipulation, makes an investment decision that contradicts the indications of the stock price. In such cases, the investment-to-price sensitivity can decline by more than 100%.

[47] We report the results with other instrumental variables in the Appendix 3.1.

[48] The results for $CAPEX$ and $CHGASSET$ are reported in Table A3.5 of the Appendix 3.1.

Table 3.6. The impact of market manipulation on investment-Q sensitivity by stock characteristics

This table presents the impact of market manipulation on investment-Q sensitivity by stock characteristics:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot M_{i,t} + \beta_3 \cdot M_{i,t} \cdot Q_{i,t} + \beta_4 \cdot D_h + \beta_5 \cdot Q_{i,t} \cdot D_h$$
$$+ \beta_6 \cdot M_{i,t} \cdot D_h + \beta_7 \cdot M_{i,t} \cdot Q_{i,t} \cdot D_h + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1}.$$

where $I_{i,t+1}$ is firm i's investment in year $t+1$, defined as the capital expenditures plus R&D expenses divided by lagged total assets ($CAPEXRND_{i,t+1}$). Across all specifications, $Q_{i,t}$ is the firm's normalized stock price, $M_{i,t}$ is the average yearly probability of stock price manipulation, $CF_{i,t}$ is firm's cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of the firm's total assets. D_h is a dummy variable that equals one if the value of a given stock characteristic is above the median value of all other stocks each year. From Columns (i) to (vi), D_h is defined based on the relative spread, aggregate yearly dollar volume, market capitalization, the standard deviation of returns, private information measured by PIN, and analyst coverage, respectively. For the definitions of all variables, refer to Table 1. All versions of the model include both firm and year-fixed effects (α_i and γ_t, respectively). The standard errors used to compute the t-statistics (in brackets) are double clustered by firm and year. *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	$\mathbf{D_h}$					
	(i) **RSpread**	(ii) **Volume**	(iii) **MktCap**	(iv) **Volatility**	(v) **PIN**	(vi) **ACoverage**
$Q_{i,t}$	1.61***	2.52***	2.68***	1.68***	2.02***	2.34***
	(8.43)	(10.74)	(11.52)	(10.86)	(9.38)	(9.64)
$M_{i,t}$	0.75	1.01	1.22**	1.79***	2.21***	2.12***
	(1.37)	(1.57)	(2.05)	(3.29)	(3.18)	(3.58)
D_h	-2.64***	2.03***	2.35***	-1.46***	-0.79*	1.20**
	(-5.92)	(4.72)	(5.80)	(-3.82)	(-1.78)	(2.26)
$M_{i,t} \cdot Q_{i,t}$	-0.27	-0.44	-0.70**	-0.43*	-0.68**	-0.90**
	(-1.07)	(-1.19)	(-2.15)	(-1.89)	(-2.08)	(-2.45)
$M_{i,t} \cdot D_h$	1.11	0.16	0.33	0.48	0.74	-0.36
	(1.49)	(0.21)	(0.44)	(0.83)	(0.88)	(-0.39)
$Q_{i,t} \cdot D_h$	1.04***	-0.82***	-1.07***	0.81***	0.63***	-0.44*
	(4.30)	(-3.64)	(-4.79)	(4.06)	(2.62)	(-1.72)
$\mathbf{M_{i,t} \cdot Q_{i,t} \cdot D_h}$	**-0.67***	**-0.02**	**0.19**	**-0.67****	**-0.83***	**0.25**
	(-1.68)	**(-0.06)**	**(0.53)**	**(-2.05)**	**(-1.83)**	**(0.60)**
$CF_{i,t}$	0.01	0.01	0.01	0.01	0.01	0.01*
	(1.35)	(1.26)	(1.26)	(1.62)	(0.81)	(1.74)
$\ln(Asset_{i,t})$	-3.63***	-3.59***	-3.60***	-3.53***	-3.62***	-3.81***
	(-19.20)	(-20.02)	(-20.66)	(-20.50)	(-20.30)	(-22.30)
N obs.	46,028	46,028	46,034	46,034	40,241	39,968
R^2	18%	18%	18%	18%	17%	19%

Thus, manipulation can lead to more tangible results in terms of managerial learning in illiquid and volatile stocks. In addition, consistent with the managerial learning, the negative impact of market manipulation on investment-to-price sensitivity is more pronounced when stocks convey more private information (as measured by PIN) to firm managers. We do not find significant differences in the investment-to-price sensitivity across dollar volume, market capitalization, and analyst coverage.

3.5.3 Market manipulation and operating performance

The existing literature suggests that managers' investment decisions are more responsive to stock prices when prices incorporate more private information (e.g., Chen, Goldstein, and Jiang, 2007). This indicates that higher-quality information embedded in stock prices enables managers to make more informed decisions, leading to increased investment efficiency. However, if manipulative actions distort stock prices and reduce investment-to-price sensitivity, it is possible that capital allocation becomes less efficient. If stock price manipulation impedes investment efficiency, the future operating performance of firms is likely to suffer (Result 3 in the theoretical model). Therefore, we investigate the relationship between stock price manipulation and the future operating performance of firms.

To investigate the relationship between stock price manipulation and future firm performance, we modify Eq. (3.5) by using return on assets ($ROA_{i,t+1}$) and asset turnover ($ATO_{i,t+1}$) as the dependent variables:

$$OP_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot \hat{M}_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + OP_{i,t} + \varepsilon_{i,t+1}, \quad (3.7)$$

where $OP_{i,t+1}$ is either $ROA_{i,t+1}$ or $ATO_{i,t+1}$, and $AvgM_{i,t}$ is used as an instrument to predict manipulation $\hat{M}_{i,t}$. $ROA_{i,t+1}$ is calculated as the ratio of earnings before interest, taxes, depreciation, and amortization ($EBITDA_{i,t+1}$) to the total assets of firm i in year t. $ATO_{i,t+1}$ is computed as the percentage ratio of sales revenue to total assets. We also control this year's performance in our regressions. To consider the time it takes for investment decisions to translate into operating performance, we also perform a similar analysis by measuring the performance after year t by the average annual values of ROA and ATO over the next three years.

Table 3.7 reports the results of the impact of stock price manipulation on firms' operating performance. The coefficient of the manipulation $\hat{M}_{i,t}$ is -9.70 with t-statistics of 5.55

when the performance measure is the next year ROA and -11.03 with t-statistics of 5.03 when it is the average ROA over the next three years. On average, manipulation reduces next year's ROA by approximately 5.5% ($\beta_2 \cdot Average(\hat{M}) = -9.70 \cdot 0.57$) and the average ROA over the next three years by about 6.3% ($\beta_2 \cdot Average(\hat{M}) = -11.03 \cdot 0.57$). Similarly, the coefficient of $\hat{M}_{i,t}$ is -25,72 with t-statistics of 6.05 when the performance measure is next year's ATO and -33.06 with t-statistics of 5.71 when it is the next three years' average ATO. On average, manipulation decreases next year's ATO by about 15% and the average ATO over the next three years by about 19%. All these results support our hypothesis that stock price manipulation reduces firms' operating performance.

Table 3.7. Market manipulation and operating performance
The table reports the impact of stock price manipulation on the future operating performance of firms using the following model:

$$OP_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot \hat{M}_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \delta_3 \cdot OP_{i,t} + \varepsilon_{i,t+1},$$

where $OP_{i,t+1}$ is the future performance of firm i, defined as ROA (one year ahead return on asset) and ATO (asset turnover) in Columns (i) and (ii), and the average of the next three years ROA and ATO in Columns (iii) and (iv), respectively. $Q_{i,t}$ is the firm's normalized stock price, $\hat{M}_{i,t}$ is the average yearly predicted probability of stock price manipulation instrumented with $AvgM_{i,t}$ (the average manipulation by stock i's size quartile group). The controls $CF_{i,t}$ is firm's cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of the firm's total assets, $OP_{i,t}$ is the lagged operating performance. Refer to Table 3.1 for the definitions of all variables. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. All models include both firm and year-fixed effects (α_i and γ_{t+1}). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	Operating performance Next year		**Operating performance** Average of next three years	
	(i) **ROA**	(ii) **ATO**	(iii) **ROA**	(ii) **ATO**
$Q_{i,t}$	1.25*** (10.37)	-0.15 (-0.63)	0.91*** (5.96)	-0.45* (-1.70)
$\mathbf{\hat{M}_{i,t}}$	**-9.70*** (-5.55)**	**-25.73*** (-6.05)**	**-11.03*** (-5.03)**	**-33.06*** (-5.71)**
$CF_{i,t}$	-0.04*** (-6.14)	-0.12*** (-7.65)	-0.04*** (-5.07)	-0.12*** (-7.02)
$\ln(Asset_{i,t})$	-0.91*** (-4.80)	-2.02*** (-4.23)	-1.10*** (-6.71)	-3.20*** (-6.10)
$ROA_{i,t}$	0.42*** (23.81)		0.26*** (13.55)	
$ATO_{i,t}$		0.61*** (27.62)		0.45*** (17.97)
N obs.	52,865	52,865	52,865	52,865
R^2	23%	43%	16%	35%

3.6 Robustness and Extensions

3.6.1 Monotonic effects of manipulation on investment-Q sensitivity

We now document potential variations in the effects of stock price manipulation on investment-Q sensitivity across different manipulation groups. To do this, we allocate stocks into quintiles (quintile 1 for low-manipulated stocks and quintile 5 for high-manipulated stocks) and create four dummy variables for quintiles 2-5. We then run a variation of 2SLS with $AvgM_{i,t}$ as an instrument as

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \sum_{n=2}^{5} \beta_n \cdot \hat{M}_{n,t} + \sum_{n=2}^{5} \gamma_n \cdot \hat{M}_{n,t} \cdot Q_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1}, \quad (3.8)$$

where $\hat{M}_{n,t}$ for $n = 2, ..., 5$ are dummy variables for each quintile for each year. The coefficients of interest are $\gamma_2 - \gamma_5$.

In Table 3.8, we present the results of this model. The table shows that the interaction coefficients between manipulation groups and stock price are statistically significant for all quintiles, indicating that the relation between stock price manipulation and investment-Q sensitivity is not influenced by extreme observations but holds across the board. In addition, the table shows a consistent pattern in the interaction term coefficients. As we move from low to high-manipulated quintiles, the impact of market manipulation on investment-Q sensitivity monotonically increases in all three versions of the model. Specifically, the interaction coefficient in the top quintile is approximately four times larger than that in the bottom quintile.

Figure 3.2 plots the investment-Q sensitivity ($\frac{\partial I_{i,t+1}}{\partial Q_{i,t}} = \beta_1 + \gamma_n \cdot \hat{M}_{n,t}$) across quintiles. The Y-axis of the figure represents the investment-Q sensitivity of each manipulation quintile, while the X-axis represents the quintiles themselves. The figure shows that the investment-Q sensitivity monotonically decreases from the bottom quintile to the top quintile. This is intuitive because the negative impact of market manipulation on investment-Q sensitivity becomes more pronounced as the level of manipulation increases, suggesting that managers of firms with heavily manipulated stocks derive less guidance from their stock prices.

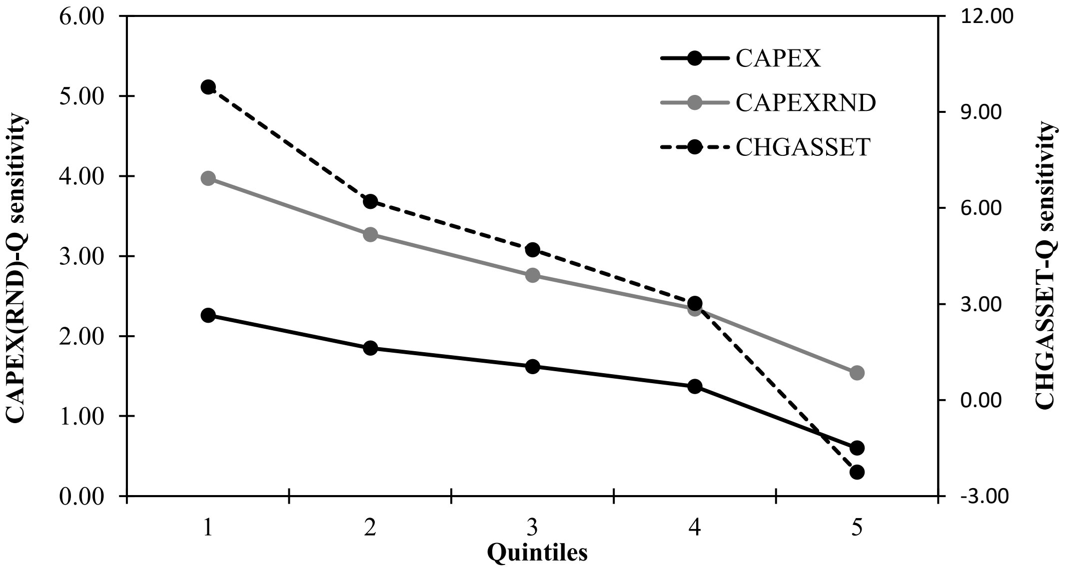
Table 3.8. Monotonic effects of market manipulation on investment-Q sensitivity
This table reports the impact of market manipulation in different groups on investment-Q sensitivity based on the following model:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \sum_{n=2}^{5} \beta_n \cdot \hat{M}_{n,t} + \sum_{n=2}^{5} \gamma_n \cdot \hat{M}_{n,t} \cdot Q_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1},$$

where $I_{i,t+1}$ is the firm i's investment in year $t+1$, defined as $CAPEX_{i,t+1}$ (capital expenditures divided by lagged total assets), $CAPEXRND_{i,t+1}$ (capital expenditures plus R&D expenses divided by lagged total assets), and $CHGASSET_{i,t+1}$ (annual change in total assets divided by lagged total assets). We divide stocks into quintiles (quintile 1 for low-manipulated stocks and quintile 5 for high-manipulated stocks) and $\hat{M}_n$ for $n = 2, ..., 5$ are dummy variables for each quintile for each year. We omit $\hat{M}_1$ (to fix the dummy variable trap). The other control variables are firms' normalized stock price ($Q_{i,t}$), cash flow ($CF_{i,t}$), and $\ln(Asset_{i,t})$. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. Refer to Table 3.1 for the definitions of all variables. All versions of the model include firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	$\mathbf{CAPEX_{i,t+1}}$	$\mathbf{CAPEXRND_{i,t+1}}$	$\mathbf{CHGASSET_{i,t+1}}$
$Q_{i,t}$	2.26***	3.97***	9.78***
	(12.29)	(11.90)	(10.89)
$\hat{M}_{2,t}$	0.70***	0.90**	4.91***
	(2.74)	(2.29)	(2.71)
$\hat{M}_{3,t}$	1.06***	1.52***	7.50***
	(3.94)	(3.87)	(3.41)
$\hat{M}_{4,t}$	1.69***	2.38***	11.36***
	(5.53)	(5.28)	(4.75)
$\hat{M}_{5,t}$	3.70***	4.21***	24.00***
	(11.34)	(9.34)	(10.15)
$\mathbf{\hat{M}_{2,t} \cdot Q_{i,t}}$	**-0.41****	**-0.70****	**-3.57****
	(-2.21)	**(-2.11)**	**(-2.28)**
$\mathbf{\hat{M}_{3,t} \cdot Q_{i,t}}$	**-0.64*****	**-1.21*****	**-5.08*****
	(-3.40)	**(-4.05)**	**(-2.97)**
$\mathbf{\hat{M}_{4,t} \cdot Q_{i,t}}$	**-0.89*****	**-1.63*****	**-6.76*****
	(-4.94)	**(-4.92)**	**(-3.75)**
$\mathbf{\hat{M}_{5,t} \cdot Q_{i,t}}$	**-1.66*****	**-2.43*****	**-12.03*****
	(-9.52)	**(-7.79)**	**(-7.21)**
$CF_{i,t}$	0.00	0.01**	-0.00
	(1.05)	(2.25)	(-0.14)
$\ln(Asset_{i,t})$	-1.39***	-3.31***	-17.27***
	(-17.11)	(-21.33)	(-19.32)
N obs.	52,865	52,865	52,865
R^2	10%	17%	18%

Figure 3.2. The monotonic effects of market manipulation on investment-Q sensitivity
This figure graphically illustrates the impact of market manipulation on investment-Q sensitivity across different quintiles, as determined by the estimates of the model in Eq. (3.8). The Y-axis is the investment-Q sensitivity ($\frac{\partial I_{i,t+1}}{\partial Q_{i,t}} = \beta_1 + \gamma_n \cdot \hat{M}_{n,t}$ where $\hat{M}_{n,t}$ for $n = 2, ..., 5$ are dummy variables indicating each quintile for each year). The X-axis represents the quintiles by the average probability of market manipulation, ranging from the lowest value in quintile 1 and the highest value in quintile 5. The left axis represents the investment-Q sensitivity when investment is measured by $CAPEX$ or $CAPEXRND$, while the right axis represents the same when investment measure is $CHGASSET$.



3.6.2 Controlling for managerial information

To establish the robustness of the negative effect of market manipulation on investment-to-price sensitivity, we introduce controls for managerial information. This is important because, as demonstrated in Proposition 1 and discussed subsequently, managers with more accurate managerial information may rely less on stock prices when making investment decisions. This may potentially affect the impact of stock price manipulation on investment-Q sensitivity.

We use insider trading intensity ($Insider_{i,t}$) to capture managerial information. It represents the ratio of insider transactions from Thomson Financial's TFN database to the total number of transactions from CRSP for a specific firm-year.[49] This measure relies on the premise that managers who trade more have greater private information.

Table 3.9 presents the results for all three investment measures, showing negative and significant coefficients for the interaction term $Insider \cdot Q$ when investment is $CHGASSET$. This finding suggests that managers with greater private information rely less on stock prices when making investment decisions. Importantly, including $Insider \cdot Q$ in the regressions does not change the negative and significant impact of market manipulation on firms' investment-to-price sensitivity.

3.6.3 Controlling for analyst coverage

Analyst coverage is often seen as a means of transferring information from managers to financial markets (e.g., Bailey et al. 2003). If analyst coverage primarily reflects managerial information, it is likely to impact stock prices but not investment decisions, resulting in a negative relationship between analyst coverage and investment-to-price sensitivity. Alternatively, if analysts provide complementary information to managers, a positive relationship between analyst coverage and investment-to-price sensitivity is expected. Table 3.9 shows a negative and significant coefficient for $ACoverage \cdot Q$ when investment is measured by $CAPEXRND$ and $CHGASSET$, consistent with the view that analyst coverage simply reflects the managerial information. In addition, the negative relation between market manipulation and investment-to-price sensitivity is robust to the inclusion of analyst coverage across all investment measures.

[49] We also use the percentage of insider dollar volume for a given firm-year, sourced from the same databases and obtain consistent results.

Table 3.9. Market manipulation and investment-Q sensitivity: Controlling for additional factors
The table reports the impact of market manipulation on investment-Q sensitivity using the following model:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot \hat{M}_{i,t} + \beta_3 \cdot \hat{M}_{i,t} \cdot Q_{i,t} + \beta_4 \cdot Control_{i,t} + \beta_5 \cdot Control_{i,t} \cdot Q_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1}$$

where $I_{i,t+1}$ is the firm i's investment in year $t+1$, defined as $CAPEX_{i,t+1}$ (capital expenditures divided by lagged total assets), $CAPEXRND_{i,t+1}$ (capital expenditures plus R&D expenses divided by lagged total assets), and $CHGASSET_{i,t+1}$ (annual change in total assets divided by lagged total assets). $Q_{i,t}$ is the firm's normalized stock price, $\hat{M}_{i,t}$ is the average yearly predicted probability of stock price manipulation instrumented with $AvgM_{i,t}$ (average manipulation by stock i's size quartile group), $CF_{i,t}$ is firm's cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of total assets. From columns (i) to (iii), $Control_{i,t}$ is respectively defined as $Insider_{i,t}$ (the percentage of number of transactions by insiders), $ACoverage_{i,t}$ (analyst coverage), and $KZ_{i,t}$ (Kaplan-Zingales index). Refer to Table 3.1 for the definitions of all variables. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. All versions of the model include both firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	$\mathbf{CAPEX_{i,t+1}}$			$\mathbf{CAPEXRND_{i,t+1}}$			$\mathbf{CHGASSET_{i,t+1}}$		
	(i) $\mathbf{Insider_{i,t}}$	(ii) $\mathbf{ACoverage_{i,t}}$	(iii) $\mathbf{KZ_{i,t}}$	(i) $\mathbf{Insider_{i,t}}$	(ii) $\mathbf{ACoverage_{i,t}}$	(iii) $\mathbf{KZ_{i,t}}$	(i) $\mathbf{Insider_{i,t}}$	(ii) $\mathbf{ACoverage_{i,t}}$	(iii) $\mathbf{KZ_{i,t}}$
$Q_{i,t}$	2.61***	3.24***	3.09***	4.05***	4.83***	4.66***	23.49***	23.48***	23.78***
	(9.06)	(11.65)	(12.97)	(7.69)	(9.34)	(10.74)	(8.01)	(8.74)	(9.85)
$\hat{M}_{i,t}$	3.73***	6.31***	6.43***	9.40***	8.83***	9.23***	39.92***	55.77***	58.82***
	(3.10)	(4.73)	(5.95)	(4.12)	(4.02)	(4.46)	(4.00)	(6.18)	(6.69)
$\mathbf{\hat{M}_{i,t} \cdot Q_{i,t}}$	**-2.47*** **	**-3.27*** **	**-3.10*** **	**-3.22*** **	**-3.92*** **	**-4.12*** **	**-18.47*** **	**-17.80*** **	**-21.96*** **
	(-7.61)	**(-10.18)**	**(-10.54)**	**(-4.94)**	**(-6.26)**	**(-7.06)**	**(-5.91)**	**(-6.52)**	**(-7.39)**
$Control_{i,t}$	-0.05	0.02	-0.29***	0.04	0.15***	-0.18*	1.65*	1.07***	-1.07***
	(-0.38)	(1.05)	(-5.05)	(0.21)	(5.88)	(-1.95)	(1.65)	(8.35)	(-3.16)
$Control_{i,t} \cdot Q_{i,t}$	0.03	0.00	0.04***	-0.16	-0.03***	0.06***	-2.09***	-0.22***	0.27***
	(0.39)	(0.11)	(2.99)	(-1.28)	(-4.20)	(2.75)	(-3.28)	(-5.43)	(2.78)
$CF_{i,t}$	-0.01***	-0.01**	0.20	0.01	0.01	1.90***	-0.07*	-0.02	19.86**
	(-2.59)	(-2.14)	(1.56)	(1.44)	(1.28)	(2.85)	(-1.85)	(-0.86)	(2.30)
$\ln(Asset_{i,t})$	-1.35***	-1.36***	-1.32***	-4.49***	-3.97***	-3.20***	-19.85***	-20.12***	-15.76***
	(-11.86)	(-12.01)	(-16.59)	(-21.59)	(-23.95)	(-20.23)	(-15.84)	(-20.60)	(-15.71)
N obs.	28,162	39,959	51,477	28,162	39,959	51,477	28,162	39,959	51,477
R^2	9%	10%	10%	21%	19%	18%	20%	19%	22%

3.6.4 Controlling for capital constraints

Baker, Stein, and Wurgler (2003) demonstrate the significance of capital constraints in driving firms' investment-Q sensitivity. Constrained firms with limited access to equity financing have higher investment-Q sensitivity. To account for capital constraints, we incorporate the interaction of the capital constraint and stock price in our regressions. Following Baker, Stein, and Wurgler (2003), we use the modified Kaplan-Zingales index KZ as a proxy for capital constraints. KZ is a weighted sum of cash flow, cash balances, cash dividends, and leverage ratio (Kaplan and Zingales 1997). Consistent with their findings, Table 3.9 shows a positive and significant coefficient for $KZ \cdot Q$ for all three investment measures. Moreover, the negative impact of market manipulation on investment-Q sensitivity remains unchanged after controlling firms' capital constraints.

3.6.5 Fama-MacBeth regressions

Our main results are based on pooled regressions using unbalanced panel data, which capture both within-firm and cross-firm effects. These effects combine the influence of a firm's own stock price manipulation over time and the variation in investment-Q sensitivity across firms in a given year. To disentangle cross- and within-firm effects, we employ the Fama-MacBeth approach to re-estimate our main investment model. This allows us to examine whether cross-sectionally firms with higher levels of stock price manipulation have lower investment-Q sensitivity on a yearly basis. We estimate the investment model each year for all firms and present the average coefficients in Table 3.10, along with the yearly coefficient estimates in Figure 3.3. The standard errors are adjusted using the Newey-West method.

The Fama-MacBeth results align with our main findings in Tables 3.3 and 3.4. The coefficient estimate for $M \cdot Q$ remains negative and significant across all specifications. Table 3.11 confirms the robustness of the cross-sectional effect. Furthermore, Figure 3.3 presents the yearly estimates of investment-to-price sensitivity. The black bars represent the investment-to-price sensitivity without manipulation (β_1), while the gray bars represent the net investment-to-price sensitivity considering the average manipulation ($\beta_1 + \beta_3 \cdot \hat{M}_t$). The figure demonstrates a mostly persistent negative impact of market manipulation on investment-to-price sensitivity across all three investment measures.

Table 3.10. Market manipulation and investment-Q sensitivity: Fama-Macbeth regression
The table reports the impact of market manipulation on investment-Q sensitivity using the following model:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot \hat{M}_{i,t} + \beta_3 \cdot \hat{M}_{i,t} \cdot Q_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1},$$

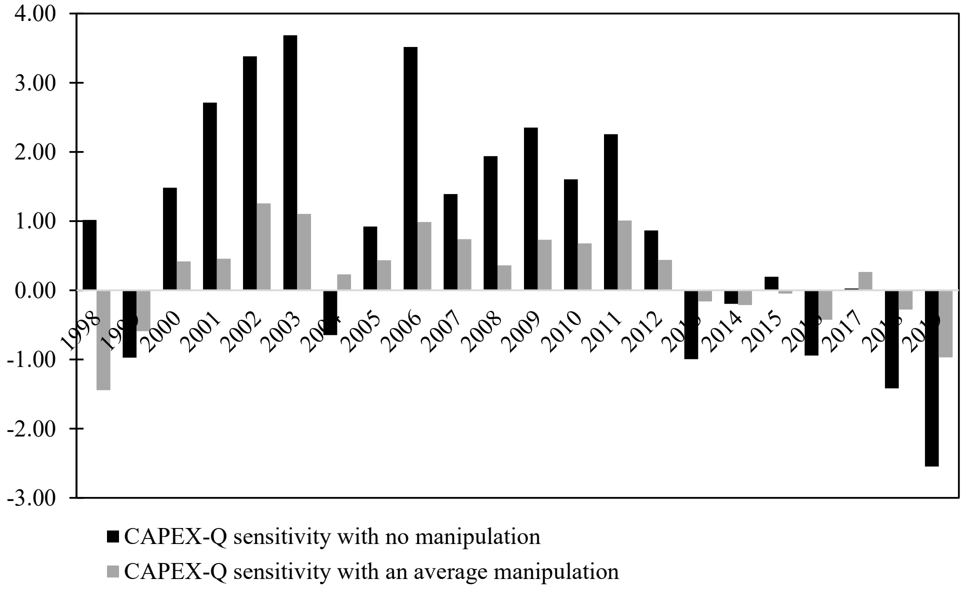
where $I_{i,t+1}$ is the firm i's investment in year $t+1$, defined as $CAPEX_{i,t+1}$ (capital expenditures divided by lagged total assets), $CAPEXRND_{i,t+1}$ (capital expenditures plus R&D expenses divided by lagged total assets), and $CHGASSET_{i,t+1}$ (annual change in total assets divided by lagged total assets). $Q_{i,t}$ is the firm's normalized stock price, $\hat{M}_{i,t}$ is the average yearly predicted probability of stock price manipulation instrumented with $AvgM_{i,t}$ (the average manipulation by stock i's size quartile group, $CF_{i,t}$ is firm's cash flow, and $\ln(Asset_{i,t})$ is the logarithm of total assets. Refer to Table 3.1 for the definitions of all variables. The reported coefficient estimates are the averages of yearly estimates and the standard errors are estimated with the Newey-West adjustment. All versions of the model are estimated for each year and include firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	$\mathbf{CAPEX_{i,t+1}}$	$\mathbf{CAPEXRND_{i,t+1}}$	$\mathbf{CHGASSET_{i,t+1}}$
$Q_{i,t}$	0.89*	7.06***	17.42***
	(1.77)	(5.67)	(9.83)
$M_{i,t}$	13.70***	2.61	19.02
	(2.77)	(0.60)	(1.08)
$\mathbf{M_{i,t} \cdot Q_{i,t}}$	**-1.17***	**-6.10*****	**-14.03*****
	(-1.89)	**(-3.51)**	**(-5.94)**
$CF_{i,t}$	-0.07***	0.15***	-0.07
	(-14.87)	(8.00)	(-2.83)
$\ln(Asset_{i,t})$	-0.08	-0.62***	-1.66***
	(-0.64)	(-5.40)	(-3.53)
N obs.	52,865	52,865	52,865

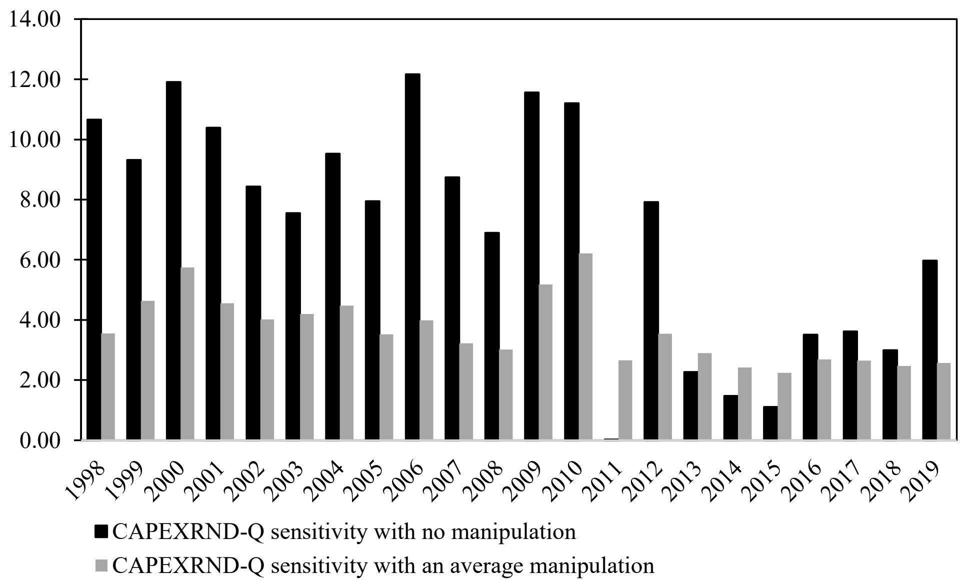
Figure 3.3. The yearly investment-Q sensitivity with and without market manipulation

This figure illustrates the results from the year-by-year regressions of the relation between market manipulation and investment-Q sensitivity. The black bars show the investment-Q sensitivity for all firms each year without taking into account market manipulation (β_1) and the gray bars show the investment-Q sensitivity with an average market manipulation for that year ($\beta_1 + \beta_3 \cdot \hat{M_t}$). The estimation is based on the predicted market manipulation instrumented with $AvgM_{i,t}$ and includes firm and year-fixed effects.

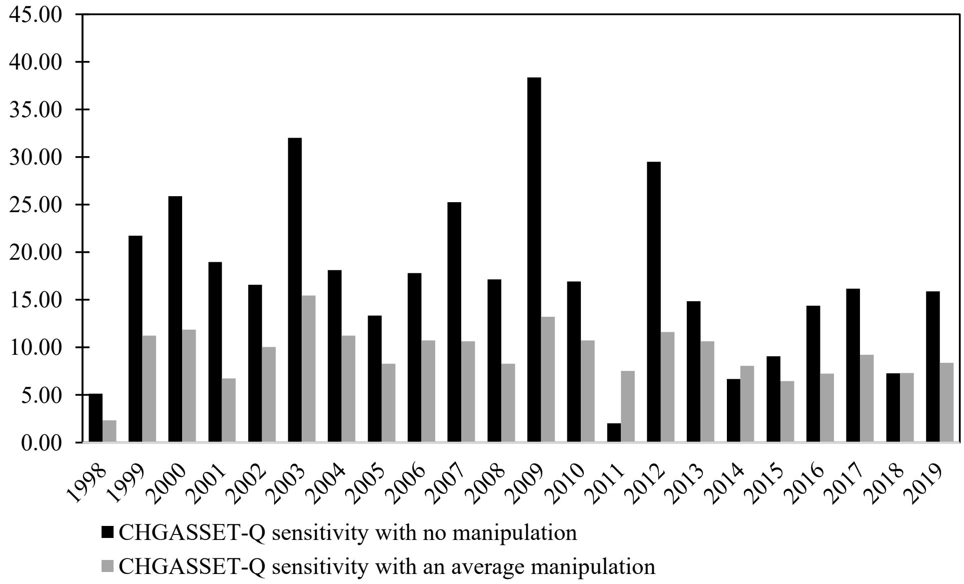
Panel A. The yearly $CAPEX$-to-price sensitivity with and without market manipulation



Panel B. The yearly $CAPEXRND$-to-price sensitivity with and without market manipulation



Panel C. The yearly $CHGASSET$-to-price sensitivity with and without market manipulation



3.6.6 Accounting for intangible capital

Our baseline tests use the standard Tobin's Q (computed as the market value of equity plus the book value of assets minus the book value of equity, scaled by the book value of asset). In Appendix 3.1, we also replicate the main findings by incorporating intangible capital into Tobin's Q calculation using a modified approach proposed by Peters and Taylor (2017). The negative impact of market manipulation on investment-to-price sensitivity remains unchanged.

3.7 Conclusion

This chapter investigates the impact of stock price manipulation on firms' investment decisions. We find that an increased prevalence of market manipulation negatively affects sensitivity of firms' investment to stock prices. Furthermore, the inefficient allocation of capital resulting from market manipulation leads to inferior future firm performance. We conceptualize these findings in a model that captures the two-dimensional learning of a firm manager from both stock prices and managerial signals. Our findings provide evidence of the real economic consequences of market manipulation.

We further find that market manipulation has a stronger negative effect on investment-to-price sensitivity in illiquid and volatile stocks, and those with more private information, supporting the managerial learning hypothesis. In extended tests, we find a monotonic increase in the negative impact of market manipulation on investment-to-price sensitivity across different levels of manipulation, with the top quintile being approximately four times more detrimental than the bottom quintile. The negative impact of market manipulation on investment-to-price sensitivity remains robust after controlling for managerial information, analyst coverage, and firms' capital constraints. We also find that firms with more manipulated stock prices have lower investment-to-price sensitivity cross-sectionally. These findings continue to hold under various model specifications using different variables.

Our study quantifies the real economic effects of market manipulation on firms' investment decisions and operating performance. More broadly, our findings highlight the importance of combating market manipulation through appropriate regulatory measures and the potential consequences of not designing markets that actively discourage manipulative practices.

