PhD-FDEF-2024-009
The Faculty of Law, Economics and Finance

DISSERTATION

Defense held on 07/03/2024 in Luxembourg
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG
EN Finance

by

Patricia PAIVA GOMES CRUZ
Born on 22 May, 1975 in Rio de Janeiro, Brazil

ESSAYS IN BLOCKHOLDER ACTIVISM

Dissertation defence committee

Dr Ulf Von Lilienfeld, dissertation supervisor
Professor, Université du Luxembourg

Dr Michael Halling, Chairman
Professor, Université du Luxembourg

Dr Roman Kräussl
Professor, Bayes Business School

Dr Christos Koulovatianos
Professor, Université du Luxembourg

Dr Jan Schnitzler
Professor, Grenoble École de Management

I hereby declare that this thesis and the work reported herein was composed by and originated entirely from me. Information derived from the published and unpublished work of others has been acknowledged in the text and references are given in the list of sources.

Patricia Paiva Gomes Cruz (2024)

Introduction

This thesis contains three chapters, each consisting of an essay related to Blockholder Activism.

Within the realm of research, particularly in studies on Governance, Shareholder Activism has garnered considerable attention both in both Law as well as in Finance Scholarship. However, in the last decade, the emergence of big data, open-source software, sustainable finance, and research ethics, including reproducibility, has opened new avenues for research, often linked to aspects previously explored but now framed within a fresh perspective. A case in point is the revival of the somewhat old debate about the purpose of companies (Zingales et al. (2020), Rajan et al. (2022)). It is in the light of these changes that I write this thesis. In the following chapters, I explore the Activist Blockholders setup in three quite different studies: from causal inference of ownership stakes (Chapter 1), to abnormal returns related to objectives linked to sustainability using natural language processing (NLP) (Chapter 2), to detection of problems on datasets currently used for research on this topic, presenting corresponding solutions for obtaining free quality data that leads to more reliable results and reproducibility (Chapter 3).

Governance scholarship has identified 3 main elements in the corporate governance framework: Board of Directors (BoD), shareholders' participation (including proposals, voting and private engagement with management and BoD) and executive compensation. This thesis focuses on shareholders' participation.

First, I examine whether the initial stakes of activist blockholders change as function of market movements. Empirical research has found that activist investors can discipline management and reorient company strategy both directly (Brav et al. (2010), 2018, Barry et al. (2020), Lilienfeld-Toal and Schnitzler (2020)) and indirectly, through spillovers (Gantchev (2013)).

Those findings are based on clear improvement of financial and operational metrics after a company is targeted and, importantly, those positive results are persistent on the long run.

Despite those findings, criticism against blockholder activists remains. Public figures, from Larry Fink to Hillary Clinton, blame activists to pressure companies towards "managerial myopia" to maximize short term profits in detriment of long term sustainable earnings. Scholarship on corporate finance points to another direction: activist blockholders are beneficial and any claims to the contrary is based on anecdotal evidence (Bebchuk et al. (2013), Edmans, 2014, Brav et al., 2022). Given this background, a question emerges: even in face of evidence that activist blockholders are *systematically* beneficial to targeted companies, should we remain silent about the ones that indeed focus on short term gains just because they are not the majority?

I do not try to answer that question in this thesis as it involves a moral dilemma. Still, I recognize that blockholders have power, this power can be used for good or for bad and the consequences of using this power for *bad* might have deleterious consequences for the smaller group of targeted companies and its stakeholders. Rather than chose a side for one of these two antagonistic views presented so far, I leave that important debate aside and focus on the factors that lead activist blockholders to obtain larger stakes in targeted companies. This is an important mechanism because the larger the stake owned by an activist, the more power he will have to influence corporate decisions.

A better understanding of the what determines activist blockholders initial stakes is relevant no matter the side one stands (pro or against activist investors), but for opposing reasons. Activists arguably gather additional power cheaply, after reaching the regulatory threshold, but before market adjusts the stock price upwards once learning about the activist block acquisition. Those against activists will lobby for shrinkage or total elimination of the interval to reduce activist power. Activist advocates, on the other hand, will claim the interval is beneficial, because higher stakes translates into being better equipped to exercise influence that will, in their view, boost governance and benefit other shareholders and society.

In Chapter 1, I study one mechanism used by activist investors to increase its power through amassing a larger stake: the grace period between the date the block activist reaches the regulatory ownership threshold (5%) and the date he needs to make this information public.

As this paper was being written, the US Securities and Exchange Commission (US SEC) has updated the regulation concerning disclosure of blockholders beneficial ownership. Our paper adds to the discussion that has been revived on the policy level, mainly regarding the disclosure interval.

I take advantage of the activist blockholders setting to answer a quite different question in Chapter 2: whether market participants value sustainability. Note that here, I acknowledge that claims for sustainability business case are prone to endogeneity because good companies with healthy financial are the ones most likely to invest in sustainability. I do not try to circumvent that problem; instead I reinforce that the possible question I could try to answer here is whether the market assigns a premium to companies targeted by blockholder activists when their investment objectives are aligned with sustainability goals.

Empirical work on activist blockholders[1] documents the role of activists as drivers of change on targeted companies. Multiple studies have documented abnormal returns once the presence of the activist becomes public. The credible explanation for the abnormal returns is that market participants anticipate future positive financial and operational outcomes that will be driven by the activist intervention (Albuquerque et al. (2022)). Our paper takes advantage of activist investors data to perform an event study around announcement date, similar to the ones found in the academic literature, but with a twist: I incorporate information about activist objective. I extract the objective (a mandatory item) from activist filings and get its similarity score with UN Sustainable Development Goals (SDG)s using NLP techniques. As activists investors drive change, this setup provides a way to identify a premium linked to sustainability. I indeed find this positive link, acknowledging that this intervention is not likely to be an exogenous shock, so there is no claim of causality.

Finally, Chapter 3 is a treatise on a frequently overlooked subject: sourcing data for governance studies. Despite being the last chapter, it serves as essential background for the empirical work in Chapters 1 and 2, as I establish the groundwork for improving identification in event study regressions. In this chapter, I provide a comprehensive outline of the data that serves as the primary source for Blockholder Activism studies. I then discuss considerations

[1] see Edmans and Holderness (2017) for a general literature review on blockholders (theoretical and empirical) and to Brav et al., 2022 for a in-depth coverage of hedge fund activism, including more recent findings.

when using such data, demonstrating, through regressions, how retaining non-core data points can distort research conclusions. This chapter also includes an extensive appendix with practical guidelines for researchers to extract data and identify non-core datapoints.

References

Albuquerque, R., Fos, V., & Schroth, E. (2022). Value Creation in Shareholder Activism. *Journal of Financial Economics*, *145*, 153–178. https://doi.org/10.1016/j.jfineco.2021.09.007

Barry, J., Brav, A., & Jiang, W. (2020). Hedge Fund Activism Updated Tables and Figures. *Working paper*. https://people.duke.edu/~brav/HFactivismMarch2019.pdf

Bebchuk, L. A., Brav, A., Jackson, R. J., & Jiang, W. (2013). Pre-Disclosure Accumulations by Activist Investors: Evidence and Policy. *Journal of Corporation Law*, *39*, 1–34. http://www.sec.gov/news/speech

Brav, A., Jiang, W., & Kim, H. (2010). Hedge Fund Activism: A Review. *Foundations and Trends® in Finance*, *4*, 185–246. https://www.nowpublishers.com/article/Details/FIN-026

Brav, A., Jiang, W., & Li, R. (2022). Governance by Persuasion: Hedge Fund Activism and Market-based Shareholder Influence. *ECGI working paper*. https://papers.ssrn.com/sol3/papers.cfm?abstractid=3955116

Brav, A., Jiang, W., Ma, S., & Tian, X. (2018). How Does Hedge Fund Activism Reshape Corporate Innovation? *Journal of Financial Economics*, *130*, 237–264. https://doi.org/10.1016/j.jfineco.2018.06.012

Edmans, A. (2014). Blockholders and Corporate Governance. *Annual Review of Financial Economics*, *6*, 23–50. https://www.annualreviews.org/doi/abs/10.1146/annurev-financial-110613-034455

Edmans, A., & Holderness, C. G. (2017). *Blockholders: A Survey of Theory and Evidence*. https://www.sciencedirect.com/science/article/abs/pii/S2405438017300029

Fink, L. (2015). *Letter to CEOs*. https://www.blackrock.com/corporate/investor-relations/2015-larry-fink-ceo-letter

Gantchev, N. (2013). The costs of shareholder activism: Evidence from a sequential decision model. *Journal of Financial Economics*, *107*, 610–631. https://doi.org/10.1016/j.jfineco.2012.09.007

Lilienfeld-Toal, U., & Schnitzler, J. (2020). The anatomy of block accumulations by activist shareholders. *Journal of Corporate Finance*, *62*. https://www.sciencedirect.com/science/article/abs/pii/S092911992030064X

Rajan, R., Ramella, P., & Zingales, L. (2022). What Purpose Do Corporations Purport? Evidence from Letters to Shareholders. *NBER Working paper*. https://www.nber.org/papers/w31054

Rosenfeld, E. (2015). *Clinton lashes out at activists in Apple*. https://www.cnbc.com/2015/07/24/clinton-lashes-out-at-activists-in-apple-others.html2/8

U.S. SEC. (2022). Modernization of Beneficial Ownership Reporting - public consultation. https://www.sec.gov/rules/proposed.shtml

Zingales, L., Kasperkevic, J., & Schechter, A. (Eds.). (2020). *Milton Friedman 50 Years Later*. Stiegler Center. https://www.promarket.org/wp-content/uploads/2020/11/Milton-Friedman-50-years-later-ebook.pdf

Acknowledgements

My sincere thanks go to the University, particularly to the Department of Finance, for providing the resources necessary to complete this thesis. I extend my appreciation to the professors and staff in the Department of Finance and Economics, with special acknowledgment to Professor Dr. Ulf von Lilienfeld, Professor Dr. Roman Kräussl, and Professor Dr. Christos Koulovatianos. My sincere gratitude to Dr. Zsofia Kräussl, whose contribution, although not directly involved in my research, is indispensable; without her, I would not be writing this text.

I also want to acknowledge the support provided by Anja Lenninger, Céline Lecarpentier, and Lise Vilar-Tena, from the BED office, as well as to non-academic staff of FDEF — Roswitha Glorieux, Stéphanie Metry, Noémie Courtois and Sophie Lux. I am grateful for having your assistance on my doctoral journey.

I want to express my sincere gratitude to all the students who attended my courses at the University of Luxembourg, both at the Undergraduate and Master's levels. From the top-performers to those who found the content challenging; from the ones with a solid background to those who needed a helping hand with foundations in Maths; from the super-committed ones to those that needed something extra to be motivated; from the highly self-confident students to those shy enough to avoid the taking his turn on the microphone, you all brought substance to much of what I was working on. The interaction with each one of you greatly benefited the present work. I want to thank in particular Jacques Dablah, who from an attentive student came to be a friend, that brought encouragement and thoughtful words when I needed them most. Thank you, Jacques.

My doctoral trajectory led me to the place I truly desired — it opened my perspectives to academic research, not without certain trade-offs. I am grateful for all the opportunities to learn and develop new skills in recent years. None of these would be possible without genuine help from individuals who, in different moments, were there by my side. Thank you Nikita Gaponiuk, Alexander Ermakov, Zaruhi Zakobian, and Bilal Kchouri, my dear friends who shared the burden of the initial years with me and brought freshness to my cohort of companions. More towards the end of my studies, I had the support of friends who literally pushed me to the very end: Ana Cristina Castro, Genuino Cristino, Adriane Buchner, Alexandre Gartner, Cynthia Mary Anne, and Marcelo Guidugli. Thanks you Alexandra Merlino and François Baratte for your ongoing interest and friendship, and for adding moments of joy in the last years. A special thanks to Maria Lucia Fortes, who helped me understand so many things, but most of all, what

I truly aspired to achieve. And dear Lu Dorigo, who, while putting her fingers to the bone, was regularly following each new chapter, successes and failures, class prep, and expectations.

To Fabio Veiga, Maria Eduarda, Maria Luisa, and Isabela Gomes da Veiga; Cristiane Paiva Gomes Cruz, Ana Narumi, and Luiz Koji Yamashita; Ricardo, Ricardo Filho, Marcia, and Ana Paula Kalaidjian; José, Loide, and José Ricardo Kalaidjian, my deepest thanks for all your good wishes and support. This work took a big toll on our time together, and I am eagerly awaiting a time when we can celebrate. My husband Paulo, and our sons, Antonio and Eduardo Bilezikjian, have been and continue to be my greatest inspiration - this all started with them in mind. Though I have been robbed of valuable time with them, the doctorate has shaped into a better person, and I am confident that this will benefit the way I am a mother and wife in the years to come.

Lastly, none of this would have been possible without the support, education, and love of my parents, Celia Maria Paiva Gomes Cruz, and Antonio Carlos Gomes Cruz. Mamãe e papai, although you are not here to witness the fruits borne by the unconditional love and hard work you both instilled in me, your presence lies in every line written here.

To Maria.

"Let's be clear: the work of science has nothing whatever to do with consensus. Consensus is the business of politics. Science, on the contrary, requires only one investigator who happens to be right, which means that he or she has results that are verifiable by reference to the real world. In science consensus is irrelevant. What is relevant is reproducible results."

Michael Crichton in "Aliens Cause Global Warming"

Acronyms

ADU advanced document understanding

AGM annual general meeting

AI artificial intelligence

API application programming interface

ASCII American Standard Code for Information Interchange

ATE average treatment effects

AUC-ROC Area Under the Receiver Operating Characteristic Curve

BED Bureau des Études Doctorales

BECCS biomass emission with carbon capture and storage

BoD Board of Directors

CalPERS California Public Employees' Retirement System

CAPM Capital Asset Pricing model

CCS carbon capture and storage

CECL Current Expected Credit Loss

CIK central index key

CRSP Center for Research in Security Prices

CUSIP Committee on Uniform Security Identification Procedures

CSR corporate social responsibility

CSS cascade style sheets

DEI diversity, equity, and inclusion

DOM Document Object Model

EDA exploratory data analysis

Edgar Electronic Data Gathering, Analysis, and Retrieval

EMH efficient market hypothesis

ESEF European Single Electronic Format

ESG Environmental, Social and Governance

ESMA European Securities and Markets Authority

EU European Union

EW equal-weighted

FASB Financial Accounting Standards Board

FDEF Faculté de Droit, d'Économie et de Finance

FF3 Fama-French 3 factors model

FFM Fama-French 3 factors + momentum model

FF5 Fama-French 5 factors model

FUQ'd Fundamentally Unidentified Questions

GHG greenhouse gas

GT ground truth

GUI graphical user interface

GVKEY Compustat global company key identifier

HF hedge funds

HTML hypertext markup language

IFRS International Financial Reporting Standards

IV instrumental variable

JSON javascript object notation

JTM Just Transition Mechanism

LBO leveraged buyout

ML machine learning

NLP natural language processing

NGO Non-Governmental Organization

OCR optical character recognition

PERMNO CRSP permanent number

PRI UN Principles for Responsible Investment

regex regular expression

RSS really simple syndication

SDG Sustainable Development Goals

SDT Signal Detection Theory

SEC US Securities and Exchange Commission

SEC Edgar Electronic Data Gathering, Analysis, and Retrieval

SIC standard industry classification

SQL structured query language

SRI Social and Responsible Investment

URL uniform resource locator

US SEC US Securities and Exchange Commission

VW value-weighted

WRDS Wharton Research Data Services

XBRL extensible business reporting language

XML Extensible Markup Language

shrcd security code identifier

siccd SIC identifier

2SLS two-stages least square

Contents

Introduction **3**

Acknowledgements **9**

1 Activist investors stakes and market movements **26**

1.1 Introduction 27

1.1.1 Beneficiary ownership, raiders and activists 31

1.1.1.1 Raiders 32

1.1.1.2 Regulation mechanism 33

1.1.1.3 From raiders to activists 34

1.1.2 Regulation amendment 35

1.1.3 A polarized debate 37

1.1.4 Theoretical framework 39

1.1.5 Empirical work on blockholder activism 42

1.1.6 Contribution, results and paper overview 44

1.2 Data 45

1.2.1 Data extraction 45

1.2.2 Firm-related control variables 48

1.2.3 Event related abnormal returns and turnover 48

1.2.4 Ownership stakes 53

1.2.5 Market-related variables 54

1.2.5.1 Daily market returns: *10-lagging days* 56

1.2.5.2 Market trend (absolute value): *10-lagging days* 57
1.2.5.3 Deviation from market trend 58
1.2.5.4 Deviation from market trend: quintile breaks 61
1.3 Results 62
1.3.1 Ownership stakes 63
1.3.1.1 Effects of daily market returns on ownership 64
1.3.1.2 Regression on *deviation from market trend* 66
1.3.1.3 Regression on *deviation from market trend* ***quintiles*** 67
1.3.2 Abnormal returns 70
1.3.2.1 Regression on *deviation from market trend* 70
1.3.2.2 Regression on ownership with *deviation from market trend* as instrument 71
1.4 Conclusion 73

Appendix A Additional Tables and Figures 80

2 Is there a premium for sustainable development goals? The case of activist investors 107

2.1 Introduction 108
2.2 The UN SDG similarity measure 115
2.2.1 Natural language processing (NLP) 116
2.3 Data 120
2.4 Similarity measure 121
2.4.1 Textual content 121
2.4.2 Vectorization 124
2.5 Results 124
2.6 Conclusion 127

Appendix A Additional tables and figures 132

Appendix B UN Sustainable Development Goals **145**

Appendix C Embeddings **146**

3 Unveiling Non-Core Activist Blockholder Events **150**

3.1 Introduction . . . 151

3.1.1 Data extraction shifts across time . . . 153

3.1.2 Activist datasets: parsing and event identification . . . 155

3.1.3 Enhancing data extraction . . . 188

3.2 Data and Methodology . . . 192

3.2.1 Activist data . . . 194

3.2.2 Multiple fillings . . . 198

3.2.3 Other SEC Electronic Data Gathering, Analysis, and Retrieval (Edgar) sources for *event categorization* . . . 201

3.2.4 Market and fundamental data . . . 204

3.2.5 Descriptive statistics . . . 205

3.3 Empirical results . . . 220

3.3.1 Abnormal returns: effect of mergers, bankruptcies, insiders . . . 221

3.3.2 Ownership stakes: effect of multiple filings . . . 227

3.4 Conclusion . . . 236

Appendix A Enhancing Transparency and Reproducibility in *13D filings* Data Extraction **244**

A.1 Data . . . 245

A.2 EDGAR API access . . . 247

A.2.1 Starting point . . . 247

A.3 13D scraping . . . 249

A.3.1 Get *crawler.idx* . . . 249

A.3.2 Single filing data . . . 250

A.3.3 Scrape *index.htm* files . . . 251
A.3.4 Geocoding: drop non-US companies and some insiders . . . 254
A.3.5 Access *13D filing documents* . . . 255
A.3.6 General challenges to scrape *13D filings* . . . 258
A.3.7 CUSIP, investment objective, event date and ownership stakes . . . 259
A.4 Information from other filings (not-13Ds) . . . 270
A.4.1 20-F, 40-F or 6-K . . . 270
A.4.2 8Ks: bankruptcy and notice of delistings . . . 271
A.4.3 Merger-related proxy filings . . . 272
A.4.4 Form 4: mapping insiders to *targeted companies* . . . 273
A.5 Conclusion . . . 273

Appendix B Additional Tables and Figures **275**

List of Tables

1.1 Targeted firms fundamentals before trigger date 49

1.2 Cumulative abnormal returns ±20 days around trigger date 53

1.3 Ownership stakes - descriptive statistics 53

1.4 Daily market returns: *10-lagging days* after *event date* 56

1.5 Market trend - destriptive statistics 57

1.6 Deviation from market trend: quintile breaks 62

1.7 Regression: ownership *(dollar log)* over lagging daily market returns 65

1.8 Regression: ownership stake *(log dollars)* over market trend *(absolute value and mean daily deviation) - value weighted not controlled for size* 68

1.9 Regression: ownership (dollar log) over quintiles (deviation of market trend - vw) 69

1.10 IV Regression: abnormal returns on ownership stake *(natural logarithm of dollar amounts)* 71

1.11 Regression: abnormal returns *(CAPM)* over market trend *(absolute value and mean daily deviation) - value weighted* 74

A.1 Regression: abnormal returns *(CAPM)* over market trend *(absolute value and mean daily deviation) - equal weighted* 81

A.2 Regression: abnormal returns *(CAPM)* over market trend *(absolute value and mean daily deviation) - S&P 500* 82

A.3 Regression: abnormal returns *(Fama-French 3 factors)* over market trend *(absolute value and mean daily deviation) - equal weighted* 83

A.4 Regression: abnormal returns *(Fama-French 3 factors)* over market trend *(absolute value and mean daily deviation) - value weighted* . . . 84
A.5 Regression: abnormal returns *(Fama-French 3 factors)* over market trend *(absolute value and mean daily deviation) - S&P 500* . . . 85
A.6 Regression: abnormal returns *(FF 3 factors + momentum)* over market trend *(absolute value and mean daily deviation) - equal weighted* . . . 86
A.7 Regression: abnormal returns *(FF3 factors + momentum)* over market trend *(absolute value and mean daily deviation) - value weighted* . . . 87
A.8 Regression: abnormal returns *(FF3 factors + momentum)* over market trend *(absolute value and mean daily deviation) - S&P 500* . . . 88
A.9 Regression: Abnormal returns *(Fama-French 5 factors)* over market trend *(absolute value and mean daily deviation) - equal weighted* . . . 89
A.10 Regression: abnormal returns *(Fama-French 5 factors)* over market trend *(absolute value and mean daily deviation) - value weighted* . . . 90
A.11 Regression: abnormal returns *(Fama-French 5 factors)* over market trend *(absolute value and mean daily deviation) - S&P 500* . . . 91
A.12 Regression: ownership *(dollar log)* over average deviation from market trend - ew 92
A.13 Regression: ownership *(dollar log)* over average deviation from S&P500 trend . . 93
A.14 Regression: ownership *(dollar log)* over quintiles *(deviation market trend - ew)* . . 94
A.15 Regression: ownership *(dollar log)* over quintiles *(deviation of S&P500 trend)* . . . 95
A.16 Regression: ownership *(dollar log)* over average deviation *(deviation of market trend - vw)* . . . 96
A.17 Regression: ownership *(dollar log)* over quintiles *(deviation of market trend - vw)* . 97
A.18 Regression: ownership *(dollar log)* over lagging daily market returns . . . 98

2.1 Descriptive statistics: target fundamentals . . . 122
2.2 Descriptive statistics: SDG similarity scores . . . 125
2.3 Descriptive statistics: abnormal returns *(market model)* centered around *filing date* 126
2.4 Regression: Abnormal return over SDGs similarity *(market model)* . . . 128

A.1 Descriptive statistics: abnormal returns centered around *filing date* 133
A.2 Descriptive statistics: abnormal returns centered around *event date* 134
A.3 Regression: Abnormal return over SDGs similarity *(reference textual content: SDG goals + targets)* 135
A.4 Regression: Abnormal return over SDGs similarity *(reference textual content: SDG goals)* 136
A.5 Regression: Abnormal return (CAPM) over SDGs similarity (targets) - full table 137
A.6 Regression: Abnormal return (Fama-French 3 factors model) over SDGs similarity (targets) - full table 138
A.7 Regression: Abnormal return (Fama-French 3 factors + momentum) over SDGs similarity (targets) - full table 139
A.8 Regression: Abnormal return (Fama-French 5 factors model) over SDGs similarity (targets) - full table 140
A.9 Regression: Abnormal return (CAPM) over SDGs similarity (goals) - full table . 141
A.10 Regression: Abnormal return (Fama-French 3 factors model) over SDGs similarity (goals) - full table 142
A.11 Regression: Abnormal return (Fama-French 3 factors + momentum) over SDGs similarity (goals) - full table 143
A.12 Regression: Abnormal return (Fama-French 5 factors model) over SDGs similarity (goals) - full table 144

3.1 Target firms fundamentals: firm-months for all years *(1994-2023)* 206
3.2 Targeted firms fundamentals before trigger date 207
3.3 Target firms fundamentals before 13D trigger date 209
3.4 Summary statistics: targeted stocks cumulative abnormal returns 216
3.5 Summary statistics: average market daily returns 217
3.6 Summary statistics: ownership stakes *(% market capitalization)* by year 218
3.7 Summary statistics: ownership stakes *(log dollars)* by year 219
3.8 Regression: abnormal return over flags 222
3.9 Regression: ownership stake (percentage) over flags 229

3.10 Regression: logarithm of ownership stake (dollars) over flags 230

B.1 Regression: abnormal return over flags *(full table)* 276

2.3 Regression: logarithm of ownership stake (dollars) over flags *(full table)* 277

2.4 Regression: logarithm of ownership stake (dollars) over flags *controled for size - (full table)* 278

2.5 TOP 100 13D filers with filing amounts 280

2.6 TOP 100 13D filers in alphabetical order 281

List of Figures

1.1 Cumulative abnormal returns around trigger date 51
1.2 Abnormal turnover around trigger date 51
1.3 Examples of deviations from market trend 59

2.1 Unique Events, Filers and Subjects 121

C.1 Embedding: Parameters' Heatmap 147
C.2 Embeddings: investment objective and SDGs 148

3.1 Time deltas of 13D Events 197
3.2 Unique events, filers and targeted companies 198
3.3 Flags on pre-13D windows (-3/-6/-12mos) 203
3.4 Insider flags: interval between insider filing and 13D filing 204
3.5 Target companies fundamental data by market decile 211
3.5 Target companies fundamental data by market decile - continued 212

A.1 Number of records - 13D filings 251
A.2 Spatial Distribution of Filers and Targeted Companies 255
A.3 13D initial filings counts - before/after preliminary cleaning 256
A.4 Time delta between filings with coinciding CUSIP/filer 261

2.1 Events for the top 5 13D filers by year 279

Chapter 1

Activist investors stakes and market movements

Abstract

This paper investigates two questions related to activist blockholders: i) What drives their purchase decisions? ii) Is a higher share of activist holdings value increasing? We employ information disclosed in regulatory filings (SEC *Schedule 13D*) to address these inquiries as follows. The period between the *event date* (the day the investor reaches the regulatory ownership threshold, also known as *trigger date*) and the day of its public disclosure (*filing date*) is referred to as *"pre-disclosure accumulation period"* or *"grace period"*. We find that activist investors do indeed use the *grace period* to adjust their stakes, taking advantage of market-wide price swings. When the market experiences an upward deviation (last quintile) from its trend within the *10 days* following the *trigger date*, initial stakes in dollars are 19% lower. Though activist investors use the *pre-disclosure accumulation period* strategically, we do not find evidence that this (plausibly exogenous) variation in ownership has an impact on the overall valuation of the firm, at least within the scope of our dataset and specification.

JEL Classification: G14, G23, G30, G32

Keywords: large shareholders, blockholders, activist investors, activism, corporate governance, corporate finance, voice

1.1 Introduction

Activist investors are those that use their positions to exert influence on the companies in which they invest,[1] often by monitoring or directly engaging with management. Although investors can have an active role alongside investee companies regardless of their *ownership stake*,[2] academic studies on activism commonly concern only large investors. Consequently, in the academic context, the term *activist* is frequently employed as a synonym for *blockholder activist*. Throughout this paper we adopt the same practice: the terms *blockholders* or *activists* will be used interchangeably to specifically refer to *blockholder activists*.

There are two main reasons for researchers to focus on holders of large blocks, rather than considering activists of all sizes. First, larger stakes confer greater power over the investee company. Second, regulations mandate blockholders to disclose their *ownership stake*, the *event date*[3] and the *investment purpose*[4] if they have the intention to influence the investee's business.

The disclosure must occur within a span of *10-calendar days*.[5] Hence, the regulatory framework brings forth a new *informational element* – the *filing date*, the *official date* when the

[1]This definition can be expanded to include influence that extends beyond the company level (e.g., industry, government, regulators, non-governmental organizations). Nevertheless, market-level activism falls outside the scope of this paper.

[2]While there are eligibility criteria for shareholders to sponsor proposals (e.g. shareholders must have continuously held a minimum of $2,000 in market value or 1% of the company's securities for at least one year, as per SEC's Rule 14a-8), these thresholds are notably low; in practical terms, they are essentially negligible. What significantly restrains the activism of small investors is their limited capacity to exert influence. Small shareholders have less potential to persuade, as noted by Brav et al. (2022). However, there are instances of small sponsors successfully driving campaigns, such as Engine No.1. Despite holding only 0.02% of ExxonMobil, the investment firm led a successful campaign to elect two board members, advocating for the company's transition to cleaner energy (CNBC, 2021).

[3]*Event date* is the date the investor reaches a certain threshold (e.g., 5% in the US). It is also known as *trigger date*, because this is the point at which the obligation to disclose beneficial ownership via public regulatory filings is triggered.

[4]This includes plans or proposals for buying or selling securities, involvement in significant corporate transactions (e.g. mergers), selling substantial assets, changing the board of directors or management, modifying the issuer's capitalization or dividend policy, plans for changes in the issuer's business or structure, and taking actions that may limit management control. If the aim is to gain control, the activist must state its plans or proposals for liquidation, selling assets, mergers, or making major changes in the issuer's business or structure.

[5]As we are writing the current paper, there is now a *5-calendar days* requirement, reduced from the previous *10-calendar days*, following a recent regulatory update (U.S. SEC, 2023). However, when we refer to the *grace period* throughout this paper, unless explicitly noted otherwise, we are specifying a 10-calendar day interval in line with the regulation as it stood during the period covered by our sample, which extends only up to 2022.

presence of the activist becomes publicly known.[6] The time between crossing the ownership threshold and publicly disclosing this information gives rise to another feature: the *"grace period"*, also referred in the literature as the *"pre-disclosure accumulation period"*.

The combination of two features, namely the ability to influence company affairs, together with the rich documentation provided by the regulatory filings, make blockholder activist's *events* a particularly attractive setting for empiricists to investigate questions within Corporate Finance, such as the potential impact on company performance when executives are subjected to closer monitoring. This paper is motivated by the perception that, as activist investors amass larger *ownership stakes*, they exert increasing influence over their investee companies (Brav et al., 2022). Therefore, understanding the mechanisms that contribute to the accumulation of larger stakes, as well as exploring the consequences of such accumulation, constitutes a meaningful addition to the body of literature on Activist Investors.

Specifically, we ask two essential questions: i) Do activists respond to market-wide variations in the stock price during the *grace period*? ii) Does this response impact the overall value of the firm? Concerning the first question, a topic of debate in the literature (see Part 1.1.3), we obtained novel, clean results with a relatively simple research design, as we will discuss shortly. As for the second question, establishing a causal relationship requires specific methods for proper identification, as *ownership* is likely to be an endogenous variable. For instance, activists may actively acquire larger *stakes* in companies they believe have the potential for higher returns.

Nonetheless, as shown in Angrist and Pischke, 2008, we can still claim causality in the presence of endogeneity if we devise a proper experiment using randomized controlled trial. The ideal experiment in this particular case would consist into randomly assigning different *ownership stakes* to a plausibly large number of activist investors and then examine whether positive (negative) changes drive increases (decreases) in the company's value. Unfortunately, conducting such an experiment is not feasible,[7] so we need an alternative solution.

[6] We refer to the *filing date* as the *official* disclosure date to acknowledge that the information may have become public earlier, through means other than those specified in the regulation.

[7] Angrist and Pischke (2008) note that though ideal experiments are usually not feasible, it is always prudent to outline them. The primary purpose of that exercise is to serve as a sanity check: if we cannot conceive any such experiment, the questions being asked are likely to be, in the authors' words, Fundamentally Unidentified Questions (FUQ'd).

Our proposed identification strategy consists into devising an instrument[8] that explores the *random-walk* nature of the stock market prices[9] during the activist *grace period*. That instrument serves as a source of exogenous variation for *activist ownership*, bringing our setting closer to the ideal experiment.

We can characterize the main idea behind our strategy by considering two hypothetical, yet independent, activist *events*.[10,11] Assume that, when each of those events is triggered, stock markets are at a common level (i.e. market index coincides). Let's denote the index at the beginning of their respective *grace period* as *index₀*. In addition, assume that the market index also coincides upon both events disclosure (*index₁*). Now consider that $index_0 = index_1$, meaning the total market return is zero over the *grace period* for both events. The description so far refers to a scenario in which, on average, we would expect the values of the two individual stocks to remain constant.

Now, let's incorporate information about the *dynamics* of market prices within the *grace period*. The only boundary conditions we have imposed up to here, are minimal: the prices at the *start* and at the *end* of the interval are given, and they happen to be the same. These conditions, together with the *random walk*,[12] leaves infinite degrees of freedom for the price dynamics. In simpler terms, there are infinitely many possible paths price can take, that would satisfy those conditions. Now, let's assume uncertainty is resolved and we observe the realized paths for the two events. For the first event, we observe that the overall stock market initially rose and later, declined; for the second event, we assume the opposite: the overall stock market first decreased and then increased. So, despite the net change being zero in both cases, the

[8]Among the methods for proper identification in the presence of endogeneity, a solution is employing an instrumental variable (IV), a tool that isolates the exogenous variation in the independent variable. An IV must meet two criteria to be valid: *relevance*, being correlated to the endogenous variable, and *orthogonality*, being uncorrelated with the regression error term. In other words, an IV should only affect the dependent variable through the endogenous variable channel.

[9]The *random walk* theory posits that future prices depend solely on the current prices plus a random shock, and is mathematically expressed as $P_{t+1} = P_t + \epsilon_{t+1}$. Notice that this is a discrete time model. The equivalent model in continuous time is $dP_t = \mu dt + \sigma dW_t$, where dW_t, the stochastic term, is also known as Wiener process or Brownian motion, a stochastic process characterized by random movements in which each increment is independent and normally distributed. The parameters are μ, the drift, and σ is the volatility. Fama, 1995 gives a concise, non-mathematical overview of *market random walks* and their connection to the efficient market hypothesis (EMH). For an academic reference using discrete time, see Cochrane, 2001. For comprehensive coverage, including continuous-time, please refer to Munk, 2013.

[10]Further explanation on the methodology to construct this instrument is given in Part 1.2.5.2.

[11]Neither event refers necessarily to the same *targeted security* nor occurs on the same *event/filing dates*.

[12]Or, alternatively, stochastic process, if we are considering continuous time.

trajectories of market prices between the *start* and *end* dates followed notably distinct paths.

Under this assumed setting, we hypothesize that *activists* purchase more stocks in the second case rather than in the first. The rationale is that acquiring shares during the *down-up* trajectory would likely be cheaper, as long as the target stock prices follow market dynamics.[13] If this hypothesis is correct, it implies that the purchasing behavior of activists is influenced by market movements.

Indeed, we find evidence supporting this hypothesis, a result that is not only interesting in itself, but is also promising for addressing the likely endogenous nature of our second research question: how *abnormal returns* respond to marginal changes in *ownership stakes*. To establish a causal relationship, we need candidates for instruments, and the last result suggests that *market movements* qualifies, at least, in terms of *relevance*.[14] In this setup, for the instrumental variable (IV) to be valid, *market movements* (exogenous variable) should only influence *abnormal returns* (dependent variable) through the *ownership* (endogenous variable) channel.

Hence, in an attempt to answer the second question, we employ *market-wide variations in stock prices* as an instrument for *activist ownership* in two-stages least square (2SLS) regressions. We detail this approach in Section 1.3, just before presenting our empirical findings. In connection to that, despite observing a significant and robust first stage, the obtained results for the impact of *(instrumented) ownership* on firm value, as measured by *abnormal returns*, were not statistically significant. This implies that we did not have findings that contribute to answering the second question. While this may seem disappointing, our efforts have not been entirely in vain. After all, we have identified an IV that has proven robust in the first stage, suggesting it holds potential in alternative specifications wherever *ownership* endogeneity is anticipated. We resume possible next steps in Section 1.4.

Note that our identification strategy relies on the assumption that blockholders exploit the regulatory *grace period*, during which they can trade in secrecy. Given its central role in our empirical work, the rest of this introductory section will cover topics related to the *grace period*

[13] This is a reasonable assumption, on average, given that, during the *grace period*, the presence of the activist is not public information.

[14] Concerning *orthogonality*, there is no indication that market *random walks* might be systematically correlated with those actions or interventions targeted at individual companies, that end up producing abnormal returns. Furthermore, any market-wide effects should be neutralized in the computation of *abnormal returns*, as they are accounted for by the pricing models employed in our analysis.

and, to some extent, to their respective associated regulatory aspects. In Part 1.1.1, we provide a brief historical context concerning the introduction of the *SEC SC 13D*. Subsequently, we explore the shifts in *13D filers'* profiles in the last decades. Initially designed to prevent *raiders* from operating covertly, undetected by stakeholders, the disclosure requirement seems to have contributed to reducing *raiders'* activity. Today, instead of *raiders*, *activists* predominantly consist of *minority non-controlling investors*. Next, in Part 1.1.3 we explore the divergent perspectives regarding activist investors. This debate gained prominence again in 2022/2023 due to the proposal for regulatory change that underwent public review, a topic we briefly address in Part 1.1.2. Following, we analyze the interests of agents involved using a theoretical model in Part 1.1.4, exposing their antagonistic roles. Finally, in Part 1.1.5, we provide a quick overview of the empirical literature on activism to contextualize the contributions of this paper.

1.1.1 Beneficiary ownership, raiders and activists

In this Part, we provide a backdrop for understanding the origins, functioning and practical impact of *Regulation 13D*. We begin by characterizing *raiders*, as the beneficiary ownership regulation was largely motivated to inform about their presence to smaller investors. The rationale for disclosure lies in the fact that *raiders'* interventions often resulted in substantial gains for themselves, at the expense of minority shareholders. This naturally leads us to distinguish *raiders* from the contemporary concept of *activists*. Such differentiation serves a practical purpose since advocates and detractors of *activists*, which we will characterize in Part 1.1.3 and Part 1.1.4, frequently employ these terms, occasionally distorting meanings to further their respective interests. We then elaborate on the functioning of the regulation mechanism with the aim to show its contribution in reducing the number of *raiders' episodes*. Overtime, this led to a change in the profile of beneficial owners, compared to those initially targeted by the regulation at the time of its inception.

1.1.1.1 Raiders

Corporate raiders are described in the literature as those who use aggressive tactics in pursuing takeovers and restructuring acquired companies. They are often associated with an image of opportunism and are considered detrimental to the broader business environment. *Raiders* are accused of prioritizing short-term gains over long-term stability to the detriment of employees, communities, and the overall economy. Additionally, as they often use leveraged buyout (LBO), they are deemed to leave acquired companies financially vulnerable.[15] In this paper, we refer to *"raiders"* as investors who aim to extract value from the acquired company using aggressive, detrimental tactics.

The rationale for alerting investors about the presence of *raiders* has to be put into historical context. In the past, there were many episodes in which *raiders* were successful in employing strategies that could significantly disadvantage smaller shareholders. For instance, in a practice known as *greenmail*, a *raider* would purchase a significant stake in a company, and upon threatening a hostile takeover, negotiate a premium for selling their shares back in a private transaction. Another tactic was to force *two-tiered tender offers*, structured to provide more favorable terms to themselves, e.g. larger stakes receive higher prices. Another example consisted in compelling the targeted company to acquire other businesses at *higher purchase prices*, leading to fees or direct gains from the difference between the inflated price and the correct, fair one.

Since then, the regulatory landscape has evolved, governance practices have been refined, and companies are now subjected to rules that mandate enhanced transparency. Very importantly, almost half a century forward, thankfully, moral standards have also advanced. So, for an observer today, corporate practices as described above not only appear inappropriate, but are also considered morally unacceptable. Moreover, these practices would not make it through in most jurisdictions, including Delaware.[16]

Therefore, the three examples provided here, wherein the adverse aspects of *raiders'* in-

[15] *Raiders* are not synonymous with *hostile bidders*; while some *raiders* may also be *hostile bidders*, not all *hostile bidders* are necessarily *raiders*.

[16] Note that generally, these actions are not expressly referenced in legal or regulatory provisions in most jurisdictions. Nevertheless, they are practically unfeasible in the prevailing business and legal landscape. At the minimum, they typically involve breaching the fiduciary duty of company executives and might also contravene other general regulations. Moreover, while in the past *raiders'* actions would be kept undisclosed, facilitating their activities, under current rules that enforce better governance, that would be likely not the case.

terventions are clear, would, as of today, expose both companies and executives to legal and financial liabilities, and evidently, reputational harm. Consequently, the mechanisms in place today protect the interests of smaller shareholders much more directly, safeguarding them from most of the detrimental strategies that *raiders* used to employ in the past.

1.1.1.2 Regulation mechanism

In the US, the disclosure of large beneficial owners is regulated by Section 13(d) and 13(g) of the Williams Act, officially known as the *"Securities Exchange Act of 1968"*. These provisions were enacted to better inform shareholders (primarily individual investors at that time) about changes in control that could materially impact their investments. The typical case at the time the regulation was conceived related to the acquisition of large blocks by *corporate raiders*.

The original text mandated that investors, upon reaching a *5% ownership stake* in a publicly traded security of a company with the intent to influence its operations, disclose their intentions by filing a *Schedule 13D* within *10 days*.[17] This implied there was a period between the day of reaching the threshold (*event date* or *trigger date*) and the *filing date*, during which the blockholder's presence was not publicly disclosed. We refer to this regulatory interval as the *grace period*, also known in the literature as the *pre-disclosure accumulation period*.

While blockholders can still increase their stakes after the initial filing, and they often do so, as documented by filing amendments (*Schedule 13D/A*), trading in the pre-disclosure interval is potentially advantageous (Shleifer and Vishny (1986)). This is true if *public knowledge of activist presence* triggers an upwards adjustment of the stock price, driven by market participants anticipating gains from activist oversight. In fact Albuquerque et al. (2022) estimate that almost 75% of price increase upon activists announcement refer to *expected* value creation.[18]

The potential exploitation of the *grace period* by blockholders as an opportunity to trade cheaply is undisputed across both critics and supporters of blockholder activists. However, this

[17] The time frame was amended to *5 days* at the conclusion of 2023, effective from February 2024.

[18] The authors document 6.34% average return for blocks acquired by activist investors, specifically those disclosed in *13D filings*. In contrast, for blocks related to passive investment (Schedule 13G filings), the average is 0.59%. We present their numbers to evidence the contrast between activists and non-activists returns, but note that those figures refer to their sample and are not comparable to the ones found in our study. Due to differences in time coverage and composition, *13D's* average returns are much larger for our dataset, as we make numerous adjustments to remove events we consider non-core (Cruz, 2023), as presented in Section 1.2.

consensus leads to divergent viewpoints: critics call for shortening or eliminating the interval, whereas supporters argue that there are social gains that would justify non-immediate disclosure. We will further explore this debate shortly in Part 1.1.3.

1.1.1.3 From raiders to activists

The 1980s represented the height of the *raiders'* era. However, the rise of anti-takeover measures, notably poison pills, coupled with successive cases where Delaware Courts ruled favorably for companies, has subsequently curbed their activities.

Regulation 13D appears to have played a role in this decline. As disclosure was required once relatively small stakes were acquired *(5% ownership)*, company management became informed about the *raider's* intentions at an early stage, as documented in Mikkelson and Ruback (1985). This early awareness empowered management to implement more comprehensive anti-takeover measures or consider other counter strategies. The obligation to disclose, coupled with jurisprudence that discouraged *raiders*, led to the near disappearance of *raiders episodes*. However, we should not overstate the relevance of *Regulation 13D* alone in the decline of *raiders*. Other factors, notably the changing business landscape and evolving moral standards, have contributed to this phenomenon.

Consequently, in the last couple of decades, those that file a *Schedule 13D* are not *raiders*. Instead they are mainly *activists*, as laid out in Brav et al., 2022:

> *"Activist hedge funds also differ from corporate raiders that operated in the 1980s, as they tend to accumulate strict minority equity stakes and do not seek direct control. Activists differ from raiders not only in terms of the size of their stakes but also in how they interact with companies. While they may sponsor some corporate battles, over time, they have increasingly leaned toward working collaboratively with management."*

In conclusion, unlike *raiders*, who seek controlling stakes, *activists* generally acquire only minority shares and need to persuade fellow shareholders to gain influence in company matters.

1.1.2 Regulation amendment

The original *10-days* deadline was established when the regulation was introduced in 1968. Back then, *10 days* seemed adequate for investors to complete paperwork, validate its contents and either send them by regular mail or deliver it physically to the SEC. In today's context, however, such long window is unwarranted, as submissions are now sent electronically, and information is promptly aggregated and processed.

In 2022, SEC issued a proposal for public comment, suggesting amendments to modernize regulations governing beneficial ownership (U.S. SEC (2022)). An important element in the proposal was the review of the *10-day* disclosure deadline to better achieve the main objective of the regulation, i.e., to provide timely information for both the *targeted company* and the general public:[19]

> *"In reassessing whether or not the current 10-day deadline still serves the primary purposes of Section 13(d), which are to provide information to the public and the subject issuer about accumulations of a covered class by persons who had the potential to change or influence control of such issuer and to regulate rapid accumulations of beneficial ownership that occurred within a short period of time, we have determined that an amendment to Rule 13d-1(a) is needed to adequately support those regulatory objectives."*

The SEC's proposal recognized how technological advancements had not only made it feasible for investors to disclose block acquisitions within a shorter interval, but has completely reshaped the investment landscape, leading market participants to seek timely access to information:

> *"We believe the 10-day filing deadline for the initial Schedule 13D filing should be revised in light of advances in technology and developments in the financial markets. Our proposal to shorten the initial filing deadline for Schedule 13D is consistent with previous Congressional and Commission efforts to accelerate public disclosures of*

[19]The original regulation text was designed to safeguard smaller shareholders from actions that could harm their investments. As discussed in Part 1.1.1, over time, the regulation evolved into an important tool used by targeted companies to prevent *raider* activity. The modernized amendments recognize this aspect, incorporating the *public issuer* as the interested party in the disclosure, not only *shareholders* as in the original version.

> *material information to the market. (...) significant technological advances over the last three decades have both increased the market's demand for more timely corporate disclosure and the ability of companies to capture, process and disseminate this information."*

Finally, towards the close of 2023, slightly over a year after the release of the proposal, the SEC formally issued the amendment. The deadline was reduced to *5 calendar days*,[20] which were the terms set forth in the proposal and, like most other provisions, was scheduled to take effect in February 2024.[21]

In conclusion, the regulation currently specifies a *5-day* window; however, the historical archives of *13D filings* up to the end of 2023 correspond to a period where *10 days* was the norm. Since our dataset extends only up to 2022, throughout the other section of the paper, that covers data and empirical results (Section 1.2, Section 1.3 and Section 1.4), references to the *grace period* will consistently imply a *10-day* duration, in accordance with the legal requirement in place at that time.

While there was a general agreement that the initially specified *10-day period* for disclosing block acquisitions, as outlined in the 1968 original text, was unwarranted due to technological advancements, there was a fierce debate around how long it should be. Supporters of activists favored the *5-day* grace period, as proposed, which ended up being adopted in the amended text effective from Feb/2024. In contrast, opponents argued for even shorter intervals. For instance, WRLK proposed a 1-day disclosure window, along with a moratorium mechanism restricting additional securities acquisition for 2 days after filing the *Schedule 13D*. We will explore some of these arguments in the next, in Part 1.1.3.

[20]The original text was not specific on how to count days, leaving the reference schedule ambiguous – whether *calendar days*, *business days*, or another reference. This information remained as a response within a Q&A published by SEC, clarifying some aspects of beneficiary ownership regulation. As it was not incorporated, for example, via amendment in the main regulatory text, it was less effective in informing the general public. In the 2023 amendment, the revised deadline is explicitly *five calendar days*, offering a clear guideline.

[21]Specifically, the amendment mandates the submission of *13D filings* in a machine-readable form. Such technical changes typically grants a lengthier compliance period, offering those impacted sufficient time to prepare and adapt their processes to accommodate these modifications.

1.1.3 A polarized debate

In Part 1.1.1 and 1.1.2, we pointed out the existence of two distinct perspectives on activists. *Advocates* or *proponents* are those who endorse activists; and *detractors* or *opponents* are those considering them detrimental. In this Part, we offer a concise overview of the participants in each group and bring some arguments they have presented in the literature over recent decades.

Advocates find representation among Corporate Finance scholars and investors, while *opponents* are primarily constituted by Law scholars and incumbent managers. *Advocates* argue activists play a role in disciplining companies, and their activity produces social gains. Therefore, they favor activists' business should be encouraged. Hence due care should be taken to avoid hindering their profits, as these are viewed as an economic incentive for them to continue in that business. Conversely, *opponents* argue activists act as value destructors, pursuing gains in a detrimental way. In their view, they not only harm the targeted company but also impact other investors negatively because, so they say, the disclosure regulation is flawed, creating information asymmetry.

Corporate Finance scholars claim activists consistently produce positive outcomes for the companies they target, as substantiated in the academic literature (Edmans and Holderness (2017), Brav et al. (2022)). They reason *activists* contribute to disciplining investee companies, and their interventions enforce good corporate governance, ultimately maximizing social welfare. Consequently, caution should be exercised so to avoid reducing incentives driving their activities, as this would diminish their inclination to invest time and resources in monitoring efforts.

A case in point is limiting activists' potential gains through early disclosure. *Advocates* defended the maintenance of *grace period* that is relatively long in the context of current technology. Within that timeframe, activists can accumulate larger positions before prices are adjusted upwards by other market players, in anticipation of their intervention. The extra profits derived from trading in secrecy should be seen as a motivation for activists to continue with their activity as, so *proponents* claim, it ultimately results in positive social outcomes.

On the opposite side, there are some Law scholars, corporate management incumbents and law firms that represent them, that are joined by some public figures.[22] They accuse activists[23] of *"managerial myopia"*, forcing companies' management to take decisions that will benefit shareholders on the short run in detriment of long term sustainable earnings.

Wachtell, Lipton, Rosen & Katz (WRLK[24]), a law firm specialized in corporate and business law, typically representing incumbent corporate directors, is a prominent voice among this group. They have actively campaigned to limit blockholders' potential to acquire larger stakes, including advocating for a shorter *grace period*. In 2011, they initiated a campaign that involved public media and a SEC petition, urging the tightening of the *10 days* window, but without immediate[25] success. On many other occasions, they have published reports *"denouncing"* activists of acting undercover during the *grace period*. They argued this mechanism allowed activists to amass large blocks at the expense of uninformed investors. Once holding these blocks, so they claim, activists forced corporate decisions aimed at short-term gains in detriment of sustainable long-term operations.

The conflicting viewpoints of these two groups can be reconciled by distinguishing between *activists* and *raiders*. As discussed in Part 1.1.1, *raiders* are inherently detrimental. And *opponents* essentially equates *activists* with *raiders*, which is evidently a misconception. Adding to this debate, scholars in Corporate Finance argue that instances where activists cause harm to investee companies (resembling *raider episodes*) do exist, but are isolated, so the evidence presented by critics should be viewed as anecdotal.[26]

Assuming the majority of activists, as of today, have a positive role in the targeted companies, why would *opponents* characterize them overall as negative? Moreover, why would they even

[22]Public figures that have expressed concerns about short-term focus and disruptive tactics employed by some activist investors includes Larry Fink (CEO BlackRock), Hillary Clinton, Warren Buffet (CEO of Berkshire Hathaway) and Jamie Dimon (CEO of JP Morgan).

[23]Notable hedge fund activists are Carl Icahn, Daniel Loeb, William Ackman, Nelson Peltz and Paul Singer. They are founders of the investment companies Icahn Enterprises, Third Point Management, Pershing Square Capital Management, Trian Fund Management and Elliott Management Corporation, respectively.

[24]The "L" in WRLK stands for Martin Lipton, who is the creator of poison pills — a defensive tactic used by a company's management to deter hostile takeovers. When a potential acquirer accumulates a certain percentage of the target company's shares (usually around 10% or 20%), the poison pill is activated. Once activated, existing shareholders, except the triggering bidder, are given the opportunity to purchase additional shares at a significant discount. This mechanism dilutes the bidder's ownership stake and increases the overall number of outstanding shares, making it more challenging and costly for the acquirer to gain control.

[25]Although there was no change in the regulation for many years to come, their campaign surely influenced the reform of Schedule 13D, which was presented in Part 1.1.2.

[26]Notably, Bebchuk et al. (2013) specifically counter the arguments advanced by the WRLK campaign mentioned earlier.

create legal artifacts such as *"anti-activist poison pills"*, as studied in Eldar et al. (2023)? Well, activists are indeed inconvenient for incumbent executives.[27] As activists exert closer monitoring, incumbents have their freedom curtailed – decisions are scrutinized and have to be justified, which can be occasionally uncomfortable and somewhat limiting. That being said, on the other hand, the proposition by some Corporate Finance scholars explicitly suggesting that regulations should maintain certain indirect incentives for activists, such as a lengthy *grace period*, also seems a bit of a stretch.

These differing opinions reflect the vested interests of various stakeholders. Next, in Part 1.1.4, we explore a theoretical model that helps understand the dynamics at play among the various players involved in activist events and how abnormal returns are influenced by the existence of a *grace period*.

1.1.4 Theoretical framework

The effects of two disclosure thresholds, namely *ownership stake* and the *time span of the grace period*, on interactions involving activists, incumbent managers, and other market participants was examined in a model proposed by Ordóñez-Calafi and Bernhardt (2022). Their fundamental assumption is that activists' interventions enhance value, as increased monitoring serves to mitigate agency problems.[28] In their model, activists are informed investors who benefit from pre-disclosure trading, i.e. trading in secrecy, because the market maker adjusts prices upwards only after learning about their presence. Consequently, more stringent regulatory measures such as lower *ownership thresholds* and shorter *grace periods* operate, at first glance, as disincentives, constraining the potential gains of activists.[29] However, as we will see later, the behavior of all the agents in the model creates a dynamic that counteracts these disincentives, and in turn, there is an optimum for *thresholds* that lead to maximum social benefit.

Investors other than activists are considered uninformed because they are unaware of the

[27] And consequently for those that represent their interests.

[28] As we have just seen in Part 1.1.3, this is prevalent perception among Corporate Finance scholars.

[29] Though in this model, the authors evaluate two different thresholds, we are only interested in what concerns the *time span of the grace period*. The other aspect studied, *ownership stake*, is considered as given at 5%, as we are not interested in studying the marginal effects of its variation.

activist's presence. Though uninformed, most of them benefit from the activist's intervention,[30] as this leads to improved governance in targeted companies, resulting in higher valuations. Nevertheless, a small subset of shareholders incurs trading losses by unknowingly selling their positions in the presence of activists, thus at discounted prices – a direct consequence of information asymmetry driven by the, perhaps unwarrantedly large, *grace period*.

Internally, company management is discouraged from exploiting the business for private benefit, as practices that erode value attract activist blockholders. From the perspective of incumbent executives, the presence of activists represent less autonomy in decision-making due to their close oversight. Moreover, numerous initiatives that are potentially detrimental to incumbents (e.g. changes in management composition, cost saving measures that reduce executive's perks)[31] would not be on the agenda if it was not for the activists involvement.[32] These considerations encourage incumbents to avoid significantly deviating from the company's best interests, a mechanism referred to as *"discipline through spillovers"*, as explored in Gantchev and Giannetti (2021).

An interesting insight derived from this model is that activists do not benefit if the spillover mechanism effectively disciplines management. Consequently, although less stringent thresholds might initially appear advantageous for activists leading to easier pre-disclosure trading profits; counteracting forces, notably the reinforcement of the *discipline through spillover* mechanism, can neutralize that potential advantage. Conversely, with stricter thresholds, such as a shorter *grace period*, it would be harder and less profitable for potential activists to acquire larger stakes. Recognizing this, corporate incumbents, perceiving a reduced likelihood of being targeted, would be motivated to institute various agency costs within the firm. However, if, contrary to their expectations, they become a target, activists stand to achieve significantly greater profit in this particular scenario.