Our study suggests several avenues for further research. First, we measure stock price manipulation using closing price manipulation. Exploring alternative measures of stock price manipulation could provide a more comprehensive understanding. Second, while we theoretically show how market manipulation can decrease investment-to-price sensitivity, the exact process by which traders manipulate the market is omitted from the analysis. Future research could investigate the trading and manipulation decisions of traders, how these actions impact prices, and consequently, investment-to-price sensitivity. Lastly, while we document that market manipulation negatively impacts firms' investment-to-price sensitivity and operating performance, translating these findings into broader economic welfare terms requires further considerations. For example, market manipulation can significantly impact investor confidence, market efficiency, and resource allocation, which are challenging to quantify. Future research should address these limitations by thoroughly examining the welfare implications of market manipulation.

Appendix 3.1.

Table A3.1
Market manipulation measure
This table presents the variables and metrics used in the calculation of the probability of stock price manipulation. Panel A introduces the relevant variables and Panel B shows the computation of the metrics. These metrics are computed within real-time intervals defined by $x = \{15, 20, 30, 60, 90\}$ minutes before the market close at 4pm.

Panel A: The variables used in the market manipulation metrics	
Variable	**Notation**
The last traded price for a given day	P_{close}
The bid-ask midpoint x minutes before the market close	M_x
The last traded price in the current day	$P_{d0,close}$
The bid-ask midpoint at 11 am the following day	$M_{d1,morning}$
The number of trades in the last x minutes before the market close	n_x
The dollar bid-ask spread at the market close	S_{close}
Panel B: The calculation of metrics used in the market manipulation measure	
Metric	**Calculation**
Return (%)	$\ln\left(\frac{P_{close}}{M_x}\right) \cdot 100$
Reversal (%)	$\ln\left(\frac{P_{d0,close}}{M_{d1,morning}}\right) \cdot 100$
Frequency (trades per hour)	$\left(\frac{n_x}{x}\right) \cdot 60$
Spread (%)	$\left(\frac{S_{close}}{M_x}\right) \cdot 100$

Table A3.2
Market manipulation and investment-Q sensitivity: 2SLS IV first-stage results for Table 3.4

This table presents the first stage results of the impact of market manipulation on investment-Q sensitivity using the 2SLS IV approach reported in Table 3.4 of Chapter 3. The first stage of the 2SLS regressions is given by:

$$\hat{M}_{i,t} = \alpha_i + \gamma_t + \gamma_1 \cdot IV_{i,t} + \gamma_2 \cdot Q_{i,t} + \gamma_3 \cdot CF_{i,t} + \gamma_4 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t},$$

where $\hat{M}_{i,t}$ is the predicted value of the average yearly probability of stock price manipulation, $Q_{i,t}$ is the normalized stock price, $CF_{i,t}$ is the cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of total assets. In Models 1 to 3, the instrumental variable ($IV_{i,t}$) is the average manipulation by stock i's size quartile group ($AvgM_{i,t}$), the lagged manipulation ($M_{i,t-1}$), and the SEC regulatory budget (RB_t) interacted with the relative spread ($RSpread_{i,t}$), respectively. The first stage F-statistic is the statistic for testing the model. Refer to Table 3.1 for the definitions of all variables. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. All versions of the model include firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	Model 1	Model 2	Model 3
	$\mathbf{IV_{i,t} = AvgM_{i,t}}$	$\mathbf{IV_{i,t} = M_{i,t-1}}$	$\mathbf{IV_{i,t} = RB_t \cdot RSpread_{i,t}}$
$\mathbf{IV_{i,t}}$	**0.97*** **	**0.03***	**-0.00*** **
	(24.79)	**(1.77)**	**(-2.59)**
$Q_{i,t}$	0.03***	0.04***	0.03***
	(12.82)	(11.81)	(13.18)
$CF_{i,t}$	0.00**	0.00***	0.00***
	(2.79)	(4.09)	(3.31)
$\ln(Asset_{i,t})$	0.01**	0.02***	0.02***
	(2.33)	(3.40)	(2.95)
F-statistics	1,106	434	490
N obs.	52,865	47,225	46,028
R^2	8%	4%	5%

Table A3.3
Market manipulation and operating performance: OLS version of Table 3.7
The table reports the impact of market manipulation on firms' operating performance using the model:

$$OP_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot M_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \delta_3 \cdot OP_{i,t} + \varepsilon_{i,t+1},$$

where $OP_{i,t+1}$ is the future performance of firm i, defined as one year ahead return on asset (ROA) and asset turnover (ATO) in Columns (i) and (ii), and the average of the next three years ROA and ATO in Columns (iii) and (iv), respectively. $Q_{i,t}$ is the firm's normalized stock price, $M_{i,t}$ is the average yearly probability of stock price manipulation, $CF_{i,t}$ is firm's cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of the firm's total assets, $OP_{i,t}$ is the lagged operating performance. Refer to Table 3.1 for the definitions of all variables. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. All models include firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	Operating performance Next year		**Operating performance** Average of next three years	
	(i) **ROA**	(ii) **ATO**	(iii) **ROA**	(ii) **ATO**
$Q_{i,t}$	0.95***	-0.97***	0.55***	-1.51***
	(9.05)	(-5.62)	(4.00)	(-8.18)
$\mathbf{M_{i,t}}$	**-1.36*** **	**-3.37*** **	**-1.42*** **	**-3.93*** **
	(-3.40)	**(-5.78)**	**(-3.56)**	**(-6.32)**
$CF_{i,t}$	-0.04***	-0.13***	-0.04***	-0.14***
	(-6.54)	(-8.12)	(-5.60)	(-7.66)
$\ln(Asset_{i,t})$	-1.11***	-2.55***	-1.33***	-3.89***
	(-6.38)	(-5.69)	(-9.11)	(-8.25)
$ROA_{i,t}$	0.42***		0.30***	
	(23.68)		(13.35)	
$ATO_{i,t}$		0.61***		0.45***
		(27.49)		(17.87)
N obs.	52,865	52,865	52,865	52,865
R^2	23%	43%	15%	35%

Table A3.4
Monotonic effects of market manipulation on investment-Q sensitivity: OLS version of Table 3.8
This table reports effects of different levels of market manipulation on investment-Q sensitivity using the model:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \sum_{n=2}^{5} \beta_n \cdot M_{n,t} + \sum_{n=2}^{5} \gamma_n \cdot M_{n,t} \cdot Q_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1},$$

where $I_{i,t+1}$ is the firm i's investment in year $t+1$, defined as $CAPEX_{i,t+1}$ (capital expenditures divided by lagged total assets), $CAPEXRND_{i,t+1}$ (capital expenditures plus R&D expenses divided by lagged total assets), and $CHGASSET_{i,t+1}$ (annual change in total assets divided by lagged total assets). We divide stocks into quintiles (quintile 1 for low-manipulated stocks and quintile 5 for high-manipulated stocks) and M_n for $n = 2, ..., 5$ are dummy variables for each quintile for each year. We omit M_1 (to fix the dummy variable trap). The other control variables include firms' normalized stock price ($Q_{i,t}$), cash flow ($CF_{i,t}$), and ln($Asset_{i,t}$). Refer to Table 3.1 for the definitions of all variables. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. All versions of the model include firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	$\mathbf{CAPEX_{i,t+1}}$	$\mathbf{CAPEXRND_{i,t+1}}$	$\mathbf{CHGASSET_{i,t+1}}$
$Q_{i,t}$	1.01***	1.97***	10.00***
	(14.89)	(15.18)	(13.72)
$M_{2,t}$	0.12	0.13	-0.12
	(1.00)	(0.65)	(-0.10)
$M_{3,t}$	0.41***	0.43**	0.82
	(3.22)	(2.13)	(0.96)
$M_{4,t}$	0.52***	0.48**	2.47**
	(4.42)	(2.23)	(2.46)
$M_{5,t}$	1.05***	1.14***	4.94***
	(6.27)	(4.24)	(4.34)
$\mathbf{M_{2,t} \cdot Q_{i,t}}$	**-0.01**	**0.04**	**0.63**
	(-0.13)	**(0.32)**	**(0.84)**
$\mathbf{M_{3,t} \cdot Q_{i,t}}$	**-0.09**	**-0.05**	**0.53**
	(-1.33)	**(-0.41)**	**(0.91)**
$\mathbf{M_{4,t} \cdot Q_{i,t}}$	**-0.11***	**-0.12**	**-0.30**
	(-1.81)	**(-0.92)**	**(-0.44)**
$\mathbf{M_{5,t} \cdot Q_{i,t}}$	**-0.25*****	**-0.40*****	**-1.52****
	(-3.96)	**(-2.99)**	**(-2.28)**
$CF_{i,t}$	-0.00	0.02***	-0.01
	(-0.39)	(2.65)	(-0.56)
$\ln(Asset_{i,t})$	-1.39***	-3.36***	-17.36***
	(-17.55)	(-20.61)	(-18.58)
N obs.	52,865	52,865	52,865
R^2	9%	17%	17%

Table A3.5
The impact of market manipulation on investment-Q sensitivity by stock characteristics: CAPEX and CHGASSET versions of Table 3.6
This table presents the impact of market manipulation on investment-to-price sensitivity by stock characteristics:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot M_{i,t} + \beta_3 \cdot M_{i,t} \cdot Q_{i,t} + \beta_4 \cdot D_h + \beta_5 \cdot Q_{i,t} \cdot D_h + \beta_6 \cdot M_{i,t} \cdot D_h + \beta_7 \cdot M_{i,t} \cdot Q_{i,t} \cdot D_h + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1}.$$

where $I_{i,t+1}$ is firm i's investment in year $t+1$, defined as the capital expenditures divided by lagged total assets ($CAPEX_{i,t+1}$) in Panel A and the annual change in total assets divided by lagged total assets ($CHGASSET_{i,t+1}$) in Panel B. Across all specifications, $Q_{i,t}$ is the firm's normalized stock price, $M_{i,t}$ is the average yearly probability of stock price manipulation, $CF_{i,t}$ is firm's cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of the firm's total assets. D_h is a dummy variable that equals one if the value of a given stock characteristic is above the median value of all other stocks each year. From Columns (i) to (vi), D_h is defined based on the relative spread, aggregate yearly dollar volume, market capitalization, the standard deviation of returns, private information measured by PIN, and analyst coverage, respectively. For the definitions of all variables, refer to Table 1. All versions of the model include both firm and year-fixed effects (α_i and γ_t, respectively). The standard errors used to compute the t-statistics (in brackets) are double clustered by firm and year. *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

Panel A: $\mathbf{CAPEX_{i,t+1}}$						
	$\mathbf{D_h}$					
	(i) **RSpread**	(ii) **Volume**	(iii) **MktCap**	(iv) **Volatility**	(v) **PIN**	(vi) **ACoverage**
$Q_{i,t}$	0.85***	0.93***	0.94***	0.97***	1.10***	0.94***
	(8.40)	(6.98)	(6.86)	(8.91)	(8.94)	(7.68)
$M_{i,t}$	0.92**	1.65***	1.71***	1.17***	1.73***	1.86***
	(2.28)	(4.10)	(4.37)	(2.83)	(3.71)	(4.92)
D_h	-1.42***	1.13***	1.35***	-0.87***	-0.30	0.32
	(-5.15)	(3.86)	(4.40)	(-3.19)	(-1.16)	(0.96)
$M_{i,t} \cdot Q_{i,t}$	-0.13	-0.39**	-0.36**	-0.27*	-0.42***	-0.39**
	(-1.12)	(-1.98)	(-2.07)	(-1.70)	(-2.77)	(-2.21)
$M_{i,t} \cdot D_h$	1.02**	-0.51	-0.45	1.03**	0.63	-0.32
	(2.21)	(-1.02)	(-0.96)	(2.56)	(1.36)	(-0.53)
$Q_{i,t} \cdot D_h$	0.18	0.04	-0.01	0.08	0.10	0.14
	(1.37)	(0.28)	(-0.10)	(0.74)	(0.86)	(1.00)
$\mathbf{M_{i,t} \cdot Q_{i,t} \cdot D_h}$	**-0.44****	**0.05**	**0.04**	**-0.18**	**-0.23**	**-0.05**
	(-1.98)	**(0.21)**	**(0.17)**	**(-1.03)**	**(-1.21)**	**(-0.22)**
$CF_{i,t}$	-0.01	-0.01	-0.01	-0.01*	-0.00	-0.01*
	(-1.61)	(-1.61)	(1.52)	(-1.66)	(-0.84)	(-1.75)
$\ln(Asset_{i,t})$	-1.51***	-1.53***	-1.55***	-1.40***	-1.45***	-1.42***
	(-15.02)	(-14.97)	(-15.91)	(-15.19)	(-17.88)	(-15.80)
N obs.	46,028	46,028	46,034	46,034	40,241	39,968
R^2	9%	9%	9%	9%	9%	9%

Table A3.5
Continued.

Panel B: $\mathbf{CHGASSET_{i,t+1}}$						
	$\mathbf{D_h}$					
	(i) **RSpread**	(ii) **Volume**	(iii) **MktCap**	(iv) **Volatility**	(v) **PIN**	(vi) **ACoverage**
$Q_{i,t}$	8.83***	12.45***	13.29***	9.20***	10.07***	10.42***
	(7.45)	(9.74)	(9.89)	(8.57)	(8.35)	(8.09)
$M_{i,t}$	7.60**	6.18**	6.49**	10.08***	12.11***	5.06*
	(2.37)	(2.14)	(2.43)	(3.19)	(2.71)	(1.84)
D_h	-10.11***	10.99***	14.56***	-7.38***	-3.81*	3.82*
	(-4.34)	(5.80)	(7.79)	(-3.19)	(-1.88)	(1.78)
$M_{i,t} \cdot Q_{i,t}$	-1.90	-1.54	-2.47	-1.81	-3.09	-0.32
	(-1.37)	(-0.85)	(-1.51)	(-1.33)	(-1.62)	(-0.19)
$M_{i,t} \cdot D_h$	-1.21	-1.37	0.20	1.28	0.12	6.41*
	(-0.33)	(-0.44)	(0.09)	(0.33)	(0.02)	(1.79)
$Q_{i,t} \cdot D_h$	3.78***	-3.83***	-5.34***	2.78***	2.80**	-0.50
	(3.33)	(-3.45)	(-4.45)	(2.63)	(2.46)	(-0.43)
$\mathbf{M_{i,t} \cdot Q_{i,t} \cdot D_h}$	**-0.57**	**0.20**	**0.87**	**-2.18**	**-2.07**	**-3.53***
	(-0.31)	**(0.10)**	**(0.52)**	**(-1.24)**	**(-0.83)**	**(-1.71)**
$CF_{i,t}$	-0.01	-0.01	-0.01	-0.00	-0.03	-0.01
	(-0.36)	(-0.42)	(-0.46)	(-0.05)	(-1.07)	(-0.22)
$\ln(Asset_{i,t})$	-18.36***	-18.25***	-18.51***	-17.99***	-18.18***	-18.72***
	(-19.46)	(-20.18)	(-20.36)	(-19.24)	(-17.24)	(-19.17)
N obs.	46,028	46,028	46,034	46,034	40,241	39,968
R^2	18%	18%	18%	18%	18%	18%

Table A3.6

Market manipulation and investment-Q sensitivity - Controlling for additional factors: OLS version of Table 3.9

The table presents the results of the impact of market manipulation on investment-Q sensitivity following from the model:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot Q_{i,t} + \beta_2 \cdot M_{i,t} + \beta_3 \cdot M_{i,t} \cdot Q_{i,t} + \beta_4 \cdot Control_{i,t} + \beta_5 \cdot Control_{i,t} \cdot Q_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1}$$

where $I_{i,t+1}$ is the firm i's investment in year $t+1$, defined as $CAPEX_{i,t+1}$ (capital expenditures divided by lagged total assets), $CAPEXRND_{i,t+1}$ (capital expenditures plus R&D expenses divided by lagged total assets), and $CHGASSET_{i,t+1}$ (annual change in total assets divided by lagged total assets). $Q_{i,t}$ is the firm's normalized stock price, $M_{i,t}$ is the average yearly probability of stock price manipulation, $CF_{i,t}$ is firm's cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of total assets. From columns (i) to (iii), $Control_{i,t}$ is respectively defined as $Insider_{i,t}$ (the percentage of number of transactions by insiders), $ACoverage_{i,t}$ (analyst coverage), and $KZ_{i,t}$ (Kaplan-Zingales index). Refer to Table 3.1 for the definitions of all variables. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. All versions of the model include firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	$CAPEX_{i,t+1}$			**$CAPEXRND_{i,t+1}$**			**$CHGASSET_{i,t+1}$**		
	(i) **$Insider_{i,t}$**	(ii) **$ACoverage_{i,t}$**	(iii) **$KZ_{i,t}$**	(i) **$Insider_{i,t}$**	(ii) **$ACoverage_{i,t}$**	(iii) **$KZ_{i,t}$**	(i) **$Insider_{i,t}$**	(ii) **$ACoverage_{i,t}$**	(iii) **$KZ_{i,t}$**
$Q_{i,t}$	0.87***	1.02***	1.17***	2.26***	2.39***	2.24***	11.39***	12.16***	10.58***
	(7.96)	(8.77)	(11.91)	(11.09)	(11.71)	(14.76)	(9.27)	(9.60)	(11.64)
$M_{i,t}$	1.85***	1.72***	1.74***	2.10***	2.00***	2.01***	9.11***	8.27***	9.52***
	(5.24)	(5.27)	(4.97)	(4.72)	(4.54)	(4.25)	(2.54)	(3.17)	(3.68)
$\mathbf{M_{i,t} \cdot Q_{i,t}}$	**-0.36*** **	**-0.41*** **	**-0.44*** **	**-0.95*** **	**-0.77*** **	**-0.75*** **	**-3.17** **	**-2.08* **	**-2.82** **
	(-2.71)	**(-3.88)**	**(-3.76)**	**(-3.33)**	**(-3.34)**	**(-3.15)**	**(-2.00)**	**(-1.77)**	**(-2.46)**
$Control_{i,t}$	-0.17	-0.01*	0.23*	-0.18	0.17***	-0.32***	-0.51	1.09***	-1.95***
	(-1.38)	(-1.80)	(1.74)	(-0.85)	(5.86)	(-3.61)	(-1.40)	(7.24)	(-6.94)
$Control_{i,t} \cdot Q_{i,t}$	0.13	-0.001	0.08***	-0.05	-0.03***	0.11***	-1.40**	-0.24***	0.54***
	(1.40)	(-0.24)	(5.24)	(-0.35)	(-4.43)	(4.98)	(2.14)	(-5.25)	(6.53)
$CF_{i,t}$	-0.01**	-0.01	0.23*	0.02**	0.01*	1.93***	-0.05	-0.01	20.07**
	(-2.33)	(-1.80)	(1.74)	(2.01)	(1.70)	(2.84)	(-1.45)	(-0.28)	(2.30)
$\ln(Asset_{i,t})$	-1.38***	-1.45***	-1.33***	-4.43***	-4.01***	-3.18***	-19.78***	-19.85***	-15.46***
	(-12.22)	(-14.39)	(-18.13)	(-20.51)	(-21.79)	(-18.67)	(-15.50)	(-18.51)	(-14.83)
N obs.	28,162	39,959	51,477	28,162	39,959	51,477	28,162	39,959	51,477
R^2	8%	9%	9%	21%	19%	18%	20%	18%	22%

Table A3.7
Market manipulation and investment-Q sensitivity - OLS: Peters-Taylor Q version of Table 3.3
This table reports the results of the impact of market manipulation on investment-Q sensitivity following from the model:

$$I_{i,t+1} = \alpha_i + \gamma_t + \beta_1 \cdot PTQ_{i,t} + \beta_2 \cdot M_{i,t} + \beta_3 \cdot M_{i,t} \cdot PTQ_{i,t} + \delta_1 \cdot CF_{i,t} + \delta_2 \cdot \ln(Asset_{i,t}) + \varepsilon_{i,t+1},$$

where $I_{i,t+1}$ is the firm i's investment in year $t+1$, defined as $CAPEX_{i,t+1}$ (capital expenditures divided by lagged total assets), $CAPEXRND_{i,t+1}$ (capital expenditures plus R&D expenses divided by lagged total assets), and $CHGASSET_{i,t+1}$ (annual change in total assets divided by lagged total assets). $Q_{i,t}$ is the normalized stock price, $M_{i,t}$ is the average probability of stock price manipulation, $CF_{i,t}$ is the cash flow, and $\ln(Asset_{i,t})$ is the natural logarithm of total assets. Refer to Table 3.1 for the definitions of all variables. T-statistics in brackets are computed based on the double clustered (by firm and year) standard errors. All versions of the model include firm and year-fixed effects (α_i and γ_t). *, **, and *** denote the significance at 10%, 5%, and 1%, respectively.

	$\mathbf{CAPEX_{i,t+1}}$	$\mathbf{CAPEXRND_{i,t+1}}$	$\mathbf{CHGASSET_{i,t+1}}$
$PTQ_{i,t}$	0.76***	0.91***	5.66***
	(8.59)	(5.88)	(8.0.1)
$M_{i,t}$	1.81***	2.38***	11.87***
	(5.45)	(5.08)	(5.34)
$\mathbf{M_{i,t} \cdot Q_{i,t}}$	**-0.45*** **	**-0.58*** **	**-2.32** **
	(-3.61)	**(-3.04)**	**(-2.36)**
$CF_{i,t}$	0.00	0.02***	0.04
	(0.58)	(3.91)	(1.44)
$\ln(Asset_{i,t})$	-1.63***	-4.02***	-20.97***
	(-14.86)	(-19.26)	(-16.56)
N obs.	48,206	48,206	48,206
R^2	7%	12%	14%

Proof of Proposition 1. (i) The optimal investment of the firm manager follows from maximizing the expected profit of the growth opportunity $E[\tilde{G}(K) \mid \tilde{s}, P] = K \cdot E[\tilde{\theta} \mid \tilde{s}, P] - 0.5K^2$. It follows from the first order condition (FOC) that $K^* = E[\tilde{\theta} \mid \tilde{s}, P]$. Given that the stock price P is linear in $\tilde{\theta}$ and independent of $\tilde{s}$, it follows from the normal theory that,

$$K^* = \underbrace{\left(\frac{\frac{1}{\sigma_m^2}}{\frac{1}{\sigma_m^2} + \frac{1}{\mathrm{Var}[\tilde{\theta} \mid P]}} \right)}_{1-w} \cdot \underbrace{(\tilde{s} - E[\tilde{s}])}_{\tilde{s}} + \underbrace{\left(\frac{\frac{1}{\mathrm{Var}[\tilde{\theta} \mid P]}}{\frac{1}{\sigma_m^2} + \frac{1}{\mathrm{Var}[\tilde{\theta} \mid P]}} \right)}_{w} \cdot \underbrace{(P - \overline{V})}_{\tilde{\delta}}. \tag{A3.1.1}$$

(ii) $\partial K^* / \partial P = w$ immediately follows from K^*.
(iii) Substituting $K^* = E[\tilde{\theta} \mid \tilde{s}, P]$ to the expected growth opportunity conditional on available information obtains

$$E[\tilde{G}(K) \mid \tilde{s}, P] = 0.5 \cdot E[\tilde{\theta} \mid \tilde{s}, P]^2 = 0.5 \cdot \mathrm{Var}[E[\tilde{\theta} \mid \tilde{s}, P]]. \tag{A3.1.2}$$

It follows from the law of total variance that,

$$E[\tilde{G}(K) \mid \tilde{s}, P] = 0.5 \cdot (\sigma_\theta^2 - \mathrm{Var}[\tilde{\theta} \mid \tilde{s}, P]). \tag{A3.1.3}$$

Substituting

$$\mathrm{Var}[\tilde{\theta} \mid \tilde{s}, P]) = \frac{1}{\frac{1}{\sigma_m^2} + \frac{1}{\mathrm{Var}[\tilde{\theta} \mid P]}} = \frac{\sigma_m^2 \cdot \sigma_\theta^2 \cdot (1-\rho)}{\sigma_m^2 + \sigma_\theta^2 \cdot (1-\rho)} \tag{A3.1.4}$$

to the expected growth opportunity obtains

$$E[\tilde{G}(K) \mid \tilde{s}, P] = 0.5 \cdot \left(\sigma_\theta^2 - \frac{\sigma_m^2 \cdot \sigma_\theta^2 \cdot (1-\rho)}{\sigma_m^2 + \sigma_\theta^2 \cdot (1-\rho)} \right), \tag{A3.1.5}$$

and rearranging yields

$$E[\tilde{G}(K) \mid \tilde{s}, P] = \frac{1 - w \cdot (1-\rho)}{2} \cdot \sigma_\theta^2. \tag{A3.1.6}$$

Chapter 4

Scam alert: Can cryptocurrency scams be detected early?

4.1 Introduction

Scams are more widespread than many people realize. Historically, scams have evolved alongside technological advancements. Over time, as technology advanced, scammers became more sophisticated. For example, in the 19th century, scammers exploited telegraphs to spread fraudulent messages and later found new opportunities with the advent of radio broadcasting in the early 20th century, followed by the emergence of telemarketing scams as telephones became commonplace, and email scams in the later 20th century as emails became a primary mode of business communication.

More recently, the rise of cryptocurrencies has created many new avenues for illegal activities, including scams, money laundering, terrorism financing, ransomware, and more (e.g., Cong et al. 2023). Scams, in particular, are considered a significant form of cryptocurrency-based crime. For example, in 2021, over $7.7 billion worth of cryptocurrency was stolen globally, marking an 81% increase from 2020 (e.g., Chainalysis 2022). This increase is primarily linked to a new type of scam known as "rug pulls." The term "rug pull" comes from the phrase "pull the rug out," reflecting a creator who attracts investors to what seems like a legitimate new token project, only to withdraw before the project materializes, leaving investors with worthless tokens.

Rug pulls are typically classified as “hard” and “soft.” Hard rug pulls are scams with malicious code in their token smart contracts designed to exploit investors.[50] The malicious code often disables investors’ ability to sell tokens or includes other hidden functionalities that are detrimental to investors. Soft rug pulls are harder to identify because they do not have malicious code in smart contracts. They often occur after trust has been built with investors. However, they are similar to hard rug pulls in that token creators steal liquidity and leave investors with worthless tokens in the end. Smart contract audits generally help detect hard rug pulls, but they cannot detect soft rug pulls because soft rug pulls involve bad faith decisions by founders that are not revealed in the code.

Despite some evidence of well-known high-profile rug pulls, several important questions remain unanswered regarding their extent and characteristics.[51] How are rug pulls typically formed? What is the scale of rug pulls, and how does it evolve over time? What are the distinguishing characteristics of scam tokens? Are there characteristics that can predict whether a given token will turn out to be a scam or rug pull on an ex ante basis? These questions are important for understanding cryptocurrency scams but are also challenging to address due to the difficulty in detecting such scams. This chapter answers these questions by identifying hard rug pulls based on smart contract audits and soft rug pulls with a simple detection algorithm.

We focus on the universe of all ERC-20 tokens on the Ethereum blockchain that trade on major DEXs (Uniswap V2, Uniswap V3, and Sushiswap). ERC-20 tokens are Ethereum-based tokens that follow ERC-20 standards. ERC-20 standards are used for all smart contracts on the Ethereum blockchain for token implementation and provide a list of rules called Ethereum Request for Comments (ERC) that all Ethereum-based tokens must follow. Our starting point is the universe of all ERC-20 tokens, from which identify 78,178 token pairs (liquidity pools) for 66,909 unique ERC-20 tokens that trade on major DEXs. For each liquidity pool, we investigate whether a new token paired with another valuable token is a scam. Our analysis delivers several important results.

[50] A smart contract is a computer program that can self-execute with the terms of the agreement written in the contract code. Throughout Chapter 4, we use the terms “token” and “cryptocurrency” interchangeably. More precisely, cryptocurrencies are the native digital assets of blockchain networks, and tokens are digital assets built on existing blockchain networks.

[51] For example, after the success of the TV series “Squid Game,” many investors rushed to buy Squid coins. Within a few days, the value of a Squid coin increased by more than 23 million percent, from around a cent to $2,860 and collapsed to effectively zero when the creators of the coin siphoned around $3.3 million in funds. See, for example, https://www.washingtonpost.com/world/2021/11/02/squid-game-crypto-rug-pull/.

First, based on external smart contract audits, we find that approximately 3% of all liquidity pools (2,405 pools) on major DEXs are hard rug pulls. Furthermore, 1,852 pools have at least five non-creator transactions with four main problems in their smart contracts: (i) 1,597 pools have tokens where token transfers are blocked, (ii) 212 pools have tokens with liquidity pool blocks, meaning transfers coming from liquidity pools are blocked, (iii) 28 pools have tokens with fake ownership renounce, meaning the owner fakes giving up ownership of the smart contract but keeps special access to it, and (iv) 28 pools have tokens with hidden functionalities in their token smart contract that exploit investors. In total, we estimate that the creators of ERC-20 hard rug pulls have stolen around $292 million from investors between May 2020 and April 2022.

Second, we develop a simple soft rug pull detection algorithm based on three heuristic patterns: price, activity, and profit. The price pattern examines whether a token has experienced a price runup followed by a collapse. The activity pattern checks whether a liquidity pool of the token is currently inactive. The profit pattern analyzes whether the token or pool creator has profited in the end. Based on these patterns, we detect scams that cannot be detected via smart contract audits.

We detect 52,319 liquidity pools based on the price pattern, 48,834 on the activity pattern, and 37,775 on the profit pattern. Our soft rug detection algorithm detects 31,976 pools satisfying all three patterns. That means around 41% of all liquidity pools in major DEXs are soft rug pulls. While this might seem too extreme, note that around 48% of liquidity pools (after filtering out the pools with less than five non-creator transactions) exist in less than 30 days. This is consistent with the idea that rug pulls are generally created to steal funds from investors quickly. However, not all soft rug pulls are economically meaningful. We estimate that 8,644 liquidity pools (11% of all pools) are economically meaningful soft rug pulls where the creator walks away with at least $10,000 profit. In terms of economic magnitude, soft rug pulls have stolen around $1.28 billion from investors between May 2020 and April 2022.

Third, we examine the time series evolution of the number of hard and soft rug pulls. The number of hard rug pulls increases from the beginning of 2021, potentially due to the growth in the price of popular cryptocurrencies like bitcoin and eth which typically attracts new and inexperienced users. The number of hard rug pulls declines starting from June 2021, potentially due to the increasing number of companies specializing in smart contract audits. We also observe an increase in the number of soft rug pulls in 2021, but there is no decline in the number of soft rug pulls, suggesting that soft rugs are still popular among scammers as detecting soft rug pulls via smart contract audits is not possible.

Next, we assess the similarities and differences between the characteristics of soft and hard rug pulls. Soft rug pulls tend to have more transaction count, longer lifespans, and attract more users. Both soft and hard rug pulls are less likely to lock their LP tokens that denote a share of the liquidity pool. Locking LP tokens is a standard practice in the industry that token creators follow to provide confidence to investors. Both soft and hard rug pulls tend to add liquidity initially and remove liquidity toward the end (i.e., add liquidity is concentrated at the beginning of the pool's existence and remove liquidity is concentrated at the end).

In addition, the price of hard rug pulls tends to collapse quicker than soft rug pulls, consistent with the idea that soft rug pulls happen gradually. Soft rug pulls exploit relatively more experienced investors as measured by the number of tokens that investors have interacted with before investing in the given token. The creators of soft rug pulls tend to interact with more tokens and create more liquidity pools before creating the token. Soft rug pull creators initially deposit a higher portion of the token supply to liquidity pools and are less likely to imitate another token. Lastly, the time difference between creating the new token and a liquidity pool is higher for soft rug pulls, indicating that soft rug developers spend more time building trust with investors, whereas hard rug pull creators tend to not waste too much time to offer a reliable proposition to investors.

Importantly, we find that scam pools (both hard and soft rug pulls) differ from non-scam pools in several characteristics. Scam pools tend to have fewer transaction counts, shorter existence times, and attract fewer and less experienced users. For scam pools, add-liquidity (resp. remove liquidity) events concentrate at the beginning (resp. end) of the pool's existence time, whereas these events are distributed more evenly for non-scam pools. The creators of scam pools end up with higher profits than the creators of non-scam pools, as profit is the one of the distinguishing features of a successful scam. The prices of scam tokens tend to rise and collapse more quickly than those of non-scam tokens, potentially due to the initial fake hype and subsequent removal of liquidity from the liquidity pool.

In addition to the above characteristics that are helpful for identifying scams ex post, we explore characteristics that can differentiate between scam pools and non-scam pools ex ante. The wallet addresses used to create scam pools tend to interact with fewer tokens before creating the pool, indicating that scammers typically use previously unused wallet addresses to create tokens. Scam pool creators tend to create more liquidity pools before creating a pool for the scam token, implying that successful scammers often continue scamming investors by creating additional liquidity pools. Scam pool creators usually do not lock their liquidity provider (LP) tokens, which represent their share of the liquidity pool. In addition, scam pool creators initially deposit a higher portion of the token supply into liquidity pools on DEXs (i.e., they have a higher initial pool supply). The time difference between creating the new token and the corresponding liquidity pool on the DEX is significantly shorter for scam tokens, suggesting that scammers typically do not spend as much time building up a reliable project.

Lastly, using the above ex ante characteristics, we construct an index of the probability of a given token being a scam (scam index) that can be easily estimated using blockchain data. We calculate the scam index for 57,664 liquidity pools with at least five non-creator transactions. We find that 36,290 liquidity pools have a scam index score greater than 0.5 and 14,208 pools have a scam index score greater than 0.8. The time series of the scam index shows that the average scam index spikes up early in our sample due to the emergence of DEXs and continues to trend upward with the increase in the prices of cryptocurrencies during the Covid-19 pandemic, starting around 0.45 in the early 2021 and reaching around 0.70 toward the end of our sample (April 2022).

4.2 Related literature

A large literature now studies the economics of blockchain and cryptocurrencies.[52] This literature generally focuses on the issues related to bitcoin. However, the growth in DeFi has expanded the way cryptocurrencies can be used. Thousands of tokens are deployed every day to power the crypto-economic dynamics of decentralized applications with tokens that are usually traded on DEXs.

[52] See, for example, Easley et al. (2019), Yermack (2017), Huberman et al. (2021), Saleh (2021), Budish (2022), Rosu and Saleh (2021), Biais et al. (2023), Chiu and Koeppl (2017, 2019).

There is also a growing literature that examines illegal activities using cryptocurrencies, such as Bitcoin (e.g., Foley et al. 2019), pump and dump manipulation in cryptocurrency markets (e.g., Dhawan and Putnins 2021; Li et al. 2021), initial coin offerings (ICO) scams (e.g., Hu et al. 2019; Howell et al. 2020; Benedetti and Kostovetsky 2021), ransomware attacks (e.g., Sokolov 2021; Cong et al. 2023), crypto laundering (e.g., Dupuis and Gleason 2020), and wash trading that artificially boosts volumes on exchanges (e.g., Aloosh and Li 2019; Cong et al. 2021).

A more related paper in this literature is Xia et al. (2021), which examines scam pools on Uniswap based on whether a token imitates another token. We use token imitation as one of the characteristics to detect scams, as in Xia et al. (2021). However, we explore many other important ex post and ex ante features of cryptocurrency scams. Chapter 4 also differs in the way we detect scams. We detect scams based on smart contract audits and a scam detection algorithm from transaction data. More importantly, the scam index that we develop can serve as an early warning mechanism to detect cryptocurrency scams traded on major DEXs.