In summary, the optimal threshold level that encourages investors to assume an activist role lies in a balance between being accommodating enough to allow gains from cost-effective

[30] It is reasonable to consider that most uninformed investors stand to benefit from activist events because, as seen in Table 1.3, they represent around 10% (median) and 15% (mean) ownership. Hence, on average, more than 80% of market capitalization is in the hands of uninformed holders.

[31] Note that activists being detrimental to incumbent management does not equate to being detrimental to the company's overall business. An apt example is how increased oversight might address agency problems.

[32] This seems to align with the observed targeted companies' abnormal returns found in empirical studies.

trading prior to disclosure and maintaining a certain degree of stringency to encourage a certain level of *mismanagement* among corporate incumbents.

As in any model, the simplifying assumptions are chosen either to emphasize aspects of interest to the modeler or for tractability reasons. While this model embodies the main actors and those dimensions discussed thus far in this introduction, it clearly reflects the perspective of Corporate Finance Scholarship, one that attributes exclusively favorable outcomes to activist intervention.[33] Nonetheless, it furnishes insights into the forces at play that drives the action of each actor when *activists* are small enough not to seek control (they are not *raiders*).[34]

The discussion up to this point evidences the conflicting positions of different economic agents. Note that the entire debate relies on the assumption that activists exploit the pre-disclosure period as a means to bolster their holdings. But do they genuinely employ this strategy? Moreover, what is the marginal effect of increasing *ownership stakes*? Interestingly, there is not many studies that explore these questions empirically. Next, we give an overview of the existing empirical literature to situate how this paper fills the gap regarding those same question.

[33] This premise, as argue the authors, is justified by empirical findings.

[34] Goshen and Steel (2022) challenges the conventional notion that *activists* yield more favorable outcomes for companies compared to *raiders*. They posit that *raiders*, as majority stakeholders, are inherently vested in the company's interests, thus having a substantial *"skin in the game"*. Conversely, *activists* are suggested to have greater latitude for pursuing more assertive strategies, owing to their comparatively less committed engagement in the company's operations. Once again, in our perspective, those claims would benefit from a proper characterization of the terms *raiders* and *activists*. What the authors label as *raiders* does not align with the definition adopted in this paper. In our terminology, this term is reserved for blockholders who act solely in their own interest, to the detriment of other stakeholders. Using the conventions adopted in our paper, we would rephrase the authors' argument as *"controlling stakeholders are generally aligned with the best interests of the companies they control, while minority blockholders have more leeway to pursue aggressive strategies in search of outstanding returns"*. Still using our terminology, their argument suggests that if minority stakeholders' risky strategies go awry, they would not incur as much loss as those controlling a company. While there is merit in considering *skin in the game* for controlling parties, there are some problems in this argument. First, it is somewhat naive to believe that opportunities for controllers to effectively appropriate gains not shared with other shareholders or that might end up being detrimental for companies in the long run do not exist. Similarly, it is naive to assume that minority blockholders will always act in the most aggressive and detrimental way. Moreover, considering stakes that represent at least 5% of a company's as not having enough *skin in the game* than those of controllers is quite a simplification.

1.1.5 Empirical work on blockholder activism

Early research on activism mainly explored topics related to proxy fights, often investigating the efficacy of activists in successfully passing their proposals in proxy battles.[35,36] Later, Brav et al. (2008) inaugurated a new strand of literature, where they limited the events to those initiated by hedge funds (HF). Their main finding is that hedge fund activism is linked to improved financial, operational and market performance[37] of the targeted companies, a result that holds for both the short and the long term. In the wake of their seminal work, dozens of articles explored hedge fund activism, either connecting it to other positive outcomes (e.g. better capital allocation, more innovation, higher productivity) or updating previous results by including more recent data points (Brav et al. 2010, 2015, 2018, Barry et al. (2020), Brav et al. (2022)).[38]

Hedge funds are at the epicenter of the debate introduced in Part 1.1.3, given the distinctive attributes that empower them to assume an especially proactive role as activists.[39] So the impact of activism on performance has been particularly scrutinized for this class of investors. On the other hand, little attention has been given to the mechanisms that lead to variations in *ownership stakes*. One of the rare studies that explores this aspect is Bebchuk et al. (2013). Among the results presented by those authors, they specifically examine *ownership* (the outcome variable) to address two questions: whether there was an increase in *ownership stakes* over time and whether HF hold larger *stakes* than other categories of activists.

Regarding the first question, for the period from 1994 to 2006,[40] they neither find evidence of a positive trend nor of an increase that could be observed in the last five years of their sample.

---

[35] See Gillan and Starks (2007) for a review of studies until 2006.

[36] More recently, there has been a resurgence of interest topics involving proxy fights as investors are increasingly seen as stewards due to their fiduciary duty.

[37] The later as measured by abnormal returns.

[38] We refer the reader to the later for a rich literature review for blockholder's activism, including the cases for which causality is claimed.

[39] Unlike some other institutional investors, such as pension funds and mutual funds, hedge funds operate independently and do not face the same investment limitations and potential conflicts of interest. Among hedge funds, there are even some that prioritize activism as their core investment strategy. They specialize in researching various companies to identify potential targets and acquiring minority stakes in those they believe can be enhanced and create value through their influence. As a result, among all categories of activist investors, hedge funds are often seen as the most representative one.

[40] Bebchuk et al., 2013 dataset is recycled from Brav et al. (2008). That is why, although the paper is published in 2013, the period covered for the main empirical results stops in 2006.

For the second question, they restricted the study to a subsample[41] of events randomly selected from a *single* month,[42] June 2011, which they claim was also chosen randomly. Regressing ownership over a dummy for HF class, they concluded that HFs acquire lower stakes than other activist investor classes.[43]

Although our papers both investigate blockholders' pre-disclosure accumulations within the regulatory *10-day* period, the similarities end there. Our study differs from theirs in all other aspects, including research questions, design, and dataset construction.[44] First, we are neither interested in ownership evolution through time[45] nor in the relative size of HF's ownership stakes: our research questions (see first part of this Section 1.1) are distinct. And as such, given the entirely unrelated nature of our inquire, there is no basis for a comparison of research designs.

Second, in terms of our dataset construction, ours differs substantially from theirs in several aspects. To begin with, although our dataset mainly comprises hedge funds, this was not an explicit choice but rather a consequence of our methodology to include only *core events*, as detailed in Section 1.2.[46] Note, however, that in our sample, there are instances of other investor classes as well, such as mutual funds, pension funds, and banks. The differences do not stop there; even if we were to consider only their hedge fund sample, the methodology we used to exclude *non-core events* is relatively restrictive, probably eliminating some of events that made it into their sample.

Another study that investigates, how accumulation occurs is Collin-dufresne and Fos (2015). Those authors manually collected the (multiple) days and respective amounts when *filers* have traded on stocks of the *targeted companies*, for the periods 1994-2010.[47] They found evidence

[41]This subsample contains events initiated by 20 activist hedge funds and 154 by non-hedge funds.

[42]They reduced the time span of observations because they had to collect extra data - activist events initiated by non-hedge funds - to answer this question. While basic information of activist events is easy to collect, extraction of ownership stakes is time-consuming and prone to errors.

[43]See Bebchuk et al. (2013) *Table 6* for descriptive statistics of ownership and *Table 7* for regression over the hedge fund dummy.

[44]Besides the core distinctions outlined in the main text, our study spans different time periods as well. 2013 recycled the dataset from Brav et al. (2008), with the last data point reaching 2006. On the other hand, our dataset starts in 2006 due to a methodological constraint (availability of data for identifying non-core events). Hence, the coverage of our datasets is totally distinct, with only a single year of overlap.

[45]For example, we are neither concerned about the temporal evolution of ownership stakes, as evidenced by the inclusion of time fixed effects in all our regressions.

[46]It just happens the core events are largely comprised by those that are initiated by hedge funds.

[47]The authors mention that their dataset includes only purchases with *exhibits* showing trading activity. We

that activists trade around 1% of outstanding shares on the *event date* and around 0.1% and 0.15% on the day before and after the *event date*, respectively. Their findings suggest activists prefer limit orders over market orders and have a tendency to buy stocks during low liquidity and negative market conditions.

Our study differs from theirs in that, while we also integrate market movement and liquidity into our regressions, our research aims to address a more specific question. We do not primarily investigate the market conditions that prompt activist investors to acquire blocks initially; rather, our focus is on whether, having already acquired a block, market conditions influence an increase in activist *ownership stakes*.

1.1.6 Contribution, results and paper overview

Our paper contributes to the limited literature investigating determinants of pre-disclosure ownership stakes, complementing the works of Bebchuk et al. (2013) and Collin-dufresne and Fos (2015).

We formally test the hypothesis that blockholders use the days after the *grace period* to increase their stakes strategically, contingent on market movements. Our results empirically substantiate this hypothesis, presenting a novel finding that complements Bebchuk et al. (2013). While those authors suggests pre-disclosure accumulations are predominantly carried out until reaching the threshold, our study finds systematic accumulation beyond the threshold, which is curtailed if market prices rise to the level of the last quintile.

In a similar vein, we complement the findings in Collin-dufresne and Fos (2015), where the authors conducted a detailed study about the individual trades used by the activist to acquire the initial block. While we have some specifications where we study daily changes, our main findings concern the *grace period* as a whole, and in our setting we introduce exogenous variation that allows us to address additional questions.

Finally, as discussed in the first part of this Introduction (Section 1.1), we did not find

have some concerns regarding this choice because events corresponding to single-day purchases frequently lack accompanying exhibits. The rationale for it is, if all stocks referring to the event to be disclosed were acquired on a single day, which is obviously the same that figures in the regulatory filing as *event date*, there is no need for an additional detailed exhibit. Hence, we believe that results in 2015 might potentially be biased towards events triggered by multiple-day purchases.

conclusive evidence about the positive effects of larger stakes. However, we see a lot of space for further investigation that seems promising, particularly over the long run, which we have not explored.

The rest of this paper is dedicated to empirical analysis. In Section 1.2, we follow the methodology outlined in Cruz, 2023 to construct our dataset and present descriptive statistics for the control variables, as well as for variables used in the different specifications of our regressions: abnormal returns, ownership stakes, and market movement. Regarding the latter, we detail how this variable, which is used as an instrument in our study, is constructed and provide some illustrative examples. In Section 1.3, we present our results. We conclude in Section 1.4, where we explore possible future research directions that could benefit from the research design developed in this paper.

1.2 Data

1.2.1 Data extraction

In this part, we give an overview of how we obtained our datasets. We adopted the same methodology described in Cruz, 2023,[48] so here we just give a concise coverage. For a comprehensive understanding of data cleaning and processing, along with extensive dataset statistics, readers should refer to the original work.

We gathered data from three sources: Electronic Data Gathering, Analysis, and Retrieval (SEC Edgar), the primary repository for filings and disclosures mandated by the US SEC; market data from Center for Research in Security Prices (CRSP); and fundamental company information from Compustat.[49]

The data extraction and cleaning processes were entirely algorithmic and followed a systematic

[48] This approach offers a notable advantage as it does not depend on commercial databases, which often requires costly annual subscriptions. Moreover, contrary to the ordinary perception, commercial datasets require additional cleaning steps. The combination of high costs and the need for supplementary work encourages researchers to choose the easier path of reusing outdated datasets. We address these issues by employing an algorithmic methodology, making it convenient to access up-to-date data and effectively identify *non-core events*.

[49] The latter two datasets are available in the basic subscription of Wharton Research Data Services (WRDS).

sequence of steps. Initially, we retrieved *Schedule 13D* records[50] from the Edgar application programming interface (API). We eliminated duplicate entries, as well as those related to non-US companies, and removed insiders based on coincident company addresses for *filer* and *target*.[51] We then fetched *13D main documents* referring to those records that remained after that pre-cleaning, and parsed them to extract *event dates*, *securities CUSIP*, and *investment objectives*.[52]

The next step was the conversion of these *filings* into *events*. This included tasks such as consolidating *filings* that have been likely submitted individually by members of a group. We created dummies to flag those *events* that were used later, on our regressions, to neutralize systematic errors incorporated due to the consolidation. In addition, we identified/removed cases involving another *"type"* of duplication, those concerning *initial filings* that were later corrected (i.e. after some days).[53]

At this point, the *events dataset* was merged with market and fundamental data from CRSP and Compustat. The pairs CUSIP-*event dates* formed the keys used to join the datasets. Subsequently, we excluded *events* related to *targeted companies* classified as utilities or financial firms. We also excluded those involving securities that were neither *common stocks* nor issued by *US incorporated companies.*

Next, we implemented a procedure to identify *non-core events* – those that do not align with the typical activist context, such as *pre-merger announcements*, *insider transactions*, *bankruptcies*, and *reorganizations*. For this process, we used various datasets, which can be essentially categorized into two groups. The first group consisted of data obtained from SEC regulatory filings, other than *13Ds* (e.g., *Form 4*, proxy filings related to mergers, and *8Ks*). We dropped those *events* for which the *targeted company* had submitted any of those *non-13D*

[50]These records are organized on a structural way, where each entry has basic information from filings and the uniform resource locator (URL) segment that points to the document itself. The document is only retrieved on a subsequent step.

[51]This pre-selection is efficient because it reduces the number of filing documents that need to be parsed, by early-removing records that that would only be discarded later.

[52]This is rather a complex task for *13D main documents* archives, as these are presented as semi-structured, non-standardized text.

[53]These are essentially *imperfectly* duplicated records referring to the same *event*. In the first *initial filing*, either some elements are missing, or there is incorrect information. The filer subsequently submits a second *initial filing*, which is the correct version. Note that, although the latter serves as a correction, it is still submitted as a *initial filing*, not as an amendment. Amendments, designated under a different code *(SC 13D/A)*, are used for updates, such as changes in investment position or objectives, rather than for correcting errors.

filings within a specified *pre-event window* (e.g., *20 days*, *3 months*, *6 months*).

The second group was derived from observed patterns concerning those filings that *sequentially targeted the same company* but were not characterized as *group filings*.[54] We then added dummy variables to indicate, among those filings, whether it referred to the *first episode* of the sequence or to a *subsequent* one. For the latter, we differentiated between those *under 6 months apart* and those that took longer to occur.[55] These dummies were later used as controls, along with the dummy that indicates *events obtained via consolidation*. They are particularly relevant for regressions in which *ownership* is the dependent variable.[56]

Finally, we eliminated any remaining *events* for which percentage ownership felt outside the 5%-50% interval.

Note that, while SEC data spans from 1994 to the present, empirical studies on activism typically limit the *event series* from 2001 onwards. By that time, electronic filing had become standard practice, and submissions contained far fewer errors than those observed in the early days of the system. However, in our paper, the *body of events* studied starts with those initiated in 2006, five years beyond the usual practice. We discard data points for those initial years due to a methodological constraint: information on *"filing topic"* in *8K filings*, necessary for flagging *non-core events*, is conveniently accessible only from mid-2005 onward. Note, though, that for dataset construction, we did collect and process data since 1994 to capture the historical *sequence of events* leading up to 2006. We trim the start of the dataset to 2006 only after assigning dummies related to *sequential filings*.

[54]Previously, we consolidated multiple filings referring to the same targeted company by indicating they belonged to a *group filing*. The remaining cases might, or might not, refer to *group filings*.

[55]These control variables consist of dummies for *events* obtained by aggregating multiple entries, a dummy for the *first event* in which a company is targeted (tracking starting from 1994, the beginning of the Edgar dataset), and dummies for subsequent *events* involving the same company, for periods both *below* and *above 6 months*. Therefore, the base case corresponds to *events* where companies are targeted only once and were not obtained by aggregation.

[56]As demonstrated in the original paper (Cruz, 2023), *ownership* is a variable prone to upward bias. While these controls do not likely eliminate all biases in the dataset, they do capture systematic variations that, as the author suggests, may result from double-counting or potential instances of seller-initiated episodes (e.g., private placements) that necessarily differ from buyer-initiated instances we are interested into.

1.2.2 Firm-related control variables

We incorporated *firm-specific controls* into our regressions that are standard in the literature. Summary statistics for those are provided in Table 1.1 for two distinct periods: *2006-2022* and *2010-2019*. All statistics presented, excluding the *Amihud Illiquidity Measure*, are drawn from financial data of the latest *reporting period* before the *13D event*, sourced from Compustat.[57] The *Amihud illiquidity measure* was computed using *trading volume* retrieved from CRSP for the leading *100-trading days* window preceding the *event*. Additional information on the computation for all *firm-level* statistics can be found in the table's caption.

Before proceeding, we mention two implications of the data extraction methodology used in this paper. First, even though we did not explicitly restrict *activist events* to those initiated by hedge funds, a quick examination following the removal of *non-core* cases reveals that most of the remaining *events* are filed by investors from that class. Second, despite applying exclusions theoretically equivalent to those described in reference papers[58] – removing corporations, non-US companies, reorganizations/bankruptcies, and insiders – our statistics differ noticeably from the ones reported in those studies. Although we do not explicitly investigate this discrepancy, overall, we obtained *abnormal returns* that appear more substantial than those documented in previous literature. Cruz (2023) interprets these larger figures as suggestive of a successful approach in detecting *non-core events*, compared to traditional, ad-hoc, manual methods.

1.2.3 Event related abnormal returns and turnover

As noted earlier, our statistics on *targeted stocks* differ markedly from those in the literature. Hence, to facilitate comparison, we present plots with a similar format to those found in other studies, displaying *daily averages* for both *excess returns* and *abnormal trading volume*.

Figure 1.1 depicts *daily abnormal returns* over a window of ± 20 *trading days*, centered around the *trigger date* for two periods, *2006-2022* (Panel A and B) and *2010-2019* (Panel C and D). For each period, we present equal-weighted (EW) and value-weighted (VW) *abnormal returns*

[57]The time lag between the conclusion of the reporting period and the *event date* varies, ranging on average from one fiscal year prior to less than a month, contingent on the data availability.

[58]Such as those conducted by Brav and coauthors (Brav et al. (2008); 2010; 2015; 2018; 2022.).

Table 1.1: Targeted firms fundamentals before trigger date

	2006-2022				*2010-2019*			
	count	*mean*	*50%*	*std*	*count*	*mean*	*50%*	*std*
ln market capitalization	2574	5.607	5.541	1.651	1391	5.649	5.573	1.675
book-to-market	2574	0.596	0.492	0.581	1391	0.609	0.501	0.610
tobin's Q	2574	1.838	1.443	1.269	1391	1.811	1.409	1.270
sales growth	2377	0.544	0.046	14.890	1298	0.832	0.038	20.052
ROA	2436	0.028	0.079	0.246	1321	0.035	0.079	0.229
cash flow	2432	-0.013	0.047	0.278	1317	-0.005	0.045	0.284
market leverage	2574	22.612	13.782	24.927	1391	22.744	13.476	24.975
book leverage	2574	36.510	25.192	45.966	1391	36.877	25.313	47.030
cash-to-assets	2574	24.148	14.842	25.050	1391	23.731	15.377	23.935
dividend yield	2574	0.621	0.000	1.928	1391	0.572	0.000	1.722
payout	2574	2.143	0.042	4.357	1391	2.005	0.080	3.994
profit margin	2499	-55.804	7.744	419.109	1357	-24.064	7.964	310.028
amihud illiquidity measure	2574	0.310	0.113	0.511	1391	0.293	0.108	0.483

This table shows summary statistics for the targeted companies fundamentals before the 13D trigger event (variables taken from the window -13 months to -1 month). Variables are winsorized at 1% and 99% levels. *Market capitalization* is in millions of dollars; *book-to-market* is *book value of equity/market value of equity*; *tobin's Q* is (*book value of debt + market value of equity)/(book value of debt + book value of equity)*; ROA is *EBITDA/lagged assets*; *cashflow* is *(net income + depreciation and amortization)/lagged assets*; *market leverage* is *total debt/(total debt + market value of equity)*; *book leverage* is *total debt/(total debt + book value of equity)*; *cash* is *cash + cash equivalents* scaled by assets; *dividend yield* is *common dividend/market value of equity*; *payout ratio* is *(common dividend + share repurchases)/market value of equity. Amihud illiquidity measure* is the yearly average (using daily data) of $10,000\sqrt{|return|/(dollar\ trading\ volume)}$.

using various pricing models as reference. Capital Asset Pricing model (CAPM), Fama-French 3 factors model (FF3), Fama-French 3 factors + momentum model (FFM) are all represented in the plot by **thin gray lines**; and Fama-French 5 factors model (FF5) is shown as a **thick blue line**.[59] *Factor loadings for the pricing models*,[60] as well as *average trading volume*, are computed over the leading *100-trading days* windows that precedes event windows ($t - 120$ to $t - 20$ days, where t is the trigger date).

As observed in Figure 1.1, *cumulative average abnormal returns* for the period 2006-2022 are in the range 12.1%-12.5% EW (Panel A) and 6% VW (Panel B). For the shorter period, 2010-2019, which excludes crises, *abnormal returns* are lower, 10.8% EW (Panel C) and 3% VW (Panel D). Note that the majority of the price uptick takes place on the *trigger date* ($t = 0$) and on the subsequent day ($t + 1$), which corresponds, respectively, to the date when the threshold is reached and the following day.

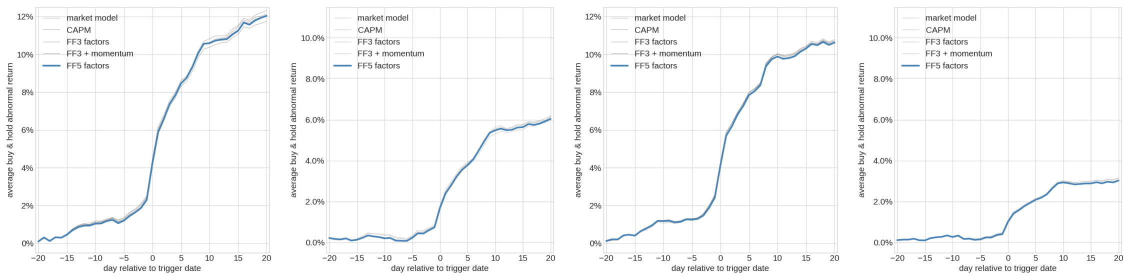
Figure 1.2 illustrates *abnormal turnover* around the *trigger date*.[61] The peaks therein align with the ones just observed; those relative to changes in prices. To be specific, they are more pronounced on the *event date* and on the next *trading day* ($t + 1$). Average *turnover* on those days, are approximately *5 times* and *4 times*, respectively, the average observed for the reference window.

As expected, VW figures, both for *abnormal return* and *abnormal turnover*, are much smaller than the corresponding EW ones. For example, when the averages are value-weighted (VW) *abnormal returns* are 6% and 3% for the periods 2006-2022 and 2010-2019, respectively. These values are significantly lower than the corresponding simple averages (12.5% and 10.8% EW). While other factors may contribute to this difference, lower VW *abnormal returns* are consistent with a well-documented characteristic of *activist events*. On average, *events* related to *larger* targeted companies, as measured in terms of *market capitalization*, tend to yield *lower abnormal*

[59] In the figure, we only observe minimal differences among distinct pricing models. However, note that the data points within the plots represent *averages*. These *averages* are derived from *raw abnormal returns* measured around different *event dates*, spanning a substantial temporal interval. Consequently, although variations can be substantial for individual *events*, they are mechanically smoothed out by *averaging*.

[60] Daily *stock prices* and Fama-French *factors* are sourced from WRDS.

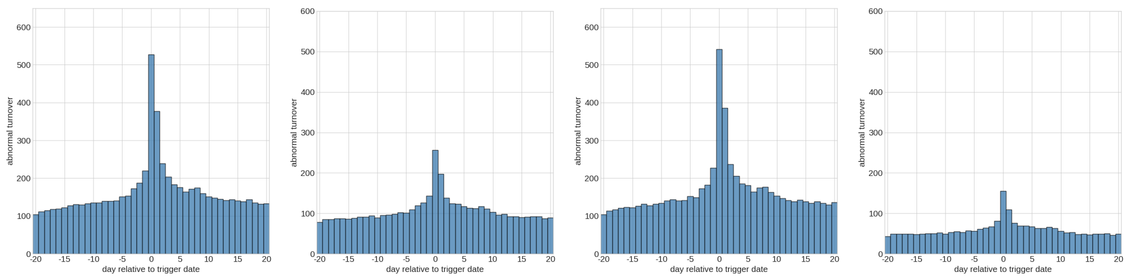
[61] In Table 1.2, we follow the same presentation structure used in Table 1.1. The first two panels refer to data spanning over 2006-2022, and the last two refer to data over 2010-2019. For each interval, we present EW and then VW figures.



(a) 2006-2022 (EW) (b) 2006-2022 (VW) (c) 2010-2019 (EW) (d) 2010-2019 (VW)

Figure 1.1: Cumulative abnormal returns around trigger date

This figure shows *abnormal returns* for the period 2006-2022 (Panels A and B) and for the period 2010-2019 (Panel C and D), centered around *event dates*. For each period we present plots for equal weighted (EW) levels (Panel A and C) and for value weighted (VW) levels (Panel B and D). Market capitalization to compute weights on Panel B and D have been winsorized to 97% level. Mean *abnormal returns* are shown for different pricing models: the **thick blue line** represents Fama French 5 factors (FF5) model. All the other models, see in the legend, are represented by **thin grey lines**. Loadings for each pricing model were computed using data of the *leading 100 trading* days window that precedes the event window (i.e. from $t - 120$ to $t - 20$).



(a) 2006-2022 (EW) (b) 2006-2022 (VW) (c) 2010-2019 (EW) (d) 2010-2019 (VW)

Figure 1.2: Abnormal turnover around trigger date

This figure shows *abnormal turnover* for the period 2006-2022 (Panels A and B) and for the period 2010-2019 (Panel C and D), centered around *event date*. For each period we present plots for equal weighted (EW) levels (Panel A and C) and for value weighted (VW) levels (Panel B and D). Market capitalization to compute weights on Panel B and D have been winsorized to 97% level. The reference for the abnormal turnover refers to the *leading 100 trading* days window that precedes the event window (i.e. from $t - 120$ to $t - 20$).

return and upticks in *daily turnover* are less pronounced.[62] Therefore, weighting returns based on value will lower the average figures.

In summary, Figures 1.1 and 1.2 are qualitatively similar to those documented in the literature for *excess returns* and *turnover* around *trigger dates*, as they display analogous patterns. However, the *events* in our sample command, on average, superior gains. As noted in Part 1.2.2, these larger figures are somewhat expected, as the data extraction methodology used in this paper is likely to be more stringent in excluding *non-core events* than those used in other studies.

Next, in Section 1.3, we use *abnormal returns* as dependent variable in event study regressions for different specifications. In all of them, we control for *firm-level* variables (as indicated in Part 1.2.2). Note that these controls include, among others, company size (i.e. log market capitalization). Additionally, we incorporate time fixed effects. Therefore, while Figure 1.1 helps illustrate *average abnormal returns'* daily progression around the *event date*, to interpret our regressions we take values from descriptive statistics tables.

In that regard, in concluding this section, we refer to Table 1.2, presenting statistics for *cumulative abnormal returns* over the ±20 *trading days* around the *trigger date*.[63] The *mean* figures, which can be also observed as the last EW data points (day +20) on Panels A and C of Figure 1.1, are notably larger than respective medians, indicating a highly negative skewness – instances of *lower abnormal returns* are more frequent. In addition, there is a notable dispersion, with the *standard deviation* being nearly half of the corresponding *mean*.

[62](Brav et al. (2022)) presents plots of *abnormal returns* for three groups of companies separated by market capitalization, and these *abnormal returns* move inversely with respect to the size groups.

[63]Once again, results are provided for four different pricing models (CAPM, FF3, FFM, FF5).

Table 1.2: Cumulative abnormal returns ±20 days around trigger date
descriptive statistics

		2006-2022				*2010-2019*			
pricing model	*variable*	*count*	*mean*	*50%*	*std*	*count*	*mean*	*50%*	*std*
CAPM	*capm_ar*	2362	0.125680	0.055241	0.531167	1288	0.107571	0.057662	0.320641
Fama-French 3 factors	*ff3_ar*	2362	0.124333	0.056644	0.500133	1288	0.107718	0.065130	0.323354
FF3 + momentum	*ffm_ar*	2362	0.121480	0.058345	0.494621	1288	0.108848	0.064939	0.325942
Fama-French 5 factors	*ff5_ar*	2362	0.123839	0.060077	0.506610	1288	0.107810	0.063755	0.326791

This table shows descriptive statistics for *cumulative abnormal returns* of targeted companies, using as a reference four different pricing models. Loadings for computing *abnormal returns* refer to the period $t-120$ to $t-20$ (*leading 100 trading days* window, that precedes the evaluation period).

1.2.4 Ownership stakes

Ownership stakes, as extracted from regulatory filings, are prone to biases and errors.[64] In an effort to mitigate distortions, we adopt the methodology outlined by Cruz, 2023, which incorporates *dummies* to the dataset, that are later used as controls to capture systematic variations that might have been introduced during data extraction (see Part 1.2.1).

Table 1.3 presents descriptive statistics for *ownership*, both in terms of *percentage of market capitalization* and *dollar values*. *Dollar ownership stakes* were computed by multiplying *ownership percentages* by the closing stock price on the *event date*. These computed values are approximations as, evidently, trades do not occur exclusively at closing prices. Moreover, for many *events*, acquisitions take place on multiple dates.

Table 1.3: Ownership stakes - descriptive statistics

		2006-2022				*2010-2019*			
unit	*variable*	*count*	*mean*	*50%*	*std*	*count*	*mean*	*50%*	*std*
% market cap	*pct*	2362	15.315329	9.900000	11.452902	1288	15.048373	9.900000	11.153149
log US dollars	*own_stake*	2362	8.095220	7.999047	1.825156	1288	8.124757	8.041968	1.820396

This table presents descriptive statistics for two measures of *ownership*. The first line refers to statistics for ownership as *percentage of market capitalization (pct)*, extracted from *13D filings* (some figures result from the aggregation of multiple filings into *single events*, as detailed in Part 1.2.1). The second line displays statistics for ownership in *US dollar (log)*, obtained by multiplying *percentage ownership* by *closing stock price* on the *event date* (source: CRSP).

[64] See Dlugosz et al. (2006) and Cruz (2023).

At first blush, considering that stock prices of targeted companies, on average, increase during the accumulation period there are two possible scenarios. These are contingent on whether most of the accumulation occurs before of after the *event date*: *events* for which most accumulation occurs before (after) *event date*, *dollar ownership stakes* are likely to be understated (overstated). This is evidently a point of attention, as systematic effects on *ownership* linked to our instrumental variable would be disastrous. In particular this might appear worrisome as one of the findings in this paper is that *ownership stakes* are influenced by market dynamics. However worries are not warranted and we explain the simple rationale for it in what follows.

First we should not confound two different variables: *market trend (absolute value)* and *deviation from market trend*, which will be characterized respectively in Part 1.2.5.2 and 1.2.5.3. For now it suffices to say, that the former, which is likely to have a direct role into biases into *ownership stakes* if these are acquired after the *event date*, is in fact used as control. So when we regress ownership against *deviation from market trend*, our instrument, those potential bias are already well addressed by those controls.

Now, besides the bias issue above,[65] there is another source of upwards bias that affect not only *dollar values*, but also *percentage ownership stakes*. This is the effect observed in Dlugosz et al., 2006 and likely to be, at least partially addressed by the aforementioned dummies related to multiple filings which are adopted on our regressions as well.

In conclusion, we acknowledge that ownership stakes are likely upward biased but we assume most of the bias is addressed. Now for any remaining biases, these should not impact the statistical significance of regression coefficients, though caution is needed when interpreting the magnitude of the results.

1.2.5 Market-related variables

Up to this point, we have solely presented descriptive statistics for variables related to *targeted companies* or to their respective *events*.[66] Those variables referred to data collected

[65]Related to bias related to the period of accumulation with respect to *event date*.

[66]The later refers to dummies created to address potential biases/errors related to *multiple filings* targeting the same company. They indicate whether *events* were obtained through filings aggregation or if they are preceded or followed by other filings targeting the same company that happen not to be characterized as group filings outright.

either within intervals around the *event date*, such as *abnormal returns* and *ownership stakes*, or referred to the latest information available, preceding the *event date*, as was the case for *firm-specific controls*.

Now we abstract from individual stocks and turn to *market returns*. Not only is the nature of the variables in this Part distinct (market instead of individual stocks), but also, although these observations are still collected relative to the *event date*, they no longer refer to its *leading days* or to intervals *centered* around it. Instead, they only refer to *lagging days*.

Limiting the observation interval of market-related variables to *lagging days* conforms to the specific questions we ask in this paper. For the sake of clarity, let's revisit our *first* research question, this time with a slight rephrasing so we can use it as an example:[67] Does activist *ownership*, as reported in *13D initial filings*, vary with *marginal changes* in *market prices* on the days following the *event date*? Stating the obvious, that question explicitly refers to market-related variables observed throughout the *lagging period*. The same applies as well to the original version of our *first* research question and to the *second* question too (i.e., whether changes in ownership cause higher returns), which in our setting is answered in a two-stage approach; the first stage precisely addresses the *first* question.[68]

Besides the lagging observation window, we also extend the *random walk assumption of market prices* to all the other market-related variables discussed in this paper. As a reminder, in Section 1.1, we introduced this assumption as the foundation for the validity of our instrumental variable. If this assumption holds, the specifications presented in next Section, 1.3, are likely to address potential endogeneity problems.

In what follows, we begin by presenting descriptive statistics for simpler measures of variation in market prices used in our specifications: *daily changes in market returns* and *market trend (absolute value)*. These cases set the stage for the construction of the instrumental variable, *deviation from market trend*, which is the first to last topic covered in this Section. We close with the quintile breaks for *deviation from market trend* which are used to construct dummies.

[67]Strictly speaking, the wording makes it slightly distinct from the question we are truly asking, but this is only to make the explanation simpler.

[68]This is done to manage the likely endogeneity issues with *abnormal returns* and *ownership*.

1.2.5.1 Daily market returns: *10-lagging days*

Earlier in this paper, we hypothesized that market dynamics influence activist's accumulation during the pre-disclosure (i.e., grace) period. The first approach we employ in our investigation is a simple setup where we regress *ownership* over *market returns* on each day comprising that lagging period. Table 1.4 displays descriptive statistics for *market returns* on each individual lagging day following the *event date* for the two distinct periods under consideration in our study.

Table 1.4: Daily market returns: *10-lagging days* after *event date*

	2006-2022				*2010-2019*			
	count	*mean*	*50%*	*std*	*count*	*mean*	*50%*	*std*
t+1	2576	0.000845	0.000920	0.013537	1391	0.000499	0.000748	0.009629
t+2	2576	0.000303	0.000985	0.014300	1391	0.000676	0.001120	0.010474
t+3	2576	0.000291	0.000624	0.013428	1391	0.000830	0.000719	0.009626
t+4	2576	0.000286	0.000861	0.014046	1391	0.000455	0.000760	0.010115
t+5	2576	0.000515	0.000854	0.013607	1391	0.000679	0.000805	0.009632
t+6	2576	0.000074	0.000555	0.014019	1391	0.000137	0.000579	0.009694
t+7	2576	-0.000546	0.000629	0.014000	1391	0.000456	0.000727	0.009929
t+8	2576	0.000364	0.000831	0.014299	1391	0.000360	0.000789	0.009719
t+9	2576	0.000298	0.000696	0.013535	1391	0.000254	0.000547	0.009780
t+10	2576	0.000378	0.000593	0.014334	1391	0.000657	0.000631	0.009614

This table shows descriptive statistics for daily market returns on *10-lagging days*, following the *trigger date*, for 3 different market references: CRSP universe value weighted (*vwretd*), CRSP universe equal weighted (*ewretd*) and S&P500 (*sprtrn*). Values are shown for the periods 2006-2022 and 2010-2019, the same intervals for which we run regressions.

1.2.5.2 Market trend (absolute value): *10-lagging days*

Next, we compute *market trends* over the entire *10-lagging days* interval.[69] Panel A of Table 1.5 presents descriptive statistics for their corresponding *absolute values*. We explicitly designate this variable as *"absolute value"*, to distinctively name it in contrast to the instrument introduced shortly, the later referred to as ***deviation from market trend***.

Table 1.5: Market trend - destriptive statistics

		2006-2022				*2010-2019*			
	variable	*count*	*mean*	*50%*	*std*	count	*mean*	*50%*	*std*
Panel A: *market trend (absolute value)*									
Equal weighted	*mkttrend_ew*	2362	0.003050	0.006676	0.042948	1288	0.005090	0.006770	0.028746
Value weighted	*mkttrend_vw*	2362	0.002396	0.007159	0.038202	1288	0.005061	0.007685	0.027189
S&P500	*mkttrend_sp*	2362	0.001819	0.006485	0.036966	1288	0.004587	0.006983	0.026711
Panel B: *Daily deviations form market trend*									
Equal weighted	*avg_dev_ewretd*	2362	0.000428	0.000158	0.013295	1288	0.000327	0.000332	0.009898
Value weighted	*avg_dev_vwretd*	2362	0.000450	0.000188	0.013521	1288	0.000397	0.000344	0.010392
S&P500	*avg_dev_sprtrn*	2362	0.000429	0.000310	0.013398	1288	0.000364	0.000289	0.010241

This table shows descriptive statistics for *market trend (absolute values)* (Panel A) and for *average daily deviation from market trend* (Panel B) using three different market references: CRSP universe value-weighted *(vwretd)*, CRSP universe equal-weighted *(ewretd)*, and SP500 *(sprtrn)*. Trends were computed over the *10 lagging trading days* following the *trigger date*, t (from t to $t + 10$). Please note that these statistics specifically refer to *market* returns, as opposed to the previous tables within this section that referred to abnormal returns of the *targeted companies*.

As we have seen, the assumption that *market prices* follow a *random walk* has implications for the two *market-related* variables discussed up to this point, *daily lagging market returns* and *market trend (absolute values)*. Although this implies that the distribution for those variables are random (each observation is a random shock), which is somewhat useful in exploring *ownership* and *abnormal returns* dynamics, they are too limited for addressing our research questions. In fact, we will use them in regressions to evaluate whether the outcomes we are interested in might

[69] This results in a single variable for each *event*, rather than ten, for daily market returns.

be influenced by variations in them. However, they mostly serve, particularly in *market trend*, as controls for *market trend levels*, as well as a stepping stone to construct a robust variable that credibly brings exogenous variations to activists' decisions, as we will explore next.

1.2.5.3 Deviation from market trend

In Section 1.1, we explored the motivation for creating a variable that could serve as a source of exogenous variation for *ownership*: the ex-post[70] *deviation from market trend*. In this part, we describe how this variable was constructed, how we interpret the values obtained, provide some examples and, finally, explain the mechanism that makes it an interesting candidate for source of exogenous variation in *ownership*.

Variable computation

The initial step in the computation is to obtain *market trend (absolute value)*, the variable introduced in Part 1.2.5.2. Remember we measure all *market-related variables* over the *10-lagging days interval* for each *event*; hence, *market trend* follows the same protocol. Then, for *each day*, we calculate the difference between the *observed cumulative return of the market* and the corresponding *cumulative return implied by the trend*, both using as reference the *event date*. The resulting difference represents the **deviation** from *observed market return* from the *trend* over that given day. We perform this calculation for all the *10-lagging days* within the ex-post period and subsequently aggregate these deviations (i.e. Riemann integral). We refer to the resulting variable simply as *deviation from market trend*. For convenience, we compute the daily average, so as to obtain a value with dimensions compatible with, and hence more easily comparable to, daily returns.

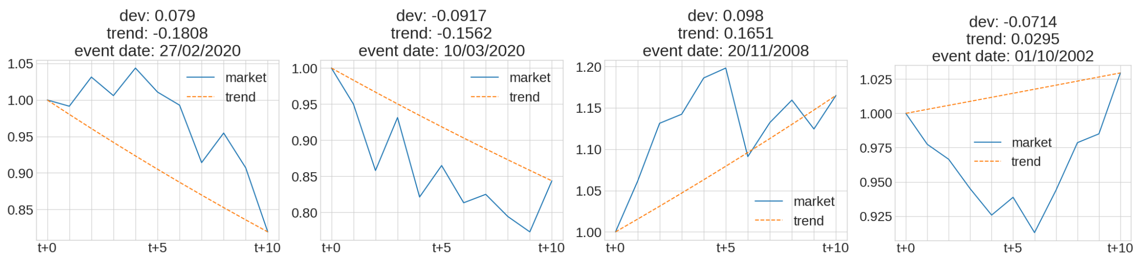
Variable interpretation

We interpret the values assumed by this variable as reflective of market dynamics on those *lagging days*. Its *sign* indicates, for *trend* as benchmark, whether *market prices* initially went up and then down (positive values) or first down and then up (negative values). Regarding its

[70]Ex-post with respect to the event date, meaning it refers to lagging observations.

absolute value, it is zero either when market prices do not deviate from the *trend* or when they oscillate around the *trend.* Large *absolute values* capture substantial deviations from the *trend* in single or multiple days, which are not reversed with similar intensity in the opposite direction on the following days.

We illustrate these concepts in Figure 1.3, both graphically and numerically, relying on four events drawn from our sample. The *events* therein were chosen so we could exemplify the fact that *deviation from market trend* can take on negative or positive *sign*, regardless of the *sign* of the *trend.* As evidenced in the figure, the *sign* obtained is determined by the relative path of *observed cumulative return* with respect to the correspondent *implied trend.* If the path observed for *market prices* is mostly above (below) the one implied by the *trend*, *market deviation from trend* assumes a positive (negative) value.



(a) (b) (c) (d)

Figure 1.3: Examples of deviations from market trend

This figure illustrates, for selected *events*, the *10-lagging days market trend* (dashed, orange lines) alongside the corresponding *daily market observed returns* (solid blue line). Computed values for *trend* and the average *deviation from trend* are displayed on top of each plot. This picture exemplifies that both deviation *sign* and *absolute value* result from market performance relative to the *10-lagging days* implied *trend.* Deviations can take a *positive sign* when *trends* are either positive or negative, as observed in Panels A and B. Similarly, deviations with *negative signs* can result either for positive or negative *market trends*, as seen in Panels C and D. Market in these plots corresponds to *CRSP universe* and are value-weighted.

Variable properties

As mentioned in Section 1.1, *market deviation from trend* corresponds to exogenous shocks that are plausibly unrelated to targeted stock ***abnormal*** *returns.* This is a fundamental assumption in our empirical design; without it, the instrument would lack validity for use in

2SLS regressions. The exogeneity assumption of *abnormal returns* is reasonably supported by the combined effect of the *random walk assumption* along with the fact that all models used to compute *abnormal returns* in our study incorporate *market returns* as a factor (including CAPM and the three Fama-French variations). For the sake of clarity, we emphasize that this would not be the case if, instead of ***abnormal*** *returns*, we were using ***absolute*** *returns*, as the later are highly correlated to *market returns*.[71]

Now, the other desirable feature of this variable is to be a credible source of exogenous variation for *ownership*. Once again, we lean on the *random walk assumption* of market prices to develop a rationale, as discussed to some length in 1.1. However, *random walk* alone does not make the case. For example, under our assumptions, the variable *market trend (absolute value)* is a *random walk sum*, so by the properties of sum of independent normal distributions it is a *random walk* as well. However, *trend* is too simple of a mechanism. Though its marginal change is likely to affect *ownership* to some extent, it lacks a compelling rationale to be considered as a exogenous source of variation. On the contrary, we have reasons to use, instead, the *trend* as a *control variable for trend levels* in our specifications.[72]

A stronger candidate for assuming the role of *source of exogenous variation* is the *deviation of market trend*. This variable in not only likely exogenous with respect to stocks *abnormal returns*, but it captures, as just discussed, the dynamics of stocks prices. It indicates if stocks are either cheap or expensive over the *10-lagging days* after the *event date*. We assume it as a reasonable source of exogenous variation, just like discussed in Section 1.1, because when activists perceive stocks as cheaper (more expensive), they are likely to increase (do not increase) stakes.

Finally, although at the risk of belaboring the point, we have a concluding consideration on exogeneity. We find it reasonable to assume that there are no systematic components in the error terms when regressing *ownership* against *market deviation from trend*. At least we can easily refute *simultaneity* and *reverse causality*.[73]

[71] This is a well-established fact in the economic finance literature. It has been consistently demonstrated over decades that the primary factor influencing stock prices, on average, is their co-movement with market prices. Alternatively, to put it simply, market prices are the main factor in *factor pricing models*.

[72] We have briefly discussed this case in Part 1.2.4. In that occasion we have argued that *market trends* are likely to have some implication on ownership dollar stakes, given the approximations we do to compute the later.

[73] As a sanity check, first and foremost, walks are obviously random. Second, it is reasonable to assume, as

Variable descriptive statistics and applications

We provide descriptive statistics for *deviations from market trend* over *10 lagging days*, in Panel B of Table 1.5. Like in other tables, we calculate statistics for the periods *2006-2022* and *2010-2019*. Additionally, we compute these figures for the same three market references (value-weighted, equal-weighted, and SP500), as we have also done for *absolute values*.[74]

As detailed shortly in Section 1.3, we use the variable *deviations from market trend* in two main ways. First, we investigate the hypothesis that activists' accumulation before their presence becomes common knowledge is influenced by pre-disclosure market dynamics. Indeed, our results support it, as we will see shortly. Subsequently, based on these results, we employ this variable as an instrument to address a classical problem of endogeneity found in activist studies, specifically concerning the relationship between *ownership* and *abnormal returns*.

1.2.5.4 Deviation from market trend: quintile breaks

In our investigation, we also employ alternative specifications by assigning dummies corresponding to quintiles of the *deviation from market trend*. Table 1.6 displays the quintile breaks used for constructing those dummies.

implied by the *random walk theory*, that the stock market as a whole is not affected by a single individual trade, including those related to *activism*. Though trading does affect the targeted stock own price, particularly for activist events, as evidenced in Figure 1.1.

[74]Shortly, in Section 1.3, we use figures from Panel A and Panel B concurrently for certain specifications; where *absolute values* is used as control for *market trend levels*, and the coefficient of interest is the one for *deviation from market trend*.

Table 1.6: Deviation from market trend: quintile breaks

	count	*0%*	*20%*	*40%*	*60%*	*80%*	*100%*
Panel A: *2006-2022*							
avg dev vwretd	2576	-0.091729	-0.007349	-0.001820	0.002291	0.008626	0.098033
avg dev ewretd	2576	-0.096285	-0.006827	-0.001711	0.002453	0.007785	0.117877
avg dev sprtrn	2576	-0.089300	-0.007447	-0.001682	0.002329	0.008419	0.093900
Panel B: *2010-2019*							
avg dev vwretd	1391	-0.047170	-0.006443	-0.001336	0.002072	0.006571	0.075993
avg dev ewretd	1391	-0.046876	-0.006039	-0.001247	0.002288	0.006477	0.076336
avg dev sprtrn	1391	-0.045093	-0.006123	-0.001293	0.001957	0.006390	0.070219

This table shows the quintile breaks for average daily deviations from the market trend for 3 different market references: CRSP universe value weighted (*vwretd*), CRSP universe equal weighted (*ewretd*) and S&P500 (*sprtrn*). All figures were computed for the *10-lagging days* after *event date*, t (from $t+1$ to $t+10$).*Panel A* covers *events* that took place over (2006-2022) while *Panel B* is a subsample of *Panel A* (2010-2019), that excludes crisis years.

1.3 Results

In this Section, we employ various regression specifications to explore our datasets in light of the research questions proposed in this paper and discuss the results in terms of both statistical and economic relevance. We start by examining how *market return* and *its relative movement with respect to the implied trend* are related to *ownership stakes*. Subsequently, we extend this evaluation for *abnormal returns*. Finally, we employ the 2SLS setup to address endogeneity concerns.

All tables presented in this Section display results for two distinct periods: *2006-2022* and *2010-2019*. The first period consists in the most extensive timeframe during which we could conveniently access information referring to some controls used in our regressions (see Part 1.2.1). The second period is a subset of the first, excluding initial and final years corresponding to economic crisis.

1.3.1 Ownership stakes

In this Part, we discuss our main findings regarding how *markets-related variables* impact blockholder activist *ownership stakes*.[75] Before jumping to the results, we discuss the rationale for selecting a *10-day lag* window for those, including *daily returns*, *trend*, as well as *its relative movement with respect to trend*. While we could have introduced this topic in the previous section, we opted to defer its discussion to just before presenting results, because we use those *market-related variables* as regressors. In particular, this delayed treatment draws attention to the importance of those coefficients that refer to the initial days of the *10-lagging days* interval, compared to the later ones.

Referring back to Section 1.2, we indicated that our data extraction followed the methodology proposed by Cruz, 2023. Given the primary nature of this paper is analytical (to investigate the proposed research questions) rather than descriptive, we only provided a general overview of the data collection process. For those interested in further details, we referred them to the original paper. However, the number of days in the *pre-disclosure accumulation period* deserves some additional attention, which we address in what follows.

In Section 1.1, we noted that the amended version of *Regulation 13D* (refer to Part 1.1.2) explicitly defined a *5-calendar days* deadline for filing activist *events*. However, the original text, the one in effect until *February 2024*, lacked the same level of clarity. The wording therein did not offer explicit guidance on how to count days. The issue was clarified later, on subsequent communication from SEC, part of a Q&A document, that specified that counting days should be in accordance with *calendar days*. Evidently, clarification on separate documents, as the case in point here, is not as effective as explicit reference in the regulation text.

In fact, the text ambiguity led to various interpretations on how to count those *(then) 10 days*. This resulted in significant variations in the number of days investors effectively waited till filing. Cruz (2023) illustrated this variation and revealed that, based on the sample therein, which spanned up to 2022 (inclusive), the average interval between the *event date* and the *filing date*

[75]The results presented herein are for ownership stakes *in dollars (log)*. As discussed in Part 1.2.4, *ownership stakes as a percentage of market capitalization* is a variable bounded both by left and right. The *dollar* alternative, especially when log-transformed, exhibits a smoother distribution, mitigating the impact of boundedness on coefficient estimation. Anyhow, we do incorporate regression results for *ownership as percentage of outstanding shares* on the Appendix A, for reference.

consisted of *8 trading days*. Note, though, that some filings presented *event dates* that postdates their respective *filing dates* (resulting in negative intervals), while a substantial number of cases exhibited intervals extending well beyond *20 trading days*. Those cases could correspond either to error, or to other alternative determinants; for example negative intervals could be associated to derivatives. However such conclusions could only be reached upon additional investigation. Given the exposed, the author opted for a *20-lagging trading days* cutoff, meaning that either *events* corresponding to *negative filing delays* or those that took over *20 trading days* to be filed, were dropped from the sample. The rationale was to incorporate only those *events* filed within a sound interval from the *event date*, given the prevalent information at that point.

The aforementioned debate surrounding the counting of the regulatory *10 days* was documented, for example, in Bebchuk et al., 2013. What emerged from their exposition is that the *de facto* deadline for disclosure considered by investors was unlikely to be *10 consecutive days*, but more. Given that additional evidence, we opted to extend the observation interval for *market-related variables* to *10-lagging* ***trading*** *days* after the *event date*. For consistency, we adopted the same interval to be the timeframe in which *market moves* plausibly influence activists' decisions to purchase additional stocks at a lower cost before their presence becomes widely known.

1.3.1.1 Effects of daily market returns on ownership

In Table 1.7, we present regression results for *dollar ownership stakes (log)* over daily *absolute market returns* for the *10-lagging days* after the *event date*. Coefficients that are statistically significant at the 1% level are found for selected *lagging days* on both periods studied (*2006-2022* and *2010-2019*). The results also hold when controls are included. Note that the obtained coefficients exhibit the expected *sign*; when the market goes up after the *event date*, it exerts an opposite effect on *ownership stakes* (they become smaller). However, the *lagging days* with significant coefficients are not the same for the two periods studied ($t + 3$ for the larger interval and $t + 1$ and $t + 5$ for the second one).

To evaluate the economic relevance of these effects, we refer back to Table 1.4, which provides descriptive statistics for these regressors. Without loss of generality, we will interpret only

Table 1.7: Regression: ownership *(dollar log)* over lagging daily market returns

Dependent variable: *ownership stake (log dollars)* *not controlled for size*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
intercept	8.0952***	8.0968***	7.6618***	8.6164***	8.3987***	8.1248***	8.1404***	7.6873***	8.9279***	8.7127***
	(0.0376)	(0.0729)	(0.0142)	(0.1782)	(0.1813)	(0.0507)	(0.1045)	(0.0187)	(0.1060)	(0.1293)
t+1		-0.6891	0.0712	0.9064	0.9576		-12.1198***	-12.0293***	-8.1898***	-8.3338***
		(2.7575)	(2.8017)	(2.7653)	(2.7976)		(3.2561)	(3.2264)	(2.9071)	(2.8738)
t+2		0.6362	1.1016	0.8276	1.1744		-4.6506	-3.6646	-4.0597	-3.8441
		(2.2275)	(2.0194)	(2.2741)	(2.2588)		(5.1956)	(5.2198)	(3.4396)	(3.2974)
t+3		-4.2211*	-3.5363	-4.4734**	-3.9788**		-3.0292	-3.3002	-2.2645	-2.5063
		(2.4657)	(2.4203)	(1.8078)	(1.8138)		(5.2493)	(5.2590)	(3.0955)	(2.9112)
t+4		-1.9351	-1.3849	-2.4458	-2.1709		-0.7939	-0.5569	1.1467	0.7095
		(1.9428)	(1.9068)	(2.0166)	(1.9775)		(4.4744)	(4.3771)	(3.2682)	(3.3085)
t+5		-0.0201	1.4979	1.1158	1.5584		-12.1023*	-9.4600*	-11.1424**	-9.8596**
		(3.9148)	(3.4508)	(3.1637)	(2.9987)		(6.4478)	(5.6229)	(4.9934)	(4.5722)
t+6		5.5753*	6.1990**	2.3822	2.7294		2.8835	3.6134	-0.8038	-0.3769
		(2.9111)	(2.7631)	(2.2710)	(2.3908)		(4.9570)	(4.3567)	(2.2557)	(2.4028)
t+7		1.7459	2.1607	1.3317	2.2012		0.8560	1.4425	0.2378	0.1635
		(2.5032)	(2.3575)	(1.7271)	(1.8097)		(5.3994)	(4.3699)	(4.0072)	(3.8193)
t+8		-0.1925	0.9618	2.0172	2.4569		-3.0696	-2.2515	3.9113	3.6604
		(1.9308)	(1.6615)	(1.8591)	(1.7313)		(5.4602)	(5.4478)	(4.6862)	(4.6741)
t+9		-1.5314	-0.3155	1.7046	2.0585		-2.1179	-2.2436	-3.7397	-3.8858
		(3.2907)	(2.6973)	(2.5176)	(2.3369)		(3.9043)	(3.7774)	(3.3397)	(3.1405)
t+10		3.3971	4.2536	3.5253**	3.9557**		6.3287	6.4494	3.8370	3.4754
		(3.8608)	(3.6400)	(1.7277)	(1.9022)		(5.6417)	(5.3320)	(2.3361)	(2.4184)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
controls	N	N	N	Y	Y	N	N	N	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0037	0.0339	0.3925	0.4006	0.0000	0.0110	0.0398	0.4282	0.4358
R-squared adj.	0.0000	-0.0005	0.0232	0.3854	0.3894	0.0000	0.0032	0.0254	0.4159	0.4195
number of observations	2362	2362	2362	2362	2362	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership (in dollar (log) over the lagging daily market returns (value weighted), with reference to the trigger date (day which activist investor passes the 5% threshold).
Columns 1 to 5 refers to the full period for which we have extracted flags from *8K filings (2006-2022)*. Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

the economic meaning for the $t+1$ coefficient for the interval *2010-2019*. On average, for the interval *2010-2019*, when the market goes up on the day following the trigger date ($t+1$) by one standard deviation (around 120 bps), the average dollar ownership stake decreases by 1.2%. The other two meaningful results ($t+3$ for *2006-2022* and $t+5$ for *2010-2019*) hold similar interpretations. Note that we are regressing variables with very low standard deviations against logs.