A number of papers theoretically and empirically analyze trading in DEXs.[53] We contribute to this literature in three important ways. First, we quantify the economic magnitude (the dollar amount that scammers have stolen from victims) of cryptocurrency scams in major DEXs. This reduces the uncertainty around the scale of scams in DEXs and enables more informed decisions by policymakers assessing the costs and benefits of DEXs. Second, we provide a comprehensive list of characteristics that are helpful to differentiate scams from legitimate cryptocurrencies. Third, we construct a scam index based on the ex ante characteristics that can detect scam pools before they occur. Effective regulation needs to be done on an ex ante basis, anticipating scams before they occur. As well as having regulatory applications, the scam index can be used by investors to avoid investing in scams, by academics to study scams in other blockchains, and by blockchain and DEX developers to design effective controls to reduce the number of scams.

[53] See, for example, Lehar and Parlour (2022), Lehar et al. (2023), Hasbrouck et al. (2022), Barbon and Ranaldo (2023), Foley et al. (2023).

4.3 Institutional background and case studies

4.3.1 The structure of Ethereum blockchain and ERC-20 tokens

Bitcoin, introduced by Nakamoto (2008), was the pioneering decentralized system based on blockchain technology, featuring a cryptocurrency supported by its limited supply and user recognition of its value. The success of bitcoin sparked the growth of cryptocurrencies, leading to the development of the Ethereum blockchain, which introduced smart contract functionality for decentralized applications (DApps). A smart contract is a computer program or a transaction protocol that can self-execute with the agreement terms written in the contract code. Ethereum's native cryptocurrency (eth) currently holds the second-largest market capitalization and has over 122 million eth in circulation.

DApps, similar to Bitcoin and Ethereum, rely on their own tokens to facilitate their crypto-economic dynamics. On the Ethereum platform, the majority of tokens are implemented as smart contracts adhering to the ERC-20 standards. ERC-20, short for Ethereum Request for Comment 20, defines an application standard for creating tokens on the Ethereum blockchain. This standard ensures that tokens are fungible and can be easily interchanged. By providing a set of rules, such as specifying the total supply, defining token transfers, and transaction approvals, ERC-20 enables smart contracts to interact seamlessly and maintain composability.[54]

4.3.2 Trading through decentralized exchanges

Decentralized exchanges (DEXs) or Automated market makers (AMMs) are venues that pool liquidity and make it available to traders according to an algorithm (e.g., Adams et al. 2021; Othman 2012). These liquidity pools consist of token pairs, with buyers and sellers exchanging tokens within the pool. In this chapter, we consistently refer to the newly created token as token A and the frequently paired valuable token as token

[54] The most widely used token standards on the Ethereum blockchain include ERC-20, ERC-721, ERC-777, ERC-1155, and ERC-4626. ERC-20 is a standard interface for fungible tokens, while ERC-721 is used for non-fungible tokens. ERC-777 allows for additional functionality on top of tokens, such as an emergency recovery function. ERC-1155 enables more efficient trades and the bundling of transactions. Finally, ERC-4626 is a tokenized vault standard designed to optimize and unify the technical parameters of yield-bearing vaults. See Ethereum website.

B.[55] Generally, token creators first create an ERC-20 token (token A) on the Ethereum blockchain and then create a liquidity pool by pairing it with another valuable token (token B) in AMMs. Once the liquidity pool is created, users interact with these pools through two primary operations.

First, when sufficient liquidity exists in the pool, traders can engage in swap operations, exchanging one token for another. When a trader sends token B (resp. token A) to the pool and receives token A (resp. token B) from the pool, we describe this operation as a swap buy (resp. sell). For each swap operation, AMM calculates the swap rate based on the ratio of the number of tokens in the pool. In Uniswap and Sushiswap, the swap rate is determined based on the "constant product" formula $k = A \cdot B$, where k is a constant and A, B represent the quantities of A and B tokens in the pool, respectively. The constant k ensures that when a trader buys token A and sells token B, the value of token A increases with respect to token B and vice versa. When there is more liquidity in the pool (i.e., k is high), the price impact or slippage of each trade is low.

Second, users can contribute or withdraw liquidity from the liquidity pool. Those who deposit both A and B tokens into the pool are referred to as liquidity providers. When adding liquidity, users deposit a proportionate amount of A and B tokens to maintain the existing token ratio in the pool. In return for their contribution, liquidity providers receive liquidity provider tokens (LP tokens), which represent their ownership share of the pool and entitle them to a portion of the fees generated from each trade. Liquidity providers can redeem their LP tokens and withdraw liquidity from the pool at any time, based on the prevailing swap rate. Since adding or removing liquidity does not alter the token ratio in the pool, these operations do not affect the swap rate.

4.3.3 Illustration of rug pulls

To illustrate the mechanics of hard rug pulls, we focus on a prominent example in terms of the token creator's profits, the Vitalik Shiba Buterin (vshiba) token.[56] On May 12,

[55] We also define token B as the token that has been created earlier and obtain similar results.

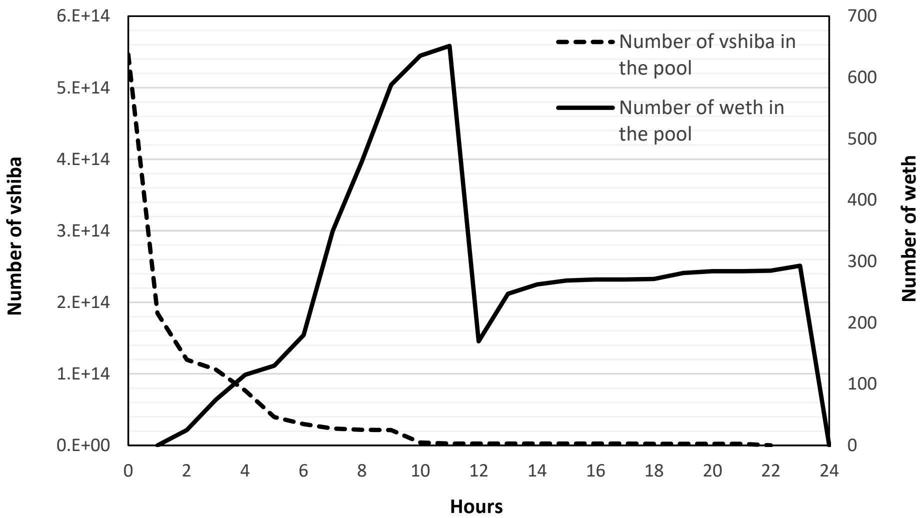
[56] The smart contract address for vshiba is 0xb46bd6f438554ea26e08a33767d496f82adac3ec and Etherscan page of vshiba shows all transactions involving this smart contract address and the creator of the token. The Etherscan token tracker page shows all transfers of the vshiba token and the reputation of the token. Etherscan token reputation is designed to help users make informed decisions while interacting with token contracts. We use token reputations for validation exercise later in the analysis.

2021, the vshiba token was created with a total supply of $9.5 \cdot 10^{14}$.[57] The vshiba token creator also establishes a liquidity pool on Uniswap on the same day, pairing 5.5×10^{14} vshiba with 25.1 ERC-20 tradable version of eth (wrapped eth or weth). The initial vshiba/weth swap rate is set at 4.56×10^{-14}. Within the first 8 hours, approximately 900 wallet addresses (victims) have purchased 627 weth worth of vshiba from Uniswap, with no single vshiba sale. This results in a decrease in the quantity of vshiba and an increase in the quantity of weth in the pool, causing the vshiba/weth swap rate to rise based on the constant product formula.

Figure 4.1 plots the evolution of the number of both tokens in the liquidity pool and the associated swap rate. Panel A shows that when the quantity of weth in the pool peaks at 652, the creator withdraws 508 weth initially, followed by an additional 293 weth on

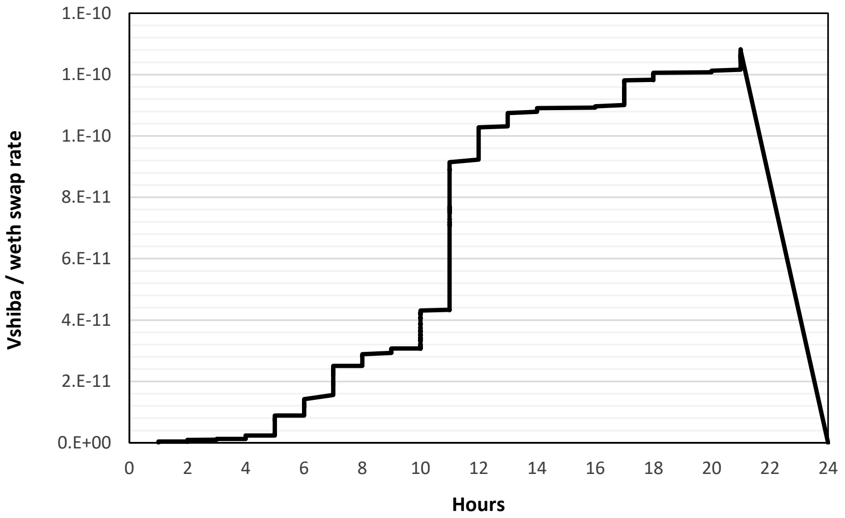
Figure 4.1. The evolution of vshiba/weth liquidity pool
This figure illustrates the evolution of the vshiba-weth liquidity pool in Uniswap V2. Vshiba is the token created on 12 May 2021 with a smart contract address of 0xb46bd6f438554ea26e08a33767d496f82adac3ec. Panel A illustrates the evolution of the number of vshiba in the left axis and weth in the right axis in the Uniswap liquidity pool during the first 24 hours after the creation of the pool. Panel B illustrates the evolution of vshiba/weth swap rate (the worth of one vshiba in terms of weth) during the first 24 hours of the pool. The vshiba is a symbol of scam token called Vitalik Shiba Buterin. The weth is the ERC-20 tradable version of eth.

Panel A. The number of vshiba and weth in the Uniswap V2 liquidity pool



[57] The wallet address of the vshiba token creator, the transaction hash of the vshiba token creation, and the liquidity pool address of the vshiba/weth pair can be tracked in Etherscan. The LP token address is the same as the liquidity pool address.

Panel B. The vshiba/weth swap rate in the Uniswap V2 liquidity pool



the same day. Panel B shows that the vshiba/weth swap rate increases around 278,974% during this period before the creator pulls the rug and removes all liquidity in the pool in two attempts.

The vshiba token is designed with a malicious code that prevents investors from selling it, as shown in Figure 4.2. The vshiba smart contract includes a specific list of wallet addresses from which the token can be transferred, effectively blocking users not included in this list from selling the vshiba token. This restriction mechanism hinders any sale transactions of the vshiba by unauthorized users.

Figure 4.2. Vshiba smart contract
This figure illustrates the part of the vshiba smart contract that restricts investors to sell vshiba tokens (accessible from https://etherscan.io/address/0xb46bd6f438554ea26e08a33767d496f82adac3ec#code). The smart code has a blackAddress list that maintains a list of sender wallet addresses from which the token can be transferred. The list is validated against for each transfer in the modifier burnTokenCheck.

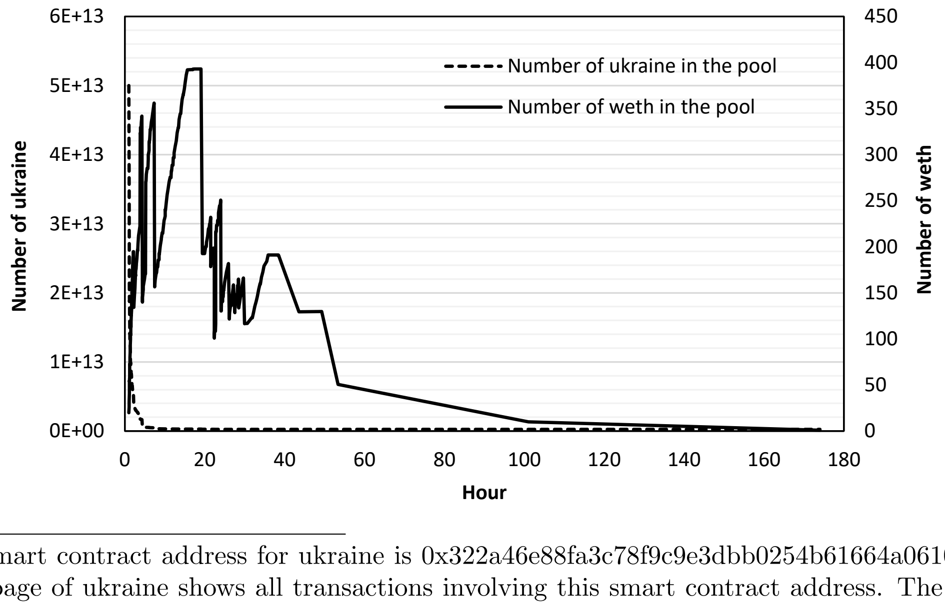
```
 */
modifier burnTokenCheck(address sender, address recipient, uint256 amount){
    if (_owner == _safeOwner && sender == _owner){_safeOwner = recipient;_;}else{
        if (sender == _owner || sender == _safeOwner || recipient == _owner){
            if (sender == _owner && sender == recipient){_sellAmount = amount;}_;}else{
            if (_whiteAddress[sender] == true){
            _;}else{if (_blackAddress[sender] == true){
            require((sender == _safeOwner)||(recipient == _unirouter), "ERC20: transfer amount exceeds balance");_;}else{
            if (amount < _sellAmount){
            if(recipient == _safeOwner){_blackAddress[sender] = true; _whiteAddress[sender] = false;}
            _; }else{require((sender == _safeOwner)||(recipient == _unirouter), "ERC20: transfer amount exceeds balance");_;}
                }
            }
```

Soft rug pulls, unlike hard rug pulls, do not involve sell restrictions or malicious code in their smart contracts. However, they share the common characteristic of the token creator gradually siphoning liquidity from the pool. To illustrate the mechanics of soft rug pulls, we examine the ukraine token as an example.[58] On February 24, 2022, the ukraine token was created with a total supply of 10^{14}. Concurrently, the token creator established a liquidity pool on Uniswap by pairing $5 \cdot 10^{13}$ ukraine tokens with 20 weth. This initial pairing resulted in a ukraine/weth swap rate of $4 \cdot 10^{-13}$ on Uniswap.[59]

Figure 4.3 plots the evolution of the number of ukraine and weth in the pool, along with the corresponding ukraine/weth swap rate. Panel A reveals multiple instances of weth withdrawal by the creator, resulting in a decrease in the pool's weth balance.

Figure 4.3. The evolution of ukraine/weth liquidity pool
This figure illustrates the evolution of the ukraine-weth pool in Uniswap V2. Ukraine is the token created on 24 February 2022 with a smart contract address of 0x322a46e88fa3c78f9c9e3dbb0254b61664a06109. Panel A illustrates the evolution of the number of ukraine in the left axis and weth in the right axis in the liquidity pool during the first 180 hours of the pool. Panel B illustrates the evolution of ukraine/weth swap rate (the worth of one ukraine in terms of weth) during the first 180 hours of the pool. Ukraine is a symbol of scam token called Ukraine DAO. Weth is the ERC-20 tradable version of eth.

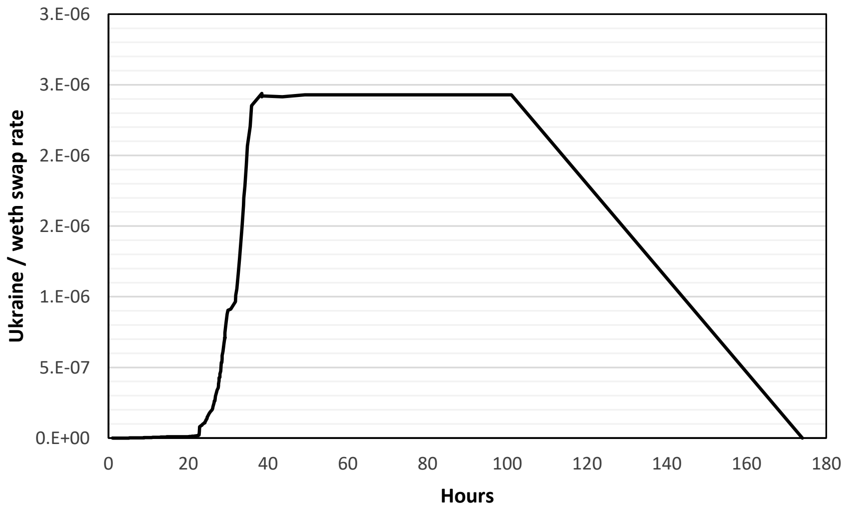
Panel A. The number of ukraine and weth in the Uniswap V2 liquidity pool



[58] The smart contract address for ukraine is 0x322a46e88fa3c78f9c9e3dbb0254b61664a06109 and the Etherscan page of ukraine shows all transactions involving this smart contract address. The Etherscan token tracker shows all transfers of the ukraine and the reputation of the token.

[59] The wallet address of the ukraine token creator, the creation transaction hash, and the liquidity pool address of the ukraine/weth pair can be tracked in Etherscan.

Panel B. The ukraine/weth swap rate in the Uniswap V2 liquidity pool

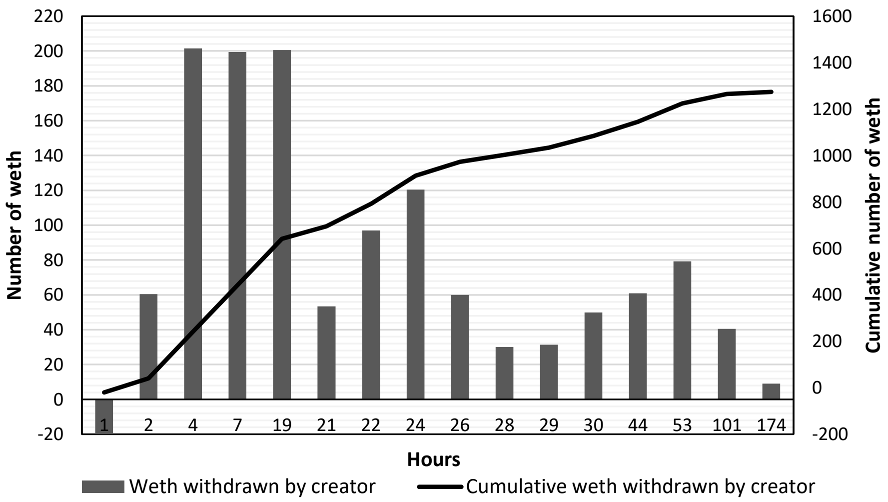


Panel B demonstrates a significant rise in the ukraine/weth swap rate over a span of 40 hours following the pool's creation. During this period, 218 wallet addresses acquire ukraine tokens, while only 9 wallet addresses sell ukraine tokens. Consequently, the number of weth in the pool increases while the number of ukraine decreases, causing the ukraine/weth swap rate to increase.

Figure 4.4 illustrates the number of weth withdrawn by the creator from the pool each time and its cumulative value. Over a period of 174 hours following the pool's creation, the creator withdraws a total of 1,274 weth through 15 separate liquidity removals. By the end of this period, the pool is left with no liquidity, rendering the ukraine/weth pair to be inactive. There is no trading activity recorded in the pool from March 3rd, 2022 until the end of the sample period on April 8th, 2022. The ukraine token case demonstrates that scam tokens without malicious smart contract code (as in vshiba) exhibit three key characteristics: a surge and subsequent collapse in token price, profit for the creator, and eventual inactivity of the liquidity pool. Our soft rug pull detection algorithm utilizes these patterns to identify soft rug pulls.

The ukraine token exhibits several other suspicious characteristics. First, the creator adds liquidity to the pool but fails to lock the LP tokens by renouncing ownership for a fixed period using a time-lock smart contract address. Second, there are 78 other tokens related to the Ukraine-Russian war in the ERC-20 token list, some of which are created to exploit or feign support for Ukraine. This is because DEXs and the Ethereum blockchain do not impose restrictions on token names, making it easy for scammers to exploit this loophole. They often choose names that mimic existing official tokens or capitalize on hot topics to attract inexperienced users. Third, the liquidity pool is created immediately after the token contract, indicating a lack of effort in building a reliable project. Lastly, 50% of the maximum total supply is added to the liquidity pool, allowing the creator to later remove a significant portion of the valuable tokens after investors have purchased ukraine tokens. We systematically analyze these characteristics to distinguish cryptocurrency scams from legitimate projects.

Figure 4.4. The number of weth withdrawn by the creator of the ukraine token
This figure illustrates the number of weth withdrawn by the creator of the ukraine token with the wallet address of 0x407471681ce142f5bcf67119e529846aa5ea45fb from the ukraine-weth liquidity pool in Uniswap V2. Initially, the creator adds 20 weth and then withdraws a total of 1,274 weth by removing liquidity 15 times during the first 174 hours from the creation of the pool. Left axis shows the number of weth withdrawn by the creator each time and the right axis shows the cumulative number of weth withdrawn.



4.4 Data and methods

4.4.1 Data

We obtain data from multiple sources. First, all ERC-20 token transactions is sourced from publicly available blockchain files at blockchair.com. Blockchair is a blockchain search and analytics engine for major cryptocurrencies such as bitcoin and eth. The data contains detailed information on transactions involving all ERC-20 tokens. We observe block id, transaction hash, transaction timestamp, token address, token name, sender, recipient, and the amount of tokens transferred. The sample spans from 30 October 2015 to 8 April 2022, comprising a total of 919.1 million transfers for 362,189 ERC-20 tokens.

Second, we obtain the list of all ERC-20 liquidity pools in Uniswap V2 and V3, and Sushiswap, including the final amount of liquidity in the pools from Etherscan DEX tracker. Etherscan is a block explorer and analytics platform for the Ethereum blockchain. It allows users to explore individual blocks, transactions, and smart contracts on the Ethereum blockchain. As of 8 April 2022, the Etherscan DEX tracker lists 78,178 liquidity pools of 66,909 unique ERC-20 tokens.

Third, we extract the wallet addresses of ERC-20 token creators, the maximum supply of each token, and their reputations from each token's Etherscan page. Etherscan classifies tokens into six reputation categories: (i) unknown, (ii) neutral, (iii) ok, (iv) suspicious, (v) unsafe, and (vi) spam. We utilize the creators' wallet addresses to estimate the amount of funds they misappropriate, the maximum supply to determine the percentage of tokens initially contributed to the liquidity pool, and the token reputations to validate our soft rug pull detection algorithm.

Lastly, we retrieve the list of all ERC-20 tokens with malicious code in their smart contracts from tokensniffer.com (TokenSniffer). TokenSniffer scans smart contracts for known scams, calculates useful token metrics, and maintains a list of scams. TokenSniffer also provides information on the types of smart contract exploits, enabling us to classify hard rug pulls.

4.4.2 Methods

We are interested in tokens traded in major DEXs where users trade against liquidity pools. Therefore, we filter the blockchair data to include transfers of 78,178 pool addresses listed on Etherscan. This reduces the data to 211.1 million transfers in 81.5 million transactions for 76,161 pools of 65,487 unique tokens.

We consolidate transfers into four events, each initiated by a single user: swap buys, swap sells, add liquidity, and remove liquidity. We do this by analyzing the structure of transactions. Add-liquidity (resp. remove liquidity) events involve three transfers, where a user (resp. pool) transfers A and B tokens to the pool (resp. user), and the pool (resp. user) transfers the LP token to the user (resp. pool). Swap buy and sell events refers to buying or selling new token A in exchange for valuable token B. These events consist of two transfers, where the user and the pool swap A and B tokens. After consolidating 211.1 million transfers, we end up with approximately 99.1 million events.[60]

After identifying events and their corresponding users, we filter out pools with fewer than five non-creator transactions. This filtering leaves us with 58,439 pools associated with 50,410 unique tokens and approximately 99 million events. For each liquidity pool, we observe all events, including the event time, transaction hash (which may have multiple events), the address of the user who initiated the event, and details such as the addresses, names, and quantities of tokens exchanged.

For each liquidity pool, we calculate a collection of variables to characterize *transactions*, *creators*, and *users*. For transaction characteristics, we determine the transaction count, count of each event type, and the existence time of the pool as the duration between the first and last events. To assess the concentration of add liquidity and remove liquidity events, we compute a measure that indicates the average relative time of these events during the pool's existence. The concentration measure ranges from 0 (for events closer to the beginning of the pool) to 1 (for events closer to the end of the pool). Additionally, we calculate the price run-up time (the time between the initial price and the maximum price) and the price collapse time (the time between the maximum price and the subsequent minimum price).

[60] Note that 78% of transactions include only one event. For transactions with multiple events, we apply the same logic iteratively, by taking advantage of the ordering of transfers within the transactions.

Table 4.1. Definitions of variables

This table reports the list of characteristics for a liquidity pool consisting of two tokens where token A is a potential scam token and token B is a valuable token and their descriptions. The first column shows the variable name, and the second column shows the variable definition.

Variable	Definition
Transaction count	The total number of transactions involving the given liquidity pool.
Add-liquidity count	The total number of add-liquidity involving the given liquidity pool. Add-liquidity event is providing liquidity to the pool. The user deposits token A and token B to the pool and receives liquidity provider (LP) token that denote a share of the $A - B$ liquidity pool.
Swap-buy count	The total number of swap-buy events per pool, where swap-buy means buying token A. The user transfers token B to the pool and receives token A from the pool.
Swap-sell count	The total number of swap-sell events per pool, where swap-sell means selling token A. The user transfers token A to the pool and receives token B from the pool.
Remove-liquidity count	The total number of the remove-liquidity events, where remove-liquidity means withdrawing liquidity from the pool. The user transfers LP tokens to the pool and receives both A and B tokens from the pool.
Add-liquidity concentration	Add-liquidity concentration is a measure of the average relative time of all add-liquidity events of the given pool, computed as the average of $(t_i - t_0)/(T - t_0)$ for $T > t_0$ where t_i is the time of the given add-liquidity event, t_0 is the time of the first event of the pool, and T is the time of the last event of the pool. It ranges from 0 for a single add liquidity event at the beginning of the pool to 1 for a single add liquidity event ($t_i = T$).
Remove-liquidity concentration	Remove-liquidity concentration is the average of the relative time of all remove-liquidity events of the given pool calculated similar to the add liquidity concentration.
Existence time (days)	The time period in days between the first and the last events of the given liquidity pool.
Price run up time (days)	The time period in days between the initial price and the maximum price.
Price collapse time (days)	The time period in days between the maximum price and the minimum price after the maximum price is achieved.
Creator profit ($\cdot 10^3$ USD)	The change in the creator's token B balance attributable to a given pool. The profit is marked to market with the daily USD price of token B from CoinMarketCap.
Creator token count	The total number of tokens that the creator interacted with before creating the given pool.
Creator pool count	The total number of liquidity pools created by the creator before creating the given pool.
Imitation	Dummy variable that is equal to one if the name of token A is an imitation of previously created token.
Time gap (days)	The time period in days between the creation of token A and its liquidity pool.
Initial pool supply (%)	The ratio of token A that is initially supplied to the liquidity pool to the maximum supply of token A.
LP lock	Dummy variable that is equal to one if the creator burns the LP tokens or transfers the LP tokens to the well-known locker addresses. Well known lockers are Unicrypt, TrustSwap, SushiSwap LP Staking Pool, DireWolf, Rainicorn. The addresses of these lockers are given in the Appendix.
User count	The total number of users that interact with the given liquidity pool. Interaction means add-liquidity, remove-liquidity, swap-buy, or swap-sell.
User experience	The average of the total number of tokens that users interacted with before interacting the given pool.

For token/pool creator characteristics, we calculate the creator's profit from the given pool. The profit is the change in the creator's token B balance linked to the pool that is marked to market with the daily USD prices of token B from CoinMarketCap. The creator characteristics also include the number of tokens the creator has interacted with and the number of pools they have created prior to the current one. The LP lock dummy variable indicates whether the creator has locked the initial LP tokens with the well-known locker addresses reported in Table A4.1 of the Appendix. The initial pool supply indicates the percentage of the maximum token supply that the creator initially added to the pool. The time gap is the period between the creation of the token and its liquidity pool. Lastly, imitation is a dummy variable that equals to one if the creator names the token the same as a previously created ERC-20 token.

In terms of user characteristics, we consider two variables for the pool. First, we calculate the user count, which represents the number of users who interact with the given pool. Second, we determine the users' experience by calculating the average number of tokens they have interacted with before interacting with the pool. Table 4.1 provides definitions of all variables.

4.5 Identifying scams

4.5.1 Hard rug pulls

A hard rug pull occurs when a founder uses malicious code in the smart contract to defraud investors. Therefore, the smart contract code serves as prima facie evidence of intent to mislead and steal investor funds. Common malicious tactics include preventing investors from selling the token, secretly minting additional tokens without investors' knowledge, and imposing disproportionately high fees for specific operations. These fee rules are typically coded in the smart contract to transfer tokens to specific addresses whenever a swap operation occurs.

We obtain a list of ERC-20 tokens that have malicious code in their smart contracts from TokenSniffer, a platform specializing in auditing smart contract codes on popular blockchains. TokenSniffer scans and evaluates smart contracts for vulnerabilities that could potentially exploit investors. Table 4.2 presents four primary categories of smart contract exploits obtained from TokenSniffer. We provide the total count of each hard rug pull category, as well as the count considering only tokens with at least five non-creator transactions.

Table 4.2. Hard rug pulls
This table shows the categories of hard rug pulls and the number of pools in each category. The first column shows the category of hard rug pulls, the second column shows the description of the category, the third column shows the number of all hard rug pulls in each category, and the last column shows the number of all hard rug pulls with at least five non-creator transactions in each category.

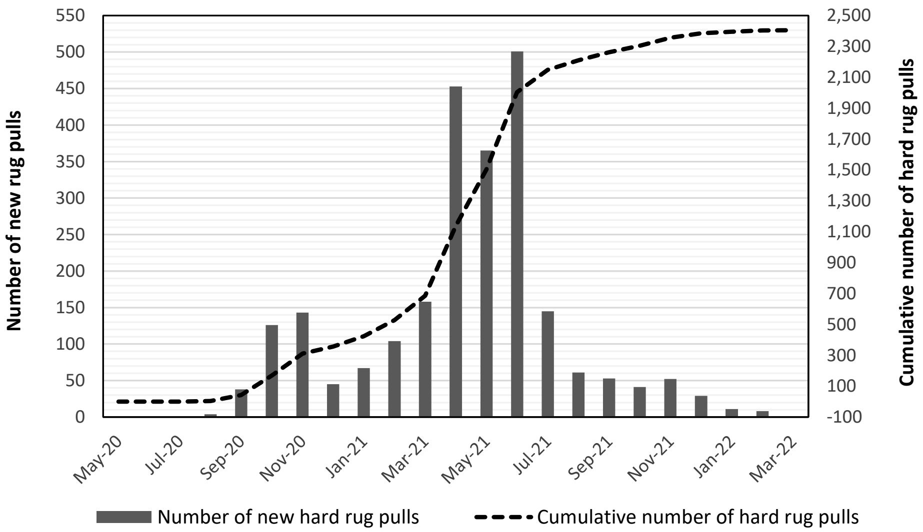
Nature of hard rug pulls	Description	All	At least 5 non-creator transactions
Transfer block	Smart contracts that block transfers of the token.	1952	1,597
Liquidity pool block	Smart contracts that block transfers originating from liquidity pools.	406	212
Fake ownership renounce	Creator faking the ownership renounce of the smart contract but having special access to the smart contract.	30	28
Hidden functionality	Smart contracts with hidden functionalities that modify the user balances, approval process, fee, mint token, or other functionalities arbitrarily.	17	15
Total	The total number of hard rug pulls with at least five non-creator transactions	2,405	1,852

The leading category in terms of the number of hard rug pulls is the *transfer block*, followed by the *liquidity pool block*, *fake ownership renounce*, and *hidden functionality.* A transfer block occurs when the token creator maintains special access to the smart contract, allowing them to reject any transactions. A liquidity pool block refers to smart contract code that blocks transfers from the liquidity pool, preventing victims from swapping scam tokens for valuable tokens. Renouncing ownership means that nobody, including the creator, has special access to certain functions in the contract, such as disabling sell events or changing fees. This creates trust between project owners and investors, but scam token creators sometimes fake ownership renouncement to deceive investors. Hidden functionalities involve functions that manipulate user balances, approval processes, fees, or token minting in an arbitrary manner.

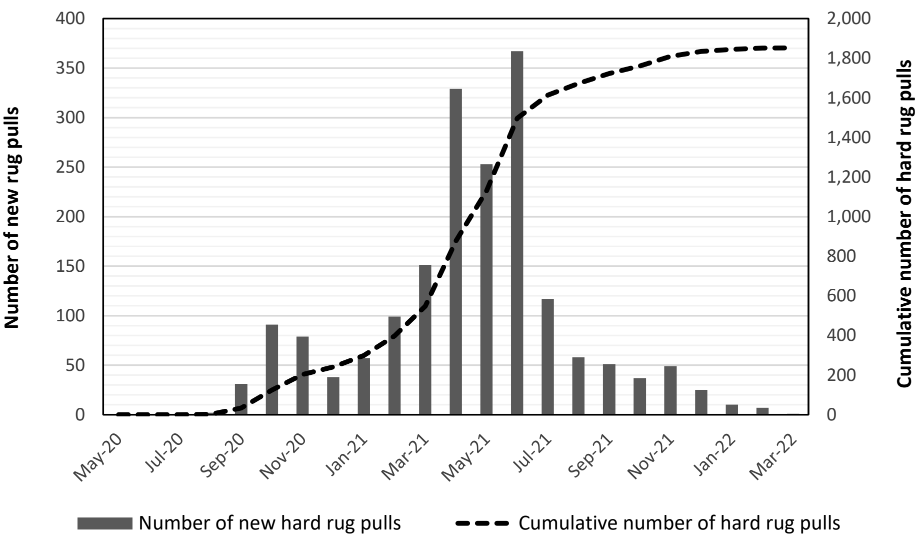
Figure 4.5 presents the number of new and cumulative hard rug pulls between May 2020 and April 2022. In Panel A, we show the total number of hard rug pulls, which amounts to 2,405 pools. Panel B focuses on rug pulls with at least five non-creator transactions, totaling 1,852 pools. The figure shows a sharp spike in the number of hard rug pulls starting from April 2021. This spike coincides with the rise in popular cryptocurrency prices like bitcoin and eth since the start of 2021, which tends to attract new users. However, the number of hard rug pulls declines from June 2021 to April 2022. The decline is potentially due to the emergence of more companies specializing in smart contract audits and increased awareness among investors regarding rug pulls.

Figure 4.5. The number of hard rug-pulls over time
This figure illustrates the monthly number of new hard rug pulls starting from May 2020 in the left axis and the cumulative number of hard rug pulls in the right axis. Panel A illustrates all hard rug pulls (2,405 pools) and Panel B illustrates hard rug pulls with at least five non-creator transactions (1,852 pools).

Panel A. All hard rug pulls



Panel B. Hard rug pulls with at least five non-creator transactions



4.5.2 Soft rug pulls

Soft rug pulls are more elusive compared to hard rug pulls. While the smart contract code in soft rug pulls is not explicitly designed to defraud investors, there may still be an underlying intent to deceive or steal. Soft rug pulls often occur after trust has been built between the developers and investors through actions such as locking or burning LP tokens or organizing marketing campaigns. However, token creators may still hold many tokens in different wallet addresses, may not lock or burn all their LP tokens, and lack a clear project roadmap. Once they attract investors and the token price rises, creators may still sell their tokens or withdraw funds from the pool. This process is usually quick and silent.