These results suggest that activists exhibit price sensitivity; given marginal decrease in stock prices, they tend to purchase more targeted stocks. As discussed earlier, this effect becomes more pronounced during the initial days of the *grace period*, when it is more likely that most activists have not yet submitted the *13D filing* and might be benefiting from target stock prices that were not updated. However, regressing ownership against market return for each individual lagging day is a bit audacious. The observed timing of pre-disclosure accumulation can vary considerably; some activist might use the first, second, third day, and so forth. Eventually, some will use all of those days to increase their stakes, while others may not increase their stakes at all.

This regression against daily market returns was more of an exploratory exercise. As we have discussed in depth in Section 1.2, the variable we have devised that is likely to be a credible *source of exogenous variation* for *ownership* is *deviation from market trend*. We see the results for using those variables as regressors next.

1.3.1.2 Regression on *deviation from market trend*

Now we use *deviation from market trend* in our empirical setting to investigate whether activists respond to market movements during the *grace period*. Remember we have discussed about the properties of this variable, and we have claimed that it can indicate whether market returns went first up and then down, or vice-versa, during the 10-lagging days after the event.

Table 1.8 shows regression results of ownership stakes in dollar (log) over *deviations from market trend*, for market reference are CRSP stocks, and returns are value weighted. For the period that includes crisis *(2006-2022)*, when *average daily market returns* goes up by 1 standard deviation (around 130 bps), the *dollar size of the activist ownership stake* reduces by 1.34%

(opposite direction). This result is statistically significant and is robust to periods with or without crisis and different market references. The table presented here is for *value weighted market returns*, but we provide in the Appendix A tables with regressions for CRSP universe (equal weighted) and for S&P500, both yielding similar conclusion.

Now, for the economic effect, at first blush it might seem small. A simple calculation using the coefficient of the regression constant from column 1 of Table 1.8 and reverting the log, gives an average *ownership stake* of approximately USD 3 million.[76] If we take 1.34% of it, it means for 1 standard deviation on average *deviation from market trend* gives USD 34.000 dollars effect on the opposite direction, a quite small amount. *Deviation from market trend* leads to economically relevant amount only for large deviations - from up to 5 standard deviations up.

Hence it is reasonable to expect in order to change their pre-disclosure accumulation behaviour (increase block size or give up acquiring extra shares during the *grace period*) that only large changes in market prices, not the ones of the size of 1 standard deviation, would reasonably affect the prices of the target shares to be material. In other words, given larger deviations with respect to market trend, the pre-disclosure accumulation could change considerably. This is exactly what we investigate next by regressing *ownership* against deviation quintiles.

1.3.1.3 Regression on *deviation from market trend quintiles*

For each period studied, we aggregate the *deviation from market trend* into quintiles, and assign each event to one of those bins. We refer back to Table 1.6 for the quintile breaks used for creating quintile bins.

Table 1.9 shows the regression of *dollar ownership stakes (log)* over quintiles of *deviation from market trend*, using CRSP value weighted returns. The reference quintile is the first one (dropped dummy). Notice that quintiles 2, 3, and 4 are not statistically significant. However for larger positive deviations, the ones on the fiftieth and last quintile, are large and statistically significant. The interpretation of this result is the following: if stock prices go up, with respect to the market trend during the pre-disclosure period, activist investors give up adding more shares to their blocks. This effect is not only statistically significant but also economically

[76]this value is not at constant dollars

Table 1.8: Regression: ownership stake *(log dollars)* over market trend
(absolute value and mean daily deviation) - value weighted
not controlled for size

Dependent variable: ownership stake (log dollars) **not controlled for size**

Panel A: *2006-2022*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	8.0952***	8.7254***	9.5464***	8.7413***	9.6053***	8.7233***	9.5090***	8.7398***	9.5610***
	(0.0376)	(0.1967)	(0.1520)	(0.1980)	(0.1534)	(0.1959)	(0.1415)	(0.1977)	(0.1441)
avg_dev_vwretd				-4.8766***	-5.1350***			-5.2950***	-5.6910***
				(1.2954)	(1.2724)			(1.2905)	(1.3083)
mkttrend_vw						0.6831	0.9324	0.9678	1.2639
						(0.8499)	(0.9407)	(0.8272)	(0.9318)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.3907	0.4065	0.3920	0.4080	0.3909	0.4069	0.3924	0.4086
R-squared Adj.	0.0000	0.3861	0.3961	0.3871	0.3972	0.3860	0.3962	0.3872	0.3976
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	8.1248***	9.0200***	8.9338***	9.0307***	8.9403***	9.0323***	8.9576***	9.0399***	8.9601***
	(0.0507)	(0.1032)	(0.1989)	(0.1055)	(0.2062)	(0.1010)	(0.2004)	(0.1029)	(0.2058)
avg_dev_vwretd				-5.7351**	-5.5001**			-5.0604**	-4.7863**
				(2.4502)	(2.2311)			(2.4532)	(2.2516)
mkttrend_vw						-2.0218*	-2.1548**	-1.7336	-1.8793*
						(1.1920)	(1.0939)	(1.2185)	(1.1271)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.4218	0.4371	0.4229	0.4381	0.4227	0.4381	0.4235	0.4388
R-squared Adj.	0.0000	0.4136	0.4223	0.4142	0.4228	0.4141	0.4229	0.4144	0.4231
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership stake (log dollars) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table 1.9: Regression: ownership (dollar log) over quintiles
(deviation of market trend - vw)

Dependent variable: *ownership stake (log dollars)* *not controlled for size*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	8.0952***	8.1368***	7.7261***	8.7518***	8.5346***	8.1248***	8.2878***	7.8158***	8.9765***	8.7637***
	(0.0376)	(0.1281)	(0.0746)	(0.1778)	(0.1821)	(0.0507)	(0.1787)	(0.0839)	(0.1466)	(0.1600)
quintile[2]		0.0211	0.0154	-0.1450*	-0.1103		-0.0960	-0.0524	-0.0464	-0.0251
		(0.1229)	(0.1183)	(0.0836)	(0.0879)		(0.1666)	(0.1300)	(0.0955)	(0.0767)
quintile[3]		-0.1076	-0.1168	-0.1842**	-0.1568**		-0.2337	-0.2049	-0.0650	-0.0457
		(0.1134)	(0.1178)	(0.0733)	(0.0793)		(0.1548)	(0.1384)	(0.0870)	(0.0831)
quintile[4]		0.0067	-0.0123	-0.0683	-0.0410		-0.1904	-0.1863	-0.0033	0.0093
		(0.1229)	(0.1207)	(0.0893)	(0.1016)		(0.1597)	(0.1347)	(0.1048)	(0.0912)
quintile[5]		-0.1319	-0.1076	-0.2205***	-0.2139***		-0.3014**	-0.2577*	-0.2417**	-0.2104*
		(0.1024)	(0.0992)	(0.0669)	(0.0655)		(0.1441)	(0.1426)	(0.1216)	(0.1231)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
controls	N	N	N	Y	Y	N	N	N	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0012	0.0308	0.3916	0.3992	0.0000	0.0034	0.0334	0.4228	0.4309
R-squared Adj.	0.0000	-0.0005	0.0225	0.3862	0.3896	0.0000	0.0003	0.0235	0.4132	0.4173
number of observations	2362	2362	2362	2362	2362	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership (in dollar (log) over the deciles of average daily deviation from market trend, using value weighted market returns. The trend is computed using 10 lagging trading days from trigger date.

Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

significant. Coefficients vary from -0.21 to -0.24, depending on the period considered and on the controls used. For the coefficient -0.21, when activist *triggering event* are followed by days where market raises considerably before disclosure (5th quintile), are on average, 18.94% smaller on absolute dollar values than the ones on the first quintile. These results are robust for alternative market reference (CRSP equal weighted or S&P500) and to periods with or without crisis. Additional tables with these robustness checks are provided in Appendix A.

1.3.2 Abnormal returns

So far, we have shown that activists are price sensitive and respond to market movements. In particular, they purchase more targeted stocks if market prices initially decrease and later increase, with respect to market trends.

The next question follows naturally: do changes in *ownership* impact firm value? A plausible hypothesis would be that it does impact positively, as with higher stakes involved, activists might have more power to persuade either the targeted company directly, or to aggregate fellow stockholders in their best be more engaged with the company and thus increase firm value more than if they purchase lower stakes.

Now, we benefit from the setting of using market deviations from the trend. Ideally, we compare targets in which the stock market remained flat over the ten days. The only difference is whether prices first went up or first went down. Then, any difference in firm value of target company can be assigned to the action of the activist.

1.3.2.1 Regression on *deviation from market trend*

We now explore *abnormal returns* in relation to *market movements*. Initially, we examine whether *abnormal returns* relate to deviations form market trend using simple regressions. Columns 4, 5, 8 and 9 of Table 1.11.[77] Not surprisingly, the coefficients for market fluctuations lack statistical significance, which contrasts starkly with our earlier findings concerning ownership stake as the dependent variable. While the absence of significant coefficients does not conclusively prove that market movements do not impact*abnormal returns*, it does aligns with a plausible explanation: investor reactions to activist presence is unaffected by whether market move first up and then down or vice versa. Moreover, these results are compatible with the research design we intend to use - if we find results for the instrumented case, this is strong evidence of marginal changes in ownership do affect the targeted stock valuation.

[77] In that table, we show, for reference columns *1* and *2* for which the only regressor is the constant, with and without controls respectively; on columns *6* and *7* we use as regressor only *market trend*, that is of interest as it is used as control for regression against market deviation, on columns *8* and *9*.

Table 1.10: IV Regression: abnormal returns on ownership stake
(natural logarithm of dollar amounts)

Dependent variable: *abnormal return (CAPM), ± 20 days, t_0=event date*

	*equal weighted**				*value weighted**				*S&P500**			
	2006-2022		*2010-2019*		*2006-2022*		*2010-2019*		*2006-2022*		*2010-2019*	
	2^{nd} stage	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*
intercept	-1.3948	9.5759	-0.4390	8.9837	-1.5776	9.5610	0.0117	8.9647	-1.4898	9.5413	0.0950	8.9563
	(-0.7825)	(67.260)	(-0.4314)	(51.261)	(-0.9603)	(69.032)	(0.0071)	(51.409)	(-0.9073)	(69.630)	(0.0450)	(51.573)
ownership stake (log US$)	0.1317		0.0536		0.1521		0.0031		0.1430		-0.0064	
	(0.7166)		(0.4637)		(0.8981)		(0.0166)		(0.8434)		(-0.0267)	
deviation from trend*		-6.1175		-6.6633		-5.6910		-4.7863		-5.2386		-4.0406
		(-3.7366)		(-3.3200)		(-4.5263)		(-2.2727)		(-4.0388)		(-1.6877)
market trend*	0.3096	1.3136	-0.4471	-2.4056	0.0440	1.2639	-0.5932	-1.8793	0.0249	1.2955	-0.5550	-1.6592
	(0.8047)	(1.4884)	(-0.9032)	(-2.1747)	(0.1331)	(1.4113)	(-0.9764)	(-1.7827)	(0.0779)	(1.3977)	(-0.8487)	(-1.6827)
R-squared	-0.0868	0.409	-0.0299	0.4401	-0.1280	0.4086	0.0427	0.4388	-0.1092	0.4083	0.0455	0.4383
Adj. R-squared	-0.1070		-0.0586		-0.1490		0.0159		-0.1298		0.0188	
Partial R-squared		0.0032		0.0022		0.0029		0.0013		0.0024		0.0009
F-statistic	1.76e+15	13.963	-6.399e+14	11.022	7.735e+14	20.487	-1.134e+15	5.1653	-1.66e+15	16.312	-8.409e+14	2.8482
P-value (F-stat)	0.0000	0.0002	1.0000	0.0009	0.0000	6.004e-06	1.0000	0.023	1.0000	5.373e-05	1.0000	0.0915
number of observations	2362	2362	1288	1288	2362	2362	1288	1288	2362	2362	1288	1288

This table presents IV regression results for abnormal returns regressed over ownership stakes (natural logarithm of dollar amounts); coefficients' t-statistics are presented in parentheses. The dependent variable is the cumulative ±20 abnormal return around the event date, with reference to the market model. Reference returns are the averages over the 100 trading days window ($t-121$ to $t-21$) preceding the evaluation window. Table **??** in the Appendix additionally includes coefficients and standard errors for the controls, that were omitted from this table, as well as for four other commonly used pricing models. The results are grouped based on market returns calculated using different methodologies (*). From left to right, they are CRSP equal-weighted, CRSP value-weighted, and S&P500. These market returns are used to compute the lagging 10-trading days market trend* (used as a control variable) and deviations from the trend* (the instrumental variable). The provided results cover each of the three market return categories and the two periods studied (2006-2022 and 2010-2019). For each period, both the second stage and the first stage of two-stage least squares regressions (2SLS) are included. The endogenous variable is ownership, and the instrument is the deviation from the market trend. The regressions are controlled for market trend, industry and time fixed effects, firm-level variables (pre-event), and variables related to the characteristics of the event (to control for biases in ownership data extraction).

1.3.2.2 Regression on ownership with *deviation from market trend* as instrument

The uncorrelatedness of market movement and *abnormal returns* is particularly interesting in conjunction with results in the previous sections for answering our next question: whether*abnormal returns* are influenced by the *ownership stake* acquired by the activist investor. As we saw in our previous discussion in the introduction, scholars affirm that the larger the stakes the more influence activists have over company decisions. So it is reasonable to expect that larger *stakes* should drive higher*abnormal returns* - understanding that these are composed of a component of buying pressure exerted by activist, but also as market reaction due to the presence of the activist. This problem, though, suffers from endogeneity as explained in the introductory section, 1.1: investors might buy larger stakes in companies they them to have better odds to increase value.

So far, we have seen that the relative[78] path of observed *market returns* on lagging days during the *grace period* matter for predicting the dollar stake of stock holdings. A natural

[78]Relative with respect to *market trend*.

next question to ask is whether increased investments meaningfully affect the power exerted by activists and thereby affect the value of the firm. To do this, we use *deviations from the market trend* as an instrument. Table 1.10 depicts the main results. For reference, please refer back to Table 1.5 for descriptive statistics of the regressors.

Results of IV regressions in Table 1.10 should be read as follows. The first stage corresponds to ownership regressed over deviations from market trend. As we have just seen, we find statistically significant results for these regressions. Then the second stage, regresses *abnormal return*, computed in ± 20 day window, centered around the activist *event* over teh residuals of the first stage. Abnormal returns are computed using CAPM, observed prices in $t-120$ to $t-20$ as reference to compute loadings.[79]

Notice that *abnormal returns* are estimated in a second stage where the value of the ownership stake measured in log US$ are used as a control and constitutes the variable of interest. For the first model, the estimated impact is 0.1. Importantly, the ownership stake measured in log US$ is an estimated value and the outcome of the first stage regression reported in the second column. For example, the value of -6.1175 implies that the log dollar value of the investments decreases by -6.1175 if the deviation from the market return increases by one unit.

Here, we use *deviation from market trend* as an instrument for *log dollar stake.* Accordingly, the control variable deviation from market return cannot (and does not) occur in the second stage regression. A first and important question is whether we have a sufficiently strong instrument to avoid the pitfalls of weak instruments. We rely on an analysis of the F-statistic in the first stage and observe that the rule of thumb, namely an F-statistic exceeding 10 is well fulfilled in our first stage regressions.

Turning to the interpretations of our results, we first see that *deviation from market trend* does a good job in predicting the *log dollar value of ownership stakes.* However, this increase in the dollar holdings of the activist investor does not imply an increase in firm value.

We have some possible interpretations for it. Though inconclusive results do not equal proving there is no relationship, once we did not find evidence it is reasonable to consider as a possibility that there is no effect indeed. This would be consistent to a theory where marginal

[79]We have reported also results for Fama-French 3, 4, and 5 factor models in the Appendix A, and they do not differ meaningfully from those presented here.

changes in ownership stakes do not matter, what really mater is to be targeted or not. This hypothesis appears even more plausible considering that we are assessing market value over the short term, specifically within ±20 days. It might well be that anticipation alone in the short run, do not respond to marginal changes in ownership. Whereas, a similar exercise might find significant results for long-term abnormal returns. Though in the later case, long run abnormal returns incorporates observable, tangible evidence that the activist is adding value to the targeted company.

Finally, eventually what worked well for the first stage might be inadequate for the second stage. Remember we decided for using *dollar stakes* earlier, as these variables are less affected by boundedness. Now, it sounds reasonable that *percentage ownership* is better suited to evaluate whether marginal increases in ownership affect prices. Hence we would have to change our investigation, by first using ownership in *percentage of market capitalization* for the first stage, which in turn would need adjustments to address the boundedness of these regressors.

x

1.4 Conclusion

In this paper, we study determinants and implications of activist *ownership*. We show that market movements on the trading days after the event, plausibly during the *grace period* (between *trigger date* and *filing date*) impact dollar value of ownership holdings.

We found evidence that active investors use the *grace period* to increase their dollar stakes, conditional on market movement. Using various specifications we showed that when market prices deviates considerably from the trend during the period following trigger date (last quintile), dollar ownership stakes are 20% lower. These results are robust to different market references (equal weighted, value weighted and S&P500) and for periods with and without financial crisis and the pandemics. We also found that for the period *2010-2019*, market absolute returns on day $t+1$ and $t+5$, when t is the trigger date, affects *dollar ownership* - but for those results to be economically relevant the market returns on those days has to be considerably high.

These findings are at odds with 2013 who concludes that activist blockholders do not

Table 1.11: Regression: abnormal returns *(CAPM)* over market trend
(absolute value and mean daily deviation) - value weighted

Dependent variable: *abnormal return (CAPM), ± 20 days, t_0=event date*

Panel A: *2006-2022*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1257***	0.0671	-0.1152	0.0715	-0.1039	0.0671	-0.1219	0.0717	-0.1109
	(0.0109)	(0.0766)	(0.0876)	(0.0744)	(0.0832)	(0.0766)	(0.0974)	(0.0739)	(0.0911)
avg_dev_vwretd				-0.7347	-0.7689			-0.7630	-0.8745
				(0.8605)	(0.8236)			(0.9754)	(1.0009)
mkttrend_vw						0.0239	0.1869	0.0651	0.2382
						(0.3099)	(0.4389)	(0.3560)	(0.4966)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0233	0.0337	0.0236	0.0341	0.0233	0.0339	0.0237	0.0343
R-squared Adj.	0.0000	0.0154	0.0162	0.0153	0.0161	0.0150	0.0159	0.0149	0.0160
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1076***	0.2086***	0.1348	0.2105***	0.1359	0.2161***	0.1464	0.2169***	0.1466
	(0.0089)	(0.0719)	(0.1027)	(0.0738)	(0.1041)	(0.0716)	(0.1016)	(0.0730)	(0.1024)
avg_dev_vwretd				-0.4418	-0.3212			-0.2343	-0.0820
				(0.9843)	(0.9021)			(1.0763)	(0.9662)
mkttrend_vw						-0.5533*	-0.6419**	-0.5401	-0.6372**
						(0.2895)	(0.2730)	(0.3393)	(0.3161)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0356	0.0460	0.0358	0.0461	0.0378	0.0487	0.0378	0.0487
R-squared Adj.	-0.0000	0.0212	0.0201	0.0206	0.0194	0.0226	0.0221	0.0219	0.0213
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (CAPM) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

systematically use the **grace period** to increase their stakes.

We credit our success in showing such relationship for two main reasons. First we use a restrictive methodology to identify events that reasonably can be associated to activists, while the cited paper do not discriminate among various *13D filings*. It is surely the case that many of those do not represent activists' events as understood by scholars. As their datasets includes improper datapoints, there is no surprise their results were inconclusive. Second, we devise a variable *deviation from market trend*, that has shown to be a reasonably strong source of variation on ownership, hence leading to a clean research design.

Among numerous possibilities for future research, those that explore *alternative dependent variables* as (i) *long-term abnormal returns*[80] and (ii) *operational performance*, sound promising. Activist's literature presents results for those variables, without claiming causal relationships, making them particularly interesting to be explored in our setting. At the most fundamental level, the drivers for *short-term abnormal returns*, the variable for which our instrument yielded inconclusive results, differ significantly from those influencing the *alternative variables* we intend to examine.[81] This distinction gives some hope that our identification strategy may yield conclusive results for these alternative specifications.

Since we did not conduct alternative experiments, there is not much left to explore beyond discussing potential, unobserved outcomes. Unfortunately, we cannot rule out the possibility that these exercises yield inconclusive results as well. But hopefully, instead, inferences for these alternative specifications will lead to findings, given economic and statistical significance.

Now, unlike our analysis for *short-term abnormal returns*, for which we had an intuition about the expected sign of the *instrumented ownership* coefficient (positive), the possibilities for *long-term outcomes* are not unequivocal. Marginal changes in *ownership*, a priori, might lead to various outcomes under the long-term perspective: from *positive* abnormal returns, as assumed in the Corporate Finance literature, to *negligible* abnormal returns (where marginal effects are economically insignificant but coefficients remain statistically relevant), or even *negative*

[80] Though we have explored different specifications, such as alternative *pricing models* and *market benchmarks*, regarding *abnormal returns*, we have not experimented beyond those computed over the ± 20 days window.

[81] While the former is significantly shaped by activists' own buying activity and market reactions anticipating their intervention, long-term outcomes are likely to reflect impact of concrete initiatives and observed performance resulting from activist interventions within the firm.

abnormal returns (value-destruction), as argued by detractors of activist investors. Given the opposing interests involved[82] and the uncertainty around the sign of marginal effects, this is indeed fertile ground for study.

[82] We elaborate on the contrasting perspectives on activists in Part 1.1.3 and 1.1.4.

References

Albuquerque, R., Fos, V., & Schroth, E. (2022). Value Creation in Shareholder Activism. *Journal of Financial Economics*, *145*, 153–178. https://doi.org/10.1016/j.jfineco.2021.09.007

Angrist, J., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. www.degruyter.com/document/doi/10.1515/9781400829828/html

Barry, J., Brav, A., & Jiang, W. (2020). Hedge Fund Activism Updated Tables and Figures. *Working paper*. https://people.duke.edu/~brav/HFactivismMarch2019.pdf

Bebchuk, L. A., Brav, A., Jackson, R. J., & Jiang, W. (2013). Pre-Disclosure Accumulations by Activist Investors: Evidence and Policy. *Journal of Corporation Law*, *39*, 1–34. http://www.sec.gov/news/speech

Brav, A., Jiang, W., & Kim, H. (2010). Hedge Fund Activism: A Review. *Foundations and Trends® in Finance*, *4*, 185–246. https://www.nowpublishers.com/article/Details/FIN-026

Brav, A., Jiang, W., & Kim, H. (2015). The Real Effects of Hedge Fund Activism: Productivity, Asset Allocation, and Labor Outcomes. *Review of Financial Studies*, *28*, 2723–2769. https://doi.org/10.1093/rfs/hhv037

Brav, A., Jiang, W., & Li, R. (2022). Governance by Persuasion: Hedge Fund Activism and Market-based Shareholder Influence. *ECGI working paper*. https://papers.ssrn.com/sol3/papers.cfm?abstractid=3955116

Brav, A., Jiang, W., Ma, S., & Tian, X. (2018). How Does Hedge Fund Activism Reshape Corporate Innovation? *Journal of Financial Economics*, *130*, 237–264. https://doi.org/10.1016/j.jfineco.2018.06.012

Brav, A., Jiang, W., Partnoy, F., & Thomas, R. (2008). Hedge Fund Activism, Corporate Governance, and Firm Performance. *Journal of Finance*, *63*, 1729–1775. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.2008.01373.x

CNBC. (2021). Engine no. 1 wins at least 2 exxon board seats as activist pushes for climate strategy change. https://www.cnbc.com/2021/05/26/engine-no-1-gets-at-least-2-candidates-elected-to-exxons-board-in-win-for-the-activist.html

Cochrane, J. H. (2001). Princeton University Press. %7Bhttps://press.princeton.edu/books/hardcover/9780691121376/asset-pricing%7D

Collin-dufresne, P., & Fos, V. (2015). Do Prices Reveal the Presence of Informed Trading? *Source: The Journal of Finance*, *70*, 1555–1582. https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12260

Cruz, P. P. G. (2023). Unveiling Non-Core Activist Blockholder Events. *Working Paper*.

Dlugosz, J., Fahlenbrach, R., Gompers, P., & Metrick, A. (2006). Large Blocks of Stock: Prevalence, Size, and Measurement. *Journal of Corporate Finance*, *12*, 594–618. https://doi.org/10.1016/j.jcorpfin.2005.04.002

Edmans, A., & Holderness, C. G. (2017). *Blockholders: A Survey of Theory and Evidence*. https://www.sciencedirect.com/science/article/abs/pii/S2405438017300029

Eldar, O., Kirmse, T., & Wittry, M. D. (2023). The Rise of Anti-Activist Poison Pills. *ECGI working paper*. http://ssrn.com/abstractid=4198367

Fama, E. F. (1995). Random walks in stock market prices. *Financial Analysts Journal*, *51*, 75–80. https://www.jstor.org/stable/4479810

Fink, L. (2015). *Letter to CEOs*. https://www.blackrock.com/corporate/investor-relations/2015-larry-fink-ceo-letter

Gantchev, N., & Giannetti, M. (2021). The Costs and Benefits of Shareholder Democracy: Gadflies and Low-Cost Activism. *Review of Financial Studies*, *34*, 5629–5675. https://doi.org/10.1093/rfs/hhaa128

Gillan, S., & Starks, L. (2007). The Evolution of Shareholder Activism in the United States. *Journal of Applied Corporate Finance*, *19*, 55–73. https://onlinelibrary.wiley.com/doi/10.1111/j.1745-6622.2007.00125.x

Goshen, Z., & Steel, R. (2022). Barbarians Inside the Gates. *The Yale Law Journal*, *132*, 411–486. https://www.yalelawjournal.org/article/barbarians-inside-the-gates

Klein, A., & Zur, E. (2009). Entrepreneurial Shareholder Activism: Hedge Funds and Other Private Investors. *Journal of Finance*, *64*, 187–229. https://doi.org/10.1111/j.1540-6261.2008.01432.x

Lilienfeld-Toal, U., & Schnitzler, J. (2020). The anatomy of block accumulations by activist shareholders. *Journal of Corporate Finance*, *62*. https://www.sciencedirect.com/science/article/abs/pii/S092911992030064X

Lipton, M. (2007). Some Thoughts for Boards of Directors in 2008. *AEI Leagal Center - Briefly*, *11*. https://ssrn.com/abstract=1090970

Mikkelson, W., & Ruback, R. (1985). An empirical analysis of the interfirm equity investment process. *Journal of Financial Economics*, *14*, 523–553. https://www-sciencedirect-com.proxy.bnl.lu/science/article/abs/pii/0304405X8590025X

Munk, C. (2013). *Financial Asset Pricing Theory*. Oxford University Press.

Ordóñez-Calafi, G., & Bernhardt, D. (2022). Blockholder Disclosure Thresholds and Hedge Fund Activism. *Journal of Financial and Quantitative Analysis*, *57*, 2834–2859. https://doi.org/10.1017/S0022109022000059

Shleifer, A., & Vishny, R. (1986). Large shareholders and corporate control. *Journal of Political Economy*, *94*, 461–488. https://www.jstor.org/stable/1833044

U.S. SEC. (2009). Questions and Answers of General Applicability for Exchange Act Sections 13(d) and 13(g) and Regulation 13D-G Beneficial Ownership Reporting (2009, November 16). http://www.sec.gov/divisions/corpfin/guidance/reg13d-interp.htm

U.S. SEC. (2022). Modernization of Beneficial Ownership Reporting - public consultation. https://www.sec.gov/rules/proposed.shtml

U.S. SEC. (2023). Modernization of Beneficial Ownership Reporting - Final Rule and guidance. https://www.sec.gov/files/rules/final/2023/33-11253.pdf

U.S. SEC. (n.d.). Rules and Regulations Under the Securities Exchange Act of 1934. https://www.ecfr.gov/current/title-17/chapter-II/part-240/subpart-A

Appendix A

Additional Tables and Figures

Table A.1: Regression: abnormal returns *(CAPM)* over market trend
(absolute value and mean daily deviation) - equal weighted

Dependent variable: *abnormal return (CAPM), ± 20 days, t_0=event date*

Panel A: *2006-2022*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1257***	0.0671	-0.1152	0.0701	-0.1047	0.0673	-0.1332	0.0714	-0.1201
	(0.0109)	(0.0766)	(0.0876)	(0.0742)	(0.0810)	(0.0751)	(0.1003)	(0.0720)	(0.0900)
avg_dev_ewretd				-0.5211	-0.5812			-0.6965	-0.8156
				(0.9080)	(0.9263)			(1.0693)	(1.1571)
mkttrend_ew						0.3444	0.4451	0.3754	0.4850
						(0.4148)	(0.5178)	(0.4582)	(0.5745)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0233	0.0337	0.0235	0.0339	0.0241	0.0349	0.0244	0.0353
R-squared Adj.	0.0000	0.0154	0.0162	0.0151	0.0160	0.0157	0.0170	0.0156	0.0170
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1076***	0.2086***	0.1348	0.2122***	0.1374	0.2177***	0.1510	0.2200***	0.1522
	(0.0089)	(0.0719)	(0.1027)	(0.0731)	(0.1040)	(0.0719)	(0.1002)	(0.0727)	(0.1011)
avg_dev_ewretd				-0.7796	-0.6851			-0.5946	-0.4479
				(0.8169)	(0.7428)			(0.8907)	(0.7877)
mkttrend_ew						-0.5032**	-0.6480***	-0.4758	-0.6265**
						(0.2541)	(0.2346)	(0.2915)	(0.2695)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0356	0.0460	0.0362	0.0464	0.0376	0.0490	0.0379	0.0492
R-squared Adj.	-0.0000	0.0212	0.0201	0.0210	0.0197	0.0224	0.0224	0.0220	0.0218
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (CAPM) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.2: Regression: abnormal returns *(CAPM)* over market trend
(absolute value and mean daily deviation) - S&P 500

Dependent variable: *abnormal return (CAPM), ± 20 days, t_0=event date*

Panel A: *2006-2022*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1257***	0.0671	-0.1152	0.0707	-0.1076	0.0670	-0.1208	0.0708	-0.1138
	(0.0109)	(0.0766)	(0.0876)	(0.0747)	(0.0844)	(0.0765)	(0.0962)	(0.0743)	(0.0920)
avg_dev_spretd				-0.6300	-0.6615			-0.6407	-0.7570
				(0.8073)	(0.7592)			(0.9131)	(0.9282)
mkttrend_sp						-0.0128	0.1646	0.0242	0.2121
						(0.2968)	(0.4294)	(0.3406)	(0.4853)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0233	0.0337	0.0235	0.0340	0.0233	0.0338	0.0235	0.0342
R-squared adj.	0.0000	0.0154	0.0162	0.0152	0.0160	0.0149	0.0159	0.0148	0.0158
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1076***	0.2086***	0.1348	0.2099***	0.1355	0.2145***	0.1434	0.2150***	0.1434
	(0.0089)	(0.0719)	(0.1027)	(0.0737)	(0.1038)	(0.0716)	(0.1017)	(0.0730)	(0.1024)
avg_dev_spretd				-0.3610	-0.2421			-0.1728	-0.0311
				(1.0537)	(0.9708)			(1.1494)	(1.0392)
mkttrend_sp						-0.5127*	-0.5807**	-0.5032	-0.5790*
						(0.2894)	(0.2704)	(0.3426)	(0.3165)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0356	0.0460	0.0358	0.0460	0.0374	0.0482	0.0374	0.0482
R-squared adj.	-0.0000	0.0212	0.0201	0.0205	0.0194	0.0222	0.0215	0.0215	0.0208
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (CAPM) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.3: Regression: abnormal returns *(Fama-French 3 factors)* over market trend *(absolute value and mean daily deviation) - equal weighted*

Dependent variable: *abnormal return (Fama-French 3 factors),* ± *20 days,* t_0*=event date*

Panel A: *2006-2022*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1243***	0.0667	-0.2385***	0.0718	-0.2222***	0.0667	-0.2430***	0.0720	-0.2272***
	(0.0103)	(0.0687)	(0.0753)	(0.0663)	(0.0693)	(0.0685)	(0.0863)	(0.0656)	(0.0760)
avg_dev_ewretd				-0.8534	-0.9049			-0.8840	-0.9816
				(0.9321)	(0.9520)			(1.1161)	(1.1982)
mkttrend_ew						0.0262	0.1107	0.0655	0.1587
						(0.4147)	(0.5071)	(0.4658)	(0.5712)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0221	0.0308	0.0227	0.0313	0.0222	0.0308	0.0227	0.0315
R-squared adj.	0.0000	0.0142	0.0132	0.0143	0.0134	0.0138	0.0129	0.0139	0.0131
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1077***	0.1915**	0.1191	0.1938**	0.1207	0.2022***	0.1370	0.2033**	0.1373
	(0.0090)	(0.0787)	(0.1143)	(0.0796)	(0.1151)	(0.0784)	(0.1114)	(0.0790)	(0.1119)
avg_dev_ewretd				-0.5031	-0.4041			-0.2775	-0.1369
				(0.9134)	(0.8455)			(0.9962)	(0.8955)
mkttrend_ew						-0.5933**	-0.7124***	-0.5805*	-0.7058**
						(0.2621)	(0.2476)	(0.3034)	(0.2854)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0352	0.0437	0.0354	0.0439	0.0379	0.0474	0.0380	0.0474
R-squared adj.	0.0000	0.0207	0.0178	0.0202	0.0172	0.0227	0.0207	0.0220	0.0200
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (Fama-French 3 factors) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.4: Regression: abnormal returns *(Fama-French 3 factors)* over market trend *(absolute value and mean daily deviation) - value weighted*

Dependent variable: *abnormal return (Fama-French 3 factors), ± 20 days, t_0=event date*

Panel A: *2006-2022*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1243***	0.0667	-0.2385***	0.0727	-0.2237***	0.0667	-0.2366***	0.0724	-0.2239***
	(0.0103)	(0.0687)	(0.0753)	(0.0664)	(0.0711)	(0.0692)	(0.0839)	(0.0665)	(0.0774)
avg_dev_vwretd				-0.9910	-1.0082			-0.9313	-1.0116
				(0.8740)	(0.8475)			(1.0075)	(1.0384)
mkttrend_vw						-0.1876	-0.0517	-0.1372	0.0076
						(0.3313)	(0.4469)	(0.3830)	(0.5099)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0221	0.0308	0.0229	0.0315	0.0223	0.0308	0.0230	0.0315
R-squared adj.	0.0000	0.0142	0.0132	0.0145	0.0135	0.0140	0.0128	0.0142	0.0131
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1077***	0.1915**	0.1191	0.1919**	0.1190	0.1992**	0.1306	0.1988**	0.1300
	(0.0090)	(0.0787)	(0.1143)	(0.0803)	(0.1152)	(0.0779)	(0.1129)	(0.0791)	(0.1132)
avg_dev_vwretd				-0.1000	0.0309			0.1221	0.2743
				(1.0696)	(0.9970)			(1.1659)	(1.0645)
mkttrend_vw						-0.5715**	-0.6328**	-0.5783*	-0.6484**
						(0.2698)	(0.2491)	(0.3245)	(0.2977)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0352	0.0437	0.0352	0.0437	0.0374	0.0464	0.0375	0.0464
R-squared adj.	0.0000	0.0207	0.0178	0.0200	0.0170	0.0222	0.0197	0.0215	0.0190
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (Fama-French 3 factors) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.5: Regression: abnormal returns *(Fama-French 3 factors)* over market trend *(absolute value and mean daily deviation) - S&P 500*

Dependent variable: *abnormal return (Fama-French 3 factors), ± 20 days, t_0=event date*

Panel A: *2006-2022*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1243***	0.0667	-0.2385***	0.0716	-0.2286***	0.0665	-0.2371***	0.0711	-0.2290***
	(0.0103)	(0.0687)	(0.0753)	(0.0667)	(0.0723)	(0.0691)	(0.0829)	(0.0668)	(0.0785)
avg_dev_spretd				-0.8520	-0.8648			-0.7888	-0.8710
				(0.8097)	(0.7728)			(0.9328)	(0.9550)
mkttrend_sp						-0.1882	-0.0409	-0.1426	0.0137
						(0.3224)	(0.4433)	(0.3710)	(0.5035)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0221	0.0308	0.0227	0.0313	0.0223	0.0308	0.0228	0.0313
R-squared adj.	0.0000	0.0142	0.0132	0.0143	0.0133	0.0140	0.0128	0.0140	0.0129
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1077***	0.1915**	0.1191	0.1915**	0.1188	0.1973**	0.1272	0.1967**	0.1265
	(0.0090)	(0.0787)	(0.1143)	(0.0803)	(0.1150)	(0.0779)	(0.1131)	(0.0792)	(0.1134)
avg_dev_spretd				0.0024	0.1336			0.1949	0.3395
				(1.1369)	(1.0654)			(1.2334)	(1.1343)
mkttrend_sp						-0.5039*	-0.5460**	-0.5146	-0.5647*
						(0.2673)	(0.2442)	(0.3242)	(0.2944)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0352	0.0437	0.0352	0.0438	0.0369	0.0456	0.0369	0.0457
R-squared adj.	0.0000	0.0207	0.0178	0.0199	0.0170	0.0217	0.0190	0.0209	0.0183
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (Fama-French 3 factors) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.6: Regression: abnormal returns *(FF 3 factors + momentum)* over market trend *(absolute value and mean daily deviation) - equal weighted*

Dependent variable: *abnormal return (Fama-French 3 factors + momentum), ± 20 days, t_0=event date*

Panel A: *2006-2022*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1215***	0.0662	-0.1893***	0.0705	-0.1752***	0.0662	-0.1919**	0.0705	-0.1785***
	(0.0102)	(0.0613)	(0.0669)	(0.0591)	(0.0622)	(0.0612)	(0.0782)	(0.0585)	(0.0683)
avg_dev_ewretd				-0.7276	-0.7827			-0.7356	-0.8337
				(1.0641)	(1.0886)			(1.2455)	(1.3273)
mkttrend_ew						-0.0155	0.0647	0.0172	0.1055
						(0.3963)	(0.4873)	(0.4509)	(0.5547)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0217	0.0299	0.0221	0.0303	0.0217	0.0299	0.0221	0.0304
R-squared Adj.	0.0000	0.0138	0.0123	0.0137	0.0123	0.0133	0.0119	0.0133	0.0120
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1088***	0.1790**	0.1291	0.1817**	0.1308	0.1911**	0.1489	0.1923**	0.1494
	(0.0091)	(0.0837)	(0.1183)	(0.0848)	(0.1193)	(0.0843)	(0.1168)	(0.0849)	(0.1173)
avg_dev_ewretd				-0.5718	-0.4613			-0.3168	-0.1636
				(0.9956)	(0.9168)			(1.0755)	(0.9607)
mkttrend_ew						-0.6705**	-0.7943***	-0.6559**	-0.7864***
						(0.2660)	(0.2531)	(0.3089)	(0.2915)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0334	0.0408	0.0337	0.0410	0.0368	0.0452	0.0369	0.0453
R-squared Adj.	-0.0000	0.0189	0.0148	0.0185	0.0142	0.0216	0.0186	0.0210	0.0178
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (Fama-French 3 factors + momentum) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.7: Regression: abnormal returns *(FF3 factors + momentum)* over market trend *(absolute value and mean daily deviation) - value weighted*

Dependent variable: *abnormal return (Fama-French 3 factors + momentum),* $\pm$ *20 days,* t_0*=event date*

Panel A: *2006-2022*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1215***	0.0661	-0.1868**	0.0708	-0.1764**	0.0661	-0.1868**	0.0708	-0.1764**
	(0.0102)	(0.0617)	(0.0758)	(0.0597)	(0.0701)	(0.0617)	(0.0758)	(0.0597)	(0.0701)
avg_dev_vwretd				-0.7605	-0.8273			-0.7605	-0.8273
				(1.0374)	(1.0746)			(1.0374)	(1.0746)
mkttrend_vw		-0.1936	-0.0701	-0.1525	-0.0216	-0.1936	-0.0701	-0.1525	-0.0216
		(0.3224)	(0.4407)	(0.3754)	(0.5042)	(0.3224)	(0.4407)	(0.3754)	(0.5042)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0219	0.0299	0.0223	0.0304	0.0219	0.0299	0.0223	0.0304
R-squared Adj.	0.0000	0.0136	0.0119	0.0136	0.0120	0.0136	0.0119	0.0136	0.0120
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1088***	0.1875**	0.1416	0.1873**	0.1411	0.1875**	0.1416	0.1873**	0.1411
	(0.0091)	(0.0836)	(0.1179)	(0.0849)	(0.1184)	(0.0836)	(0.1179)	(0.0849)	(0.1184)
avg_dev_vwretd				0.0447	0.2132			0.0447	0.2132
				(1.2468)	(1.1353)			(1.2468)	(1.1353)
mkttrend_vw		-0.6275**	-0.6923***	-0.6300*	-0.7044**	-0.6275**	-0.6923***	-0.6300*	-0.7044**
		(0.2722)	(0.2603)	(0.3280)	(0.3080)	(0.2722)	(0.2603)	(0.3280)	(0.3080)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0361	0.0439	0.0361	0.0439	0.0361	0.0439	0.0361	0.0439
R-squared Adj.	-0.0000	0.0209	0.0172	0.0201	0.0164	0.0209	0.0172	0.0201	0.0164
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (Fama-French 3 factors + momentum) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.8: Regression: abnormal returns *(FF3 factors + momentum)* over market trend *(absolute value and mean daily deviation) - S&P 500*

Dependent variable: *abnormal return (Fama-French 3 factors + momentum), ± 20 days, t_0=event date*

Panel A: *2006-2022*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1215***	0.0662	-0.1893***	0.0701	-0.1814***	0.0660	-0.1876**	0.0696	-0.1813**
	(0.0102)	(0.0613)	(0.0669)	(0.0600)	(0.0647)	(0.0616)	(0.0748)	(0.0599)	(0.0709)
avg_dev_spretd				-0.6845	-0.6880			-0.6190	-0.6851
				(0.8357)	(0.8066)			(0.9652)	(0.9942)
mkttrend_sp						-0.1833	-0.0495	-0.1475	-0.0065
						(0.3121)	(0.4378)	(0.3626)	(0.4989)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0217	0.0299	0.0220	0.0302	0.0219	0.0299	0.0221	0.0302
R-squared Adj.	0.0000	0.0138	0.0123	0.0137	0.0122	0.0135	0.0119	0.0134	0.0118
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1088***	0.1790**	0.1291	0.1793**	0.1289	0.1854**	0.1379	0.1850**	0.1374
	(0.0091)	(0.0837)	(0.1183)	(0.0856)	(0.1193)	(0.0835)	(0.1180)	(0.0849)	(0.1184)
avg_dev_spretd				-0.0955	0.0523			0.1148	0.2770
				(1.2279)	(1.1494)			(1.3186)	(1.2110)
mkttrend_sp						-0.5562**	-0.6013**	-0.5625*	-0.6166**
						(0.2683)	(0.2558)	(0.3263)	(0.3049)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0334	0.0408	0.0334	0.0408	0.0355	0.0431	0.0355	0.0431
R-squared Adj.	-0.0000	0.0189	0.0148	0.0182	0.0140	0.0202	0.0163	0.0195	0.0156
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (Fama-French 3 factors + momentum) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.9: Regression: Abnormal returns *(Fama-French 5 factors)* over market trend *(absolute value and mean daily deviation) - equal weighted*

Dependent variable: *abnormal return (Fama-French 5 factors),* $\pm$ *20 days,* t_0*=event date*

Panel A: *2006-2022*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1238***	0.0767	-0.2818***	0.0820	-0.2649***	0.0767	-0.2845***	0.0821	-0.2685***
	(0.0104)	(0.0716)	(0.0774)	(0.0691)	(0.0702)	(0.0717)	(0.0870)	(0.0688)	(0.0766)
avg_dev_ewretd				-0.8981	-0.9425			-0.9022	-0.9974
				(0.9098)	(0.9274)			(1.0685)	(1.1418)
mkttrend_ew						-0.0312	0.0646	0.0089	0.1134
						(0.4137)	(0.4990)	(0.4590)	(0.5572)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0219	0.0304	0.0225	0.0310	0.0219	0.0304	0.0225	0.0310
R-squared Adj.	0.0000	0.0140	0.0128	0.0141	0.0130	0.0136	0.0124	0.0137	0.0126
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1078***	0.1857**	0.0888	0.1886**	0.0909	0.1970**	0.1070	0.1984**	0.1077
	(0.0091)	(0.0904)	(0.1290)	(0.0913)	(0.1297)	(0.0902)	(0.1263)	(0.0907)	(0.1267)
avg_dev_ewretd				-0.6182	-0.5360			-0.3834	-0.2659
				(0.8540)	(0.7954)			(0.9339)	(0.8447)
mkttrend_ew						-0.6217**	-0.7262***	-0.6041*	-0.7134**
						(0.2723)	(0.2659)	(0.3120)	(0.3029)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0361	0.0452	0.0364	0.0454	0.0390	0.0489	0.0391	0.0490
R-squared Adj.	-0.0000	0.0216	0.0193	0.0212	0.0188	0.0238	0.0223	0.0232	0.0216
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (Fama-French 5 factors) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.10: Regression: abnormal returns *(Fama-French 5 factors)* over market trend *(absolute value and mean daily deviation) - value weighted*

Dependent variable: *abnormal return (Fama-French 5 factors), ± 20 days, t_0=event date*

Panel A: *2006-2022*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1238***	0.0767	-0.2818***	0.0828	-0.2672***	0.0767	-0.2787***	0.0823	-0.2664***
	(0.0104)	(0.0716)	(0.0774)	(0.0693)	(0.0728)	(0.0722)	(0.0851)	(0.0696)	(0.0786)
avg_dev_vwretd				-0.9922	-0.9984			-0.9155	-0.9853
				(0.8795)	(0.8544)			(1.0003)	(1.0282)
mkttrend_vw						-0.2258	-0.0873	-0.1764	-0.0295
						(0.3407)	(0.4490)	(0.3877)	(0.5079)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0219	0.0304	0.0226	0.0311	0.0222	0.0304	0.0228	0.0311
R-squared Adj.	0.0000	0.0140	0.0128	0.0142	0.0131	0.0138	0.0124	0.0140	0.0127
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
Intercept	0.1078***	0.1857**	0.0888	0.1863**	0.0890	0.1936**	0.1003	0.1934**	0.0999
	(0.0091)	(0.0904)	(0.1290)	(0.0921)	(0.1299)	(0.0897)	(0.1278)	(0.0909)	(0.1282)
avg_dev_vwretd				-0.1525	-0.0433			0.0733	0.1998
				(1.0374)	(0.9748)			(1.1334)	(1.0448)
mkttrend_vw						-0.5838**	-0.6361**	-0.5879*	-0.6475**
						(0.2850)	(0.2704)	(0.3373)	(0.3177)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0361	0.0452	0.0361	0.0452	0.0384	0.0478	0.0384	0.0478
R-squared Adj.	-0.0000	0.0216	0.0193	0.0209	0.0185	0.0232	0.0212	0.0224	0.0204
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (Fama-French 5 factors) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.11: Regression: abnormal returns *(Fama-French 5 factors)* over market trend *(absolute value and mean daily deviation) - S&P 500*

Dependent variable: *abnormal return (Fama-French 5 factors),* ± *20 days,* t_0*=event date*

Panel A: *2006-2022*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1238***	0.0767	-0.2818***	0.0816	-0.2723***	0.0765	-0.2795***	0.0809	-0.2718***
	(0.0104)	(0.0716)	(0.0774)	(0.0697)	(0.0741)	(0.0721)	(0.0841)	(0.0699)	(0.0798)
avg_dev_spretd				-0.8408	-0.8422			-0.7644	-0.8349
				(0.8230)	(0.7886)			(0.9322)	(0.9524)
mkttrend_sp						-0.2164	-0.0684	-0.1723	-0.0161
						(0.3362)	(0.4492)	(0.3796)	(0.5048)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0219	0.0304	0.0224	0.0309	0.0222	0.0304	0.0225	0.0309
R-squared Adj.	0.0000	0.0140	0.0128	0.0140	0.0129	0.0138	0.0124	0.0138	0.0124
number of observations	2362	2362	2362	2362	2362	2362	2362	2362	2362

Panel B: *2010-2019*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.1078***	0.1857**	0.0888	0.1858**	0.0886	0.1916**	0.0969	0.1911**	0.0964
	(0.0091)	(0.0904)	(0.1290)	(0.0921)	(0.1298)	(0.0897)	(0.1280)	(0.0910)	(0.1283)
avg_dev_spretd				-0.0341	0.0751			0.1615	0.2812
				(1.1123)	(1.0517)			(1.2086)	(1.1230)
mkttrend_sp						-0.5143*	-0.5499**	-0.5232	-0.5655*
						(0.2833)	(0.2660)	(0.3377)	(0.3146)
year fx effects	N	N	Y	N	Y	N	Y	N	Y
controls	N	Y	Y	Y	Y	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0361	0.0452	0.0361	0.0452	0.0378	0.0471	0.0378	0.0472
R-squared Adj.	-0.0000	0.0216	0.0193	0.0209	0.0185	0.0226	0.0204	0.0219	0.0197
number of observations	1288	1288	1288	1288	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (Fama-French 5 factors) mainly over 2 variables for a variety of specifications: average deviation of daily market trend for the lagging 10 days after trigger date, and the market trend itself, computed for the same interval. Column 1 is just regression against a constant. Columns 2 and 3 are regressions only on controls (Col 3 is also controlling for time fixed effects). Following, the pairs of columns (4-5, 6-7, 8-9) shows regressions for average daily deviation from market trend, market trend (absolute value) and both variables, respectively. Each pair shows results WO/W control for time fixed effects. All standard errors are clustered on the sic level. Controls include industry SIC classification. Panel A shows results for the period 2006-2022 and Panel B for 2010 to 2019.

Table A.12: Regression: ownership *(dollar log)* over average deviation from market trend - ew

Dependent variable: *ownership stake (log dollars)* — *not controlled for size*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	8.0952***	8.0961***	7.6780***	8.6310***	8.4271***	8.1248***	8.1280***	7.6813***	8.9106***	8.7011***
	(0.0376)	(0.0732)	(0.0014)	(0.1835)	(0.1920)	(0.0507)	(0.1065)	(0.0024)	(0.1023)	(0.1221)
avg dev ewretd		-2.0592	-2.1727	-5.0649***	-5.1101***		-9.8495*	-9.1216*	-7.4867***	-6.8925***
		(3.1142)	(2.9977)	(1.7892)	(1.7564)		(5.1033)	(5.0641)	(2.2308)	(2.2743)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
controls	N	N	N	Y	Y	N	N	N	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0002	0.0301	0.3911	0.3988	0.0000	0.0029	0.0329	0.4221	0.4304
R-squared Adj.	0.0000	-0.0002	0.0231	0.3864	0.3901	0.0000	0.0021	0.0254	0.4139	0.4182
number of observations	2362	2362	2362	2362	2362	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership (in dollar (log) over the average daily deviation from market trend, using equal weighted market returns. The trend is computed using 10 lagging trading days from trigger date.

Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

Table A.13: Regression: ownership *(dollar log)* over average deviation from S&P500 trend

Dependent variable: *ownership stake (log dollars)* *not controlled for size*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	8.0952***	8.0954***	7.6773***	8.6291***	8.4266***	8.1248***	8.1277***	7.6820***	8.9057***	8.6965***
	(0.0376)	(0.0731)	(0.0016)	(0.1833)	(0.1921)	(0.0507)	(0.1061)	(0.0029)	(0.1001)	(0.1204)
avg dev sprtrn		-0.3898	-0.4098	-4.2149***	-4.2023***		-8.0095*	-7.2465*	-4.9158*	-4.3164
		(2.4252)	(2.3112)	(1.2991)	(1.2428)		(4.1678)	(4.1035)	(2.9368)	(2.9157)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
controls	N	N	N	Y	Y	N	N	N	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0000	0.0299	0.3907	0.3984	0.0000	0.0020	0.0322	0.4212	0.4296
R-squared Adj.	0.0000	-0.0004	0.0228	0.3860	0.3896	0.0000	0.0013	0.0246	0.4130	0.4174
number of observations	2362	2362	2362	2362	2362	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership (in dollar (log) over the average daily deviation from market trend, using S&P500 market returns. The trend is computed using 10 lagging trading days from trigger date.

Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

Table A.14: Regression: ownership *(dollar log)* over quintiles *(deviation market trend - ew)*

Dependent variable: *ownership stake (log dollars)* *not controlled for size*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	8.0952***	8.1336***	7.7178***	8.7276***	8.5101***	8.1248***	8.2878***	7.8158***	8.9765***	8.7637***
	(0.0376)	(0.1333)	(0.0794)	(0.1831)	(0.1824)	(0.0507)	(0.1787)	(0.0839)	(0.1466)	(0.1600)
quintile[2]		0.0059	0.0013	-0.1051	-0.0749		-0.0960	-0.0524	-0.0464	-0.0251
		(0.1492)	(0.1444)	(0.1140)	(0.1150)		(0.1666)	(0.1300)	(0.0955)	(0.0767)
quintile[3]		0.0124	-0.0003	-0.0793	-0.0529		-0.2337	-0.2049	-0.0650	-0.0457
		(0.1202)	(0.1185)	(0.0813)	(0.0890)		(0.1548)	(0.1384)	(0.0870)	(0.0831)
quintile[4]		-0.0624	-0.0883	-0.1187	-0.0960		-0.1904	-0.1863	-0.0033	0.0093
		(0.1043)	(0.1024)	(0.0805)	(0.0830)		(0.1597)	(0.1347)	(0.1048)	(0.0912)
quintile[5]		-0.1507	-0.1191	-0.2153**	-0.2059**		-0.3014**	-0.2577*	-0.2417**	-0.2104*
		(0.1354)	(0.1356)	(0.0890)	(0.0892)		(0.1441)	(0.1426)	(0.1216)	(0.1231)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
controls	N	N	N	Y	Y	N	N	N	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0012	0.0306	0.3912	0.3988	0.0000	0.0034	0.0334	0.4228	0.4309
R-squared Adj.	0.0000	-0.0005	0.0224	0.3857	0.3893	0.0000	0.0003	0.0235	0.4132	0.4173
number of observations	2362	2362	2362	2362	2362	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership (in dollar (log) over the deciles of average daily deviation from market trend, using equal weighted market returns. The trend is computed using 10 lagging trading days from trigger date.

Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

Table A.15: Regression: ownership *(dollar log)* over quintiles *(deviation of S&P500 trend)*

Dependent variable: *ownership stake (log dollars)* *not controlled for size*

	2006-2022					2010-2019				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intercept	8.0952***	8.1368***	7.7261***	8.7518***	8.5346***	8.1248***	8.2930***	7.8299***	8.9728***	8.7663***
	(0.0376)	(0.1281)	(0.0746)	(0.1778)	(0.1821)	(0.0507)	(0.1966)	(0.1093)	(0.1464)	(0.1617)
quintile[2]		0.0211	0.0154	-0.1450*	-0.1103		-0.0548	-0.0249	-0.0033	0.0103
		(0.1229)	(0.1183)	(0.0836)	(0.0879)		(0.1764)	(0.1668)	(0.0828)	(0.0708)
quintile[3]		-0.1076	-0.1168	-0.1842**	-0.1568**		-0.2772	-0.2568	-0.0980	-0.0882
		(0.1134)	(0.1178)	(0.0733)	(0.0793)		(0.1824)	(0.1735)	(0.0976)	(0.0924)
quintile[4]		0.0067	-0.0123	-0.0683	-0.0410		-0.1753	-0.1756	-0.0080	0.0006
		(0.1229)	(0.1207)	(0.0893)	(0.1016)		(0.1648)	(0.1558)	(0.1019)	(0.0944)
quintile[5]		-0.1319	-0.1076	-0.2205***	-0.2139***		-0.3324**	-0.2859**	-0.2580*	-0.2263*
		(0.1024)	(0.0992)	(0.0669)	(0.0655)		(0.1509)	(0.1439)	(0.1365)	(0.1336)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
controls	N	N	N	Y	Y	N	N	N	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0012	0.0308	0.3916	0.3992	0.0000	0.0043	0.0343	0.4229	0.4310
R-squared Adj.	0.0000	-0.0005	0.0225	0.3862	0.3896	0.0000	0.0012	0.0245	0.4133	0.4175
number of observations	2362	2362	2362	2362	2362	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership (in dollar (log) over the deciles of average daily deviation from S&P500 returns. The trend is computed using 10 lagging trading days from trigger date.

Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

Table A.16: Regression: ownership *(dollar log)* over average deviation *(deviation of market trend - vw)*

Dependent variable: *ownership stake (log dollars)* — *controlled for size*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	8.0952***	8.0956***	7.6776***	2.5502***	2.6020***	8.1248***	8.1280***	7.6822***	2.6484***	2.7070***
	(0.0376)	(0.0731)	(0.0017)	(0.1084)	(0.1140)	(0.0507)	(0.1062)	(0.0030)	(0.1297)	(0.1382)
avg_dev_vwretd		-0.8822	-0.8746	-0.9309**	-0.9557**		-8.2621*	-7.5038*	-0.7489	-0.7251
		(2.5767)	(2.4627)	(0.4589)	(0.4663)		(4.3242)	(4.2620)	(1.0529)	(1.0174)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
controls	N	N	N	Y	Y	N	N	N	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0000	0.0299	0.8567	0.8573	0.0000	0.0022	0.0323	0.8583	0.8588
R-squared Adj.	0.0000	-0.0004	0.0229	0.8556	0.8551	0.0000	0.0014	0.0248	0.8562	0.8556
number of observations	2362	2362	2362	2362	2362	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership (in dollar (log) over the average daily deviation from market trend, using value weighted market returns. The trend is computed using 10 lagging trading days from trigger date.

Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

Table A.17: Regression: ownership *(dollar log)* over quintiles
(deviation of market trend - vw)

Dependent variable: *ownership stake (log dollars)* — *controlled for size*

	2006-2022					2010-2019				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intercept	8.0952***	8.1368***	7.7261***	2.5777***	2.6217***	8.1248***	8.2930***	7.8299***	2.6581***	2.7164***
	(0.0376)	(0.1281)	(0.0746)	(0.1190)	(0.1249)	(0.0507)	(0.1966)	(0.1093)	(0.1523)	(0.1588)
quintile[2]		0.0211	0.0154	-0.0239	-0.0095		-0.0548	-0.0249	-0.0053	0.0025
		(0.1229)	(0.1183)	(0.0365)	(0.0375)		(0.1764)	(0.1668)	(0.0538)	(0.0509)
quintile[3]		-0.1076	-0.1168	-0.0236	-0.0060		-0.2772	-0.2568	-0.0019	0.0051
		(0.1134)	(0.1178)	(0.0377)	(0.0438)		(0.1824)	(0.1735)	(0.0443)	(0.0526)
quintile[4]		0.0067	-0.0123	-0.0215	-0.0044		-0.1753	-0.1756	0.0184	0.0258
		(0.1229)	(0.1207)	(0.0488)	(0.0506)		(0.1648)	(0.1558)	(0.0784)	(0.0798)
quintile[5]		-0.1319	-0.1076	-0.0648*	-0.0669*		-0.3324**	-0.2859**	-0.0537	-0.0516
		(0.1024)	(0.0992)	(0.0365)	(0.0370)		(0.1509)	(0.1439)	(0.0574)	(0.0559)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
controls	N	N	N	Y	Y	N	N	N	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0012	0.0308	0.8568	0.8574	0.0000	0.0043	0.0343	0.8584	0.8589
R-squared Adj.	0.0000	-0.0005	0.0225	0.8555	0.8551	0.0000	0.0012	0.0245	0.8560	0.8554
number of observations	2362	2362	2362	2362	2362	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership (in dollar (log) over the deciles of average daily deviation from market trend, using value weighted market returns. The trend is computed using 10 lagging trading days from trigger date.

Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

Table A.18: Regression: ownership *(dollar log)* over lagging daily market returns

Dependent variable: *ownership stake (log dollars)* *controlled for size*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	8.0952***	8.0968***	7.6618***	2.5512***	2.6044***	8.1248***	8.1404***	7.6873***	2.6572***	2.7094***
	(0.0376)	(0.0729)	(0.0142)	(0.1054)	(0.1090)	(0.0507)	(0.1045)	(0.0187)	(0.1306)	(0.1405)
t+1		-0.6891	0.0712	0.8659	0.8329		-12.1198***	-12.0293***	0.5735	0.4513
		(2.7575)	(2.8017)	(1.1064)	(1.1316)		(3.2561)	(3.2264)	(1.8770)	(1.8862)
t+2		0.6362	1.1016	-0.8078	-0.7899		-4.6506	-3.6646	-2.5401*	-2.7144*
		(2.2275)	(2.0194)	(0.9544)	(0.9934)		(5.1956)	(5.2198)	(1.4231)	(1.3963)
t+3		-4.2211*	-3.5363	-2.1906***	-2.1173***		-3.0292	-3.3002	-1.9891	-2.2173
		(2.4657)	(2.4203)	(0.7619)	(0.7715)		(5.2493)	(5.2590)	(2.0840)	(2.0960)
t+4		-1.9351	-1.3849	-0.5871	-0.5598		-0.7939	-0.5569	3.0245	2.8810
		(1.9428)	(1.9068)	(1.4224)	(1.4419)		(4.4744)	(4.3771)	(1.8409)	(1.8847)
t+5		-0.0201	1.4979	0.7811	0.7830		-12.1023*	-9.4600*	-1.0751	-1.0576
		(3.9148)	(3.4508)	(1.1944)	(1.1880)		(6.4478)	(5.6229)	(2.5186)	(2.5174)
t+6		5.5753*	6.1990**	0.4507	0.4128		2.8835	3.6134	0.3761	0.3586
		(2.9111)	(2.7631)	(1.0469)	(1.0701)		(4.9570)	(4.3567)	(1.9557)	(1.9302)
t+7		1.7459	2.1607	-0.2489	-0.1697		0.8560	1.4425	0.3053	0.0502
		(2.5032)	(2.3575)	(0.7999)	(0.8287)		(5.3994)	(4.3699)	(1.7974)	(1.8650)
t+8		-0.1925	0.9618	0.3165	0.3157		-3.0696	-2.2515	1.7831	1.5882
		(1.9308)	(1.6615)	(0.8684)	(0.7945)		(5.4602)	(5.4478)	(1.9260)	(1.9180)
t+9		-1.5314	-0.3155	-0.0468	0.0587		-2.1179	-2.2436	-0.5520	-0.4777
		(3.2907)	(2.6973)	(1.2283)	(1.2227)		(3.9043)	(3.7774)	(1.0322)	(1.1075)
t+10		3.3971	4.2536	1.9230**	1.9670**		6.3287	6.4494	3.5728***	3.3947***
		(3.8608)	(3.6400)	(0.9213)	(0.9381)		(5.6417)	(5.3320)	(0.9191)	(0.9362)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
controls	N	N	N	Y	Y	N	N	N	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0037	0.0339	0.8573	0.8578	0.0000	0.0110	0.0398	0.8594	0.8599
R-squared Adj.	0.0000	-0.0005	0.0232	0.8556	0.8551	0.0000	0.0032	0.0254	0.8563	0.8557
number of observations	2362	2362	2362	2362	2362	1288	1288	1288	1288	1288

This table shows the coefficients and standard errors (in parenthesis) for regression of ownership (in dollar (log) over the lagging daily market returns (value weighted), with reference to the trigger date (day which activist investor passes the 5% threshold).

Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

Table A.22: IV Regression full table: abnormal returns on ownership stake *(% market capitalization) - continued*

Panel A: *Dependent variable: abnormal return (CAPM), ± 20 days, t_0=event date*

	*equal weighted**				*value weighted**				*S&P500**			
	2006-2022		*2010-2019*		*2006-2022*		*2010-2019*		*2006-2022*		*2010-2019*	
	2^{nd} stage	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*
intercept	-1.2146	23.290	-0.2903	17.711	-1.5261	23.170	-0.0854	17.718	-1.6076	23.093	-0.1095	17.702
	(-0.7469)	(13.427)	(-0.3474)	(7.5944)	(-0.9403)	(13.551)	(-0.0300)	(7.6738)	(-0.7934)	(13.493)	(-0.0119)	(7.7210)
ownership stake *(%)*	0.0470		0.0255		0.0611		0.0136		0.0647		0.0148	
	(0.6720)		(0.5379)		(0.8576)		(0.0840)		(0.7261)		(0.0283)	
deviation from trend*		-17.356		-17.563		-14.318		-6.0270		-11.703		-2.0989
		(-2.0585)		(-1.1241)		(-1.4320)		(-0.3760)		(-1.1182)		(-0.1133)
market trend*	0.3681	2.4863	-0.7279	3.9758	0.0370	3.2938	-0.6717	2.5367	-0.0126	3.4727	-0.6262	3.1850
	(0.8748)	(0.4688)	(-1.5780)	(0.2723)	(0.0922)	(0.5264)	(-1.8421)	(0.1754)	(-0.0310)	(0.5401)	(-0.4442)	(0.2306)
book-to-market	0.0140	-0.0623	-0.0024	0.1776	0.0146	-0.0618	0.0002	0.1768	0.0148	-0.0633	0.0002	0.1788
	(0.4174)	(-0.1803)	(-0.1261)	(0.4040)	(0.3955)	(-0.1788)	(0.0052)	(0.4008)	(0.3944)	(-0.1827)	(0.0017)	(0.4050)
cash-to-assets	0.0037	-0.0380	0.0019	-0.0392	0.0042	-0.0380	0.0014	-0.0394	0.0043	-0.0379	0.0015	-0.0394
	(1.3103)	(-3.4479)	(1.1041)	(-2.6194)	(1.5735)	(-3.4743)	(0.2354)	(-2.6291)	(1.3237)	(-3.4760)	(0.0734)	(-2.6322)
return-on-assets	0.0994	-4.6271	-0.0500	-4.3054	0.1641	-4.6405	-0.0999	-4.2717	0.1809	-4.6362	-0.0959	-4.2572
	(0.2495)	(-2.1404)	(-0.2057)	(-1.3323)	(0.3947)	(-2.1203)	(-0.1454)	(-1.3301)	(0.3789)	(-2.1126)	(-0.0434)	(-1.3301)
log market capitalization	0.0002	-0.0506	-0.0101	-0.3253	0.0011	-0.0491	-0.0136	-0.3238	0.0013	-0.0473	-0.0130	-0.3221
	(0.0185)	(-0.2402)	(-0.5956)	(-1.9025)	(0.0874)	(-0.2339)	(-0.2595)	(-1.9011)	(0.0963)	(-0.2252)	(-0.0777)	(-1.8966)
tobins' q	-0.0038	-0.1745	0.0163	-0.0743	-0.0013	-0.1750	0.0158	-0.0683	-0.0007	-0.1749	0.0159	-0.0671
	(-0.2192)	(-0.8706)	(1.0861)	(-0.2390)	(-0.0690)	(-0.8691)	(0.9460)	(-0.2205)	(-0.0318)	(-0.8674)	(0.4194)	(-0.2174)
profit margin	0.0000	0.0002	-0.0000	0.0005	-0.0000	0.0002	-0.0000	0.0005	-0.0000	0.0002	-0.0000	0.0005
	(0.0685)	(0.7746)	(-0.3975)	(1.6023)	(-0.0210)	(0.7485)	(-0.0956)	(1.6262)	(-0.0362)	(0.7446)	(-0.0351)	(1.6296)
cashflow	-0.0717	1.1596	0.2207	1.3931	-0.0898	1.1838	0.2348	1.3741	-0.0945	1.1826	0.2340	1.3695
	(-0.3081)	(0.8212)	(1.6660)	(0.7194)	(-0.3694)	(0.8173)	(0.9168)	(0.7110)	(-0.3816)	(0.8128)	(0.3177)	(0.7089)
market leverage	0.0004	0.0004	0.0015	-0.0148	0.0004	0.0004	0.0014	-0.0147	0.0004	0.0005	0.0014	-0.0146
	(0.3757)	(0.0278)	(1.1424)	(-0.7525)	(0.3062)	(0.0284)	(0.5092)	(-0.7646)	(0.2942)	(0.0374)	(0.1758)	(-0.7652)
book leverage	-0.0002	0.0012	-0.0007	0.0076	-0.0002	0.0012	-0.0007	0.0075	-0.0002	0.0011	-0.0007	0.0074
	(-0.3047)	(0.1840)	(-1.3226)	(1.2807)	(-0.2894)	(0.1842)	(-0.4829)	(1.2872)	(-0.2824)	(0.1764)	(-0.1645)	(1.2900)
dividend yield	0.0065	-0.1244	-0.0020	-0.0310	0.0082	-0.1235	-0.0023	-0.0293	0.0086	-0.1230	-0.0022	-0.0290
	(0.7076)	(-1.2648)	(-0.3706)	(-0.1262)	(0.7590)	(-1.2499)	(-0.3649)	(-0.1187)	(0.6727)	(-1.2461)	(-0.1316)	(-0.1175)
payout ratio	-0.0026	0.0599	-0.0004	-0.0417	-0.0034	0.0599	-0.0009	-0.0415	-0.0036	0.0598	-0.0009	-0.0413
	(-0.4247)	(0.9529)	(-0.1695)	(-0.4839)	(-0.4901)	(0.9531)	(-0.1601)	(-0.4778)	(-0.4493)	(0.9541)	(-0.0427)	(-0.4750)
sales growth	-0.0005	-0.0071	-0.0003	-0.0055	-0.0004	-0.0072	-0.0004	-0.0056	-0.0004	-0.0073	-0.0004	-0.0057
	(-0.9024)	(-2.6255)	(-1.2401)	(-2.3213)	(-0.6356)	(-2.7151)	(-0.4386)	(-2.2148)	(-0.4786)	(-2.7438)	(-0.1323)	(-2.2526)
amihud liquidity measure	0.0526	-0.9518	0.0026	-0.9982	0.0665	-0.9486	-0.0086	-0.9957	0.0699	-0.9419	-0.0072	-0.9936
	(0.7560)	(-2.2070)	(0.0578)	(-1.4066)	(0.9586)	(-2.2043)	(-0.0560)	(-1.3979)	(0.8213)	(-2.1844)	(-0.0141)	(-1.3950)
flag for multiple filings	-0.4284	12.628	-0.0345	8.7503	-0.6042	12.617	0.0685	8.7602	-0.6496	12.615	0.0576	8.7677
	(-0.4660)	(6.1431)	(-0.0810)	(4.1377)	(-0.6334)	(6.1294)	(0.0483)	(4.1226)	(-0.5487)	(6.1271)	(0.0126)	(4.1142)
multiple occurrence (1^{st})	-0.0753	0.2710	-0.1171	1.4195	-0.0785	0.2696	-0.1000	1.4272	-0.0795	0.2716	-0.1017	1.4288
	(-1.3298)	(0.3665)	(-1.1055)	(1.3152)	(-1.1526)	(0.3647)	(-0.3882)	(1.3177)	(-1.0807)	(0.3676)	(-0.1318)	(1.3186)
multiple occurrence (2^{nd} within 6MO)	0.0417	-1.0620	0.0472	-2.2203	0.0570	-1.0622	0.0198	-2.2053	0.0607	-1.0577	0.0226	-2.1993
	(0.6683)	(-1.1445)	(0.4548)	(-2.5214)	(0.9210)	(-1.1464)	(0.0558)	(-2.5043)	(0.8138)	(-1.1416)	(0.0197)	(-2.4943)
multiple occurrence (2^{nd} after 6MO)	0.0533	-0.5179	-0.0315	-0.3293	0.0604	-0.5143	-0.0353	-0.3351	0.0622	-0.5126	-0.0347	-0.3361
	(0.8079)	(-0.7784)	(-1.5543)	(-0.3595)	(0.9329)	(-0.7749)	(-0.8439)	(-0.3661)	(0.9052)	(-0.7738)	(-0.2137)	(-0.3676)
flag for notice of delisting	0.0072	1.8683	-0.0850	0.6328	-0.0190	1.8620	-0.0774	0.6464	-0.0257	1.8670	-0.0784	0.6484
	(0.0905)	(2.6441)	(-1.8488)	(0.5187)	(-0.1815)	(2.6204)	(-0.7030)	(0.5288)	(-0.1847)	(2.6285)	(-0.2292)	(0.5306)
R-squared	-0.9251	0.0386	-0.6736	0.0387	-1.5984	0.0385	-0.1486	0.0384	-1.7992	0.0384	-0.1872	0.0384
Adj. R-squared	-0.9617		-0.7218		-1.6477		-0.1817		-1.8523		-0.2213	
Partial R-squared		0.0004		0.0002		0.0003		3.1e-05		0.0002		3.658e-06
F-statistic	-8.262e+15	4.2373	-7.617e+15	1.2637	2.151e+15	2.0507	-1.778e+14	0.1414	4.354e+15	1.2504	-1.008e+15	0.0128
P-value (F-stat)	1.0000	0.0395	1.0000	0.261	0.0000	0.1521	1.0000	0.7069	0.0000	0.2635	1.0000	0.9098
number of observations	2362	2362	1288	1288	2362	2362	1288	1288	2362	2362	1288	1288

This table shows IV regression results for abnormal returns regressed over ownership stake (% of market capitalization). Values in parenthesis correspond to coefficients' t-statistics.
In all 5 panels, the dependent variable is *abnormal return* but with reference to a distinct pricing models. For example, *Panel A* shows regressions for which the dependent variable is abnormal return using the *market model*, while *Panel E* the dependent variable is abnormal return using the *Fama-French 5 factors model*.
Each panel contains 3 groups of results, each corresponding market returns calculated using different methodologies (*). From left to right, they are: *CRSP equal weighted*, *CRSP value weighted*, and *S&P500*. We use market returns to compute lagging 10-trading days market trend*(used as control variable and presented here for completeness) as well as deviations from the trend*(the instrumental variable). For each of the 3 market return categories, we present results for the 2 periods studied (2006-2022 and 2010-2019). For each period, we provide results for both the second stage and the first stage of two-stage least squares regressions (2SLS).
The endogenous variable is *ownership*, and the instrument is the *deviation from market trend*. Regressions are controlled for market trend, industry and time fixed effects, firm-level variables (pre-event), and variables related to characteristics of the event (to control for biases in ownership data extraction).
Loadings for the pricing models were computed using the 100 trading days window ($t-121$ to $t-21$) that precedes the evaluation window. The evaluation window used to compute abnormal returns is the 41 days trading window from $t-20$ to $t+20$.

Table A.22: IV Regression full table: abnormal returns on ownership stake *(% market capitalization) - continued*

Panel B: *Dependent variable: abnormal return (Fama-French 3 factors)*, ± *20 days*, t_0=*event date*

	*equal weighted**				*value weighted**				*S&P500**			
	2006-2022		*2010-2019*		*2006-2022*		*2010-2019*		*2006-2022*		*2010-2019*	
	2nd stage	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*
intercept	-1.5444	23.290	0.0072	17.711	-1.8609	23.170	0.9441	17.718	-1.9479	23.093	2.9973	17.702
	(-0.8996)	(13.427)	(0.0084)	(7.5944)	(-1.0815)	(13.551)	(0.2975)	(7.6738)	(-0.8999)	(13.493)	(0.1263)	(7.7210)
ownership stake *(%)*	0.0566		0.0078		0.0707		-0.0455		0.0744		-0.1617	
	(0.7584)		(0.1605)		(0.9294)		(-0.2491)		(0.7781)		(-0.1199)	
deviation from trend*		-17.356		-17.563		-14.318		-6.0270		-11.703		-2.0989
		(-2.0585)		(-1.1241)		(-1.4320)		(-0.3760)		(-1.1182)		(-0.1133)
market trend*	0.0181	2.4863	-0.7368	3.9758	-0.2251	3.2938	-0.5330	2.5367	-0.2447	3.4727	-0.0496	3.1850
	(0.0437)	(0.4688)	(-3.9500)	(0.2723)	(-0.5376)	(0.5264)	(-0.7138)	(0.1754)	(-0.5858)	(0.5401)	(-0.0103)	(0.2306)
book-to-market	0.0193	-0.0623	0.0044	0.1776	0.0200	-0.0618	0.0145	0.1768	0.0203	-0.0633	0.0356	0.1788
	(0.5208)	(-0.1803)	(0.2502)	(0.4040)	(0.4926)	(-0.1788)	(0.2952)	(0.4008)	(0.4929)	(-0.1827)	(0.1228)	(0.4050)
cash-to-assets	0.0039	-0.0380	0.0013	-0.0392	0.0044	-0.0380	-0.0008	-0.0394	0.0045	-0.0379	-0.0054	-0.0394
	(1.3925)	(-3.4479)	(0.7937)	(-2.6194)	(1.6417)	(-3.4743)	(-0.1239)	(-2.6291)	(1.3475)	(-3.4760)	(-0.1029)	(-2.6322)
return-on-assets	0.1352	-4.6271	-0.1289	-4.3054	0.2009	-4.6405	-0.3550	-4.2717	0.2185	-4.6362	-0.8500	-4.2572
	(0.3361)	(-2.1404)	(-0.6409)	(-1.3323)	(0.4785)	(-2.1203)	(-0.5224)	(-1.3301)	(0.4456)	(-2.1126)	(-0.1524)	(-1.3301)
log market capitalization	0.0008	-0.0506	-0.0153	-0.3253	0.0015	-0.0491	-0.0320	-0.3238	0.0017	-0.0473	-0.0692	-0.3221
	(0.0669)	(-0.2402)	(-0.8301)	(-1.9025)	(0.1078)	(-0.2339)	(-0.5214)	(-1.9011)	(0.1129)	(-0.2252)	(-0.1608)	(-1.8966)
tobins' q	0.0030	-0.1745	0.0195	-0.0743	0.0055	-0.1750	0.0162	-0.0683	0.0062	-0.1749	0.0085	-0.0671
	(0.1540)	(-0.8706)	(1.4701)	(-0.2390)	(0.2567)	(-0.8691)	(1.2176)	(-0.2205)	(0.2586)	(-0.8674)	(0.1054)	(-0.2174)
profit margin	-0.0000	0.0002	-0.0000	0.0005	-0.0000	0.0002	0.0000	0.0005	-0.0000	0.0002	0.0001	0.0005
	(-0.3181)	(0.7746)	(-0.1946)	(1.6023)	(-0.3586)	(0.7485)	(0.2142)	(1.6262)	(-0.3480)	(0.7446)	(0.1135)	(1.6296)
cashflow	-0.0624	1.1596	0.2395	1.3931	-0.0810	1.1838	0.3103	1.3741	-0.0856	1.1826	0.4702	1.3695
	(-0.2838)	(0.8212)	(2.2831)	(0.7194)	(-0.3505)	(0.8173)	(1.4274)	(0.7110)	(-0.3642)	(0.8128)	(0.2832)	(0.7089)
market leverage	0.0005	0.0004	0.0012	-0.0148	0.0005	0.0004	0.0005	-0.0147	0.0005	0.0005	-0.0012	-0.0146
	(0.3607)	(0.0278)	(0.9693)	(-0.7525)	(0.3045)	(0.0284)	(0.1886)	(-0.7646)	(0.2942)	(0.0374)	(-0.0637)	(-0.7652)
book leverage	-0.0002	0.0012	-0.0006	0.0076	-0.0002	0.0012	-0.0002	0.0075	-0.0002	0.0011	0.0007	0.0074
	(-0.3015)	(0.1840)	(-0.9672)	(1.2807)	(-0.2879)	(0.1842)	(-0.1216)	(1.2872)	(-0.2812)	(0.1764)	(0.0698)	(1.2900)
dividend yield	0.0071	-0.1244	-0.0026	-0.0310	0.0087	-0.1235	-0.0040	-0.0293	0.0092	-0.1230	-0.0073	-0.0290
	(0.6925)	(-1.2648)	(-0.9395)	(-0.1262)	(0.7438)	(-1.2499)	(-0.2428)	(-0.1187)	(0.6608)	(-1.2461)	(-0.1034)	(-0.1175)
payout ratio	-0.0022	0.0599	-0.0006	-0.0417	-0.0030	0.0599	-0.0029	-0.0415	-0.0032	0.0598	-0.0077	-0.0413
	(-0.3489)	(0.9529)	(-0.3802)	(-0.4839)	(-0.4197)	(0.9531)	(-0.3783)	(-0.4778)	(-0.3875)	(0.9541)	(-0.1457)	(-0.4750)
sales growth	-0.0003	-0.0071	-0.0004	-0.0055	-0.0002	-0.0072	-0.0007	-0.0056	-0.0002	-0.0073	-0.0014	-0.0057
	(-0.5091)	(-2.6255)	(-1.3478)	(-2.3213)	(-0.3004)	(-2.7151)	(-0.6152)	(-2.2148)	(-0.2113)	(-2.7438)	(-0.1700)	(-2.2526)
amihud liquidity measure	0.0605	-0.9518	-0.0117	-0.9982	0.0739	-0.9486	-0.0640	-0.9957	0.0773	-0.9419	-0.1791	-0.9936
	(0.8117)	(-2.2070)	(-0.2567)	(-1.4066)	(0.9762)	(-2.2043)	(-0.3570)	(-1.3979)	(0.8316)	(-2.1844)	(-0.1322)	(-1.3950)
flag for multiple filings	-0.5678	12.628	0.1136	8.7503	-0.7446	12.617	0.5797	8.7602	-0.7923	12.615	1.5988	8.7677
	(-0.5752)	(6.1431)	(0.2706)	(4.1377)	(-0.7238)	(6.1294)	(0.3623)	(4.1226)	(-0.6183)	(6.1271)	(0.1337)	(4.1142)
multiple occurrence (1st)	-0.0794	0.2710	-0.0879	1.4195	-0.0831	0.2696	-0.0116	1.4272	-0.0842	0.2716	0.1544	1.4288
	(-1.2990)	(0.3665)	(-0.9604)	(1.3152)	(-1.1466)	(0.3647)	(-0.0434)	(1.3177)	(-1.0775)	(0.3676)	(0.0787)	(1.3186)
multiple occurrence (2nd within 6MO)	0.0523	-1.0620	0.0115	-2.2203	0.0670	-1.0622	-0.1068	-2.2053	0.0709	-1.0577	-0.3620	-2.1993
	(0.8085)	(-1.1445)	(0.1071)	(-2.5214)	(1.0422)	(-1.1464)	(-0.2641)	(-2.5043)	(0.9126)	(-1.1416)	(-0.1236)	(-2.4943)
multiple occurrence (2nd after 6MO)	0.0556	-0.5179	-0.0398	-0.3293	0.0626	-0.5143	-0.0575	-0.3351	0.0645	-0.5126	-0.0964	-0.3361
	(0.7878)	(-0.7784)	(-2.8777)	(-0.3595)	(0.8950)	(-0.7749)	(-1.0001)	(-0.3661)	(0.8662)	(-0.7738)	(-0.2545)	(-0.3676)
flag for notice of delisting	-0.0433	1.8683	-0.0913	0.6328	-0.0694	1.8620	-0.0568	0.6464	-0.0766	1.8670	0.0183	0.6484
	(-0.5184)	(2.6441)	(-2.1107)	(0.5187)	(-0.6312)	(2.6204)	(-0.4343)	(0.5288)	(-0.5117)	(2.6285)	(0.0208)	(0.5306)
R-squared	-1.5395	0.0386	-0.0081	0.0387	-2.4329	0.0385	-2.4043	0.0384	-2.7065	0.0384	-30.168	0.0384
Adj. R-squared	-1.5877		-0.0372		-2.4981		-2.5022		-2.7769		-31.065	
Partial R-squared		0.0004		0.0002		0.0003		3.1e-05		0.0002		3.658e-06
F-statistic	3.111e+17	4.2373	8.111e+15	1.2637	-9.404e+15	2.0507	-1.017e+15	0.1414	-4.889e+15	1.2504	-3.246e+12	0.0128
P-value (F-stat)	0.0000	0.0395	0.0000	0.261	1.0000	0.1521	1.0000	0.7069	1.0000	0.2635	1.0000	0.9098
number of observations	2362	2362	1288	1288	2362	2362	1288	1288	2362	2362	1288	1288

This table shows IV regression results for abnormal returns regressed over ownership stake (% of market capitalization). Values in parenthesis correspond to coefficients' t-statistics.
In all 5 panels, the dependent variable is *abnormal return* but with reference to a distinct pricing models. For example, *Panel A* shows regressions for which the dependent variable is abnormal return using the *market model*, while *Panel E* the dependent variable is abnormal return using the *Fama-French 5 factors model*.
Each panel contains 3 groups of results, each corresponding market returns calculated using different methodologies (*). From left to right, they are: *CRSP equal weighted*, *CRSP value weighted*, and *S&P500*. We use market returns to compute lagging 10-trading days market trend*(used as control variable and presented here for completeness) as well as deviations from the trend*(the instrumental variable). For each of the 3 market return categories, we present results for the 2 periods studied (2006-2022 and 2010-2019). For each period, we provide results for both the second stage and the first stage of two-stage least squares regressions (2SLS).
The endogenous variable is *ownership*, and the instrument is the *deviation from market trend*. Regressions are controlled for market trend, industry and time fixed effects, firm-level variables (pre-event), and variables related to characteristics of the event (to control for biases in ownership data extraction).
Loadings for the pricing models were computed using the 100 trading days window ($t-121$ to $t-21$) that precedes the evaluation window. The evaluation window used to compute abnormal returns is the 41 days trading window from $t-20$ to $t+20$.