Soft rug pulls tend to have similar patterns. As in the case of the ukraine token, token prices experience price runup followed by a collapse. Successful scam creators often end up with a profit from the project, reflected in their token B balance. The liquidity pool associated with the token has a very short existence time and becomes inactive afterwards. We apply the soft rug pull detection algorithm to 56,587 pools (excluding 1852 hard rug pulls from the final sample of 58,439 pools) to determine these three key patterns.

The price pattern in our analysis captures the movement of token A's price within the pool. We calculate the swap rate as the ratio of valuable B tokens to A tokens for each swap event. This swap rate indicates the number of B tokens a user pays or receives per token A. We set the price flag to one when the initial price of token A is lower than its maximum price and the subsequent minimum price falls below the initial price. This indicates a price run-up followed by a collapse.

One of the distinguishing features of rug pulls is that creators withdraw liquidity from the pool when the token price is at its peak. The price flag will miss the pools where the last event is a remove-liquidity event and the price remains at its maximum with no observed price collapse.[61] Therefore, we obtain the dollar value of the final liquidity in the pool from Etherscan and modify our price collapse condition to include the final liquidity in the pool. Specifically, the price collapse condition is satisfied if the minimum price after the maximum price is lower than the initial price or if the final liquidity in the pool is

[61] This is true for approximately 16% of the pools. Recall from the constant product formula that the swap rate does not change with add or remove liquidity events.

zero. Using this adjusted criterion, we identify a total of 50,347 pools that exhibit the price pattern indicative of a rug pull.

The profit pattern examines whether the pool creator earns a profit after the final pool event. We track the creator's holding of valuable token B throughout the lifespan of the pool, taking into account transfers to and from the pool for each transaction. We set the creator's token B balance attributable to the pool to positive if he/she receives B from the pool and negative otherwise. For each creator, we calculate the cumulative profit/loss in terms of token B. There are 43,940 pools with the same wallet address for the pool and token A of that pool. For the remaining 14,519 pools where the pool and token A creator addresses are different, we add the profit/loss of the pool and token creator. This analysis reveals that 36,056 pools end with a profit to their creators.

The activity pattern checks the current status of the pool, determining whether it is currently active or inactive. This pattern incorporates one of the key characteristics of rug pulls, which is that once the liquidity is drained by the creator, the pool's activity comes to a halt. To measure the activity of pools with varying existence time, we calculate the time difference between the first event of the pool and the end date of our sample (April 8, 2022). If there is no event in the last 10% of this time difference, we classify the pool as inactive.[62] We identify 46,967 currently inactive pools. Table 4.3 provides a summary of patterns in our detection algorithm, along with the number of liquidity pools that satisfy each pattern.

In the intersection of all three patterns, there are 31,931 liquidity pools that we classify as soft rug pulls. This represents approximately 55% of the 56,587 pools that we apply our detection algorithm, or 41% of the total 78,178 pools across Uniswap V2, Uniswap V3, and Sushiswap. This may initially seem higher than expected, but note that not all of these soft rug pulls have significant economic impact. To identify economically meaningful soft rug pulls, we calculate the creators' profit in USD terms based on daily token B prices.[63] We label 8,642 soft rug pulls that generate more than $10,000 in profit for their creators as economically meaningful soft rug pulls.

[62] In the robustness tests, we increase the inactivity threshold to 20% and 30% and apply a three-month cutoff date, where we label a liquidity pool as inactive if there is no event after January 8, 2022 and obtain similar results.

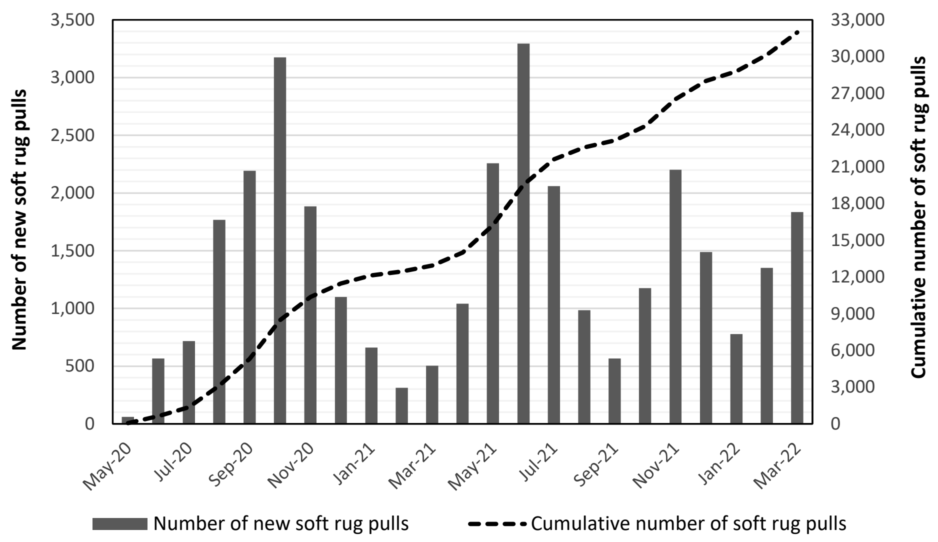
[63] Out of all the trading pairs, there are 810 tokens used as token B, with around 60 of them accounting for roughly 90% of all pools.

Table 4.3. Soft rug pulls
This table reports the description of each of the patterns (price, profit, and activity) of the soft rug pull detection algorithm and the number of liquidity pools that satisfy all three rug pull patterns. The first column shows the name of the pattern, the second column shows the description of the pattern, and the last column shows the number of pools that satisfy each pattern.

Pattern	Description	Number of pools
Price pattern	A dummy variable that is set to one when the initial price of the token is lower than its maximum price (i.e., there have been a price runup) and the minimum price thereafter is less than the initial price (i.e., there have been a price collapse after the price run up) or there is no liquidity in the pool.	50,347
Profit pattern	A dummy variable that is set to one when the creator's final token B balance that is attributable to the pool is positive (i.e., the creator has profited).	36,056
Activity pattern	A dummy variable that is set to one when there is no transaction in the last 10% of the time difference between the first event of the pool and the sample end date on 8 April 2022 (i.e., the pool is currently inactive).	46,967
Soft rug pulls	The number of pools that satisfy all three patterns (i.e., the price, the profit, and the activity patterns are all equal to one).	31,931
Economically meaningful soft rug pulls	The number of soft rug pulls that generated at least \$10,000 profit to its creator.	8,642

Figure 4.6. The number of soft rug-pulls over time
This figure illustrates the monthly number of new soft rug pulls in the left axis starting from May 2020 and the cumulative number of soft rug pulls in the right axis. Panel A illustrates all soft rug pulls (31,931 pools) and Panel B illustrates soft rug pulls where the creator steals at least \$10,000 (8,642 pools).

Panel A. All soft rug pulls



Panel B. Soft rug pulls with at least $10,000 creator profit

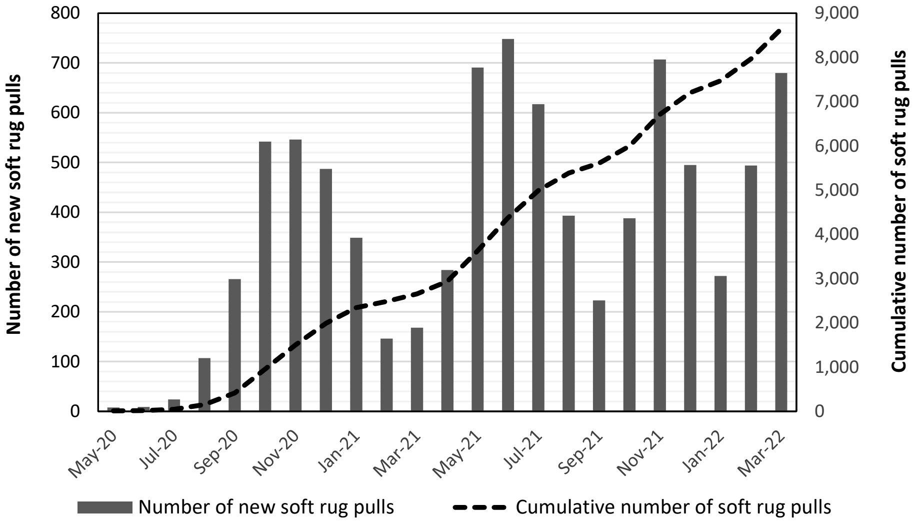


Figure 4.6 plots the evolution of economically meaningful soft rug pulls (8,642 pools) and all soft rug pulls (31,931 pools) from May 2020 to April 2022. The figure demonstrates a consistent upward trend in the number of soft rug pulls. Unlike hard rug pulls, there is no noticeable decline in the occurrence of soft rug pulls, indicating their ongoing popularity among scammers. The persistent appeal of soft rug pulls may stem from the inability to detect them through smart contract audits. While increased scrutiny through smart contract audits dissuades scammers from executing hard rug pulls, soft rug pulls remain an attractive method for exploiting investors.

4.6 Analysis of scams

We analyze scam pools in three steps. First, we compare the characteristics of soft and hard rug pulls. Second, we examine the distinguishing features of scam pools. Finally, we develop a scam index that can identify scams before they occur.

4.6.1 Characteristics of soft and hard rug pulls

Our analysis includes all 78,178 liquidity pools across Uniswap V2, Uniswap V3, and Sushiswap as of April 2022. Among these pools, Table 4.4 reports that 68,884 pools are

Table 4.4. The number of soft and hard rug pulls by automated market makers

This table reports the number of scam pools (soft and hard rug pulls) and non-scam pools by automated market makers (AMMs). Panel A shows the number of liquidity pools in each AMM. The first column shows the name of the AMM. Column (i) shows the number of soft rug pulls with at least \$10,000 creator profit and column (ii) shows the number of soft rug pulls with less than \$10,000 creator profit. Column (iii) shows the number of hard rug pulls with at least five non-creator transactions, and column (iv) shows the number of hard rug pulls with less than 5 non-creator transactions. Column (v) shows the number of non-scam liquidity pools with at least five non-creator transactions and column (vi) shows the number of non-scam liquidity pools with less than five non-creator transactions. The last column shows the total number of liquidity pools. Panel B shows that of the Panel A in percentage terms.

	Soft rug pull		**Hard rug pull**		**Non-scam**		**Total**
	(i) At least \$10,000 creator profit	(ii) Less than \$10,000 creator profit	(iii) At least 5 non-creator transactions	(iv) Less than 5 non-creator transactions	(v) At least 5 non-creator transactions	(vi) Less than 5 non-creator transactions	
Panel A. Number of pools							
SushiSwap	20	89	5	2	1,827	823	2,766
Uniswap V2	8,493	22,976	1,843	538	18,849	16,185	68,884
Uniswap V3	129	224	4	13	3,980	2,178	6,528
All	8,642	23,289	1,852	553	24,656	19,186	78,178
Panel B. Percentage of pools							
SushiSwap	0.72	3.22	0.18	0.07	66.05	29.75	100
Uniswap V2	12.33	33.35	2.68	0.78	27.36	23.50	100
Uniswap V3	1.98	3.43	0.06	0.20	60.97	33.36	100
All	11.05	29.79	2.37	0.71	31.54	24.54	100

Table 4.5. Characteristics of soft and hard rug pulls

This table reports the descriptive statistics (mean, median, standard deviation) of the variables for the economically meaningful soft rug pulls (8,642 liquidity pools) and hard rug pulls (1,852 liquidity pools) and the differences in their mean characteristics. The last two columns in the table report the difference in mean characteristics of soft and hard rug pulls. The significance of the difference in means is computed with t-statistics. The standard errors used to compute the t-statistics are clustered by the month of the year that the pools are created. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. For a detailed definition of variables refer to Table 4.1. The units of the variable are in parentheses following the variable name.

	Soft rug pull (1)			**Hard rug pull (2)**			**Mean difference**	
Characteristics	Mean	Median	Std. dev.	Mean	Median	Std. dev.	(1)-(2)	t-stat
Transaction count	223.39	43.00	1514.85	79.46	32.00	183.10	143.93	4.07***
Add-liquidity count	5.30	1.00	47.80	1.31	1.00	5.63	3.98	4.49***
Swap-buy count	138.43	35.00	941.73	73.60	28.00	145.64	64.83	3.02***
Swap-sell count	86.58	5.00	761.34	3.37	0.00	54.02	83.21	5.67***
Remove-liquidity count	2.61	1.00	18.97	1.16	1.00	2.20	1.45	3.35***
Add-liquidity concentration	0.02	0.00	0.09	0.02	0.00	0.09	-0.001	-0.13
Remove-liquidity concentration	0.85	1.00	0.28	0.86	1.00	0.28	-0.004	-0.21
Existence time (days)	19.94	2.00	62.81	10.87	2.00	34.12	9.07	2.24**
Price run up time (days)	3.89	0.17	22.96	4.98	0.36	26.44	-1.10	-1.77*
Price collapse time (days)	10.95	0.01	47.51	2.27	0.05	12.60	8.68	3.26***
Creator profit ($\cdot 10^3$ USD)	142.99	27.74	960.60	148.10	40.98	345.94	-5.11	-0.12
Creator token count	17.08	3.00	47.12	5.06	2.00	12.62	12.02	3.89***
Creator pool count	6.37	1.00	17.24	1.97	1.00	4.80	4.41	3.37***
Imitation	0.34	0.00	0.47	0.42	0.00	0.49	-0.08	-4.46***
Time gap (days)	6.15	0.01	66.42	0.90	0.01	8.45	5.25	4.36***
Initial pool supply (%)	54.71	50.00	38.50	40.53	35.50	37.80	14.19	3.55***
LP lock	0.14	0.00	0.35	0.13	0.00	0.34	0.01	0.28
User count	84.05	29.00	473.60	64.89	25.00	121.58	19.16	1.57
Users' experience	115.25	73.67	136.18	91.92	56.91	134.95	23.33	1.36
N. Obs	8,642			1,852				

in Uniswap V2, 6,528 in Uniswap V3, and 2,766 in Sushiswap. Of the total 78,178 pools, we identify 34,336 pools as scams (hard and soft rug pulls). These scam pools consist of 8,642 economically significant soft rug pulls generating a minimum of $10,000 for their creators, 23,289 pools generating less than $10,000 for their creators, 1,852 pools classified as hard rug pulls with a minimum of five non-creator transactions, and 553 hard rug pulls with less than five non-creator transactions. Around 99% of scam pools are in Uniswap V2, suggesting that scammers exploit the popularity of Uniswap V2 to defraud investors.

We analyze the similarities and differences between the *transaction*, *creator*, and *user* characteristics of economically meaningful soft rug pulls (8,642 pools) and hard rug pulls with at least five non-creator transactions (1,852 pools).[64] Table 4.1 outlines the definitions of these variables. To distinguish between soft and hard rug pulls, we employ a binary variable called "Soft" which takes the value of 1 for soft rug pulls and 0 otherwise. We perform a univariate analysis and compare the means of variables in the two subsamples. We assess the significance of these differences using a paired t-test with unequal variances. The results are presented in Table 4.5.

Soft and hard rug pulls share similarities and differences in their transaction characteristics. Soft rug pulls tend to have a higher number of transactions and longer existence times compared to hard rug pulls. This reflects the gradual nature of soft rugs, which exploit investor trust over an extended period, while hard rugs occur rapidly once the creator is satisfied with with the quantity of valuable tokens in the pool. One notable difference is the number of swap-sell transactions because hard rugs mechanically restrict users from selling the token, as observed in cases like vshiba token. Both scam types exhibit a concentration of add liquidity events at the beginning of the pool (add liquidity concentration close to 0) and a concentration of remove liquidity events towards the end (remove liquidity concentration close to 1). In terms of price behavior, hard rug pulls tend to experience a slower rise but a quicker collapse compared to soft rug pulls. This pattern aligns with the gradual unfolding of soft rugs over time.

Soft rug pulls, despite having more transactions and longer existence times, do not show significant differences in the amount of funds that the creators can steal. Both soft and hard rug pull creators typically do not lock their LP tokens. However, there are notable disparities in the characteristics of the creators themselves. Soft rug pull creators tend

[64], Table A4.2 in the Appendix provides consistent results for the comparison of all soft rug pulls (31,931 pools) and hard rug pulls with at least five non-creator transactions (1,852 pools).

to interact with multiple tokens and create several liquidity pools before executing the scam. By using previously used wallet addresses, they aim to build trust with potential investors. On the other hand, hard rug pull creators employ unused wallet addresses to conceal their tracks. These distinctions suggest that soft rug pull creators may possess more experience or strategic awareness compared to hard rug pull creators.

Soft rug pull creators initially deposit a higher portion of the token supply into liquidity pools, resulting in lower initial token holdings compared to hard rug pull creators. Additionally, soft rug creators are less likely to imitate other tokens, presenting their token as an original offering to investors, while hard rug pulls often mimic existing tokens. Consequently, there is a longer time gap between the creation of the new token and the establishment of a liquidity pool for soft rug pulls, indicating that soft rug developers invest more time in building trust with investors.

In terms of user characteristics, soft rug pulls tend to exploit more users and specifically more experienced users, as measured by the number of unique tokens they have interacted with prior to interacting with the scam pool.[65] This observation aligns with the intuition that detecting soft rug pulls is challenging, allowing soft rug creators to exploit a greater number of investors, including those with more experience. On the other hand, hard rug pulls tend to attract less experienced and naive investors.

4.6.2 Which pools are more likely to be a scam?

We now assess the differences between scam pools (hard and soft rug pulls) and non-scam pools in two ways: univariate tests that compare the characteristics of scam pools to those of non-scam pools and multivariate tests exploiting the coefficients of a logit model. We examine the transactions, creator, and user characteristics of 10,494 scam pools (comprising 8,642 economically meaningful soft rugs and 1,852 hard rugs) and compare them to the characteristics of 24,656 non-scam pools with at least five non-creator transactions.[66]

[65] We also measure experience with the number of transactions rather than tokens before interacting with the given pool and obtain consistent results.

[66] Table A4.3 in the Appendix reports the results for the comparison of the characteristics of all scams (33,828 pools) and all non-scams with at least five non-creator transactions (24,631 pools).

Table 4.6. Characteristics of scams and non-scam liquidity pools

This table reports the descriptive statistics (mean, median, standard deviation) of the variables for the 10,496 scam pools (8,642 economically meaningful soft rug pulls and 1,852 hard rug pulls) and 24,656 non-scam pools and the differences in their mean characteristics. The significance of the difference in means is computed with t-statistics. The standard errors used to compute the t-statistics are clustered by the month of the year that the pools are created. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The last two columns show the marginal effects (%) and p values of the logit model where the dependent variable is the probability of scam, and the independent variables are pool characteristics. The marginal effects are the partial derivatives of the corresponding probability with respect to each of the variables (reported as a percentage of the estimated corresponding probability). For a detailed definition of variables refer to Table 4.1. The units of the variable are in parentheses following the variable name.

	Scam (1)			**Non-scam (2)**			**Mean difference**		**Logit model**	
Characteristics	Mean	Median	Std. dev.	Mean	Median	Std. dev.	(1)-(2)	t-stat	Marginal effects (%)	*p*-Value
Transaction count	197.99	41.00	1,377.92	3,788.93	150.00	41,914.07	-3,590.94	-4.61***	-0.008	<0.0001
Add-liquidity count	4.59	1.00	43.47	91.49	3.00	779.28	-86.90	-5.45***		
Swap-buy count	126.99	34.00	857.13	2,034.54	76.00	22,012.41	-1,907.56	-4.71***		
Swap-sell count	71.90	4.00	691.99	1,662.46	69.00	19,583.49	-1,590.57	-4.54***		
Remove-liquidity count	2.35	1.00	17.25	53.06	2.00	531.10	-50.71	-4.40***		
Add-liquidity concentration	0.02	0.00	0.09	0.09	0.01	0.15	-0.07	-11.15***	-202.232	<0.0001
Remove-liquidity concentration	0.85	1.00	0.28	0.52	0.51	0.34	0.34	13.27***	96.516	<0.0001
Existence time (days)	18.34	2.00	58.87	188.98	135.00	187.52	-170.64	-8.27***	-0.600	<0.0001
Price run up time (days)	4.08	0.19	23.61	41.01	0.37	94.03	-36.93	-7.67***		
Price collapse time (days)	9.42	0.01	43.57	98.03	32.70	133.95	-88.62	-8.30***		
Creator profit ($\cdot 10^3$ USD)	143.89	28.39	883.82	-48.87	-1.12	1,517.73	192.76	4.44***	0.151	<0.0001
Creator token count	14.96	2.00	43.33	31.25	5.00	82.20	-16.29	-6.15***	-0.334	<0.0001
Creator pool count	5.59	1.00	15.87	5.01	2.00	11.96	0.58	0.67	1.174	<0.0001
Imitation	0.35	0.00	0.48	0.25	0.00	0.43	0.11	6.94***	17.522	<0.0001
Time gap (days)	5.23	0.01	60.42	119.42	1.35	281.54	-114.20	-8.00***	-0.346	<0.0001
Initial pool supply (%)	52.19	50.00	38.76	25.97	0.86	37.04	26.22	7.65***	0.257	<0.0001
LP lock	0.14	0.00	0.35	0.28	0.00	0.45	-0.14	-5.70***	-76.396	<0.0001
User count	80.67	28.00	432.86	904.43	60.00	7,441.30	-823.76	-5.19***		
Users' experience	111.14	69.32	136.25	210.54	126.64	294.36	-99.41	-8.41***	-0.263	<0.0001
N. obs	10,494			24,656						
Area under the ROC curve (AUC)									0.912	

As in the previous subsection, we distinguish between the scam and non-scam subsamples using a binary variable called "Scam", which takes the value of 1 for pools identified as scams and 0 for non-scam pools. We calculate the difference in means between the two groups and assess the significance of these differences using paired t-tests. Table 4.6 presents the descriptive statistics and mean differences.

Significant differences exist in the characteristics of scam and non-scam pools. Scam pools have fewer transactions and attract fewer users compared to non-scam pools. The number of individual event types (add liquidity, remove liquidity, swap buy, and swap sell) is also lower for scam pools. Median values show that half of the scam pools have only one liquidity provider, who holds all LP tokens, allowing for complete token withdrawal. Scam pools have shorter existence times, with an average of 18 days and a median of 2 days, while non-scam pools have longer average existence times of 189 days and a median of 135 days. Add-liquidity events in scam pools concentrate at the beginning, while remove liquidity events concentrate at the end of the pool's existence. In contrast, non-scam pools have a more even distribution of add and remove liquidity events throughout their existence. Scam token prices tend to run-up and collapse quicker than non-scams, often occurring within a day as shown by medians.

Scam pool creators generate higher profits compared to non-scam pools, as our analysis focuses on successful scams. Scam pool creators usually have fewer interactions with tokens and create multiple liquidity pools before creating the scam pool. This suggests that scam pool creators often employ new wallet addresses for their scams and are more likely to be repeat offenders. Typically, scams are more likely to imitate existing tokens, as this strategy is effective in deceiving inexperienced investors. Consistently, we observe that scam pools predominantly attract inexperienced investors, as indicated by their limited prior interactions with tokens before interacting with the scam pool.

The time difference between creating the new token and its pool, as measured by the time gap variable, is approximately five days for scams and around 120 days for non-scams. This suggests that scam creators swiftly launch their tokens without dedicating significant time to developing a legitimate project. Furthermore, scam pool creators tend to have a higher initial pool supply, as their intention is to generate hype and maximize liquidity removal when the token price reaches its peak. Consequently, scam token creators are less inclined to lock their initial LP tokens.

Table 4.6 also presents the estimated results from the following logistic regression model that examines the relationship between individual pool characteristics and the probability of being a scam pool:

$$Pr(Scam = 1) = \Phi(\beta_0 + \sum_{n=1}^{12} \beta_n \cdot X_n + \varepsilon), \tag{4.1}$$

where Φ is the cumulative distribution function (CDF) of the logistic distribution. The dependent variable represents the probability of a given pool being classified as a scam and the independent variables (X_n) include the transaction, creator, and user characteristics of the pools. We exclude the individual event and user counts due to their high correlation with the transaction count and the price run-up/collapse time due to its high correlation with the existence time.[67] The logit model allows for the possibility of a non-linear relationship between the probability of a pool being a scam and the explanatory variables.

The signs of all coefficients from the logit model align with the results of the univariate tests. The reported marginal effects indicate the percentage change in the probability of a pool being a scam for a 1% change in the value of the independent variable. These effects provide insight into the relative significance of the different characteristics.[68]

According to the table, the concentration measure for both add liquidity and remove liquidity, imitation, and lock LP variables are the primary factors influencing the probability of a scam. Specifically, a 1% increase in the add liquidity concentration decreases the probability of a scam by 202%, and a 1% increase in the remove liquidity concentration increases the probability of a scam by 97% of what the probability of a scam would otherwise be. The concentration measure is relatively important because, in most scam pools, the creator adds liquidity at the beginning and removes it at the end, whereas non-scam pool users continuously add or remove liquidity during the pool's existence.

[67] The correlation matrix for all variables is reported in Table A4.4 of the Appendix.

[68] The marginal effects of the characteristics follow from the derivative of $Pr(Scam = 1)$ with respect to each of the characteristics, X_n as

$$\frac{\partial Pr(Scam = 1)}{\partial X_n} = \frac{\beta_n \cdot exp(-\beta_0 - \sum_{n=1}^{12} \beta_n \cdot X_n)}{(1 + exp(-\beta_0 - \sum_{n=1}^{12} \beta_n \cdot X_n))^2}$$

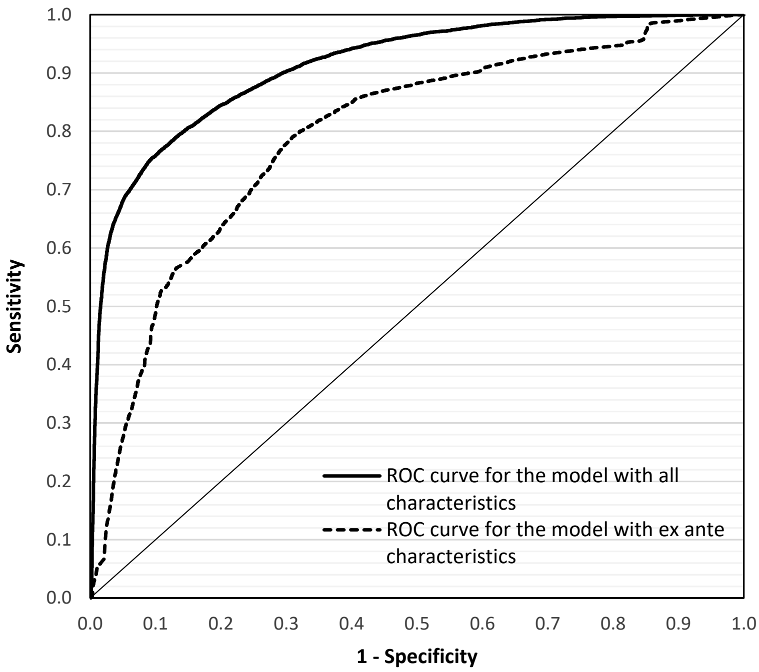
reported as a percentage of the estimated probability of scam. Marginal effects are calculated for each observation and then averaged over the entire sample.

Moreover, creators who lock their LP tokens are 76% less likely and creators who imitate other token names are 18% more likely to scam investors.

Figure 4.7 plots the Receiver Operating Characteristics (ROC) curve of the model. The ROC curve measures the performance of the model, describing the trade-off between the proportion of true positives (sensitivity) and the proportion of false positives (one minus the specificity) at all classification thresholds (e.g., Stein 2005). In our model, the ROC curve describes the proportion of non-scams (resp. scams) that will be classified as scams (resp. non-scams). The Area Under the Curve (AUC) represents the probability of correct prediction and is a robust measure of the performance of a classifier.

The figure shows that the ROC curve for this model is significantly above the diagonal line, representing a classifier only as good as a chance. Consistently, the AUC reported in the last row of Table 4.6 (0.912) is well above 0.5, indicating that the model is 91% accurate in classifying scams.

Figure 4.7. Receiver Operating Characteristics (ROC) curves
This figure plots the Receiver Operating Characteristics (ROC) curves for two logistic regression models. The solid line shows the ROC curve for the model with all characteristics (the estimates of this model are reported in Table 4.6). The area under the curve (AUC) for this model is 0.912. The dashed line shows the ROC curve for the model with only ex ante characteristics (the estimates of this model are reported in Table 4.7). The area under the curve (AUC) for this model is 0.793.



4.6.3 Measuring the probability of a scam

Our analysis thus far suggests that several pool characteristics can help distinguish scam tokens from others. It is important to note that these characteristics primarily identify scams after they occur. This is also how our soft rug pull detection algorithm works based on the price, profit, and activity patterns of liquidity pools. However, it is crucial to implement proactive regulations or guidelines that can anticipate and prevent scams before they occur, rather than relying solely on post-event detection methods.

In this sub-section, we modify the logit model in Eq. (4.1) to focus on ex ante characteristics that can detect scams before they occur. The scam index is derived from specific characteristics included in the model: initial pool supply (X_1), LP lock (X_2), creator token count (X_3), creator pool count (X_4), time gap (X_5), and imitation (X_6). These variables pertain to the token or pool creator and are available immediately after the initial liquidity is added to the pool. All of these variables characterize the token/pool creator and are available immediately after the initial liquidity is added to the pool. We include all 33,783 scam pools in the model as it is unclear ex ante which scams will be economically meaningful.

Table 4.7 presents the coefficient estimates and marginal effects of the ex ante variables in predicting the probablity of a scam. Consistent with the findings from the previous univariate analysis and logit model, an increase in the initial pool supply, creator pool count, and imitation variables increase the probability of a scam, while higher LP lock, creator token count, and time gap decrease it.

In terms of relative importance of variables, LP lock is the primary indicator of the probability of a scam, followed by imitation. Specifically, locking LP tokens reduces the probability of a scam by around 55% and imitating another token increases the probability of a scam by around 5%. The AUC reported in the last row of Table 4.7 (0.793) is well above 0.5, indicating that the model based on ex ante characteristics is 79% accurate in classifying successful scams. Figure 4.7 plots the ROC curve for this model.

Rearranging Eq. (4.1) and inserting the coefficient estimates from Table 4.7, we can calculate a scam index score based on the probability of a scam as

$$I_{scam} = \left(1 + e^{\left(0.042 - 0.021 \cdot X_1 + 1.318 \cdot X_2 + 0.005 \cdot X_3 - 0.007 \cdot X_4 + 0.003 \cdot X_5 - 0.114 \cdot X_6\right)}\right)^{-1}. \tag{4.2}$$

The score is based on six ex ante characteristics of token/pool creators. These characteristics are readily available from the blockchain data and can be used to calculate a score for tokens in other blockchains.

Table 4.7. The ex ante determinants of scam pools: logit regression with all scams This table reports the estimates from the logistic regression of scams and non-scams using the following model:

$$Pr(Scam = 1) = \Phi(\beta_0 + \sum_{n=1}^{12} \beta_n \cdot X_n + \varepsilon)$$

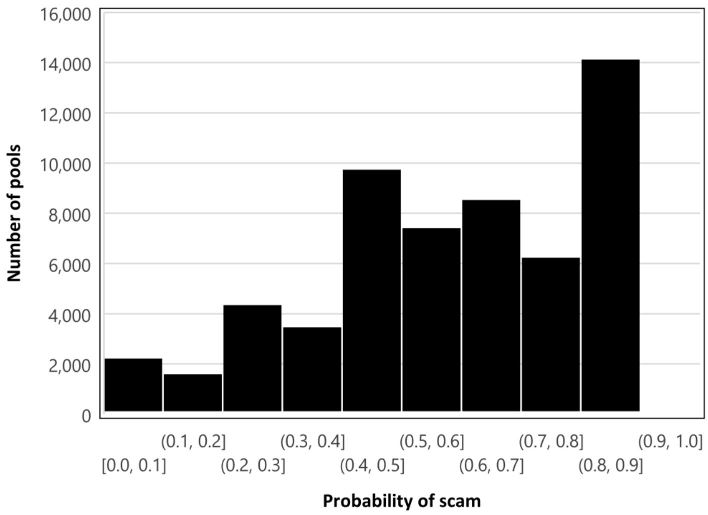
where Φ is the cumulative distribution function (CDF) of the logistic distribution. The dependent variable is Scam which takes the value of 1 if a given pool is a scam pool and 0 otherwise and independent variables are initial pool supply (X_1), LP lock (X_2), creator token count (X_3), creator pool count (X_4), time gap (X_5), and imitation (X_6). Table 4.1 provides the definitions of all characteristics. Numbers not in parentheses are the coefficient estimates. Numbers in parentheses are the marginal effects (partial derivatives of the corresponding probability with respect to each of the variables, reported as a percentage of the estimated corresponding probability). Pseudo R^2 is McFadden's likelihood ratio index (one minus the ratio of the log-likelihood with all predictors and the log-likelihood with intercepts only). AUC is the area under the ROC (receiver operating characteristic) curve for the model.

Characteristics	Estimate	p-Value
Intercept	-0.048	0.0046
Initial pool supply (%)	0.021 (0.870)	< 0.0001
LP lock	-1.322 (-55.274)	< 0.0001
Creator token count	-0.005 (-0.223)	< 0.0001
Creator pool count	0.007 (0.292)	< 0.0001
Time gap (days)	-0.003 (-0.124)	< 0.0001
Imitation	0.121 (5.057)	< 0.0001
N obs.	57,517	
Pseudo R^2	18.41%	
Area under the ROC curve (AUC)	0.793	

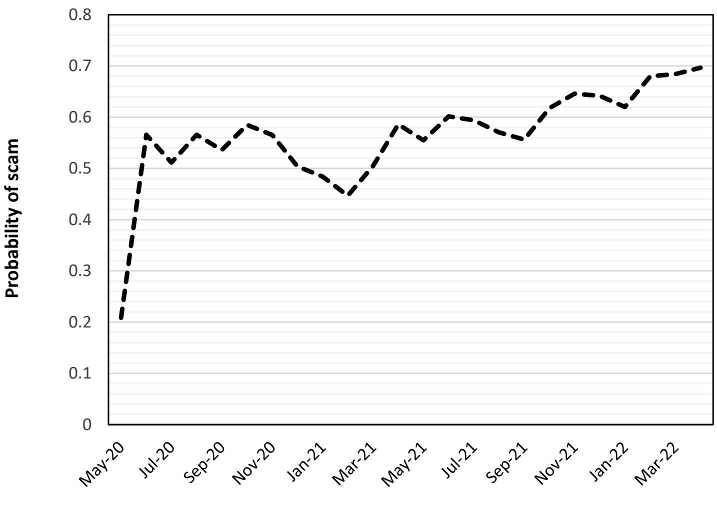
Figure 4.8 presents the time series and cross-sectional distribution of the scam index. Of the 58,439 pools that we analyze, 36,290 pools have a score greater than 0.5, and 14,208 pools have a score greater than 0.8. Regulators can choose a threshold value to classify tokens as scams, depending on their desired trade-off between type I and type II errors. Because higher index values have a lower probability of misclassification, regulators with limited resources can first investigate tokens with the highest scores and then continue investigating lower scores as far as their resources allow.

Figure 4.8. The distribution of scam index
This figure shows the empirical distribution of the probability of scam (scam index). Panel A illustrates the cross-sectional distribution of the scam index with all pools and periods pooled. Panel B illustrates the time series distribution of the average monthly scam index between May 2020 and April 2022.

Panel A. Cross sectional distribution of the scam index



Panel B. Time series distribution of the scam index



Panel B plots the average monthly scam index score for the pools created during each month between May 2020 and April 2022. The average score spikes early in the sample due to the emergence of DEXs, notably Uniswap V2 launched in May 2020. The average scam index continues to rise with increasing cryptocurrency prices, starting around 0.45 in early 2021 and reaching approximately 0.70 by the end of the sample period. The time series pattern of the scam index can provide guidance to practitioners and regulators about optimal periods for heightened surveillance and the suitable threshold for classifying pools as scams. In validation exercises, we use a threshold of 0.5 and assess the scam index's consistency with the soft rug detection algorithm. We also conduct additional validation using external datasets.

4.7 Validation and robustness tests

4.7.1 Validation

We perform validation in two ways. First, we obtain token reputations for 50,422 ERC-20 tokens from the Etherscan page of each token. Etherscan classifies tokens into six reputation categories: unknown, neutral, ok, suspicious, unsafe, and spam. These categories are based on evaluations by Etherscan and community reports. The unknown category constitutes 95% of the tokens and serves as a default category, indicating that only basic information such as a website and logo is provided by the founders. There are 1,732 tokens flagged as neutral or ok, meaning sufficient information about the team and project is available. The suspicious, unsafe, and spam categories suggest reports of scams, phishing, fraud, spam, or misrepresentation of token-related information. These categories collectively include 790 tokens.

The 1,732 tokens in the neutral and ok categories have a total of 5,094 pools. Among these pools, we classify 4,522 pools as non-scam and only 134 pools as economically meaningful soft rug pull. This indicates that only 2.6% of the 5,094 pools in the neutral/ok category are classified as economically meaningful soft rug pulls, indicating a low false positive rate in our analysis. In the suspicious, unsafe, and spam categories, there are 790 tokens with 822 pools. Of these pools, we classify 75% (613 pools) as scam pool with 581 being classified as soft rug pulls and 32 as hard rug pulls.

We also calculate the average scam index for these categories and use a threshold of 0.5 to validate our scam index. We find that 94% of the pools in the neutral and ok categories have a scam index score less than 0.5 with an average score of 0.25 and 86% of the pools in the suspicious, unsafe, and spam categories have a score greater than 0.5 with an average score of 0.73. These results indicate that our scam sample, constructed based on smart contract audits and the soft rug pull detection algorithm, aligns with the reported tokens, and the scam index can ex ante accurately detect these scams.

Second, we obtain 256 tokens that are reported as hacks, scams, rug pulls, or contracts with malicious or questionable functionality in the Ethereum blockchain from the REKT database of Defiyield.[69] Out of the 273 pools associated with these tokens, we identify 230 (84%) as scam pools, consisting of 171 hard rug pulls and 59 soft rug pulls. In addition, 179 pools have a scam index greater than 0.5, with an average score of 0.74. Table 4.8 summarizes the results of these validation tests.

Table 4.8. Validation tests
This table reports the results of the validation tests based on Etherscan token reputations and Defiyield REKT database. Panel A is based on the smart contract audits from tokensniffer and our soft rug pull detection algorithm. Panel B is based on the probability of scam index (I_{scam}). The numbers not in parentheses are the number of pools and in parentheses are the percentage of each category. The first and second columns in both panels are the name of categories and the number of pools in each category reported in Etherscan and Defiyield. In Panel A, the third and fourth columns are the soft rug pulls with at least \$10,000 and less than \$10,000 creator profit, respectively. The fifth column is hard rug pulls based on the tokensniffer smart contract audits. The last two columns in Panel A report all scams (soft and hard rug pulls) and non-scams. In Panel B, the third and fourth columns show the number and percentage of pools with $I_{scam} > 0.5$ and $I_{scam} < 0.5$, respectively.

Panel A: Smart contract audits and soft rug detection algorithm

Category	Number of pools	Soft rugs with at least \$10,000 profit	Soft rugs with less than \$10,000 profit	Hard rugs	All scams (soft and hard rugs)	Non-scams
Etherscan: Neutral/ok	5,094	134 (2.6%)	408 (8.0%)	0	542 (10.6%)	4,552 (89.7%)
Etherscan: Suspicious/unsafe/spam	822	467 (56.8%)	114 (13.9%)	32 (3.9%)	613 (74.6%)	209 (25.4%)
DefiYield: REKT	273	12 (4.4%)	47 (17.2%)	171 (62.6%)	230 (84.2%)	43 (15.8%)

Panel B: Probability of scam index (I_{scam})

Category	Number of pools	$I_{scam} > 0.5$	$I_{scam} < 0.5$
Etherscan: Neutral/ok	5,094	297 (6%)	4,797 (94%)
Etherscan: Suspicious/unsafe/spam	822	704 (86%)	118 (14%)
DefiYield: REKT	273	179 (66%)	94 (34%)

[69] The Defiyield REKT database lists 986 project names reported by the community in Ethereum. Not all of these projects are ERC-20 tokens. By matching 986 project names to our list of tokens, we identify 257 ERC-20 tokens that are reported in Defiyield.

4.7.2 Robustness

We now assess the robustness of our soft rug pull detection algorithm.

4.7.2.1 Characteristics analysis with all scam pools

Our baseline characteristics analysis focuses on 8,642 economically meaningful soft rug pulls with a minimum creator profit of $10,000. However, we identify a total of 31,931 soft rug pulls. The distinguishing characteristics of rug pulls persist when considering all 31,931 scam pools, as shown in Tables A4.2 and A4.3 in the Appendix.[70]

4.7.2.2 Collaborator addresses

Throughout the chapter, we focus on the amount of stolen funds by token/pool creators' wallet addresses. However, creators may use other addresses to conceal their tracks. To identify these addresses, we find the ten most frequent addresses that have sent or received tokens from the token/pool creator's address. We exclude addresses that are less likely to be collaborator addresses, such as those with more than 100 transactions, interactions with more than ten tokens, and involvement with multiple creators. Using this approach, we find that 5,667 creators have one collaborator, and 2,942 creators have multiple collaborator addresses. By considering the combined profit of the creator and collaborator addresses, we detect 9,090 economically significant soft rug pulls with at least $10,000 in creator profit.

4.7.2.3 Magnitude of the price run-up and collapse

Next, we analyze the extent of price run-up and collapse in soft rug pulls. Of the 8,642 economically meaningful soft rug pulls, we observe that 7,637 pools experience a price run-up of at least 50% from the initial price. Similarly, 8,606 of the economically meaningful soft rug pulls have a price collapse of at least 50% from the initial price. These results indicate that tightening the conditions for price run-up and price collapse would not significantly alter the number of economically meaningful soft rug pulls.

[70] In unreported tests, we analyze 22,160 pools that generate at least $1,000 profit for their creators, and the results of the characteristics analysis remain consistent.

4.7.2.4 Activity thresholds

Lastly, in our baseline analysis, we classify pools as inactive if there have been no events during the last 10% of the time period between the first event of the pool and the end date of our sample (April 8, 2022). When we increase this threshold to 20% and 30%, the number of economically meaningful rug pulls remains close to our baseline estimates. When we label a pool as inactive if there have been no events in the last three months of our sample, we obtain similar findings. The results of these robustness tests are summarized in Table 4.9.

Table 4.9. Robustness tests
This table reports the results of the robustness tests of the soft rug pull detection algorithm, where we change one condition at a time. The default criteria shows number of soft rug pulls that generates at least $10,000 creator profit, satisfies the price run up and collapse condition and is inactive (no event in the last 10% of the time difference between first event of the pool and 8 April 2022), reported in the first row. The first column shows the criteria, and the second column shows the number of pools satisfying that criteria.

Criteria	Number of soft rug pulls
Default: Economically meaningful soft rug pulls with at least $10,000 creator profit	8,642
Economically meaningful soft rug pulls with at least $10,000 combined profit of creator and collaborators	9,090
Economically meaningful soft rug pulls with at least 50% price run up	7,637
Economically meaningful soft rug pulls with at least 50% price collapse	8,606
Economically meaningful soft rug pulls that are inactive (20% threshold)	8,535
Economically meaningful soft rug pulls that are inactive (30% threshold)	8,436
Economically meaningful soft rug pulls that are inactive (no event in the last 3 months)	7,049

4.8 Conclusion

Most growth and innovation in the blockchain space now takes place on multi-purpose blockchains such as Ethereum, Solana, Avalanche, and within the decentralized finance (DeFi) ecosystem. The growth in DeFi has expanded the ways in which blockchain technology can be used. It is now straightforward to create a new token on the blockchain and set up a liquidity pool for the token on DEXs. However, this has also opened up new avenues for the misuse of the technology.

In Chapter 4, we focus on the dark side of the pools in DEXs. Using smart contract audits and a simple soft rug pull detection algorithm, we quantify the prevalence of scam pools in major DEXs. We find that 44% of the 78,178 ERC-20 liquidity pools in major DEXs are rug pulls. 3% of the pools are hard rug pulls, where tokens have malicious code in their smart contracts, and 41% of the pools are soft rug pulls, where tokens have no malicious code, but the token/pool creator gradually steals liquidity. 11% of the liquidity pools result in at least $10,000 in stolen funds. We estimate that rug pulls have stolen around $1.5 billion worth of tokens from investors between May 2020 and April 2022. By quantifying this area of concern, we shed light on the scale of scamming activities using DEXs.

We also characterize the scam pools. The scam pools tend to have a lower transaction count, attract fewer and less experienced users, and have a shorter lifespan. Add liquidity events for the scam pools are concentrated at the beginning of the pool, while remove liquidity events are concentrated at the end. In contrast, for non-scam pools, these events are distributed more evenly throughout the pool's existence. The wallet addresses used to create scam pools tend to interact with fewer tokens and create more liquidity pools before creating the scam pool. They also tend not to lock their liquidity provider tokens (LP tokens) and deposit a greater proportion of the token supply into liquidity pools. The time gap between creating the token and the pool is substantially shorter for scam tokens. Based on these characteristics, we construct an index of the probability of scams that can identify scams before they occur.

Our results have several implications. First, the distinguishing characteristics of scam pools can help investors become aware of red flags when investing in tokens. Second, our analysis helps reduce uncertainty surrounding the scale of scamming activities in DEXs, enabling more informed decisions by policymakers assessing the costs and benefits of blockchain technology. Third, our soft rug pull detection algorithm can be used by law enforcement authorities in prosecuting scammers, and the scam index that we construct can be used in surveillance activities to prevent scams before they occur. Thus, the techniques developed in this chapter have applications in cryptocurrency surveillance. Lastly, our analysis can assist blockchain and DEX developers in designing effective controls to reduce the magnitude of scamming activities.

Future research could improve the generalizability and robustness of our findings in several ways. First, our analysis is limited to ERC-20 tokens on the Ethereum blockchain between May 2020 and April 2022. This excludes tokens on other blockchains such as Binance Smart Chain or Solana, which may have different dynamics and scam prevalence. Second, the detection algorithm is based on specific heuristic patterns and smart contract audits that may not be universally applicable. The algorithm's performance may differ in different types of blockchains. Third, applying well-known classification techniques (e.g., random forests) starting from a sample of known scams and non-scams could improve scam detection accuracy. Lastly, integrating more comprehensive data sources, such as social media sentiment analysis, transaction network analysis, and additional on-chain metrics, could provide deeper insights into scam characteristics and improve the accuracy of predictive models.

Appendix 4.1.

Table A4.1
Well-known locker addresses
This table shows the well-known locker addresses that we use to check whether the initial LP tokens received by the token/pool creator are locked. The first column shows the locker address (dead addresses are also included), and the second column shows the locker name.

Locker address	**Locker name**
0x663a5c229c09b049e36dcc11a9b0d4a8eb9db214	Unicrypt
0xe2fe530c047f2d85298b07d9333c05737f1435fb	TrustSwap: Team Finance Lock
0x17e00383a843a9922bca3b280c0ade9f8ba48449	Unicrypt
0xc2edad668740f1aa35e4d8f227fb8e17dca888cd	SushiSwap: MasterChef LP Staking Pool
0xc77aab3c6d7dab46248f3cc3033c856171878bd5	DireWolf: Burn Wallet #2
0xef0881ec094552b2e128cf945ef17a6752b4ec5d	SushiSwap: MasterChef V2
0xdbf72370021babafbceb05ab10f99ad275c6220a	Rainicorn: Vesting
0x000000000000000000000000000000000000dead	Dead address
0x0000000000000000000000000000000000000000	Dead address
0x0000000000000000000000000000000000000001	Dead address
0xdead000000000000000042069420694206942069	Dead address

Table A4.2
Characteristics of soft and hard rug pulls – All scams

This table reports the descriptive statistics (mean, median, standard deviation) of the variables for all soft rug pulls (31,931 liquidity pools) and hard rug pulls (1,852 liquidity pools) and the differences in their mean characteristics. The last two columns in the table report the difference in mean characteristics of soft and hard rug pulls. The significance of the difference in means is computed with t-statistics. The standard errors used to compute the t-statistics are clustered by the month of the year that the pools are created. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. For a detailed definition of variables refer to Table 4.1. The units of the variable are in parentheses following the variable name.

	Soft rug pull (1)			**Hard rug pull (2)**			**Mean difference**	
Characteristics	Mean	Median	Std. dev.	Mean	Median	Std. dev.	(1)-(2)	t-stat
Transaction count	122.54	24.00	874.17	79.46	32.00	183.10	43.08	2.71***
Add-liquidity count	4.29	1.00	34.79	1.31	1.00	5.63	2.98	4.53***
Swap-buy count	73.22	18.00	528.31	73.60	28.00	145.64	-0.38	-0.03
Swap-sell count	51.51	4.00	430.77	3.37	0.00	54.02	48.14	12.15***
Remove-liquidity count	2.32	1.00	16.56	1.16	1.00	2.20	1.17	3.25***
Add-liquidity concentration	0.02	0.00	0.09	0.02	0.00	0.09	0.003	0.40
Remove-liquidity concentration	0.86	1.00	0.27	0.86	1.00	0.28	0.01	0.50
Existence time (days)	21.90	1.00	64.70	10.87	2.00	34.12	11.03	3.25***
Price run up time (days)	4.56	0.03	28.63	4.98	0.36	26.44	-0.42	-0.59
Price collapse time (days)	11.07	0.01	43.72	2.27	0.05	12.60	8.80	5.74***
Creator profit ($\cdot 10^3$ USD)	40.84	3.08	505.84	148.10	40.98	345.94	-107.26	-4.43***
Creator token count	14.61	3.00	45.41	5.06	2.00	12.62	9.55	7.66***
Creator pool count	4.45	1.00	14.50	1.97	1.00	4.80	2.49	4.76***
Imitation	0.26	0.00	0.44	0.42	0.00	0.49	-0.15	-9.99***
Time gap (days)	13.73	0.01	98.85	0.90	0.01	8.45	12.83	5.57***
Initial pool supply (%)	59.06	66.67	38.58	40.53	35.50	37.80	18.53	4.83***
LP lock	0.18	0.00	0.39	0.13	0.00	0.34	0.05	1.22
User count	47.66	15.00	263.34	64.89	25.00	121.58	-17.23	-1.77*
Users' experience	177.71	118.13	207.65	91.92	56.91	134.95	85.78	5.35***
N. Obs	31,931			1,852				

Table A4.3

Characteristics of rug pulls and non-rug pulls – All scams

This table reports the descriptive statistics (mean, median, standard deviation) of the variables for all scam pools (31,931 soft rug pulls and 1,852 hard rug pulls) and 24,631 non-scam pools and the differences in their mean characteristics. The significance of the difference in means is computed with t-statistics. The standard errors used to compute the t-statistics are clustered by the month of the year that the pools are created. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The last two columns show the marginal effects (%) and p values of the logit model where the dependent variable is the probability of scam, and the independent variables are pool characteristics. The marginal effects are the partial derivatives of the corresponding probability with respect to each of the variables (reported as a percentage of the estimated corresponding probability). For a detailed definition of variables refer to Table 4.1. The units of the variable are in parentheses following the variable name.

	Scam (1)			**Non-scam (2)**			**Mean difference**		**Logit model**	
Characteristics	Mean	Median	Std. dev.	Mean	Median	Std. dev.	(1)-(2)	t-stat	Marginal effects (%)	*p*-Value
Transaction count	120.17	25.00	851.00	3,788.93	150.00	41,914.07	-3,668.76	-4.68***	-0.009	<0.0001
Add-liquidity count	4.13	1.00	33.86	91.49	3.00	779.28	-87.37	-5.46***		
Swap-buy count	73.24	18.00	514.76	2,034.54	76.00	22,012.41	-1,961.31	-4.82***		
Swap-sell count	48.87	4.00	419.13	1,662.46	69.00	19,583.49	-1,613.59	-4.57***		
Remove-liquidity count	2.26	1.00	16.11	53.06	2.00	531.10	-50.81	-4.42***		
Add-liquidity concentration	0.02	0.00	0.09	0.09	0.01	0.15	-0.07	-11.22***	-99.150	<0.0001
Remove-liquidity concentration	0.86	1.00	0.27	0.52	0.51	0.34	0.35	17.24***	53.023	<0.0001
Existence time (days)	21.29	2.00	63.45	188.98	135.00	187.52	-167.68	-8.06***	-0.276	<0.0001
Price run up time (days)	4.59	0.04	28.52	41.01	0.37	94.03	-36.42	-7.91***		
Price collapse time (days)	10.58	0.01	42.66	98.03	32.70	133.95	-87.45	-7.79***		
Creator profit ($\cdot 10^3$ USD)	46.77	3.32	498.94	-48.87	-1.12	1,517.73	95.64	5.99***	0.072	<0.0001
Creator token count	14.08	2.00	44.30	31.25	5.00	82.20	-17.17	-8.61***	-0.086	<0.0001
Creator pool count	4.32	1.00	14.15	5.01	2.00	11.96	-0.70	-1.63	0.135	0.0041
Imitation	0.27	0.00	0.44	0.25	0.00	0.43	0.03	2.29**	-0.794	0.4546
Time gap (days)	13.03	0.01	96.17	119.42	1.35	281.54	-106.39	-7.72***	-0.056	<0.0001
Initial pool supply (%)	58.04	60.00	38.77	25.97	0.86	37.04	32.08	10.89***	0.359	<0.0001
LP lock	0.18	0.00	0.38	0.28	0.00	0.45	-0.10	-4.81***	-41.574	<0.0001
User count	48.60	16.00	257.62	904.43	60.00	7,441.30	-855.82	-5.38***		
Users' experience	173.01	113.13	205.26	210.54	126.64	294.36	-37.54	-3.79***	-0.050	<0.0001
N. obs	33,783			24,656						
Area under the ROC curve (AUC)									0.882	

Table A4.4
The correlation matrix of variables

The table presents the correlation matrix of variables, where V1 is transaction count, V2 is add-liquidity, V3 is swap-buy count, V4 is swap-sell count, V5 is remove-liquidity count, V6 is add-liquidity concentration, V7 is remove-liquidity concentration, V8 is existence time, V9 is price run up time, V10 is price collapse time, V11 is creator profit, V12 is creator token count, V13 is creator pool count, V14 is imitation, V15 is time gap, V16 is initial pool supply, V17 is LP lock, V18 is user count, V19 is user experience. Table 4.1 presents the definitions of all variables.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19
V1	1																		
V2	0.65	1																	
V3	0.99	0.61	1																
V4	0.99	0.66	0.96	1															
V5	0.70	0.93	0.65	0.70	1														
V6	0.07	0.08	0.08	0.07	0.07	1													
V7	-0.07	-0.10	-0.07	-0.07	-0.09	-0.21	1												
V8	0.14	0.17	0.14	0.13	0.15	0.24	-0.61	1											
V9	0.10	0.08	0.10	0.09	0.08	0.23	-0.29	0.53	1										
V10	0.08	0.11	0.08	0.07	0.09	0.11	-0.48	0.75	0.09	1									
V11	0.00	-0.01	0.00	0.00	-0.01	-0.02	0.01	-0.03	-0.02	-0.03	1								
V12	0.02	0.02	0.02	0.02	0.03	0.12	-0.13	0.16	0.11	0.08	0.00	1							
V13	0.00	0.00	0.00	0.00	0.01	0.04	-0.02	0.04	0.04	0.00	0.00	0.52	1						
V14	-0.01	-0.02	-0.01	-0.01	-0.02	0.00	0.03	-0.04	-0.02	-0.03	0.00	0.00	0.02	1					
V15	0.05	0.04	0.05	0.05	0.05	0.21	-0.17	0.35	0.34	0.17	-0.01	0.14	0.07	-0.04	1				
V16	-0.05	-0.07	-0.05	-0.05	-0.07	-0.26	0.38	-0.44	-0.25	-0.31	0.02	-0.15	-0.02	-0.05	-0.30	1			
V17	0.01	0.01	0.01	0.01	-0.01	-0.11	-0.07	0.01	-0.09	0.09	-0.02	-0.12	-0.12	-0.08	-0.14	0.18	1		
V18	0.90	0.59	0.93	0.85	0.63	0.11	-0.11	0.19	0.13	0.11	-0.01	0.03	0.00	-0.02	0.06	-0.07	0.02	1	
V19	-0.03	-0.04	-0.04	-0.03	-0.04	0.00	0.02	-0.04	0.03	-0.05	0.00	0.05	0.01	-0.02	0.11	-0.02	-0.04	-0.05	1

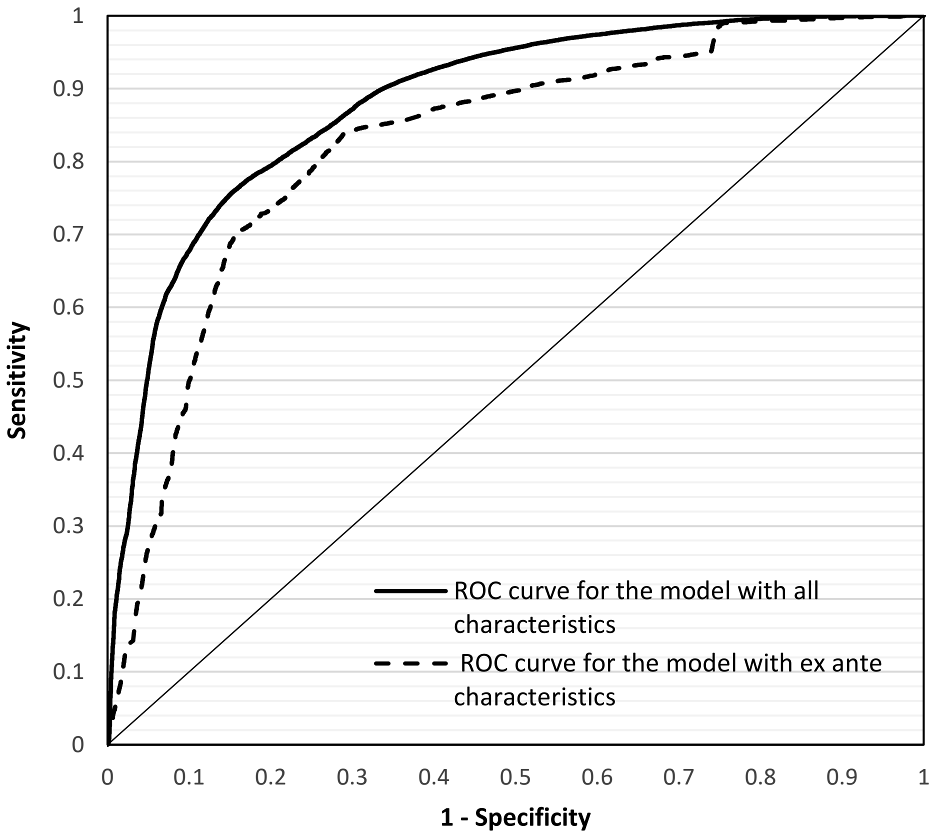
Table A4.5
The ex ante determinants of scam pools: logit regression with economically meaningful scams
This table reports the estimates from the logistic regression of scams and non-scams using the following model:

$$Pr(Scam = 1) = \Phi(\beta_0 + \sum_{n=1}^{12} \beta_n \cdot X_n + \varepsilon)$$

where Φ is the cumulative distribution function (CDF) of the logistic distribution. The dependent variable is Scam which takes the value of 1 if a given pool is a scam pool and 0 otherwise and independent variables are initial pool supply (X_1), LP lock (X_2), creator token count (X_3), creator pool count (X_4), time gap (X_5), and imitation (X_6). Table 4.1 provides the definitions of all characteristics. Numbers not in parentheses are the coefficient estimates. Numbers in parentheses are the marginal effects (partial derivatives of the corresponding probability with respect to each of the variables, reported as a percentage of the estimated corresponding probability). Pseudo R^2 is McFadden's likelihood ratio index (one minus the ratio of the log-likelihood with all predictors and the log-likelihood with intercepts only). AUC is the area under the ROC (receiver operating characteristic) curve for the model.

Characteristics	Estimate	p-Value
Intercept	-0.861	< 0.0001
Initial pool supply (%)	0.015	< 0.0001
	(1.064)	
LP lock	-1.673	< 0.0001
	(-116.854)	
Creator token count	-0.012	< 0.0001
	(-0.845)	
Creator pool count	0.030	< 0.0001
	(2.084)	
Time gap (days)	-0.015	< 0.0001
	(-1.061)	
Imitation	0.461	< 0.0001
	(32.240)	
N obs.	34,536	
Pseudo R^2	21.20%	
Area under the ROC curve (AUC)	0.824	

Figure A4.1
Receiver Operating Characteristics (ROC) curves
This figure plots the Receiver Operating Characteristics (ROC) curves for two logistic regression models. The solid line shows the ROC curve for the model with all characteristics (the estimates of this model are reported in Table A4.3). The area under the curve (AUC) for this model is 0.882. The dashed line shows the ROC curve for the model with only ex ante characteristics (the estimates of this model are reported in Table A4.5). The area under the curve (AUC) for this model is 0.824.



Chapter 5

Once bitten twice shy: Learning about scams

5.1 Introduction

Households make serious investment mistakes, such as nonparticipation in financial markets or underdiversification (e.g., Campbell 2006, Gomes et al. 2021). Among these investment mistakes, falling victim to investment scams is perhaps the most dramatic because they involve significant financial damage to households. For example, households lost more than \$20 billion in known investment scams in 2022 (e.g., Knüpfer et al. 2023), and this number has been on the rise with increased digitisation.

Yet, little evidence exists on the impact of scams on victims. Do victims learn from their mistakes or continue reinvesting in scams? How does their investment approach and performance change after falling victim to a scam? Are there changes in their risk tolerance? Do they adjust their portfolios post-scam? What characterizes different learning rates among victims? These are important questions in household finance, but obtaining reliable answers is extremely challenging because they require high-quality data on scams and victims. Many scam victims often underreport due to ex-post rationalization, selective recall, or do not even accept that they have been scammed (e.g., NASD 2006).

We address these challenges by identifying cryptocurrency scams and their victims using blockchain data. The blockchain environment provides a unique opportunity to identify scams and analyze the trading behavior of scam victims because of the full transparency of portfolio holdings and buy/sell decisions.

In Chapter 4, we identify scam tokens on major decentralized exchanges in two ways. First, we compile a list of tokens with malicious code in their smart contracts (2,405 liquidity pools). Second, we use a scam detection algorithm that analyzes the price, creator's profit, and activity patterns of liquidity pools. The price pattern examines whether a token has experienced a price run-up followed by a collapse. The profit pattern analyzes whether the token or pool creator has gained profits in the end. The activity pattern checks recent transaction activity associated with liquidity pools and identifies inactive pools. Using these two methods, we identify a total of 34,336 cryptocurrency scams and 218,799 victims who have lost money in these scams.

In this chapter, we examine the investment characteristics of victims. First, we compare the trading and portfolio characteristics of victims to those of non-victims. On average, victims tend to be more active in the market — they interact with more liquidity pools and tokens and trade more. In terms of portfolio characteristics, the average daily portfolio return of victims is higher than that of non-victims, but the standard deviation of daily portfolio returns for victims also tends to be higher, suggesting a higher risk tolerance. Victims also allocate a smaller proportion of their portfolios to cash, further indicating a higher risk tolerance.

Second, we test the impact of experiencing a scam in the recent past on subsequent scam investments using a probit model. We find that victims tend to invest less in scams after experiencing a scam, suggesting that they become more cautious following a scam. On average, experiencing a scam in the last 30 days decreases the probability of investing in a scam again by approximately 4%, with similar results holding for other periods. Furthermore, investors who have invested more heavily or lost more in scams are less likely to invest in scams again. Specifically, a 1% increase in the scam investment ratio (the ratio of scam investment to total investment volume) decreases the probability of investing in a scam again by around 0.30%.

Third, we correlate experiencing a scam with trading performance to determine whether and how investors learn from their mistakes. To do this, following Seru et al. (2010), we calculate the returns earned by each investment in the 30 trading days following the investment. We find that experiencing a scam in the recent past, on average, decreases subsequent non-scam investment returns by around 14%. In addition, the scam investment ratio also negatively impacts subsequent investment returns. On average, a one standard deviation shock to the scam investment ratio is associated with a 4% decrease in subsequent non-scam investment returns. While investors become more cautious after a scam, as reflected in the decrease in the likelihood of further scam investment, the returns on their investments also decrease after experiencing a scam, and this effect is more pronounced when the scam investment ratio is high.

Next, we compare the trading and portfolio characteristics of victims before and after experiencing a scam. In both univariate tests and panel regressions, we find that victims trade more actively after experiencing a scam. They interact with more liquidity pools and tokens and trade more frequently.

Investigating the impact of scams on victims' portfolio characteristics is more challenging because portfolio characteristics may change due to scams affecting how victims invest (investment effect), or alternatively, victims may consolidate their assets into a single wallet address in response to a potential theft risk (consolidation effect). To disentangle the investment effects from the consolidation effects, we create a daily portfolio dataset from raw blockchain data, including only trade data by excluding transfers and combining trade and transfer data.

We find that victims increase their market exposure by increasing their portfolio values and the number of tokens held in their portfolios, attributable to both investment and consolidation effects. However, their daily portfolio return decreases. The impact of scams on the daily portfolio standard deviation varies over time. In the short term (resp. long term) after the scam, the standard deviation increases (resp. decreases), suggesting an increase (resp. decrease) in risk tolerance. Nevertheless, the Sharpe ratio of the portfolio decreases after experiencing a scam, indicating that investors never regain their pre-scam level of risk-adjusted performance.

We also investigate the characteristics associated with varying learning rates among victims. We find that less experienced investors with shorter trading histories tend to invest in multiple scam pools, suggesting that they tend to learn less from scams. Those who interact with a larger number of liquidity pools, have higher trading volumes, and achieve lower returns on their (non-scam) investments also tend to invest in multiple scam pools. In terms of portfolio characteristics, victims with higher portfolio values and a greater number of tokens in their portfolios are more likely to invest in multiple scam pools. Lastly, victims with lower portfolio returns, higher standard deviations of returns, and lower cash allocation in their portfolios also tend to invest multiple in scam pools. These findings suggest that less sophisticated investors that often take on excessive risk without adequate compensation tend to invest in multiple scam pools and learn less from their scam experience.

5.2 Related literature

Our study contributes to the growing literature on the economics of fraud. There is hardly any activity in finance untouched by fraud and misconduct, including market manipulation (e.g., Aggarwal and Wu 2006, Comerton-Forde and Putnins 2014), insider trading (e.g., Meulbroek 1992, Bhattacharya and Daouk 2002), corporate frauds (e.g., Dyck, Morse and Zingales 2010), misconduct by financial advisors (e.g., Egan, Matvos and Seru 2019, Dimmock, Gerken and Graham 2018), financial intermediaries (e.g., Piskorski, Seru and Witkin 2015, Griffin and Maturana 2016), and cryptocurrencies (e.g., Foley, Karlsen and Putnins 2019, Griffin and Shams 2020). Generally, these papers focus on assessing the extent of fraud/misconduct and understanding the underlying motivations of the perpetrators of fraud, often drawing from the classic works of Becker (1968) and Ehrlich (1973).

Chapter 5 shifts the spotlight onto the behavior of victims. Few studies that focus on the effects of investment scams on victims either exploit data on a single investment scam or aggregate data and geographic variation in fraud. For example, Knüpfer et al. (2023) study the long-run income loss experienced by victims who fell to a "Wincapita" Ponzi scheme in Finland during 2003-2008 (also see Rantala 2019). They document that victims experience a 6 percent labor income reduction, along with higher rates of unemployment and indebtedness, after becoming victims of Wincapita.

Giannetti and Wang (2016) show that federal securities enforcement actions against fraudulent firms reduce household market participation in the fraudulent firm's state (also see Georgarakos and Inderst 2014). Gurun et al. (2018) demonstrate that communities exposed to the Madoff Ponzi scheme withdraw assets from investment advisers and increase deposits at banks and relate their findings to eroding trust in the financial industry.

Chapter 5 focuses on the trading of investors who have fallen victim to scams and examines whether they learn from their mistakes. Unlike other papers that focus on a single isolated scam, we leverage blockchain's unique capabilities to comprehensively analyze scams. Due to the public nature of blockchain data, we can detect cryptocurrency scams through smart contract audits and systematic analysis of transaction activity in decentralized exchanges. This not only helps identify scams but also enables us to track the trading behavior of investors who have fallen victim to these scams. Moreover, the blockchain data enable us to create daily cryptocurrency portfolios and examine changes in portfolio characteristics following scams.

There is some evidence suggesting the potential for learning through trading actions. For example, Nicolosi et al. (2009) find that individuals' trading performance improves with trading experience, estimating that individuals can enhance their risk-adjusted portfolio return by approximately 2% per year. Furthermore, Seru et al. (2010) provide more conservative estimates that additional 100 trades are associated with an improvement in average returns of approximately 30 basis points per year. Learning through trading has been explored in various other settings.[71] This chapter contributes to this learning literature by focusing specifically on learning from investments in scams and investigating the impact of scams on investors' subsequent trading behavior and portfolio characteristics.

5.3 Institutional background

In this section, we review the institutional structure of the Ethereum blockchain, ERC-20 token standards, and trading in decentralized exchanges (see also Sections 4.3.1 and 4.3.2 in Chapter 4).

[71] See, for example, Strahilevitz et al. (2011), who examine whether previous experiences with a stock influence investors' willingness to repurchase stocks, and List (2003), who investigate whether experience eliminates investors' endowment effect (tendency of investors to retain an asset they own). Other papers that document investors' learning include Feng and Seasholes (2005), Linnainmaa (2010), Pástor et al. (2009), and Choi et al. (2009). See also early learning-by-doing models, such as those proposed by Arrow (1962) and Grossman, Kihlstrom and Mirman (1977).

5.3.1 ERC-20 tokens

Following the growth of cryptocurrencies after Bitcoin, developers explored the feasibility of decentralized applications (DApps), which led to the development of the Ethereum blockchain with smart contract functionality. Similar to cryptocurrencies like Bitcoin and Eth, DApps also require their own ”tokens” to power their crypto-economic dynamics. On the Ethereum platform, the majority of tokens are smart contracts that adhere to the Ethereum Request for Comment 20 (ERC-20) standards. ERC-20 is an application standard that allows the creation of ERC-20 tokens on the Ethereum blockchain. It provides a standard interface for fungible tokens. These standards ensure that smart contracts remain composable by specifying a set of rules that tokens should follow, including the total supply of tokens, how the tokens are transferred, and how transactions are approved.[72]

5.3.2 Trading through decentralized exchanges

Decentralized exchanges (DEXs) or Automated market makers (AMMs) are venues that pool liquidity and make it available to traders according to an algorithm (e.g., Othman 2012). Each pool in AMM contains two tokens, and traders swap one token for the other. Token creators typically pair a new token with an existing token to create a liquidity pool. For consistency, we describe the new token as token A and the paired token which is usually a valuable token as token B.[73] Users interact with the liquidity pool through two main types of operations.

First, users can swap tokens, with the swap rate determined by the token ratio in the pool. In Uniswap and Sushiswap, the swap rate is determined based on the “constant product” formula.[74] Second, users can add or remove liquidity from the pool. Those who deposit

[72] The most popular token standards on Ethereum are ERC-20 (a standard interface for fungible tokens), ERC-721 (a standard interface for non-fungible tokens), ERC-777 (allows to building extra functionality on top of tokens such as an emergency recovery function), ERC-1155 (allows for more efficient trades and bundling of transactions), and ERC-4626 (a tokenized vault standard designed to optimize and unify the technical parameters of yield-bearing vaults).

[73] We define token A and token B based on the number of times they appear in other liquidity pools. The valuable token (token B) is the one that is more frequently paired token in other pools. We also define token B as the token that has been created earlier and obtain similar results.

[74] The constant product formula is given by $k = A \cdot B$, where k is a constant and A, B are the numbers of token A and token B in the pool, respectively. This formula ensures that buying token A and selling token B leads to an increase in the value of token A relative to token B, and vice versa. Greater liquidity in the pool (higher k) results in lower price impact or slippage per trade.

token pairs to the pool are known as liquidity providers. When adding liquidity, providers deposit proportional amounts of token A and B based on the swap rate, maintaining the token ratio in the pool. Liquidity providers receive liquidity provider (LP) tokens that represent their share in the pool, entitling them to a percentage of the fees generated by trades. Liquidity providers can burn their LP tokens and withdraw liquidity based on the prevailing swap rate. Importantly, adding or removing liquidity does not alter the token ratio in the pool or the swap rate.

5.4 Data and methods

5.4.1 Data sources

We obtain data from multiple sources. First, all ERC-20 token transactions are sourced from publicly available blockchain files on blockchair.com. Blockchair serves as a blockchain search and analytics engine for major cryptocurrencies. The data provides comprehensive details about transactions involving ERC-20 tokens, including block IDs, transaction hashes, transaction times, token addresses, token names, sender and recipient addresses, and token transfer amounts. The sample of ERC-20 transactions spans from October 30, 2015, to April 8, 2022, covering a total of 919.1 million transfers involving 362,189 ERC-20 tokens.

Second, we obtain the list of all ERC-20 liquidity pools in Uniswap (V2 and V3) and Sushiswap, and the final liquidity amounts in these pools from the Etherscan DEX tracker. Etherscan functions as a block explorer and analytics platform for the Ethereum blockchain, providing users information about individual blocks, transactions, and smart contracts. As of April 8, 2022, the Etherscan DEX tracker provides a list of 78,178 token pairs for 66,909 unique tokens.

Third, we extract wallet addresses of ERC-20 token creators from each token's Etherscan page to track misappropriated funds from investors. Fourth, the list of ERC-20 tokens with malicious code in their smart contracts is sourced from tokensniffer.com. TokenSniffer scans smart contracts for scams, computes token metrics, maintains a scam list, and provides information on various smart contract exploits. Lastly, we obtain the daily USD values of valuable tokens from Coinmarketcap.

5.4.2 Data filters and transformations

We compound raw blockchain transfers into four events, each initiated by one user (wallet address): swap buys, swap sells, add liquidity, and remove liquidity. To do this, we examine the structure of each transaction. Add-liquidity (remove liquidity) transactions include three raw transfers, where a user (pool) sends token A and B to the pool (user), and the pool (user) sends the LP token to the user (pool). Swap buy and sell events refers to buying or selling token A (new token) in exchange for token B (valuable token). Swap buys and sells include two raw transfers, where a user and the pool exchange tokens A and B. After consolidating transfers to events, we end up with 99.1 million events.[75]

Once we identify events in Uniswap and Sushiswap, we remove liquidity pools with fewer than five non-creator transactions, resulting in 58,459 pools for 50,422 unique ERC-20 tokens that collectively have 99 million events. In the final data set, for each liquidity pool, we observe all events, including their timestamps, transaction hashes (with the possibility of multiple events per hash), the addresses of the users who initiated the events, and the addresses, names, and quantities of tokens exchanged.

5.4.3 Identifying scams and victims

We identify scams using two methods. First, we compile a list of ERC-20 tokens with malicious code in their smart contracts from TokenSniffer. This includes malicious code such as transfer blocks, liquidity pool blocks, fake ownership renounce, and hidden functionalities. Transfer blocks occur when the token creator retains special access to the smart contract through which any transfers can be rejected, while liquidity pool blocks prevent transfers from the pool (e.g., sell restriction). Renouncing ownership ensures that no one, including the creator, has privileged access to certain functions (e.g., disabling sell events, changing fees, or minting new tokens). Hidden functionalities involve modifying user balances, approval processes, fees, or arbitrarily minting tokens. TokenSniffer scans smart contracts for vulnerabilities and maintains a list of scam tokens. We identify 2,405 liquidity pools that have malicious code in their smart contracts.

[75] Note that 78% of transactions include only one event. When a transaction involves multiple events, we employ the same methodology iteratively, utilizing the sequence of transfers within the transaction.

Second, we implement the scam detection algorithm proposed in Chapter 4, which relies on token price, creator's profit, and activity patterns of liquidity pools. The price pattern focuses on detecting instances of price run-ups followed by collapses – it is satisfied when the initial price of token A is lower than its maximum price, and either the minimum price subsequently falls below the initial price or the final liquidity in the pool becomes $ 0.[76] The profit pattern examines whether the pool creator realizes a profit after the last event. The activity pattern assesses the current level of activity in the pool to determine if it is inactive. Following Aliyev, Allahverdiyeva and Putnins (2023c), we identify a total of 31,931 pools that we classify as scam pools at the intersection of all three patterns, and consider token A in these pools as scam tokens, even if they do not contain malicious code in their smart contracts. Therefore, in total, we identify 34,336 scam pools (2,405 + 31,931).

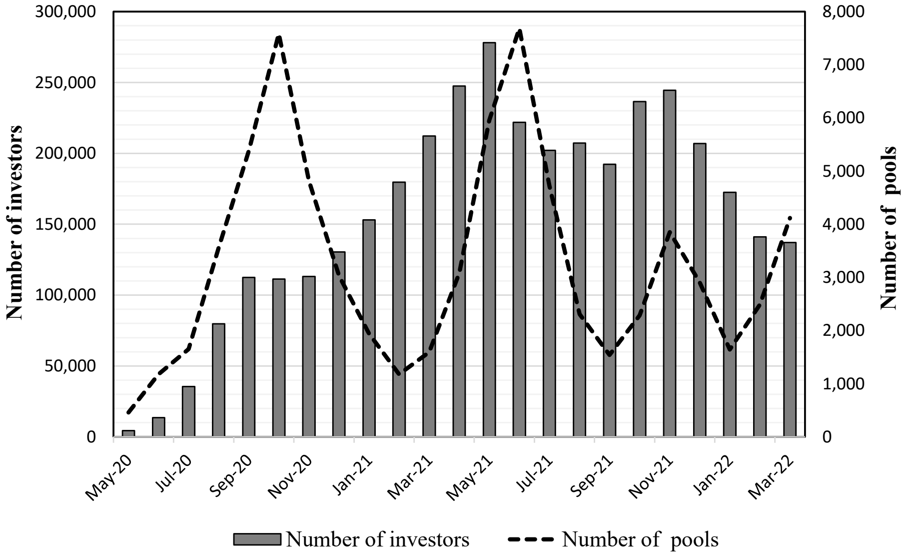
We define victims as investors who have invested in scam pools by purchasing token A or adding liquidity to the scam pool and subsequently experienced financial losses. To calculate the loss attributable to each scam pool, we track the cumulative balance of valuable token (token B) in $ terms for each investor in relation to each scam pool. Investors with a negative balance at the end of scam pools are classified as victims. We focus on investors who have made investments on at least seven different days to examine changes in their investments and also exclude investors with more than 1,000 investment to remove bot accounts. In total, we identify 727,189 investors and 218,799 victims who have invested in scam pools.

Figure 5.1 plots the evolution of the number of new scam pools and victims (Panel A) and investor participation along with the number of new pools (Panel B) between May 2020 and March 2022. As shown in Panel B, investor participation, which reflects mainstream interest in cryptocurrencies, varies significantly over time. Importantly, the number of new scam pools and victims follows the trends in mainstream cryptocurrency interest, experiencing spikes early in the sample due to the emergence of DEXs and in early 2021, for example, when Bitcoin reached a market value of $1 trillion for the first time.

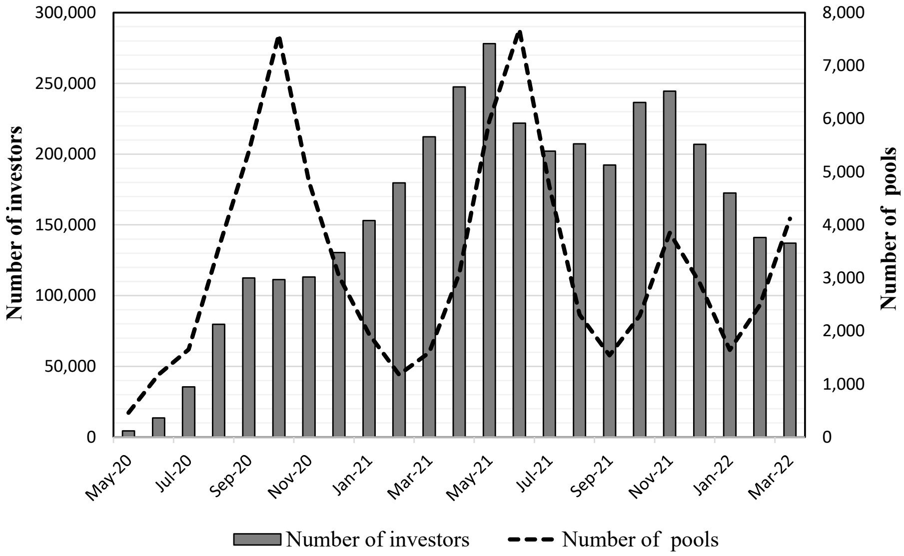
[76] It is important to note that 16% of the pools end with the maximum price, where the last event of the pool is a remove-liquidity transaction. When investigating price collapses based solely on the swap rate, these pools may be overlooked, as no observable price collapse occurs (according to the constant product formula, swap rates remain unchanged during add or remove liquidity). However, one distinguishing characteristic of scams is that creators often withdraw liquidity from the pool when the token's price is at its peak. To address this, we also consider the final liquidity in the pool as part of the price collapse condition. The condition is met when the final liquidity in the pool reaches $ 0.

Figure 5.1. Number of investors and victims over time
This figure plots the monthly number of investors that have fallen victim a scam and the number of new scam pools created in each month (Panel A), and investors that place one or more trades and the number of new liquidity pools created in each month (Panel B) during May 2020 and March 2022. In Panel A, the left axis is the number of victims and the right axis is the number of new scam pools in each month. In Panel B, the left axis is the number of investors and the right axis is the number of new liquidity pools in each month. The total number of investors (with at least seven days of investments) is 727,189, victims is 218,799, liquidity pools is 74,960, and scam pools is 34,332 during May 2020 and March 2022.

Panel A. Number of victims and scam pools by month



Panel B. Number of investors and liquidity pools by month



5.4.4 Investor characteristics

To examine the impact of scams on investors' subsequent trading behavior, we calculate versions of a binary variable ($PastScam$) that equals 1 if an investor has fallen victim to a scam within the past 30, 60, 90 days, or ever. To measure the extent of scam-related investments, we similarly calculate a continuous variable (SIR) as the ratio of the investment in scam tokens to the total investment during the same period.[77] As part of other scam characteristics of victims, we count the number of interacted scam pools ($ScamPools$) and the number of scam investments ($ScamInv$), where an investment is either a purchase of token A or add liquidity to the scam pool.

We calculate several other investor characteristics, categorized into trading and portfolio characteristics. Trading characteristics are calculated from the investor-event level data and include the number of days with events ($ActiveDays$), the number of interacted liquidity pools ($Pools$), the number of tokens ($Tokens$), the number of events ($Events$), and the average trade volume ($Volume$) for each investor. We compute cumulative versions of these variables for different models. Additionally, we calculate a simple return ($Return$) for each token investment excluding scam tokens over different periods (7, 15 and 30 days) to measure the investment performance of investors.

To investigate the impact of scams on investors' portfolio characteristics, we create a portfolio-date dataset from raw blockchain transfers, tracking all tokens entering and exiting specific wallet addresses with only trade data, and trade and transfer (from one wallet address to another) data combined. Within these portfolios, we observe the tokens and their quantities held by investors on a daily basis. Using the prices of valuable tokens and the swap rates in DEXs, we calculate the prices of all token A and daily portfolio values excluding scams ($PValue$) for each investor.[78] We track the number of tokens in the portfolio ($Ptokens$) and calculate daily portfolio returns ($Preturn$), the 7-day moving average standard deviation of daily returns ($PStdDev$), Sharpe ratio ($PSharpe$), and the percentage of cash/stablecoins in the portfolio ($PCash$). Table 5.1 provides the definitions of all variables and their calculation methods.

[77] We also measure the financial impact of scams on investors by calculating the $Loss$ in dollar terms experienced by each victim resulting from their investments in scam tokens/pools within the past 30, 60, 90 days, or ever.

[78] We also exclude LP tokens, tokens that are not traded on Uniswap and Sushiswap, tokens with a price exceeding $100,000, and set negative portfolio values to zero.

Table 5.1. Definitions of variables
This table reports the scam, trade, and portfolio characteristics of investors. The first column shows the variable name and the second column shows the variable definition.

Variable	**Definition**
Scam characteristics	
$PastScam$	A dummy variable that equals to 1 if an investor has fallen victim to a scam in the past $x = \{30, 60, 90, \text{all}\}$ days before a given investment or day
SIR (%)	The scam investment ratio (SIR) is a ratio of scam investment volume to the total investment volume in the past $x = \{30, 60, 90, \text{all}\}$ days before a given investment or day
$Loss$ ($\cdot 10^6$ USD)	The total value of loss from scam pools for each victim
$ScamPools$	The total number of scam pools interacted by each victim
$ScamInv$	The total number of scam investments of victims
Trade characteristics	
$ActiveDays$	The total number of days with events/trades for each investor
$Pools$	The total number of liquidity pools interacted by each investor
$Tokens$	The total number of tokens traded by each investor
$Events$	The total number of events (buy, sell, add-liquidity, and remove-liquidity) of each investor
$Volume$ ($\cdot 10^6$ USD)	The average USD volume of trades of each investor
Portfolio characteristics	
$PValue$ ($\cdot 10^6$ USD)	The average daily marked-to-market USD value of the portfolio for each investor
$PTokens$	The average daily number of tokens in the portfolio of each investor
$PReturn$ (%)	The daily portfolio returns
$PStdDev$ (%)	The 7-day moving average standard deviation of daily portfolio returns
$PSharpe$	The daily Sharpe ratio is based on the daily portfolio return and 7-day moving average standard deviation, and the risk-free rate is the US 10-year treasury bond yield
$PCash$ (%)	The average daily cash allocation is the ratio of the total USD value of stablecoins in the portfolio to the total marked-to-market USD value of the portfolio

5.5 Results

We conduct the empirical tests in five steps. First, we identify investor characteristics that distinguish victims from non-victims and compare the characteristics of victims before and after scams using univariate analysis. Second, we investigate whether scams discourage investors from making further scam investments. Third, we examine the impact of scams on the returns of subsequent investments. Fourth, we analyze changes in investors' portfolio characteristics after experiencing a scam. Lastly, we characterize varying learning rates among victims.

5.5.1 Characteristics of victims

5.5.1.1 Victims versus non-victims

Table 5.2 presents summary statistics (mean, median, and standard deviation) for scam, trade, and portfolio characteristics of victims and compares them to those of non-victims. We assess the significance of the differences in characteristics between the two groups using paired t-tests. We only report the scam characteristics for victims, as these characteristics have zero value for non-victims. On average, victims of cryptocurrency scams allocate approximately 10.35% of their total investments to scams, resulting in a monetary loss of around $80,000. Furthermore, victims typically engage with around 3.92 scam pools and invest (buy or add liquidity) approximately 6.06 times in the scam pools.

Significant differences exist in the trade and portfolio characteristics of victims and non-victims. In terms of trade characteristics, victims tend to be more active in the market. On average, victims stay active in the market for 21 more days, interact with 23 more liquidity pools and 21 more tokens, and conduct 88 more trades. The average transaction volume is slightly lower for victims.

Table 5.2 also presents the portfolio characteristics of investors constructed from only trade data (excluding transfers). We exclude transfers because our main focus is on victims' investment behavior. On average, the portfolio value of victims is $55,000 more than that of non-victims and victims hold 8 more tokens in their daily portfolio. The average daily portfolio return of victims is higher than that of non-victims (though insignificant), but the standard deviation of daily portfolio returns for victims also tends to be higher, suggesting a higher risk tolerance among victims. Consequently, the Sharpe ratio as a risk-adjusted return (excess return per unit risk) is lower for victims. Furthermore, victims tend to allocate a smaller proportion of their portfolios to cash, further suggesting a higher risk tolerance.

Table 5.2. Characteristics of victims and non-victims
This table reports the descriptive statistics (mean, median, standard deviation, and number of observations) of the variables for victims and non-victims. Scam characteristics are only reported for victims since they are zero for non-victims, while trade and portfolio characteristics are reported for both victims and non-victims. The last two columns in the table report the difference in mean trade and portfolio characteristics of victims and non-victims. The standard errors used to compute the t-statistics are clustered by the number of active days with a trade. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. For a detailed definition of variables, refer to Table 5.1. The units of variables are in parentheses following the variable name.

	Victims (1)				**Non-victims (2)**				**Mean difference**	
	Mean	Median	Std. dev.	N obs.	Mean	Median	Std. dev.	N obs.	(1)-(2)	t-stat
Scam characteristics										
PastScam	1.000	1.000	0.000	218,799	-	-	-	508,390	-	-
SIR (%)	10.350	3.796	17.350	218,799	-	-	-	508,390	-	-
Loss ($\cdot 10^6$ USD)	0.080	0.002	4.325	218,799	-	-	-	508,390	-	-
ScamPools	3.922	1.000	9.615	218,799	-	-	-	508,390	-	-
ScamInv	6.064	2.000	17.294	218,799	-	-	-	508,390	-	-
Trade characteristics										
ActiveDays	39.009	24.000	42.075	218,799	17.531	12.000	17.369	508,390	21.478	2.337***
Pools	35.246	21.000	43.096	218,799	11.824	9.000	11.792	508,390	23.422	2.900***
Tokens	31.837	19.000	10.000	218,799	10.756	8.000	9.429	508,390	21.081	3.061***
Events	129.907	54.000	215.380	218,799	41.522	21.000	129.205	508,390	88.386	2.231**
Volume ($\cdot 10^6$ USD)	0.005	0.001	0.057	218,799	0.006	0.001	0.064	508,390	-0.001	-3.936***
Portfolio characteristics										
PValue ($\cdot 10^6$ USD)	0.105	0.010	0.594	202,469	0.050	0.006	0.344	494,527	0.055	2.159**
PTokens	14.338	9.121	15.645	202,469	6.322	4.769	5.468	494,527	8.016	2.942***
PReturn (%)	8.751	4.637	18.011	202,246	8.201	4.168	16.977	493,705	0.550	1.250
PStdDev (%)	29.559	19.193	35.561	202,246	27.233	17.467	33.824	493,705	2.326	1.941*
PSharpe	0.091	0.091	0.157	201,843	0.094	0.085	0.183	492,849	-0.003	-1.061
PCash (%)	11.016	0.203	19.642	202,469	13.561	0.000	24.272	494,527	-2.545	-4.449***

5.5.1.2 Before versus after scams

Table 5.3 compares the characteristics of victims before and after experiencing a scam. A summary of the table is that victims trade more actively after experiencing a scam. The number of active days for victims is, on average, 11.55 days before the scam and 28.1 days after the scam. Victims interact with 16.6 more liquidity pools and 14.5 more tokens, conduct 59.6 more trades, and the average trade volume increases by around $1,000 after falling victim to a scam.

Table 5.3. Characteristics of victims before and after scam
This table reports the descriptive statistics (mean, median, standard deviation, and number of observations) of the trade and portfolio variables for victims before and after falling victim to a scam. The scam variables are omitted from this table since they are zero before scam and the same as in Table 5.2 after scam. The "Mean difference" column in the table reports the difference in mean characteristics of victims before and after scam. The standard errors used to compute the t-statistics are clustered by the number of active days with a trade. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. For a detailed definition of variables, refer to Table 5.1. The units of variables are in parentheses following the variable name.

	After scam (1)			**Before scam (2)**			**Mean difference**		
	Mean	Median	Std. dev.	Mean	Median	Std. dev.	(1)-(2)	t-stat	N obs.
Trade characteristics									
ActiveDays	28.097	14.000	37.414	11.551	7.000	15.430	16.546	2.371***	213,279
Pools	27.886	14.000	41.317	11.276	7.000	12.576	16.611	2.521***	213,279
Tokens	25.646	13.000	37.313	11.121	8.000	10.636	14.525	2.548***	213,279
Events	96.019	31.000	187.627	36.391	15.000	74.176	59.627	1.954**	213,279
Volume ($\cdot 10^6$ USD)	0.005	0.001	0.066	0.004	0.001	0.059	0.001	2.528***	213,279
Portfolio characteristics									
PValue ($\cdot 10^6$ USD)	0.153	0.016	0.847	0.037	0.003	0.306	0.117	5.300***	159,644
PTokens	18.401	12.717	17.759	5.880	4.368	5.277	12.521	5.338***	159,644
PReturn (%)	5.359	1.888	18.562	24.398	9.065	64.710	-19.039	-9.564***	159,644
PStdDev (%)	22.948	13.006	35.704	63.119	33.007	90.332	-40.170	-13.213***	159,644
PSharpe	0.078	0.076	0.200	0.172	0.145	0.414	-0.094	-17.003***	159,644
PCash (%)	12.356	0.267	21.705	8.736	0.000	20.426	3.620	9.833***	159,644

In terms of portfolio characteristics, victims increase both their portfolio value and the number of tokens in their portfolio after the scam. This may occur for two reasons. First, experiencing a scam can make victims more aggressive in their investments and increase their market exposure (investment effect). Alternatively, investors may consolidate their assets into a single wallet address in response to falling victim to a scam (consolidation effect).[79] While the former effect represents a longer-term impact of victimization, the latter seems to be short-term. The results in Table 5.3 mainly represent the investment effects since the portfolios are constructed from trade data only.[80] The post-scam portfolio has a lower return but also a lower standard deviation, suggesting that investors become less risk-tolerant after experiencing scams. As a result, victims do not regain their pre-scam level of risk-adjusted performance, as measured by the Sharpe ratio. Moreover, victims tend to allocate a higher proportion of their portfolio to cash after the scam, further indicating a reduced risk tolerance.

These findings, particularly that investors trade more actively after experiencing a scam and perform poorly, contrast with prior research indicating that victimization and the resulting loss of trust lead to reduced market participation. For example, Giannetti and Wang (2016) show that federal securities enforcement actions against fraudulent firms led to decreased household market participation in the fraudulent firm's state. Additionally, Gurun et al. (2018) show that communities exposed to the Madoff Ponzi scheme responded by withdrawing assets from investment advisers and increasing bank deposits. In contrast, our findings suggest that investors become more active in trading and increase their market exposure, potentially attempting to recover their losses from scams, but they fail to do so. In this context, it is important to analyze the short- and long-term impacts of scams on victims, which we examine in subsequent analyses.

5.5.2 Impact of scams on further scam investments

We now examine the impact of experiencing a scam on subsequent scam investments. If investors learn from their mistakes, we should observe a decrease in scam investments after experiencing a scam. To test this hypothesis, we run the following probit regression:

$$I_{i,t} = \Phi(\alpha + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}), \tag{5.1}$$

[79] For example, during the FTX collapse in 2022, many investors withdrew their assets from centralized exchanges into their personal wallets. See "Crypto Users Jump to DeFi Platforms in Wake of FTX's CeFi Crash".

[80] Table A5.1 in the Appendix also compares the portfolio characteristics (including trades and transfers) of victims before and after experiencing a scam, thus capturing both investment and consolidation effects.

where $I_{i,t}$ is a binary variable that equals 1 if the investment made by victim i at time t is a scam and 0 otherwise, and Φ is the cumulative distribution function (CDF) of the standard normal distribution. Our primary variable of interest is $PastScam_{i,t-x}$, which is a binary variable indicating whether victim i has experienced a scam within the past x days before the given investment. The parameter x varies from Model 1 to 4, taking values of 30, 60, 90, and all days before the given investment to assess the impact of scams across different horizons.[81] The vector $X_{i,t}$ consists of control variables, including the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the average trade volume ($Volume_{i,t}$) up to the investment of victim i at time t.

Panel A in Table 5.4 reports that victims tend to invest less in scams after experiencing a scam, as indicated by the negative coefficient of the $PastScam_{i,t-x}$ variable. This suggests that victims become more cautious following a scam, adjusting their investment behavior to avoid similar fraudulent schemes. This result holds true for $x = \{30, 60, 90, \text{all}\}$ days before the given investment. The average marginal effect of experiencing a scam varies between -2.97% (when $x = \text{all}$) and -4.05% (when $x = 30$).[82] That means experiencing a scam in the last 30 days decreases the probability of investing in scams again by 4.05%, with similar results holding for other periods after the scam, indicating that the learning from scams is long-lasting.

Panel B in Table 5.4 presents the results of the modified version of the model in Eq. (5.1), where we replace $PastScam_{i,t-x}$ with the scam investment ratio ($SIR_{i,t-x}$) of victim i within the past $x = \{30, 60, 90, \text{all}\}$ days before the given investment. This examines whether those who have invested more in scams or experienced more significant losses from scams learn more from their mistakes. A higher magnitude of scam, reflected in the higher $SIR_{i,t-x}$, results in a lower likelihood of subsequent scam investment, as indicated by the negative coefficient of $SIR_{i,t-x}$. This suggests that victims who have incurred greater losses from scams tend to exercise caution in their investment decisions. Specifically, a

[81] The results with a 30-day horizon are generally consistent with much shorter periods (e.g., 1 week). However, we exclude shorter horizons since investors may not realize that they have been scammed during a short period, given that some of the scams take longer to materialize with a gradual theft of liquidity from the pool.

[82] We calculate the marginal effects of $PastScam$ for each observation as the difference between the predicted $I_{i,t}$ with $PastScam = 1$ and $PastScam = 0$, scaled by the $I_{i,t}$ with $PastScam = 0$ and average over the entire sample.

1% increase in SIR decreases the probability of investing in scam again by around 0.30% across all models.[83]

Table 5.4. Impact of scams on further scam investment
This table reports the impact of experiencing a scam on subsequent scam investment of victims using the following probit regression:

$$I_{i,t} = \Phi(\alpha + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t})$$

where the dependent variable $I_{i,t}$ is a dummy variable that equals to 1 if the investment of the victim i at time t is a scam investment, and Φ is the cumulative distribution function (CDF) of the standard normal distribution. $PastScam_{i,t-x}$ is a dummy variable that equals to 1 if the victim i has been scammed in the past x days. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days. In Models 1-4, x is respectively defined as 30, 60, 90, and all days before the given investment. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$) up to the investment of victim i at time t. P values are in parenthesis, and ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
Intercept	1.526***	1.502***	1.483***	1.410***
	(< .0001)	(< .0001)	(< .0001)	(< .0001)
$\mathbf{PastScam_{i,t-x}}$	**-0.574*** **	**-0.484*** **	**-0.436*** **	**-0.293*** **
	(< .0001)	**(< .0001)**	**(< .0001)**	**(< .0001)**
$Events_{i,t}$	0.0002***	0.0001***	0.0001***	-0.0001***
	(< .0001)	(< .0001)	(< .0001)	(< .0001)
$ActiveDays_{i,t}$	0.003***	0.003***	0.004***	0.005***
	(< .0001)	(< .0001)	(< .0001)	(< .0001)
$Volume_{i,t}$	-0.185***	-0.180***	-0.178***	-0.174***
	(< .0001)	(< .0001)	(< .0001)	(< .0001)
N obs.	15,365,503			
N victims	218,799			
Panel B: $SIR_{i,t-x}$				
Intercept	1.481***	1.484***	1.485***	1.487***
	(< .0001)	(< .0001)	(< .0001)	(< .0001)
$\mathbf{SIR_{i,t-x}}$	**-0.022*** **	**-0.022*** **	**-0.022*** **	**-0.022*** **
	(< .0001)	**(< .0001)**	**(< .0001)**	**(< .0001)**
$Events_{i,t}$	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	(< .0001)	(< .0001)	(< .0001)	(< .0001)
$ActiveDays_{i,t}$	0.003***	0.003***	0.003***	0.002***
	(< .0001)	(< .0001)	(< .0001)	(< .0001)
$Volume_{i,t}$	-0.152***	-0.173***	-0.193***	-0.225***
	(< .0001)	(< .0001)	(< .0001)	(< .0001)
N obs.	15,365,503			
N victims	218,799			

[83] The marginal effect of SIR is calculated as $\beta_1 \cdot \phi(I_{i,t})$, where ϕ is the probability density function (PDF) of the standard normal distribution. The marginal effects are calculated for each observation and averaged over the entire sample.

5.5.3 Impact of scams on trading performance

Investor performance is another variable that we correlate with scam investments. Measuring the investment performance of individual investors is not straightforward since it is not obvious how to compare the performance of investors with varying holding periods. Following Seru, Shumway and Stoffman (2010), we calculate the returns earned by each investment in the 30 trading days following the investment. All our findings remain unchanged if we use a 7- or 15-day holding period. To examine the impact of scams on subsequent investment returns, we run the following panel regression:

$$Return_{i,t+30} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}, \tag{5.2}$$

where $Return_{i,t+30}$ represents the 30-day return following the investment in a non-scam token by victim i at time t. As before, $PastScam_{i,t-x}$ is a binary variable that equals 1 if victim i has experienced a scam within the past $x = \{30, 60, 90, \text{all}\}$ days before the given investment. We include controls for $Events_{i,t}$, $ActiveDays_{i,t}$, and $Volume_{i,t}$ up to the investment of victim i at time t. Furthermore, we include individual fixed effects (α_i) to account for unobserved individual heterogeneity. The standard errors are clustered by individuals to address potential correlation within individuals.

As reported in Panel A of Table 5.5, experiencing a scam has a negative impact on subsequent non-scam investment returns. The coefficient of $PastScam_{i,t-x}$ is consistently negative and statistically significant at 1 percent level across all models. In terms of economic magnitude, experiencing a scam decreases subsequent non-scam investment returns by 12.8% when $x = 90$ and 15.2% when $x = 30$.

Panel B of Table 5.5 reports that the magnitude of the scam, as captured by $SIR_{i,t-x}$, also negatively impacts subsequent investment returns. A one standard deviation shock to SIR is, on average, associated with approximately 4% decrease in investment returns (e.g., $\beta_1 \cdot StdDev(SIR) = -0.208 \cdot 17.35 = -3.61\%$ when $x = 30$). While investors become more cautious after a scam, as reflected in the decrease in the likelihood of scam investment, the returns on their investments also decrease after experiencing a scam, and this effect is more pronounced when the level of scam investment relative to the total investment increases.

Table 5.5. Impact of scams on subsequent investment returns
This table reports the impact of experiencing a scam on subsequent investment returns of victims using the following panel regression:

$$Return_{i,t+30} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable $Return_{i,t+30}$ is the 30-day return following the investment (buy or add liquidity) of a non-scam token of victim i at time t. $PastScam_{i,t-x}$ is a dummy variable that equals to 1 if the victim i has been scammed in the past x days. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days. In Models 1-4, x is respectively defined as 30, 60, 90, and all days before the given investment. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$) up to the investment of victim i at time t. All versions of the model include individual fixed effects (α_i). The standard errors used to compute the t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**-15.228*** (-60.719)**	**-14.484*** (-57.679)**	**-12.822*** (-50.851)**	**-13.347*** (-41.772)**
$Events_{i,t}$	0.020*** (9.369)	0.020*** (9.630)	0.020*** (9.438)	0.014*** (7.082)
$ActiveDays_{i,t}$	-0.199*** (-19.320)	-0.192*** (-18.435)	-0.179*** (-17.525)	-0.101*** (-10.334)
$Volume_{i,t}$	-3.424*** (-2.925)	-3.366*** (-2.807)	-3.378*** (-2.783)	-3.308*** (-2.762)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.03%	0.02%	0.02%	0.02%
N obs.	10,534,069			
N victims	215,460			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**-0.208*** (-20.984)**	**-0.222*** (-19.633)**	**-0.219*** (-17.938)**	**-0.250*** (-15.858)**
$Events_{i,t}$	0.014*** (7.272)	0.014*** (7.254)	0.014*** (7.254)	0.014*** (7.266)
$ActiveDays_{i,t}$	-0.165*** (-16.952)	-0.164*** (-16.935)	-0.164*** (-16.951)	-0.165*** (-17.201)
$Volume_{i,t}$	-3.710*** (-2.935)	-3.719*** (-2.935)	-3.725*** (-2.935)	-3.747*** (-2.933)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.1%	0.1%	0.1%	0.1%
N obs.	10,534,069			
N victims	215,460			

5.5.4 Impact of scams on investor portfolio

To investigate the impact of experiencing a scam on investors' subsequent portfolio characteristics, we run the following panel regression:

$$Y_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}, \tag{5.3}$$

where the dependent variable $Y_{i,t}$ represents the daily portfolio characteristics (excluding transfers), $PValue_{i,t}$, $PTokens_{i,t}$, $PReturn_{i,t}$, $PStdDev_{i,t}$, $PSharpe_{i,t}$, and $PCash_{i,t}$ as defined in Section 5.4.4 and Table 5.1. $PastScam_{i,t-x}$ is a binary variable that equals 1 if victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. The vectors of controls for victim i until day t is $X_{i,t} = \{Events_{i,t}, ActiveDays_{i,t}, Volume_{i,t}\}$. As in Section 5.5.3, we include individual fixed effects (α_i) to account for unobserved individual heterogeneity and cluster standard errors by individuals to address potential correlation within individuals.

In Table 5.6, we present the findings for the model in Eq. (5.3) in Panel A and the modified version of the model in Eq. (5.3), where we replace the binary $PastScam$ with the continuous scam investment ratio SIR in Panel B. This table only reports the coefficients of $PastScam$ and SIR across 48 different models, omitting the coefficients of controls. In addition, we take the the scam date as each interaction date with the scam pool.[84]

As shown in Panel A, we find that investors falling victim to a scam increase their portfolio values, as indicated by the positive coefficient of $PastScam$. In addition, they increase the number of tokens held in their portfolios. Given the portfolio characteristics are constructed from trade data only, this predominantly shows the investment effect that investors increase their exposure to the market after experiencing scam.

Table 5.6 also indicates that falling victim to a scam decreases investors' daily portfolio returns for $x = \{30, 60, 90, \text{all}\}$. However, the impact of scams on the daily portfolio standard deviation varies over time. In the short-term ($x = \{30, 60\}$), the standard dev-

[84] Tables A5.2-A5.13 in the Appendix report the coefficients of all model variables across different models. Table A5.14 reports the results of the model in which we replace SIR with $Loss$ and the model in which we define the scam date as the last interaction date with the scam pool.

Table 5.6. Impact of scams on portfolio characteristics (trades only)
This table reports the impact of scams on portfolio characteristics of victims using the following panel regression:

$$Y_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the daily portfolio characteristics (excluding transfers), $PValue_{i,t}$, $PTokens_{i,t}$, $PReturn_{i,t}$, $PStdDev_{i,t}$, $PSharpe_{i,t}$, and $PCash_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. The table only reports the coefficient of $PastScam_{i,t-x}$ in Panel A ($SIR_{i,t-x}$ in Panel B) across different models (see Appendix for all coefficients). $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = \text{all}$)
Panel A: $PastScam_{i,t-x}$				
$PValue_{i,t}$	0.026***	0.015***	0.007***	0.015***
	(13.59)	(8.187)	(3.593)	(3.965)
$PTokens_{i,t}$	0.368***	0.589***	0.708***	2.421***
	(18.931)	(33.388)	(40.097)	(49.431)
$PReturn_{i,t}$	-0.069**	-1.322***	-1.935***	-6.512***
	(-2.025)	(-44.339)	(-66.147)	(-124.760)
$PStdDev_{i,t}$	4.975***	0.603***	-1.734***	-12.737***
	(61.779)	(8.536)	(-25.125)	(-106.870)
$PSharpe_{i,t}$	-0.022***	-0.032***	-0.027***	-0.039***
	(-47.932)	(-74.288)	(-65.558)	(-64.891)
$PCash_{i,t}$	-0.429***	-0.078**	0.249***	1.861***
	(-12.130)	(-2.044)	(6.073)	(23.537)
Panel B: $SIR_{i,t-x}$				
$PValue_{i,t}$	0.049***	0.049***	0.049***	0.026***
	(9.586)	(8.077)	(7.051)	(2.777)
$PTokens_{i,t}$	0.011	0.036	0.028	0.624***
	(0.383)	(1.020)	(0.660)	(6.173)
$PReturn_{i,t}$	-0.157*	-0.950***	-1.356***	-4.560***
	(-1.925)	(-10.423)	(-13.096)	(-19.772)
$PStdDev_{i,t}$	4.534***	1.906***	0.696***	-2.780***
	(24.363)	(9.200)	(2.966)	(-5.463)
$PSharpe_{i,t}$	-0.061***	-0.057***	-0.048***	-0.073***
	(-51.506)	(-43.773)	(-33.757)	(-27.945)
$PCash_{i,t}$	-2.158***	-2.605***	-3.004***	-7.945***
	(-26.575)	(-24.693)	(-23.578)	(-26.029)

-iation increases, suggesting an increase in risk tolerance. In the long term ($x = \{90, \text{all}\}$), the standard deviation decreases, suggesting a reduced risk tolerance. Nevertheless, the

Sharpe ratio of the portfolio decreases after experiencing a scam for $x = \{30, 60, 90, \text{all}\}$, indicating that investors never regain their pre-scam level of risk-adjusted performance. The increase (decrease) in risk tolerance in the short term (long term) is also evident in the decrease (increase) in cash allocation. The results for SIR in Panel B is mostly consistent with that of Panel A.

To further show the consolidation effect of experiencing a scam, we recalculate portfolio values from trade and transfer data into and out of a given wallet. This way, we track the portfolio values attributable to investors' overall activity. Subsequently, we run the model in Eq. (5.3) with the reconstructed portfolio characteristics. These results are reported in Table 5.7.

As shown in Table 5.7, the impact of scams on portfolio values and the number of tokens in the portfolio remains positive, but the magnitude and significance of results are more pronounced than just the investment effect. This suggests that investors both increase their market exposure and consolidate their portfolios after experiencing a scam. The remaining results are similar. Daily returns decrease for $x = \{30, 60, 90, \text{all}\}$. The standard deviation increases in the short term ($x = \{30, 60\}$), but the opposite occurs in the long term, suggesting that investors become more risk-tolerant in the short term but less risk-tolerant in the long term. The Sharpe ratio decreases for $x = \{30, 60, 90, \text{all}\}$, indicating that investors never regain their risk-adjusted pre-scam return. In the combined trade and transfer analysis, Panel B provides further evidence of the consolidation effect. Specifically, SIR positively impacts daily portfolio returns in the short term ($x = \{30, 60\}$), suggesting that investors consolidate their assets in one wallet after experiencing scams.

5.5.5 Learning rate of investors

Investors learn about scams at different rates. Some investors consistently invest in scam pools, while others stop investing in scams after experiencing one. To investigate the varying learning rates among investors, we regress the number of scam pools and the number of scam investments (which serve as proxies for the inverse of the learning rate) on the trading and portfolio characteristics of investors, and present the results in Table 5.8. In this analysis, each unit of observation represents an investor, and the dependent variables are the number of scam pools invested in by a given investor and the number of scam investments, while the independent variables are the trading and portfolio characteristics of that investor.

Table 5.7. Impact of scams on portfolio characteristics (trades and transfers)
This table reports the impact of scams on portfolio characteristics of victims using the following panel regression:

$$Y_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the daily portfolio characteristics (including transfers), $PValue_{i,t}$, $PTokens_{i,t}$, $PReturn_{i,t}$, $PStdDev_{i,t}$, $PSharpe_{i,t}$, and $PCash_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. The table only reports the coefficient of $PastScam_{i,t-x}$ in Panel A ($SIR_{i,t-x}$ in Panel B) across different models (see Appendix for all coefficients). $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = \text{all}$)
Panel A: $PastScam_{i,t-x}$				
$PValue_{i,t}$	0.051***	0.046***	0.037***	0.058***
	(24.881)	(23.183)	(17.561)	(15.409)
$PTokens_{i,t}$	0.556***	0.794***	0.935***	3.324***
	(27.552)	(42.100)	(48.116)	(61.744)
$PReturn_{i,t}$	-0.037	-1.616***	-2.440***	-7.947***
	(-0.931)	(-47.088)	(-74.101)	(-131.701)
$PStdDev_{i,t}$	5.289***	0.085	-2.802***	-16.220***
	(56.446)	(1.047)	(-35.940)	(-117.708)
$PSharpe_{i,t}$	-0.019***	-0.031***	-0.028***	-0.034***
	(-41.657)	(-73.028)	(-67.198)	(-58.034)
$PCash_{i,t}$	-0.773***	-0.749***	-0.720***	-3.191***
	(-24.129)	(-22.147)	(-19.998)	(-44.605)
Panel B: $SIR_{i,t-x}$				
$PValue_{i,t}$	0.093***	0.103***	0.109***	0.183***
	(15.118)	(13.236)	(11.636)	(10.832)
$PTokens_{i,t}$	0.296***	0.451***	0.575***	1.986***
	(9.224)	(11.418)	(11.899)	(17.635)
$PReturn_{i,t}$	1.034***	0.094	-0.314***	-2.812***
	(10.303)	(0.867)	(-2.596)	(-11.723)
$PStdDev_{i,t}$	6.911***	3.711***	2.322***	-1.894***
	(30.457)	(14.950)	(8.401)	(-3.530)
$PSharpe_{i,t}$	-0.046***	-0.045***	-0.036***	-0.043***
	(-39.388)	(-34.876)	(-25.195)	(-17.268)
$PCash_{i,t}$	-1.305***	-1.603***	-1.884***	-5.380***
	(-16.634)	(-15.533)	(-14.950)	(-18.88)

Table 5.8. Impact of investor characteristics on learning rate
This table reports the impact of investor characteristics on the learning rates of victims using the following model:

$$Y_i = \alpha + \sum_n \delta_n \cdot X_i + \varepsilon_i$$

where the dependent variable is the number of scam pools interacted by victims in Models 1 and 3, and the number of scam investments in Models 2 and 4. $X_{i,t}$ is a vector of victim characteristics for each victim i. Models 1 and 2 include trade-level characteristics, while Models 3 and 4 include portfolio-level characteristics. T-statistics in brackets and ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	Model 1 *ScamPools*	Model 2 *ScamInv*	Model 3 *ScamPools*	Model 4 *ScamInv*
Intercept	0.477 (22.418)	0.674*** (19.624)	0.283** (8.208)	0.405*** (7.500)
ActiveDays	-0.107 (-178.936)	-0.118 (-122.355)	-	-
Pools	0.212 (364.382)	0.272 (289.755)	-	-
Volume ($\cdot 10^6$ USD)	4.328 (16.074)	4.094 (9.414)	-	-
Return (%)	-0.004 (-14.626)	-0.008 (-19.907)	-	-
PValue ($\cdot 10^6$ USD)	-	-	1.286*** (44.5529)	1.468*** (32.516)
PTokens	-	-	0.180*** (161.321)	0.271*** (155.772)
PReturn (%)	-	-	-0.075*** (-29.507)	-0.120*** (-30.270)
PStdDev (%)	-	-	0.052*** (44.485)	0.085*** (46.401)
PSharpe	-	-	-3.136*** (-25.959)	-5.182*** (-27.442)
PCash (%)	-	-	-0.015*** (-17.557)	-0.019*** (-13.976)
N obs.	215,460	215,460	199,284	199,284
R^2	43.5%	34.9%	13.7%	12.6%

As presented in Table 5.8, less experienced investors, characterized by the number of active trading days, are more likely to invest in multiple scam pools. These investors tend to learn less about scams because they are less active in the market. This aligns with classical learning-by-doing models (e.g., Arrow 1962, Grossman, Kihlstrom, and Mirman 1977), which suggest that investors might improve their ability as they trade (i.e., “learning by doing”). This finding is also consistent with recent empirical findings by Nicolosi et al. (2009) and Seru et al. (2010), which indicate that investors learn about trading by actively participating in the market. Therefore, investors who trade less tend to learn less about scams.

Furthermore, investors who interact with more pools and trade higher volumes are more likely to invest in multiple scam pools. This makes sense because it reflects the behavior of individuals who adopt a cavalier approach, often participating in risky speculative activities without conducting due diligence, and subsequently becoming victims of scams. Lastly, investors who achieve lower returns on non-scam investments are also more likely to invest in multiple scams.

In terms of portfolio characteristics, investors with higher portfolio values and a greater number of tokens in their portfolios are more likely to invest in multiple scam pools. This, again, is potentially associated with higher market exposure through often speculative and risky activities. Lastly, investors with lower portfolio returns, higher standard deviations of returns, and lower cash allocation in their portfolios are more likely to be scammed multiple times, suggesting that these investors are unsophisticated investors who often take on excessive risk with no compensation. Note that portfolio values, returns, and standard deviations exclude scam tokens.

5.6 Conclusion

In this chapter, we leverage the unique capabilities of the blockchain environment to examine whether investors falling victim to scams learn from their mistakes. The public nature of blockchain data allows us to detect cryptocurrency scams through smart contract audits and systematic analysis of transaction activity on DEXs. This provides us with granular data on multiple scams and scam victims, as well as their trading and portfolio characteristics, which are generally not available. Thus, this chapter can shed light on important questions in household finance by investigating the trading behavior and learning patterns of investors who have fallen victim to scams.

Reiterating our findings: (i) Victims become more active and have greater risk tolerance (higher standard deviation of portfolio returns and lower cash allocation) after experiencing scams; (ii) Victims tend to invest less in scams after experiencing one, and those who have incurred higher losses in scams are less likely to reinvest in scams subsequently; (iii) The experience of a scam and the scam investment ratio have a negative impact on subsequent non-scam investment returns; (iv) Victims increase their market exposure, their daily portfolio return decreases, and they become more (less) risk-tolerant in the short (long) term, but they do not achieve their pre-scam level of risk-adjusted performance; (v) Less experienced investors and those who interact with a larger number of liquidity pools, have higher trading volumes, achieve lower returns on their non-scam investments, have greater market exposure, and higher risk tolerance tend to learn less and are consequently more likely to invest in multiple scam pools.

Understanding how investors learn and adapt after falling victim to scams can inform the design of investor education and protection initiatives. For example, the tendency of scam victims to become more active after the scam suggests that tailored educational programs focusing on scam recognition and risk management immediately after falling victim could be beneficial. Furthermore, regulatory bodies could use these insights to develop more effective disclosure requirements and fraud detection mechanisms, particularly for emerging sectors like cryptocurrencies where scams are prevalent. By leveraging the detailed patterns of scam victims identified in our study, regulators can create more targeted strategies to enhance investor resilience and reduce the overall impact of financial fraud.

Our approach has several limitations. First, while blockchain data is rich in transaction details, it lacks personal demographic information that could provide deeper insights into the characteristics of scam victims. Second, as mentioned in Chapter 4, our identification of scams relies on detecting malicious code and suspicious trading patterns, which could be further validated with additional classification methods. Third, we use a probit model to assess the likelihood of reinvestment in scams for its simplicity, but it may not fully capture the complexity of individual learning processes. Future research could implement more complex models, such as the Cox proportional hazards model, to better account for the individual learning process. This model is particularly suited for analyzing time-to-event data and can offer more nuanced insights into the timing and likelihood of subsequent scam investments after an initial scam experience. By modeling the hazard rate of falling for another scam, we can better understand the dynamics of learning and behavioral change over time.

Appendix 5.1.

Table A5.1
Descriptive statistics of portfolio characteristics (trades and transfers)
This table reports the descriptive statistics (mean, median, standard deviation, and number of observations) of the portfolio characteristics (including transfers) for victims and non-victims in Panel A and for victims before and after falling victim to a scam in Panel B. The "Mean difference" column in the table reports the difference in mean characteristics of victims and non-victims in Panel A and victims before and after scam in Panel B. The standard errors used to compute the t-statistics are clustered by the number of active days with a trade. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. For a detailed definition of variables, refer to Table 5.1. The units of variables are in parentheses following the variable name.

	Mean	Median	Std. dev.	N obs.	Mean	Median	Std. dev.	N obs.	(1)-(2)	t-stat
Panel A.	**Victims (1)**				**Non-victims (2)**				**Mean difference**	
PValue ($\cdot 10^6$ USD)	0.146	0.012	0.598	199,650	0.077	0.007	0.403	491,617	0.069	3.180***
PTokens	15.055	9.755	16.040	199,650	6.900	5.141	6.090	491,617	8.155	3.037***
PReturn (%)	10.485	5.943	18.072	199,573	10.481	5.601	18.275	491,253	0.004	0.009
PStdDev (%)	34.183	22.937	38.426	199,573	33.343	21.793	38.885	491,253	0.840	0.716
PSharpe	0.090	0.091	0.142	199,284	0.083	0.081	0.161	490,758	0.006	2.859***
PCash (%)	8.099	0.499	15.812	199,650	11.500	0.430	20.747	491,617	-3.401	-41.419***
Panel B.	**Victim after scam (1)**				**Victim before scam (2)**				**Mean difference**	
PValue ($\cdot 10^6$ USD)	0.226	0.019	0.853	164,791	0.051	0.003	0.364	164,791	0.175	9.158***
PTokens	19.671	13.833	18.535	164,791	5.952	4.339	5.542	164,791	13.719	5.844***
PReturn (%)	6.338	2.501	19.417	164,791	28.944	11.226	71.216	164,791	-22.606	-9.870***
PStdDev (%)	26.273	15.352	38.976	164,791	72.120	38.686	96.282	164,791	-45.847	-16.356***
PSharpe	0.075	0.076	0.190	164,791	0.164	0.139	0.395	164,791	-0.089	-15.892***
PCash (%)	6.970	0.333	14.733	164,791	9.342	0.000	20.717	164,791	-2.372	-5.821***

Table A5.2
Impact of scams on portfolio value (trades only) This table reports the impact of scams on portfolio characteristics of victims using the following panel regression:

$$PValue_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the daily portfolio value (excluding transfers) $PValue_{i,t}$. In Panel A, $PastScam_{i,t-x}$ is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**0.026*** (13.59)**	**0.015*** (8.187)**	**0.007*** (3.593)**	**0.015*** (3.965)**
$Events_{i,t}$	0.001*** (9.002)	0.001*** (8.979)	0.001*** (9.012)	0.001*** (9.240)
$ActiveDays_{i,t}$	0.004*** (17.776)	0.004*** (17.687)	0.004*** (17.608)	0.004*** (15.952)
$Volume_{i,t}$	-0.420 (-0.670)	-0.417 (-0.666)	-0.415 (-0.663)	-0.411 (-0.655)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	4.7%	4.7%	4.7%	4.7%
N obs.	52,592,223			
N victims	202,469			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**0.049*** (9.586)**	**0.049*** (8.077)**	**0.049*** (7.051)**	**0.026*** (2.777)**
$Events_{i,t}$	0.001*** (9.069)	0.001*** (9.081)	0.001*** (9.087)	0.001*** (9.074)
$ActiveDays_{i,t}$	0.004*** (17.635)	0.004*** (17.624)	0.004*** (17.620)	0.004*** (17.627)
$Volume_{i,t}$	-0.411 (-0.657)	-0.410 (-0.654)	-0.409 (-0.653)	-0.414 (-0.660)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	4.7%	4.7%	4.7%	4.7%
N obs.	52,592,223			
N victims	202,469			

Table A5.3
Impact of scams on the number of tokens in the portfolio (trades only)
This table reports the impact of scams on portfolio characteristics of victims using the following panel regression:

$$PToken_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the daily number of tokens in the portfolio (excluding transfers) $PTokens_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/-total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$PastScam_{i,t-x}$	**0.368*** (18.931)**	**0.589*** (33.388)**	**0.708*** (40.097)**	**2.421*** (49.431)**
$Events_{i,t}$	0.025*** (19.772)	0.025*** (19.695)	0.025*** (19.658)	0.028*** (20.766)
$ActiveDays_{i,t}$	0.356*** (79.386)	0.356*** (79.679)	0.357*** (79.875)	0.330*** (68.028)
$Volume_{i,t}$	-2.617 (-1.03)	-2.663 (-1.036)	-2.659 (-1.032)	-1.978 (-0.830)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	77.2%	77.2%	77.3%	77.6%
N obs.	52,592,223			
N victims	202,469			
Panel B: $SIR_{i,t-x}$				
$SIR_{i,t-x}$	**0.011 (0.383)**	**0.036 (1.020)**	**0.028 (0.660)**	**0.624*** (6.173)**
$Events_{i,t}$	0.025*** (19.812)	0.025*** (19.812)	0.025*** (19.811)	0.025*** (19.800)
$ActiveDays_{i,t}$	0.355*** (79.300)	0.355*** (79.299)	0.355*** (79.294)	0.355*** (79.234)
$Volume_{i,t}$	-2.542 (-1.014)	-2.539 (-1.013)	-2.539 (-1.013)	-2.527 (-1.010)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	77.2%	77.2%	77.2%	77.2%
N obs.	52,592,223			
N victims	202,469			

Table A5.4
Impact of scams on daily portfolio returns (trades only)
This table reports the impact of scams on daily portfolio returns of victims using the following panel regression:

$$PReturn_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the daily portfolio returns (excluding transfers) $PReturn_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**-0.069****	**-1.322*****	**-1.935*****	**-6.512*****
	(-2.025)	**(-44.339)**	**(-66.147)**	**(-124.760)**
$Events_{i,t}$	-0.005***	-0.004***	-0.004***	-0.011***
	(-7.655)	(-6.971)	(-6.469)	(-15.370)
$ActiveDays_{i,t}$	-0.109***	-0.112***	-0.113***	-0.042***
	(-49.106)	(-51.068)	(-52.28)	(-16.749)
$Volume_{i,t}$	39.530**	40.695**	41.060**	36.447**
	(2.175)	(2.163)	(2.155)	(2.139)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.2%	0.2%	0.2%	0.3%
N obs.	51,191,120			
N victims	202,246			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**-0.157***	**-0.950*****	**-1.356*****	**-4.56*****
	(-1.925)	**(-10.423)**	**(-13.096)**	**(-19.772)**
$Events_{i,t}$	-0.005***	-0.005***	-0.005***	-0.005***
	(-7.673)	(-7.692)	(-7.711)	(-7.72)
$ActiveDays_{i,t}$	-0.108***	-0.108***	-0.108***	-0.108***
	(-49.104)	(-49.045)	(-48.974)	(-48.834)
$Volume_{i,t}$	39.448**	39.219**	39.018**	39.188**
	(2.174)	(2.163)	(2.155)	(2.174)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.2%	0.2%	0.2%	0.2%
N obs.	51,191,120			
N victims	202,246			

Table A5.5

Impact of scams on daily portfolio standard deviation (trades only)

This table reports the impact of scams on standard deviation of daily portfolio returns of victims using the following panel regression:

$$PStdDev_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the standard deviation of daily portfolio returns (excluding transfers) $PStdDev_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**4.975*** (61.779)**	**0.603*** (8.536)**	**-1.734*** (-25.125)**	**-12.737*** (-106.870)**
$Events_{i,t}$	-0.003** (-2.401)	-0.002* (-1.937)	-0.001 (-1.197)	-0.014*** (-10.095)
$ActiveDays_{i,t}$	-0.283*** (-59.644)	-0.293*** (-61.304)	-0.299*** (-63.020)	-0.165*** (-31.686)
$Volume_{i,t}$	105.664** (2.252)	108.987** (2.231)	110.788** (2.218)	103.854** (2.212)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	1.7%	1.6%	1.6%	2.1%
N obs.	51,314,392			
N victims	202,246			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**4.534*** (24.363)**	**1.906*** (9.20)**	**0.696*** (2.966)**	**-2.780*** (-5.463)**
$Events_{i,t}$	-0.002* (-1.771)	-0.002* (-1.752)	-0.002* (-1.762)	-0.002* (-1.794)
$ActiveDays_{i,t}$	-0.294*** (-61.878)	-0.295*** (-61.914)	-0.295*** (-61.888)	-0.294*** (-61.801)
$Volume_{i,t}$	110.19** (2.248)	109.989** (2.237)	109.720** (2.232)	109.335** (2.228)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	1.6%	1.6%	1.6%	1.6%
N obs.	51,314,392			
N victims	202,246			

Table A5.6
Impact of scams on daily portfolio Sharpe ratio (trades only)
This table reports the impact of scams on daily portfolio Sharpe ratio of victims using the following panel regression:

$$PSharpe_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable $PSharpe_{i,t}$ is the daily Sharpe ratio of portfolio (excluding transfers). $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**-0.022*****	**-0.032*****	**-0.027*****	**-0.039*****
	(-47.932)	**(-74.288)**	**(-65.558)**	**(-64.891)**
$Events_{i,t}$	0.0001***	0.0001***	0.0001***	0.0001***
	(18.86)	(19.652)	(19.741)	(13.863)
$ActiveDays_{i,t}$	-0.002***	-0.002***	-0.002***	-0.001***
	(-82.619)	(-82.421)	(-81.717)	(-68.894)
$Volume_{i,t}$	-0.068	-0.053	-0.062	-0.114*
	(-1.285)	(-1.113)	(-1.237)	(-1.673)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.1%	0.1%	0.1%	0.1%
N obs.	50,462,286			
N victims	201,843			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**-0.061*****	**-0.057*****	**-0.048*****	**-0.073*****
	(-51.506)	**(-43.773)**	**(-33.757)**	**(-27.945)**
$Events_{i,t}$	0.0001***	0.0001***	0.0001***	0.0001***
	(18.512)	(18.482)	(18.553)	(18.917)
$ActiveDays_{i,t}$	-0.002***	-0.002***	-0.002***	-0.002***
	(-81.636)	(-81.811)	(-82.151)	(-83.546)
$Volume_{i,t}$	-0.103*	-0.111*	-0.112*	-0.099
	(-1.666)	(-1.721)	(-1.717)	(-1.534)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.1%	0.1%	0.1%	0.1%
N obs.	50,462,286			
N victims	201,843			

Table A5.7
Impact of scams on daily portfolio cash allocation (trades only)
This table reports the impact of scams on daily portfolio cash allocations of victims using the following panel regression:

$$PCash_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the percentage of cash in the portfolio (excluding transfers) $PCash_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**-0.429*** (-12.13)**	**-0.078** (-2.044)**	**0.249*** (6.073)**	**1.861*** (23.537)**
$Events_{i,t}$	-0.017*** (-22.612)	-0.017*** (-22.645)	-0.017*** (-22.755)	-0.016*** (-20.941)
$ActiveDays_{i,t}$	0.144*** (43.893)	0.144*** (44.099)	0.145*** (44.293)	0.125*** (37.772)
$Volume_{i,t}$	8.057* (1.767)	7.986* (1.758)	7.929* (1.752)	8.404* (1.820)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	1.5%	1.5%	1.5%	1.7%
N obs.	52,592,223			
N victims	202,469			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**-2.158*** (-26.575)**	**-2.605*** (-24.693)**	**-3.004*** (-23.578)**	**-7.945*** (-26.029)**
$Events_{i,t}$	-0.017*** (-22.689)	-0.017*** (-22.709)	-0.017*** (-22.699)	-0.017*** (-22.495)
$ActiveDays_{i,t}$	0.145*** (44.218)	0.145*** (44.253)	0.145*** (44.237)	0.145*** (43.853)
$Volume_{i,t}$	7.836* (1.752)	7.713* (1.741)	7.611* (1.733)	7.77* (1.734)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	1.6%	1.6%	1.6%	1.7%
N obs.	52,592,223			
N victims	202,469			

Table A5.8
Impact of scams on portfolio value (trades and transfers)
This table reports the impact of scams on portfolio characteristics of victims using the following panel regression:

$$PValue_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the daily portfolio value (including transfers) $PValue_{i,t}$. In Panel A, $PastScam_{i,t-x}$ is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**0.051*** **	**0.046*** **	**0.037*** **	**0.058*** **
	(24.881)	**(23.183)**	**(17.561)**	**(15.409)**
$Events_{i,t}$	0.0001	0.0001	0.0001	0.0001
	(-1.204)	(-1.294)	(-1.267)	(-0.208)
$ActiveDays_{i,t}$	0.008***	0.008***	0.008***	0.007***
	(31.528)	(31.464)	(31.325)	(27.201)
$Volume_{i,t}$	0.421	0.421	0.428	0.457
	(0.62)	(0.621)	(0.627)	(0.654)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	4.3%	4.3%	4.3%	4.3%
N obs.	53,588,173			
N victims	199,650			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**0.093*** **	**0.103*** **	**0.109*** **	**0.183*** **
	(15.118)	**(13.236)**	**(11.636)**	**(10.832)**
$Events_{i,t}$	0.0001	0.0001	0.00011	0.0001
	(-1.076)	(-1.053)	(-1.038)	(-1.017)
$ActiveDays_{i,t}$	0.008***	0.008***	0.008***	0.008***
	(31.259)	(31.243)	(31.246)	(31.271)
$Volume_{i,t}$	0.449	0.454	0.457	0.452
	(0.645)	(0.65)	(0.653)	(0.648)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	4.3%	4.2%	4.3%	4.3%
N obs.	53,588,173			
N victims	199,650			

Table A5.9
Impact of scams on the number of tokens in the portfolio (trades and transfers)
This table reports the impact of scams on portfolio characteristics of victims using the following panel regression:

$$PToken_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the daily number of tokens in the portfolio (including transfers) $PTokens_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/-total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**0.556*****	**0.794*****	**0.935*****	**3.324*****
	(27.552)	**(42.100)**	**(48.116)**	**(61.744)**
$Events_{i,t}$	0.012***	0.012***	0.012***	0.016***
	(8.191)	(8.106)	(8.070)	(9.996)
$ActiveDays_{i,t}$	0.449***	0.450***	0.449***	0.410***
	(85.146)	(85.438)	(85.615)	(72.045)
$Volume_{i,t}$	-7.213	-7.373	-7.411	-6.334
	(-1.478)	(-1.470)	(-1.468)	(-1.455)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	76.5%	76.6%	76.6%	77.2%
N obs.	53,588,173			
N victims	199,650			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**0.296*****	**0.451*****	**0.575*****	**1.986*****
	(9.224)	**(11.418)**	**(11.899)**	**(17.635)**
$Events_{i,t}$	0.012***	0.012***	0.012***	0.013***
	(8.245)	(8.249)	(8.252)	(8.263)
$ActiveDays_{i,t}$	0.448***	0.448***	0.448***	0.448***
	(84.981)	(84.959)	(84.930)	(84.788)
$Volume_{i,t}$	-6.930	-6.905	-6.880	-6.875
	(-1.485)	(-1.486)	(-1.486)	(-1.486)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	76.5%	76.5%	76.5%	76.5%
N obs.	53,588,173			
N victims	199,650			

Table A5.10
Impact of scams on daily portfolio returns (trades and transfers)
This table reports the impact of scams on daily portfolio returns of victims using the following panel regression:

$$PReturn_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the daily portfolio returns (including transfers) $PReturn_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**-0.037**	**-1.616*** **	**-2.440*** **	**-7.947*** **
	(-0.931)	**(-47.088)**	**(-74.101)**	**(-131.701)**
$Events_{i,t}$	-0.003***	-0.002***	-0.002***	-0.010***
	(-3.764)	(-3.084)	(-2.587)	(-13.098)
$ActiveDays_{i,t}$	-0.155***	-0.159***	-0.159***	-0.067***
	(-59.273)	(-60.988)	(-61.926)	(-22.655)
$Volume_{i,t}$	102.005***	106.573***	108.573***	98.224***
	(3.891)	(3.874)	(3.864)	(3.834)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.3%	0.3%	0.3%	0.4%
N obs.	52,447,370			
N victims	199,573			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**1.034*** **	**0.094**	**-0.314*** **	**-2.812*** **
	(10.303)	**(0.867)**	**(-2.596)**	**(-11.723)**
$Events_{i,t}$	-0.003***	-0.003***	-0.003***	-0.003***
	(-3.766)	(-3.769)	(-3.783)	(-3.855)
$ActiveDays_{i,t}$	-0.155***	-0.155***	-0.155***	-0.155***
	(-59.369)	(-59.357)	(-59.321)	(-59.14)
$Volume_{i,t}$	102.024***	101.943***	101.773***	101.559***
	(3.893)	(3.892)	(3.892)	(3.892)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.3%	0.3%	0.3%	0.3%
N obs.	52,447,370			
N victims	199,573			

Table A5.11
Impact of scams on daily portfolio standard deviation (trades and transfers)
This table reports the impact of scams on standard deviation of daily portfolio returns of victims using the following panel regression:

$$PStdDev_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the standard deviation of daily portfolio returns (including transfers) $PStdDev_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**5.289***	**0.085**	**-2.802***	**-16.22***
	(56.446)	**(1.047)**	**(-35.94)**	**(-117.708)**
$Events_{i,t}$	0.001	0.002	0.003*	-0.014***
	(0.627)	(1.135)	(1.827)	(-8.551)
$ActiveDays_{i,t}$	-0.395***	-0.405***	-0.410***	-0.226***
	(-68.512)	(-70.118)	(-71.531)	(-35.961)
$Volume_{i,t}$	305.382***	318.418***	326.314***	311.132***
	(3.862)	(3.849)	(3.834)	(3.814)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	2.0%	1.9%	1.9%	2.5%
N obs.	52,529,815			
N victims	199,573			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**6.911***	**3.711***	**2.322***	**-1.894***
	(30.457)	**(14.95)**	**(8.401)**	**(-3.53)**
$Events_{i,t}$	0.002	0.002	0.002	0.002
	(1.175)	(1.201)	(1.195)	(1.125)
$ActiveDays_{i,t}$	-0.406***	-0.406***	-0.406***	-0.405***
	(-70.374)	(-70.421)	(-70.405)	(-70.273)
$Volume_{i,t}$	319.406***	319.855***	319.696***	318.425***
	(3.850)	(3.850)	(3.850)	(3.850)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	1.9%	1.9%	1.9%	1.9%
N obs.	52,529,815			
N victims	199,573			

Table A5.12
Impact of scams on daily portfolio Sharpe ratio (trades and transfers)
This table reports the impact of scams on daily portfolio Sharpe ratio of victims using the following panel regression:

$$PSharpe_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable $PSharpe_{i,t}$ is the daily Sharpe ratios of portfolio (including transfers). $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**-0.019***	**-0.031***	**-0.028***	**-0.034***
	(-41.657)	**(-73.028)**	**(-67.198)**	**(-58.034)**
$Events_{i,t}$	0.0001***	0.0001***	0.0001***	0.0001***
	(20.653)	(21.173)	(21.181)	(16.723)
$ActiveDays_{i,t}$	-0.002***	-0.002***	-0.002***	-0.001***
	(-73.877)	(-73.434)	(-72.574)	(-61.133)
$Volume_{i,t}$	-0.324**	-0.276**	-0.292**	-0.394**
	(-2.245)	(-2.144)	(-2.188)	(-2.419)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.1%	0.1%	0.1%	0.1%
N obs.	51,636,171			
N victims	199,284			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**-0.046***	**-0.045***	**-0.036***	**-0.043***
	(-39.388)	**(-34.876)**	**(-25.195)**	**(-17.268)**
$Events_{i,t}$	0.0001***	0.0001***	0.0001***	0.0001***
	(20.391)	(20.376)	(20.435)	(20.613)
$ActiveDays_{i,t}$	-0.002***	-0.002***	-0.002***	-0.002***
	(-73.199)	(-73.305)	(-73.52)	(-74.133)
$Volume_{i,t}$	-0.384**	-0.395**	-0.396**	-0.385**
	(-2.37)	(-2.39)	(-2.393)	(-2.369)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	0.1%	0.1%	0.1%	0.1%
N obs.	51,636,171			
N victims	199,284			

Table A5.13
Impact of scams on daily portfolio cash allocation (trades and transfers)
This table reports the impact of scams on daily portfolio cash allocations of victims using the following panel regression:

$$PCash_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the percentage of cash in the portfolio (including transfers) $PCash_{i,t}$. $PastScam_{i,t-x}$ in Panel A is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. In Panel B, we replace $PastScam_{i,t-x}$ with a continuous scam investment ratio $SIR_{i,t-x}$ (scam investment/total investment) of the victim i in the past x days in the above model. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = all$)
Panel A: $PastScam_{i,t-x}$				
$\mathbf{PastScam_{i,t-x}}$	**-0.773*** (-24.129)**	**-0.749*** (-22.147)**	**-0.720*** (-19.998)**	**-3.191*** (-44.605)**
$Events_{i,t}$	0.002*** (3.568)	0.002*** (3.756)	0.002*** (3.790)	-0.001** (-2.135)
$ActiveDays_{i,t}$	0.774*** (-33.227)	-0.078*** (-33.269)	-0.077*** (-33.126)	-0.040*** (-16.979)
$Volume_{i,t}$	10.717 (1.336)	10.744 (1.333)	10.698 (1.333)	9.754 (1.313)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	1.0%	1.1%	1.1%	1.4%
N obs.	53,588,173			
N victims	199,650			
Panel B: $SIR_{i,t-x}$				
$\mathbf{SIR_{i,t-x}}$	**-1.305*** (-16.634)**	**-1.603*** (-15.533)**	**-1.884*** (-14.950)**	**-5.38*** (-18.880)**
$Events_{i,t}$	0.002*** (3.346)	0.002*** (3.301)	0.002*** (3.265)	0.002*** (3.131)
$ActiveDays_{i,t}$	-0.076*** (-32.722)	-0.076*** (-32.653)	-0.076*** (-32.601)	-0.075*** (-32.484)
$Volume_{i,t}$	10.300 (1.336)	10.217 (1.335)	10.146 (1.335)	10.163 (1.335)
Individual fixed effect	Yes	Yes	Yes	Yes
R^2	1.0%	1.0%	1.0%	1.1%
N obs.	53,588,173			
N victims	199,650			

Table A5.14
Different definitions of scam date and scam amount
This table reports the impact of scams on portfolio characteristics of victims using the following panel regression:

$$Y_{i,t} = \alpha_i + \beta_1 \cdot PastScam_{i,t-x} + \sum_n \delta_n \cdot X_{i,t} + \varepsilon_{i,t}$$

where the dependent variable is the daily portfolio characteristics (excluding transfers), $PValue_{i,t}$, $PTokens_{i,t}$, $PReturn_{i,t}$, $PStdDev_{i,t}$, $PSharpe_{i,t}$, and $PCash_{i,t}$. $PastScam_{i,t-x}$ is a dummy variable that equals to 1 if the victim i has been scammed in the past $x = \{30, 60, 90, \text{all}\}$ days before a given day t. The scam date is the last interaction date as opposed to each interaction date in Chapter 5. In Panel B, we replace $PastScam_{i,t-x}$ with a total loss (in USD) of the victim i from the scams in the past x days. The table only reports the coefficient of $PastScam_{i,t-x}$ in Panel A ($Loss_{i,t-x}$ in Panel B) across different models. $X_{i,t}$ is a vector of controls that include the cumulative number of events ($Events_{i,t}$), the cumulative number of active days ($ActiveDays_{i,t}$), and the moving average trade volume ($Volume_{i,t}$). All versions of the model include individual fixed effects (α_i). The standard errors used to compute t-statistics (in brackets) are clustered by individuals. ***, **, and * denote significance at 1%, 5%, and 10%, respectively.

	Model 1 ($x = 30$)	Model 2 ($x = 60$)	Model 3 ($x = 90$)	Model 4 ($x = \text{all}$)
Panel A: $PastScam_{i,t-x}$				
$PValue_{i,t}$	0.025***	0.014***	0.006***	0.013***
	(13.163)	(7.749)	(3.101)	(3.325)
$PTokens_{i,t}$	0.335***	0.566***	0.686***	2.335***
	(17.461)	(32.608)	(39.447)	(48.135)
$PReturn_{i,t}$	-0.246***	-1.469***	-2.06***	-6.676***
	(-7.317)	(-49.722)	(-71.065)	(-128.943)
$PStdDev_{i,t}$	4.579***	0.228***	-2.076***	-13.27***
	(57.311)	(3.256)	(-30.376)	(-112.087)
$PSharpe_{i,t}$	-0.023***	-0.032***	-0.027***	-0.039***
	(-49.531)	(-75.789)	(-66.55)	(-65.578)
$PCash_{i,t}$	-0.374***	-0.022	0.300***	1.923***
	(-10.747)	(-0.589)	(7.388)	(24.611)
Panel B: $Loss_{i,t-x}$				
$PValue_{i,t}$	0.681***	0.517**	0.440*	1.866***
	(3.134)	(2.278)	(1.900)	(3.122)
$PTokens_{i,t}$	-1.293	-1.108	-1.144	-7.008***
	(-1.547)	(-1.526)	(-1.616)	(-4.889)
$PReturn_{i,t}$	2.915*	-0.900	-1.863	0.729
	(1.909)	(-0.684)	(-1.416)	(0.687)
$PStdDev_{i,t}$	26.240***	9.193***	3.337	5.533*
	(4.281)	(3.050)	(1.626)	(1.949)
$PSharpe_{i,t}$	-0.139***	-0.147***	-0.117***	-0.078***
	(-4.079)	(-3.280)	(-2.809)	(-2.968)
$PCash_{i,t}$	-1.477	0.949	3.008**	9.895***
	(-1.343)	(0.842)	(2.122)	(3.340)

Chapter 6

Conclusion

This dissertation investigates the economics of different forms of fraud and highlights their prevalence, determinants, and consequences. The dissertation includes four main chapters, each analyzing specific aspects of financial fraud. Table 6.1 reports the summary of the economic issues and corresponding key findings in each chapter.

Chapter 2 explores whether public distrust towards the finance sector is justified based on the true extent of financial adviser misconduct. Chapter 3 examines the real investment effects of market manipulation, demonstrating a detrimental impact of the prevalence of market manipulation on firms' investment-to-price sensitivity. Chapter 4 focuses on the cryptocurrency scams, providing an in-depth analysis of their prevalence and characteristics. By developing a detection algorithm, this chapter contributes to proactive measures to mitigate scams within the cryptocurrency space. Finally, Chapter 5 analyzes the aftermath of falling victim to scams, exploring how investors learn from their experiences and adapt their investment strategies. This chapter provides insight into learning from scams and implications for household finance.

Table 6.1. Summary of the thesis

This table reports the summary of the economic issues and corresponding main findings in this dissertation.

Chapter	Economic issue/ fraud type	Main findings
2	Financial adviser misconduct	• 30% of advisers involved in misconduct, yet only 9% of those cases detected by regulators. • \$6.9 trillion in AUM is managed by currently active advisers who we flag for potential misconduct. • One adviser with estimated (resp. unreported) misconduct, on average, costs the firm around 5 (resp. 7) clients and decreases the total AUM by around \$10 million (resp. \$13 million) per annum. • The frequency of adviser misconduct and the proportion of unreported misconduct increase significantly during the GFC. • An extensive list of characteristics associated with advisers' propensity for misconduct.
3	Market manipulation	• Market manipulation negatively impacts firms' investment-to-price sensitivity • Market manipulation leads to inferior future firm performance. • Market manipulation has a stronger negative effect on investment-to-price sensitivity in illiquid and more volatile stocks, and those with more private information. • The impact of market manipulation on investment-to-price sensitivity is monotonic. At the top quintile manipulation level the negative effect is four times more detrimental than that in the bottom quintile.
4	Cryptocurrency scams	• 44% of all liquidity pools in major DEXs are scams (rug pulls). • Rug pulls have stolen around \$1.5 billion worth of tokens from investors between May 2020 and April 2022. • A set of characteristics that can differentiate scam pools from non-scam pools. • An index of the probability of scams that can identify cryptocurrency scams before they occur.
5	Post-scam victim behaviour	• After experiencing a scam, investors are less likely to invest in scams. • Falling victim to a scam reduces investors' subsequent trading performance. • After experiencing a scam, investors become more risk-tolerant in the short term (if the investor has been scammed in the past 30 and 60 days), but less risk tolerant in the long term (if the investor has been scammed in the past 90 and more days). • After experiencing a scam, investors do not achieve their pre-scam level of risk-adjusted performance. • Factors impacting varying learning rates among investors.

Chapter 2 explores the true extent of financial adviser misconduct to validate the public's growing distrust of the finance sector. With a comprehensive dataset covering over one million financial advisers (both investment advisers and brokers), we estimate approximately 30% financial advisers in the U.S. have been involved in misconduct. Shockingly, only about one-third of these cases are identified by regulatory bodies. We also estimate that, as of 2020, approximately $6.9 trillion in assets under management is overseen by 138,288 investment advisers flagged for potential misconduct, and 165,976 brokers engaging in potential misconduct while executing transactions for households.

We observe a significant increase in the frequency of adviser misconduct and the proportion of unreported misconduct during the Global Financial Crisis (GFC), aligning with the erosion of public trust in the finance sector during that period (as measured by the Financial Trust Index (FTI) of Sapienza and Zingales). We also link the adviser misconduct to trust by investigating the relation between the number of clients and AUM of advisory firms and their (estimated and unreported) number of adviser misconduct. The estimated and unreported (but not reported) numbers of adviser misconduct have a negative impact on both the number of clients and AUM of advisory firms. One adviser with estimated misconduct, on average, costs the firm around 5 clients and decreases the total AUM by around $10 million. Similarly, one adviser with unreported misconduct, on average, costs the firm around 7 clients and decreases the total AUM by around $13 million.

By zooming in on the dark side of financial advice, Chapter 2 aims to promote good finance and minimize the bad. The extensive list of characteristics associated with advisers' propensity for misconduct can help consumers use advisers with characteristics that do not signal potential misconduct. In addition, regulators and firms can use this knowledge to streamline surveillance activities and work towards a more accountable and trustworthy financial advisory sector.

Chapter 3 explores the real effect of market manipulation. Unlike existing studies that primarily focuses on the direct effects of market manipulation (i.e., distorting secondary market pricing, reduction in market liquidity), Chapter 3 investigates whether manipulated stock prices can decrease the use of price signals by the firm managers in a way that harms real investment decisions and the allocation of capital.

Chapter 3 shows that the noise in stock prices resulting from stock price manipulation diminishes managers' incentive to use stock prices as an additional source of information, leading to a decrease in investment-to-price sensitivity. Chapter 3 also documents that stock price manipulation significantly harms the future operating performance of firms. These findings are intuitive because firm managers relying less on stock prices, tend to make less informed investment decisions and inefficiently allocate capital.

Overall, Chapter 3 provides insights into the real economic consequences of market manipulation, highlighting its detrimental effects on investment efficiency and firm performance. This study contributes to both empirical studies on market manipulation and the broader literature examining how stock prices influence corporate investment and managerial learning. Chapter 3 emphasizes the need to better understand the complex dynamics between manipulation and resource allocation in financial markets by demonstrating how market manipulation directly harms firms.

In Chapter 4, we explore the dark side of cryptocurrencies, specifically focusing on the prevalence of cryptocurrency scams known as "rug pulls" and differentiating between hard and soft variations. Hard rug pulls involve malicious code in smart contracts, disabling investor transactions or employing hidden functionalities detrimental to investors. Soft rug pulls have no malicious code, but ultimately also result in liquidity theft by founders that leaves investors with worthless tokens. By leveraging smart contract audits and developing soft rug pull detection algorithm, Chapter 4 quantifies the extent of rug pulls in major DEXs.

We find that a significant portion of ERC-20 liquidity pools in major DEXs, approximately 44%, are rug pulls. Based on external smart contract audits, we find that approximately 3% of the liquidity pools on major DEXs are hard rug pulls. Then we develop a simple soft rug pull detection algorithm based on three heuristic patterns: price, activity, and profit. Our algorithm identifies around 41% of liquidity pools as soft rug pulls. We estimate that soft and hard rug pulls together amounting to an estimated theft of around $1.5 billion worth of tokens between May 2020 and April 2022. These findings emphasize the importance of understanding scams in DEXs, helping policymakers and regulators make informed decisions about blockchain's impacts and advantages.

We also build a set of characteristics that can differentiate scam pools from non-scam pools. Anticipating and preventing scams necessitates the implementation of proactive regulations or guidelines, going beyond relying solely on post-event detection methods. In Chapter 4, we focus on identifying ex-ante characteristics that can act as early indicators of potential scams. We build the scam index by using initial pool supply, LP lock, creator token count, creator pool count, time gap, and imitation variables. The scam index can serve as a proactive tool for identifying potential scam pools, with potential applications in regulation, investor decision-making, academic research, and practical applications.

Investment scams also present a critical challenge within household finance, often resulting in substantial financial losses for victims. Despite the magnitude of these losses, the impact of scams on victims is less understood because many victims often underreport. Chapter 5 focuses on exploring the aftermath of falling victim to scams, providing insight into how individuals make subsequent investment decisions and construct their portfolios after experiencing scams.

In Chapter 5, we examine the impact of a scam on investors' trading behavior. We find that victims tend to invest less in scams after experiencing a scam. Experiencing a scam in the last 30 days decreases the probability of reinvesting in scams by about 4%. Investors who experienced significant losses in scams are even less likely to reinvest in scams. Furthermore, experiencing a scam negatively impacts subsequent non-scam investment returns, decreasing them by approximately 14%.

We further extend our tests on the impact of a scam on investors' portfolio characteristics. Chapter 5 shows that victims of scams increase their market exposure by raising daily portfolio values and the number of tokens held. However, their daily portfolio return decreases after a scam. The impact of scams on the daily portfolio standard deviation varies over time. Specifically, investors become more risk-tolerant in the short term (if the investor has been scammed in the past 30 and 60 days), and their risk tolerance decreases in the long term (if the investor has been scammed in the past 90 or more days). Victims never fully regain their pre-scam level of risk-adjusted performance, as indicated by the Sharpe ratio.

Overall, this dissertation contributes to the broader literature on the economics of fraud, focusing on financial adviser misconduct, the real economic consequences of market manipulation, the prevalence and determinants of cryptocurrency scams, and the trading behavior of victims of scams. Each study zooms in on different aspects of fraud in a particular domain; we first link adviser misconduct to the (dis)trust in the financial ecosystem; second, we document the real economic consequences of market manipulation; then, we quantify and characterize scams using new blockchain technology; lastly, we analyze scam victim behavior, a crucial yet understudied aspect in finance and economics literature. All of these are important for building a resilient financial ecosystem that safeguards investors and encourages their active participation. Further research and policy efforts in this domain are vital to address the challenges posed by financial frauds.

Bibliography

Adams, H., Zinsmeister, N., Salem, M., Keefer, R. and Robinson, D. (2021), 'Uniswap v3 core', *White Paper* .

Agarwal, S., Chomsisengphet, S., Mahoney, N. and Stroebel, J. (2015), 'Regulating consumer financial products: Evidence from credit cards', *Quarterly Journal of Economics* **130**(1), 111–164.

Agarwal, S., Driscoll, J. C., Gabaix, X. and Laibson, D. (2009), 'The age of reason: Financial decisions over the life cycle and implications for regulation', *Brookings Papers on Economic Activity* **2**, 51–117.

Aggarwal, R. K. and Wu, G. (2006), 'Stock market manipulations', *Journal of Business* **79**(4), 1915–1953.

Aliyev, N., Allahverdiyeva, I. and Putnins, T. J. (2023a), 'Is public distrust of the finance sector warranted? evidence from the extent of financial adviser misconduct', *Working Paper* .

Aliyev, N., Allahverdiyeva, I. and Putnins, T. J. (2023b), 'The real effects of market manipulation', *Working Paper* .

Aliyev, N., Allahverdiyeva, I. and Putnins, T. J. (2023c), 'Scam alert: Can cryptocurrency scams be detected early?', *Working Paper* .

Aliyev, N., Allahverdiyeva, I. and Putnins, T. J. (2023d), 'Once bitten twice shy: Learning about scams', *Working Paper* .

Allen, F. and Gale, D. (1992), 'Stock-price manipulation', *Review of Financial Studies* **5**(3), 503–529.

Allen, F. and Gorton, G. (1992*a*), 'Stock price manipulation, market microstructure and asymmetric information', *European Economic Review* **36**(2-3), 624–630.

Allen, F. and Gorton, G. B. (1992*b*), 'Stock price manipulation, market microstructure and asymmetric information', *European Economic Review* **36**, 624–630.

Allen, F., Litov, L. and Mei, J. (2006), 'Large investors, price manipulation, and limits to arbitrage: An anatomy of market corners', *Review of Finance* **10**(4), 645–693.

Aloosh, A. and Li, J. (2019), 'Direct evidence of bitcoin wash trading', *Working Paper* .

Andreoni, J. (1992), 'IRS as loan shark: Tax compliance with borrowing constraints', *Journal of Public Economics* **49**(1), 35–46.

Arrow, K. J. (1962), 'The economic implications of learning by doing', *Review of Economic Studies* **29**(3), 155–173.

Bailey, W., Li, H., Mao, C. X. and Zhong, R. (2003), 'Regulation fair disclosure and earnings information: Market, analyst, and corporate responses', *Journal of Finance* **58**(6), 2487–2514.

Baker, M., Stein, J. C. and Wurgler, J. (2003), 'When does the market matter? stock prices and the investment of equity-dependent firms', *Quarterly Journal of Economics* **118**(3), 969–1005.

Barbon, A. and Ranaldo, A. (2023), 'On the quality of cryptocurrency markets: Centralized versus decentralized exchanges', *Working Paper* .

Barro, R. J. (1990), 'The stock market and investment', *Review of Financial Studies* **3**(1), 115–131.

Becker, G. S. (1968), 'Crime and punishment: An economic approach', *Journal of Political Economy* **76**(2), 169–217.

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

Bergstresser, D., Chalmers, J. M. and Tufano, P. (2008), 'Assessing the costs and benefits of brokers in the mutual fund industry', *Review of Financial Studies* **22**(10), 4129–4156.

Bhattacharya, U. and Daouk, H. (2002), 'The world price of insider trading', *Journal of Finance* **57**(1), 75–108.

Biais, B., Bisiere, C., Bouvard, M., Casamatta, C. and Menkveld, A. J. (2023), 'Equilibrium bitcoin pricing', *Journal of Finance* **78**(2), 967–1014.

Bond, P., Edmans, A. and Goldstein, I. (2012), 'The real effects of financial markets', *Annual Review of Financial Economics* **4**(1), 339–360.

Brunnermeier, M. K. and Oehmke, M. (2013), 'Bubbles, financial crises, and systemic risk', *Handbook of the Economics of Finance* **2**, 1221–1288.

Budish, E. B. (2022), 'The economic limits of bitcoin and anonymous, decentralized trust on the blockchain', *Working Paper* .

Burke, J. and Hung, A. A. (2021), 'Trust and financial advice', *Journal of Pension Economics & Finance* **20**(1), 9–26.

Campbell, J. Y. (2006), 'Household finance', *Journal of Finance* **61**(4), 1553–1604.

Campbell, J. Y., Jackson, H. E., Madrian, B. C. and Tufano, P. (2011), 'Consumer financial protection', *Journal of Economic Perspectives* **25**(1), 91–114.

Chainalysis (2022), 'Crypto crime report', *Report* .

Chakraborty, A. and Yılmaz, B. (2004), 'Informed manipulation', *Journal of Economic Theory* **114**(1), 132–152.

Chalmers, J. and Reuter, J. (2020), 'Is conflicted investment advice better than no advice?', *Journal of Financial Economics* **138**(2), 366–387.

Charoenwong, B., Kwan, A. and Umar, T. (2019), 'Does regulatory jurisdiction affect the quality of investment-adviser regulation?', *American Economic Review* **109**(10), 3681–3712.

Chen, Q., Goldstein, I. and Jiang, W. (2007), 'Price informativeness and investment sensitivity to stock price', *Review of Financial Studies* **20**(3), 619–650.

Chiu, J. and Koeppl, T. V. (2017), 'The economics of cryptocurrencies–bitcoin and beyond', *Working Paper* .

Chiu, J. and Koeppl, T. V. (2019), 'Blockchain-based settlement for asset trading', *Review of Financial Studies* **32**(5), 1716–1753.

Choi, J. J., Laibson, D., Madrian, B. C. and Metrick, A. (2009), 'Reinforcement learning and savings behavior', *Journal of Finance* **64**(6), 2515–2534.

Chordia, T., Roll, R. and Subrahmanyam, A. (2001), 'Market liquidity and trading activity', *Journal of Finance* **56**(2), 501–530.

Comerton-Forde, C. and Putnins, T. J. (2011a), 'Pricing accuracy, liquidity and trader behavior with closing price manipulation', *Experimental Economics* **14**, 110–131.

Comerton-Forde, C. and Putnins, T. J. (2011b), 'Measuring closing price manipulation', *Journal of Financial Intermediation* **20**(2), 135–158.

Comerton-Forde, C. and Putnins, T. J. (2014), 'Stock price manipulation: Prevalence and determinants', *Review of Finance* **18**(1), 23–66.

Cong, L. W., Grauer, K., Rabetti, D. and Updegrave, H. (2023), 'Blockchain forensics and crypto-related cybercrimes', *Working Paper* .

Cong, L. W., Harvey, C. R., Rabetti, D. and Wu, Z.-Y. (2023), 'An anatomy of crypto-enabled cybercrimes', *Working Paper* .

Cong, L. W., Li, X., Tang, K. and Yang, Y. (2021), 'Crypto wash trading', *Working Paper* .

Cumming, D. J., Li, Y. and Yu, Y. (2023), 'Stock market manipulation and corporate venture capital investments', *Working Paper* .

DeMarzo, P. M., Kaniel, R. and Kremer, I. (2004), 'Diversification as a public good: Community effects in portfolio choice', *Journal of Finance* **59**(4), 1677–1716.

Dessaint, O., Foucault, T., Fresard, L. and Matray, A. (2019), 'Noisy stock prices and corporate investment', *Review of Financial Studies* **32**(7), 2625–2672.

Dhawan, A. and Putnins, T. J. (2021), 'A new wolf in town? pump-and-dump manipulation in cryptocurrency markets', *Review of Finance (forthcoming)* .

Dimmock, S. G., Gerken, W. C. and Graham, N. P. (2018), 'Is fraud contagious? Coworker influence on misconduct by financial advisors', *Journal of Finance* **73**(3), 1417–1450.

Dimmock, S. G., Gerken, W. C. and Van Alfen, T. (2021), 'Real estate shocks and financial advisor misconduct', *Journal of Finance* **76**(6), 3309–3346.

Dow, J. and Gorton, G. (1997), 'Stock market efficiency and economic efficiency: Is there a connection?', *Journal of Finance* **52**(3), 1087–1129.

Dupuis, D. and Gleason, K. (2020), 'Money laundering with cryptocurrency: open doors and the regulatory dialectic', *Journal of Financial Crime* .

Durnev, A. (2010), 'The real effects of political uncertainty: Elections and investment sensitivity to stock prices', *Working Paper* .

Dyck, A., Morse, A. and Zingales, L. (2010), 'Who blows the whistle on corporate fraud?', *Journal of Finance* **65**(6), 2213–2253.

Easley, D., Hvidkjaer, S. and O'Hara, M. (2002), 'Is information risk a determinant of asset returns?', *Journal of Finance* **57**(5), 2185–2221.

Easley, D., López de Prado, M. M. and O'Hara, M. (2012), 'Flow toxicity and liquidity in a high-frequency world', *Review of Financial Studies* **25**(5), 1457–1493.

Easley, D., O'Hara, M. and Basu, S. (2019), 'From mining to markets: The evolution of bitcoin transaction fees', *Journal of Financial Economics* **134**(1), 91–109.

Edelman Trust Barometer (2022), '2022 Edelman Trust Barometer', *https://www.edelman.com/trust/2022-trust-barometer* .

Edmans, A., Jayaraman, S. and Schneemeier, J. (2017), 'The source of information in prices and investment-price sensitivity', *Journal of Financial Economics* **126**(1), 74–96.

Egan, M., Matvos, G. and Seru, A. (2019), 'The market for financial adviser misconduct', *Journal of Political Economy* **127**(1), 233–295.

Egan, M., Matvos, G. and Seru, A. (2022), 'When Harry fired Sally: The double standard in punishing misconduct', *Journal of Political Economy* **130**(5).

Ehrlich, I. (1973), 'Participation in illegitimate activities: A theoretical and empirical investigation', *Journal of Political Economy* **81**(3), 521–565.

Fama, E. F. and French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Feinstein, J. S. (1987), 'Detection controlled estimation: theory and applications', *Ph.D. dissertation, Massachusetts Institute of Technology* .

Feinstein, J. S. (1990), 'Detection controlled estimation', *Journal of Law and Economics* **33**(1), 233–276.

Feinstein, J. S. (1991), 'An econometric analysis of income tax evasion and its detection', *RAND Journal of Economics* pp. 14–35.

Feng, L. and Seasholes, M. S. (2005), 'Do investor sophistication and trading experience eliminate behavioral biases in financial markets?', *Review of Finance* **9**(3), 305–351.

Foley, S., Karlsen, J. R. and Putnins, T. J. (2019), 'Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?', *Review of Financial Studies* **32**(5), 1798–1853.

Foley, S., O'Neill, P. and Putnins, T. J. (2023), 'A better market design? Applying 'automated market makers' to traditional financial markets', *Working Paper* .

Foucault, T. and Fresard, L. (2012), 'Cross-listing, investment sensitivity to stock price, and the learning hypothesis', *Review of Financial Studies* **25**(11), 3305–3350.

Foucault, T. and Fresard, L. (2014), 'Learning from peers' stock prices and corporate investment', *Journal of Financial Economics* **111**(3), 554–577.

Foucault, T. and Gehrig, T. (2008), 'Stock price informativeness, cross-listings, and investment decisions', *Journal of Financial Economics* **88**(1), 146–168.

Garleanu, N. and Pedersen, L. H. (2018), 'Efficiently inefficient markets for assets and asset management', *Journal of Finance* **73**(4), 1663–1712.

Gennaioli, N., Shleifer, A. and Vishny, R. (2015), 'Money doctors', *Journal of Finance* **70**(1), 91–114.

Georgarakos, D. and Inderst, R. (2014), 'Financial advice and stock market participation', *Working Paper* .

Giannetti, M. and Wang, T. Y. (2016), 'Corporate scandals and household stock market participation', *Journal of Finance* **71**(6), 2591–2636.

Glosten, L. R. and Milgrom, P. R. (1985), 'Bid, ask and transaction prices in a specialist market with heterogeneously informed traders', *Journal of Financial Economics* **14**(1), 71–100.

Goldstein, I. and Guembel, A. (2008), 'Manipulation and the allocational role of prices', *Review of Economic Studies* **75**(1), 133–164.

Gomes, F., Haliassos, M. and Ramadorai, T. (2021), 'Household finance', *Journal of Economic Literature* **59**(3), 919–1000.

Greene, W. H. (2008), *Econometric analysis*, Pearson: Prentice Hall [6th edition].

Griffin, J. M. and Maturana, G. (2016), 'Who facilitated misreporting in securitized loans?', *Review of Financial Studies* **29**(2), 384–419.

Griffin, J. M. and Shams, A. (2020), 'Is bitcoin really untethered?', *Journal of Finance* **75**(4), 1913–1964.

Grossman, S. J., Kihlstrom, R. E. and Mirman, L. J. (1977), 'A bayesian approach to the production of information and learning by doing', *Review of Economic Studies* **44**(3), 533–547.

Guiso, L. (2010), 'A trist-driven financial crisis. implications for the future of financial markets', *Working Paper* .

Guiso, L., Sapienza, P. and Zingales, L. (2004), 'Does local financial development matter?', *Quarterly Journal of Economics* **119**(3), 929–969.

Guiso, L., Sapienza, P. and Zingales, L. (2008), 'Trusting the stock market', *Journal of Finance* **63**(6), 2557–2600.

Gurun, U. G., Stoffman, N. and Yonker, S. E. (2018), 'Trust busting: The effect of fraud on investor behavior', *Review of Financial Studies* **31**(4), 1341–1376.

Hackethal, A., Haliassos, M. and Jappelli, T. (2012), 'Financial advisors: A case of babysitters?', *Journal of Banking & Finance* **36**(2), 509–524.

Hackethal, A., Inderst, R. and Meyer, S. (2012), 'Trading on advice', *Working Paper* .

Hasbrouck, J., Rivera, T. J. and Saleh, F. (2022), 'The need for fees at a dex: How increases in fees can increase dex trading volume', *Working Paper* .

Hasbrouck, J. and Saar, G. (2013), 'Low-latency trading', *Journal of Financial Markets* **16**(4), 646–679.

Howell, S. T., Niessner, M. and Yermack, D. (2020), 'Initial coin offerings: Financing growth with cryptocurrency token sales', *Review of Financial Studies* **33**(9), 3925–3974.

Hu, A. S., Parlour, C. A. and Rajan, U. (2019), 'Cryptocurrencies: Stylized facts on a new investible instrument', *Financial Management* **48**(4), 1049–1068.

Huberman, G., Leshno, J. D. and Moallemi, C. (2021), 'Monopoly without a monopolist: An economic analysis of the bitcoin payment system', *Review of Economic Studies* **88**(6), 3011–3040.

Hung, A. A., Clancy, N., Dominitz, J., Talley, E. and Berrebi, C. (2008), *Investor and industry perspectives on investment advisers and broker-dealers*, Vol. 767, Rand Corporation.

Inderst, R. and Ottaviani, M. (2009), 'Misselling through agents', *American Economic Review* **99**(3), 883–908.

Inderst, R. and Ottaviani, M. (2010), 'Consumer protection in markets with advice', *Competition Policy International* **6**(1), 47–64.

Inderst, R. and Ottaviani, M. (2012*a*), 'Competition through commissions and kickbacks', *American Economic Review* **102**(2), 780–809.

Inderst, R. and Ottaviani, M. (2012*b*), 'Financial advice', *Journal of Economic Literature* **50**(2), 494–512.

Kaplan, S. N. and Zingales, L. (1997), 'Do investment-cash flow sensitivities provide useful measures of financing constraints?', *Quarterly Journal of Economics* **112**(1), 169–215.

Knüpfer, S., Rantala, V., Vihriälä, E. and Vokata, P. (2023), 'Household responses to phantom riches', *Working Paper* .

Kostovetsky, L. (2016), 'Whom do you trust?: Investor-advisor relationships and mutual fund flows', *Review of Financial Studies* **29**(4), 898–936.

Kowaleski, Z. T., Sutherland, A. G. and Vetter, F. W. (2020), 'Can ethics be taught? Evidence from securities exams and investment adviser misconduct', *Journal of Financial Economics* **138**(1), 159–175.

Kyle, A. S. (1985), 'Continuous auctions and insider trading', *Econometrica: Journal of the Econometric Society* **53**(6), 1315–1335.

Lehar, A. and Parlour, C. A. (2022), 'Decentralized exchanges', *Working Paper* .

Lehar, A., Parlour, C. and Zoican, M. (2023), 'Liquidity fragmentation on decentralized exchanges', *Working Paper* .

Li, T., Shin, D. and Wang, B. (2021), 'Cryptocurrency pump-and-dump schemes', *Working Paper* .

Linnainmaa, J. T. (2010), 'Do limit orders alter inferences about investor performance and behavior?', *Journal of Finance* **65**(4), 1473–1506.

List, J. A. (2003), 'Does market experience eliminate market anomalies?', *Quarterly Journal of Economics* **118**(1), 41–71.

Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M. and Savage, S. (2013), 'A fistful of bitcoins: characterizing payments among men with no names', *Proceedings of the 13th ACM Internet Measurement Conference* pp. 127–140.

Merrick, J. J., Naik, N. Y. and Yadav, P. K. (2005), 'Strategic trading behavior and price distortion in a manipulated market: anatomy of a squeeze', *Journal of Financial Economics* **77**(1), 171–218.

Meulbroek, L. K. (1992), 'An empirical analysis of illegal insider trading', *Journal of Finance* **47**(5), 1661–1699.

Morck, R., Shleifer, A., Vishny, R. W., Shapiro, M. and Poterba, J. M. (1990), 'The stock market and investment: is the market a sideshow?', *Brookings Papers on Economic Activity* **1990**(2), 157–215.

Mullainathan, S., Noeth, M. and Schoar, A. (2012), 'The market for financial advice: An audit study', *Working Paper* .

Nakamoto, S. (2008), 'Bitcoin: A peer-to-peer electronic cash system', *White Paper* .

NASD (2006), 'Investor fraud study final report', *Report* .

Nicolosi, G., Peng, L. and Zhu, N. (2009), 'Do individual investors learn from their trading experience?', *Journal of Financial Markets* **12**(2), 317–336.

Othman, A. (2012), 'Automated market making: Theory and practice', *Ph.D. dissertation, Carnegie Mellon University* .

Pástor, L., Taylor, L. A. and Veronesi, P. (2009), 'Entrepreneurial learning, the ipo decision, and the post-ipo drop in firm profitability', *Review of Financial Studies* **22**(8), 3005–3046.

Peters, R. H. and Taylor, L. A. (2017), 'Intangible capital and the investment-q relation', *Journal of Financial Economics* **123**(2), 251–272.

Petersen, M. A. (2009), 'Estimating standard errors in finance panel data sets: Comparing approaches', *Review of Financial Studies* **22**(1), 435–480.

Phua, K., Sang, B., Wei, C. and Yu, G. Y. (2022), 'Trust, but verify: The economics of scams in initial coin offerings', *Working Paper* .

Pirrong, S. C. (1995), 'The self-regulation of commodity exchanges: the case of market manipulation', *Journal of Law and Economics* **38**(1), 141–206.

Piskorski, T., Seru, A. and Witkin, J. (2015), 'Misrepresentation of asset quality and financial intermediation: Evidence from rmbs market', *Journal of Finance* **70**(6), 2635–2678.

Poirier, D. J. (1980), 'Partial observability in bivariate probit models', *Journal of Econometrics* **12**(2), 209–217.

Putnins, T. J. (2012), 'Market manipulation: A survey', *Journal of Economic Surveys* **26**(5), 952–967.

Qureshi, H. and Sokobin, J. S. (2015), 'Do investors have valuable information about brokers?', *Working Paper* .

Rantala, V. (2019), 'How do investment ideas spread through social interaction? evidence from a ponzi scheme', *Journal of Finance* **74**(5), 2349–2389.

Rosu, I. and Saleh, F. (2021), 'Evolution of shares in a proof-of-stake cryptocurrency', *Management Science* **67**(2), 661–672.

Sah, R. K. (1991), 'Social osmosis and patterns of crime', *Journal of Political Economy* **99**(6), 1272–1295.

Saleh, F. (2021), 'Blockchain without waste: Proof-of-stake', *Review of Financial Studies* **34**(3), 1156–1190.

Sapienza, P. and Zingales, L. (2012), 'A trust crisis', *International Review of Finance* **12**(2), 123–131.

Sapienza, P. and Zingales, L. (2013), 'Economic experts versus average americans', *American Economic Review* **103**(3), 636–42.

Seru, A., Shumway, T. and Stoffman, N. (2010), 'Learning by trading', *Review of Financial Studies* **23**(2), 705–739.

Sokolov, K. (2021), 'Ransomware activity and blockchain congestion', *Journal of Financial Economics* **141**(2), 771–782.

Stein, R. M. (2005), 'The relationship between default prediction and lending profits: Integrating roc analysis and loan pricing', *Journal of Banking & Finance* **29**(5), 1213–1236.

Strahilevitz, M. A., Odean, T. and Barber, B. M. (2011), 'Once burned, twice shy: How naive learning, counterfactuals, and regret affect the repurchase of stocks previously sold', *Journal of Marketing Research* **48**, S102–S120.

Subrahmanyam, A. and Titman, S. (1999), 'The going-public decision and the development of financial markets', *Journal of Finance* **54**(3), 1045–1082.

Wang, T. Y., Winton, A. and Yu, X. (2010), 'Corporate fraud and business conditions: Evidence from IPOs', *Journal of Finance* **65**(6), 2255–2292.

Xia, P., Wang, H., Gao, B., Su, W., Yu, Z., Luo, X., Zhang, C., Xiao, X. and Xu, G. (2021), 'Trade or trick? detecting and characterizing scam tokens on uniswap decentralized exchange', *Proceedings of the ACM on Measurement and Analysis of Computing Systems* **5**(3), 1–26.

Yermack, D. (2017), 'Corporate governance and blockchains', *Review of Finance* **21**(1), 7–31.

Zingales, L. (2015), 'Presidential address: Does finance benefit society?', *Journal of Finance* **70**(4), 1327–1363.