Table A.22:IV Regression full table: abnormal returns on ownership stake *(% market capitalization) - continued*

Panel C: *Dependent variable: abnormal return (Fama-French 3 factors + momentum)*, ± *20 days*, t_0*=event date*

	*equal weighted**				*value weighted**				*S&P500**			
	2006-2022		*2010-2019*		*2006-2022*		*2010-2019*		*2006-2022*		*2010-2019*	
	2nd stage	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*
intercept	-1.2972	23.290	-0.0106	17.711	-1.5152	23.170	0.7727	17.718	-1.5332	23.093	2.4786	17.702
	(-0.7022)	(13.427)	(-0.0115)	(7.5944)	(-0.8584)	(13.551)	(0.2433)	(7.6738)	(-0.7102)	(13.493)	(0.1264)	(7.7210)
ownership stake *(%)*	0.0480		0.0093		0.0578		-0.0354		0.0585		-0.1320	
	(0.5986)		(0.1785)		(0.7438)		(-0.1943)		(0.6164)		(-0.1185)	
deviation from trend*		-17.356		-17.563		-14.318		-6.0270		-11.703		-2.0989
		(-2.0585)		(-1.1241)		(-1.4320)		(-0.3760)		(-1.1182)		(-0.1133)
market trend*	-0.0139	2.4863	-0.8235	3.9758	-0.2119	3.2938	-0.6147	2.5367	-0.2098	3.4727	-0.1962	3.1850
	(-0.0376)	(0.4688)	(-3.7719)	(0.2723)	(-0.6042)	(0.5264)	(-1.0069)	(0.1754)	(-0.6191)	(0.5401)	(-0.0489)	(0.2306)
book-to-market	0.0195	-0.0623	0.0051	0.1776	0.0200	-0.0618	0.0138	0.1768	0.0201	-0.0633	0.0314	0.1788
	(0.6386)	(-0.1803)	(0.2858)	(0.4040)	(0.6043)	(-0.1788)	(0.3036)	(0.4008)	(0.6103)	(-0.1827)	(0.1312)	(0.4050)
cash-to-assets	0.0033	-0.0380	0.0014	-0.0392	0.0037	-0.0380	-0.0003	-0.0394	0.0037	-0.0379	-0.0041	-0.0394
	(1.1140)	(-3.4479)	(0.8135)	(-2.6194)	(1.3364)	(-3.4743)	(-0.0508)	(-2.6291)	(1.0982)	(-3.4760)	(-0.0959)	(-2.6322)
return-on-assets	0.0887	-4.6271	-0.1078	-4.3054	0.1343	-4.6405	-0.2974	-4.2717	0.1379	-4.6362	-0.7092	-4.2572
	(0.2217)	(-2.1404)	(-0.4713)	(-1.3323)	(0.3339)	(-2.1203)	(-0.4295)	(-1.3301)	(0.2969)	(-2.1126)	(-0.1548)	(-1.3301)
log market capitalization	-0.0000	-0.0506	-0.0144	-0.3253	0.0005	-0.0491	-0.0283	-0.3238	0.0005	-0.0473	-0.0592	-0.3221
	(-0.0039)	(-0.2402)	(-0.7374)	(-1.9025)	(0.0391)	(-0.2339)	(-0.4607)	(-1.9011)	(0.0399)	(-0.2252)	(-0.1656)	(-1.8966)
tobins' q	0.0049	-0.1745	0.0206	-0.0743	0.0067	-0.1750	0.0179	-0.0683	0.0069	-0.1749	0.0115	-0.0671
	(0.2516)	(-0.8706)	(1.5046)	(-0.2390)	(0.3256)	(-0.8691)	(1.4829)	(-0.2205)	(0.3029)	(-0.8674)	(0.1800)	(-0.2174)
profit margin	-0.0000	0.0002	-0.0000	0.0005	-0.0000	0.0002	0.0000	0.0005	-0.0000	0.0002	0.0001	0.0005
	(-0.4474)	(0.7746)	(-0.2240)	(1.6023)	(-0.4795)	(0.7485)	(0.1505)	(1.6262)	(-0.4574)	(0.7446)	(0.1102)	(1.6296)
cashflow	-0.0351	1.1596	0.2266	1.3931	-0.0485	1.1838	0.2858	1.3741	-0.0494	1.1826	0.4190	1.3695
	(-0.1750)	(0.8212)	(1.9227)	(0.7194)	(-0.2315)	(0.8173)	(1.2044)	(0.7110)	(-0.2328)	(0.8128)	(0.3064)	(0.7089)
market leverage	0.0005	0.0004	0.0012	-0.0148	0.0005	0.0004	0.0006	-0.0147	0.0005	0.0005	-0.0008	-0.0146
	(0.4483)	(0.0278)	(0.9110)	(-0.7525)	(0.3958)	(0.0284)	(0.2276)	(-0.7646)	(0.3954)	(0.0374)	(-0.0524)	(-0.7652)
book leverage	-0.0002	0.0012	-0.0005	0.0076	-0.0002	0.0012	-0.0002	0.0075	-0.0002	0.0011	0.0005	0.0074
	(-0.3358)	(0.1840)	(-0.8889)	(1.2807)	(-0.3237)	(0.1842)	(-0.1574)	(1.2872)	(-0.3185)	(0.1764)	(0.0605)	(1.2900)
dividend yield	0.0059	-0.1244	-0.0024	-0.0310	0.0071	-0.1235	-0.0036	-0.0293	0.0072	-0.1230	-0.0063	-0.0290
	(0.6137)	(-1.2648)	(-0.7992)	(-0.1262)	(0.6735)	(-1.2499)	(-0.2547)	(-0.1187)	(0.5824)	(-1.2461)	(-0.1078)	(-0.1175)
payout ratio	-0.0009	0.0599	-0.0006	-0.0417	-0.0014	0.0599	-0.0025	-0.0415	-0.0015	0.0598	-0.0065	-0.0413
	(-0.1442)	(0.9529)	(-0.3076)	(-0.4839)	(-0.2164)	(0.9531)	(-0.3382)	(-0.4778)	(-0.1947)	(0.9541)	(-0.1479)	(-0.4750)
sales growth	-0.0004	-0.0071	-0.0004	-0.0055	-0.0003	-0.0072	-0.0006	-0.0056	-0.0003	-0.0073	-0.0012	-0.0057
	(-0.5412)	(-2.6255)	(-1.1848)	(-2.3213)	(-0.4088)	(-2.7151)	(-0.5621)	(-2.2148)	(-0.3345)	(-2.7438)	(-0.1775)	(-2.2526)
amihud liquidity measure	0.0543	-0.9518	-0.0082	-0.9982	0.0636	-0.9486	-0.0518	-0.9957	0.0642	-0.9419	-0.1474	-0.9936
	(0.6777)	(-2.2070)	(-0.1683)	(-1.4066)	(0.8300)	(-2.2043)	(-0.2943)	(-1.3979)	(0.7019)	(-2.1844)	(-0.1320)	(-1.3950)
flag for multiple filings	-0.4632	12.628	0.1014	8.7503	-0.5853	12.617	0.4917	8.7602	-0.5950	12.615	1.3388	8.7677
	(-0.4410)	(6.1431)	(0.2229)	(4.1377)	(-0.5655)	(6.1294)	(0.3097)	(4.1226)	(-0.4741)	(6.1271)	(0.1357)	(4.1142)
multiple occurrence (1^{st})	-0.0754	0.2710	-0.0848	1.4195	-0.0780	0.2696	-0.0208	1.4272	-0.0782	0.2716	0.1172	1.4288
	(-1.3245)	(0.3665)	(-0.8787)	(1.3152)	(-1.2075)	(0.3647)	(-0.0782)	(1.3177)	(-1.1499)	(0.3676)	(0.0725)	(1.3186)
multiple occurrence (2^{nd} within 6MO)	0.0327	-1.0620	0.0124	-2.2203	0.0428	-1.0622	-0.0871	-2.2053	0.0436	-1.0577	-0.2991	-2.1993
	(0.4790)	(-1.1445)	(0.1124)	(-2.5214)	(0.6749)	(-1.1464)	(-0.2162)	(-2.5043)	(0.5657)	(-1.1416)	(-0.1234)	(-2.4943)
multiple occurrence (2^{nd} after 6MO)	0.0463	-0.5179	-0.0382	-0.3293	0.0511	-0.5143	-0.0530	-0.3351	0.0515	-0.5126	-0.0853	-0.3361
	(0.6823)	(-0.7784)	(-2.5929)	(-0.3595)	(0.7910)	(-0.7749)	(-0.9361)	(-0.3661)	(0.7472)	(-0.7738)	(-0.2692)	(-0.3676)
flag for notice of delisting	-0.0052	1.8683	-0.0877	0.6328	-0.0232	1.8620	-0.0588	0.6464	-0.0247	1.8670	0.0035	0.6484
	(-0.0536)	(2.6441)	(-2.0394)	(0.5187)	(-0.2067)	(2.6204)	(-0.4567)	(0.5288)	(-0.1689)	(2.6285)	(0.0048)	(0.5306)
R-squared	-1.1145	0.0386	-0.0359	0.0387	-1.6370	0.0385	-1.4276	0.0384	-1.6818	0.0384	-19.805	0.0384
Adj. R-squared	-1.1546		-0.0657		-1.6870		-1.4975		-1.7327		-20.404	
Partial R-squared		0.0004		0.0002		0.0003		3.1e-05		0.0002		3.658e-06
F-statistic	-3.491e+16	4.2373	-3.661e+15	1.2637	-1.259e+17	2.0507	-6.963e+15	0.1414	-1.923e+15	1.2504	-1.823e+12	0.0128
P-value (F-stat)	1.0000	0.0395	1.0000	0.261	1.0000	0.1521	1.0000	0.7069	1.0000	0.2635	1.0000	0.9098
number of observations	2362	2362	1288	1288	2362	2362	1288	1288	2362	2362	1288	1288

This table shows IV regression results for abnormal returns regressed over ownership stake (% of market capitalization). Values in parenthesis correspond to coefficients' t-statistics.
In all 5 panels, the dependent variable is *abnormal return* but with reference to a distinct pricing models. For example, *Panel A* shows regressions for which the dependent variable is abnormal return using the *market model*, while *Panel E* the dependent variable is abnormal return using the *Fama-French 5 factors model*.
Each panel contains 3 groups of results, each corresponding market returns calculated using different methodologies (*). From left to right, they are: *CRSP equal weighted*, *CRSP value weighted*, and *S&P500*. We use market returns to compute lagging 10-trading days market trend*(used as control variable and presented here for completeness) as well as deviations from the trend*(the instrumental variable). For each of the 3 market return categories, we present results for the 2 periods studied (2006-2022 and 2010-2019). For each period, we provide results for both the second stage and the first stage of two-stage least squares regressions (2SLS). The endogenous variable is *ownership*, and the instrument is the *deviation from market trend*. Regressions are controlled for market trend, industry and time fixed effects, firm-level variables (pre-event), and variables related to characteristics of the event (to control for biases in ownership data extraction).
Loadings for the pricing models were computed using the 100 trading days window ($t-121$ to $t-21$) that precedes the evaluation window. The evaluation window used to compute abnormal returns is the 41 days trading window from $t-20$ to $t+20$.

Table A.22: IV Regression full table: abnormal returns on ownership stake *(% market capitalization) - continued*

Panel D: *Dependent variable: abnormal return (Fama-French 5 factors)*, ± *20 days*, t_0=*event date*

	*equal weighted**				*value weighted**				*S&P500**			
	2006-2022		*2010-2019*		*2006-2022*		*2010-2019*		*2006-2022*		*2010-2019*	
	2nd stage	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*
intercept	-1.6068	23.290	-0.1496	17.711	-1.8608	23.170	0.6979	17.718	-1.9194	23.093	2.4786	17.702
	(-0.9850)	(13.427)	(-0.1749)	(7.5944)	(-1.1129)	(13.551)	(0.2411)	(7.6738)	(-0.9142)	(13.493)	(0.1262)	(7.7210)
ownership stake *(%)*	0.0575		0.0151		0.0688		-0.0331		0.0713		-0.1340	
	(0.8120)		(0.3195)		(0.9338)		(-0.1993)		(0.7707)		(-0.1199)	
deviation from trend*		-17.356		-17.563		-14.318		-6.0270		-11.703		-2.0989
		(-2.0585)		(-1.1241)		(-1.4320)		(-0.3760)		(-1.1182)		(-0.1133)
market trend*	-0.0295	2.4863	-0.7736	3.9758	-0.2562	3.2938	-0.5634	2.5367	-0.2639	3.4727	-0.1388	3.1850
	(-0.0668)	(0.4688)	(-2.6501)	(0.2723)	(-0.5744)	(0.5264)	(-1.0258)	(0.1754)	(-0.5915)	(0.5401)	(-0.0347)	(0.2306)
book-to-market	0.0222	-0.0623	0.0099	0.1776	0.0227	-0.0618	0.0191	0.1768	0.0229	-0.0633	0.0375	0.1788
	(0.6088)	(-0.1803)	(0.5509)	(0.4040)	(0.5796)	(-0.1788)	(0.4330)	(0.4008)	(0.5820)	(-0.1827)	(0.1550)	(0.4050)
cash-to-assets	0.0040	-0.0380	0.0017	-0.0392	0.0044	-0.0380	-0.0002	-0.0394	0.0045	-0.0379	-0.0042	-0.0394
	(1.4884)	(-3.4479)	(1.0158)	(-2.6194)	(1.6740)	(-3.4743)	(-0.0359)	(-2.6291)	(1.3675)	(-3.4760)	(-0.0967)	(-2.6322)
return-on-assets	0.1332	-4.6271	-0.1092	-4.3054	0.1864	-4.6405	-0.3141	-4.2717	0.1982	-4.6362	-0.7437	-4.2572
	(0.3371)	(-2.1404)	(-0.5387)	(-1.3323)	(0.4498)	(-2.1203)	(-0.5096)	(-1.3301)	(0.4129)	(-2.1126)	(-0.1618)	(-1.3301)
log market capitalization	-0.0010	-0.0506	-0.0138	-0.3253	-0.0004	-0.0491	-0.0289	-0.3238	-0.0003	-0.0473	-0.0612	-0.3221
	(-0.0837)	(-0.2402)	(-0.7394)	(-1.9025)	(-0.0300)	(-0.2339)	(-0.5136)	(-1.9011)	(-0.0224)	(-0.2252)	(-0.1711)	(-1.8966)
tobins' q	0.0026	-0.1745	0.0209	-0.0743	0.0046	-0.1750	0.0179	-0.0683	0.0051	-0.1749	0.0112	-0.0671
	(0.1355)	(-0.8706)	(1.4071)	(-0.2390)	(0.2231)	(-0.8691)	(1.4533)	(-0.2205)	(0.2217)	(-0.8674)	(0.1735)	(-0.2174)
profit margin	-0.0000	0.0002	-0.0000	0.0005	-0.0000	0.0002	0.0000	0.0005	-0.0000	0.0002	0.0001	0.0005
	(-0.3907)	(0.7746)	(-0.2027)	(1.6023)	(-0.4158)	(0.7485)	(0.1931)	(1.6262)	(-0.3991)	(0.7446)	(0.1181)	(1.6296)
cashflow	-0.0452	1.1596	0.2506	1.3931	-0.0608	1.1838	0.3147	1.3741	-0.0638	1.1826	0.4535	1.3695
	(-0.2000)	(0.8212)	(2.3224)	(0.7194)	(-0.2581)	(0.8173)	(1.4968)	(0.7110)	(-0.2671)	(0.8128)	(0.3316)	(0.7089)
market leverage	0.0003	0.0004	0.0013	-0.0148	0.0003	0.0004	0.0006	-0.0147	0.0003	0.0005	-0.0008	-0.0146
	(0.2325)	(0.0278)	(0.9330)	(-0.7525)	(0.2010)	(0.0284)	(0.2638)	(-0.7646)	(0.1971)	(0.0374)	(-0.0526)	(-0.7652)
book leverage	-0.0000	0.0012	-0.0005	0.0076	-0.0001	0.0012	-0.0002	0.0075	-0.0001	0.0011	0.0006	0.0074
	(-0.0647)	(0.1840)	(-0.8627)	(1.2807)	(-0.0763)	(0.1842)	(-0.1331)	(1.2872)	(-0.0775)	(0.1764)	(0.0701)	(1.2900)
dividend yield	0.0070	-0.1244	-0.0021	-0.0310	0.0083	-0.1235	-0.0033	-0.0293	0.0086	-0.1230	-0.0062	-0.0290
	(0.7262)	(-1.2648)	(-0.5851)	(-0.1262)	(0.7506)	(-1.2499)	(-0.2501)	(-0.1187)	(0.6594)	(-1.2461)	(-0.1043)	(-0.1175)
payout ratio	-0.0027	0.0599	-0.0004	-0.0417	-0.0033	0.0599	-0.0024	-0.0415	-0.0035	0.0598	-0.0066	-0.0413
	(-0.4239)	(0.9529)	(-0.2460)	(-0.4839)	(-0.4686)	(0.9531)	(-0.3620)	(-0.4778)	(-0.4257)	(0.9541)	(-0.1507)	(-0.4750)
sales growth	-0.0004	-0.0071	-0.0004	-0.0055	-0.0003	-0.0072	-0.0007	-0.0056	-0.0003	-0.0073	-0.0013	-0.0057
	(-0.6161)	(-2.6255)	(-1.4839)	(-2.3213)	(-0.4171)	(-2.7151)	(-0.6793)	(-2.2148)	(-0.3194)	(-2.7438)	(-0.1898)	(-2.2526)
amihud liquidity measure	0.0556	-0.9518	-0.0073	-0.9982	0.0664	-0.9486	-0.0545	-0.9957	0.0687	-0.9419	-0.1544	-0.9936
	(0.7647)	(-2.2070)	(-0.1555)	(-1.4066)	(0.8897)	(-2.2043)	(-0.3400)	(-1.3979)	(0.7531)	(-2.1844)	(-0.1380)	(-1.3950)
flag for multiple filings	-0.5907	12.628	0.0514	8.7503	-0.7329	12.617	0.4734	8.7602	-0.7649	12.615	1.3575	8.7677
	(-0.6299)	(6.1431)	(0.1233)	(4.1377)	(-0.7370)	(6.1294)	(0.3265)	(4.1226)	(-0.6190)	(6.1271)	(0.1372)	(4.1142)
multiple occurrence (1st)	-0.0816	0.2710	-0.0925	1.4195	-0.0846	0.2696	-0.0233	1.4272	-0.0854	0.2716	0.1207	1.4288
	(-1.3133)	(0.3665)	(-0.9579)	(1.3152)	(-1.1805)	(0.3647)	(-0.0944)	(1.3177)	(-1.1190)	(0.3676)	(0.0743)	(1.3186)
multiple occurrence (2nd within 6MO)	0.0506	-1.0620	0.0358	-2.2203	0.0623	-1.0622	-0.0715	-2.2053	0.0649	-1.0577	-0.2928	-2.1993
	(0.7590)	(-1.1445)	(0.3333)	(-2.5214)	(0.9266)	(-1.1464)	(-0.1936)	(-2.5043)	(0.8135)	(-1.1416)	(-0.1207)	(-2.4943)
multiple occurrence (2nd after 6MO)	0.0503	-0.5179	-0.0355	-0.3293	0.0559	-0.5143	-0.0515	-0.3351	0.0572	-0.5126	-0.0852	-0.3361
	(0.7413)	(-0.7784)	(-1.7968)	(-0.3595)	(0.8283)	(-0.7749)	(-1.0324)	(-0.3661)	(0.7970)	(-0.7738)	(-0.2721)	(-0.3676)
flag for notice of delisting	-0.0363	1.8683	-0.0923	0.6328	-0.0571	1.8620	-0.0611	0.6464	-0.0620	1.8670	0.0040	0.6484
	(-0.4454)	(2.6441)	(-2.2015)	(0.5187)	(-0.5429)	(2.6204)	(-0.5117)	(0.5288)	(-0.4345)	(2.6285)	(0.0055)	(0.5306)
R-squared	-1.5472	0.0386	-0.1838	0.0387	-2.2424	0.0385	-1.2351	0.0384	-2.4146	0.0384	-20.268	0.0384
Adj. R-squared	-1.5955		-0.2178		-2.3040		-1.2994		-2.4794		-20.880	
Partial R-squared		0.0004		0.0002		0.0003		3.1e-05		0.0002		3.658e-06
F-statistic	2.777e+15	4.2373	-1.14e+16	1.2637	4.415e+15	2.0507	-9.566e+14	0.1414	4.865e+14	1.2504	-7.495e+12	0.0128
P-value (F-stat)	0.0000	0.0395	1.0000	0.261	0.0000	0.1521	1.0000	0.7069	0.0000	0.2635	1.0000	0.9098
number of observations	2362	2362	1288	1288	2362	2362	1288	1288	2362	2362	1288	1288

This table shows IV regression results for abnormal returns regressed over ownership stake (% of market capitalization). Values in parenthesis correspond to coefficients' t-statistics.
In all 5 panels, the dependent variable is *abnormal return* but with reference to a distinct pricing models. For example, *Panel A* shows regressions for which the dependent variable is abnormal return using the *market model*, while *Panel E* the dependent variable is abnormal return using the *Fama-French 5 factors model*.
Each panel contains 3 groups of results, each corresponding market returns calculated using different methodologies (*). From left to right, they are: *CRSP equal weighted*, *CRSP value weighted*, and *S&P500*. We use market returns to compute lagging 10-trading days market trend*(used as control variable and presented here for completeness) as well as deviations from the trend*(the instrumental variable). For each of the 3 market return categories, we present results for the 2 periods studied (2006-2022 and 2010-2019). For each period, we provide results for both the second stage and the first stage of two-stage least squares regressions (2SLS).
The endogenous variable is *ownership*, and the instrument is the *deviation from market trend*. Regressions are controlled for market trend, industry and time fixed effects, firm-level variables (pre-event), and variables related to characteristics of the event (to control for biases in ownership data extraction).
Loadings for the pricing models were computed using the 100 trading days window ($t-121$ to $t-21$) that precedes the evaluation window. The evaluation window used to compute abnormal returns is the 41 days trading window from $t-20$ to $t+20$.

Table A.23: IV Regression full table: abnormal returns on ownership stake *(natural logarithm of dollar amounts) - continued*

Panel A: *Dependent variable: abnormal return (CAPM), ± 20 days, t_0=event date*

	*equal weighted**				*value weighted**				*S&P500**			
	2006-2022		*2010-2019*		*2006-2022*		*2010-2019*		*2006-2022*		*2010-2019*	
	2nd stage	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*
intercept	-1.3948	9.5759	-0.4390	8.9837	-1.5776	9.5610	0.0117	8.9647	-1.4898	9.5413	0.0950	8.9563
	(-0.7825)	(67.260)	(-0.4314)	(51.261)	(-0.9603)	(69.032)	(0.0071)	(51.409)	(-0.9073)	(69.630)	(0.0450)	(51.573)
ownership stake *(log US$)*	0.1317		0.0536		0.1521		0.0031		0.1430		-0.0064	
	(0.7166)		(0.4637)		(0.8981)		(0.0166)		(0.8434)		(-0.0267)	
deviation from trend*		-6.1175		-6.6633		-5.6910		-4.7863		-5.2386		-4.0406
		(-3.7366)		(-3.3200)		(-4.5263)		(-2.2727)		(-4.0388)		(-1.6877)
market trend*	0.3096	1.3136	-0.4471	-2.4056	0.0440	1.2639	-0.5932	-1.8793	0.0249	1.2955	-0.5550	-1.6592
	(0.8047)	(1.4884)	(-0.9032)	(-2.1747)	(0.1331)	(1.4113)	(-0.9764)	(-1.7827)	(0.0779)	(1.3977)	(-0.8487)	(-1.6827)
book-to-market	0.0307	-0.1463	0.0199	-0.2424	0.0334	-0.1464	0.0080	-0.2397	0.0320	-0.1469	0.0059	-0.2388
	(0.6589)	(-1.7730)	(0.6479)	(-5.5383)	(0.7145)	(-1.7625)	(0.1779)	(-5.5131)	(0.7008)	(-1.7670)	(0.1047)	(-5.4774)
cash-to-assets	0.0031	-0.0093	0.0014	-0.0071	0.0033	-0.0094	0.0010	-0.0071	0.0032	-0.0094	0.0010	-0.0071
	(1.3662)	(-5.6927)	(1.9772)	(-4.1546)	(1.5152)	(-5.6118)	(0.9089)	(-4.1537)	(1.4850)	(-5.5877)	(0.6484)	(-4.1634)
return-on-assets	-0.2193	0.7533	-0.1605	-0.0730	-0.2348	0.7474	-0.1623	-0.0749	-0.2277	0.7483	-0.1638	-0.0763
	(-0.8030)	(1.9417)	(-1.3825)	(-0.1489)	(-0.8781)	(1.9316)	(-1.5552)	(-0.1531)	(-0.8400)	(1.9251)	(-1.6105)	(-0.1564)
tobins' q	-0.0450	0.2468	0.0008	0.1857	-0.0500	0.2466	0.0106	0.1876	-0.0477	0.2467	0.0125	0.1882
	(-0.9339)	(6.1764)	(0.0346)	(5.8128)	(-1.1179)	(6.1458)	(0.2856)	(5.8361)	(-1.0630)	(6.1378)	(0.2654)	(5.8531)
profit margin	0.0000	-7.875e-05	0.0000	-4.908e-05	0.0000	-8.074e-05	-0.0000	-4.847e-05	0.0000	-8.114e-05	-0.0000	-4.85e-05
	(1.1202)	(-1.7577)	(0.0903)	(-1.0127)	(1.2100)	(-1.8357)	(-0.0586)	(-0.9833)	(1.1570)	(-1.8471)	(-0.1057)	(-0.9799)
cashflow	-0.0873	0.5236	0.1686	1.2296	-0.0995	0.5316	0.2284	1.2327	-0.0951	0.5321	0.2411	1.2359
	(-0.3864)	(1.9378)	(0.8648)	(3.4770)	(-0.4358)	(1.9832)	(0.8618)	(3.4944)	(-0.4191)	(1.9735)	(0.7474)	(3.5149)
market leverage	0.0009	-0.0034	0.0014	-0.0037	0.0010	-0.0035	0.0012	-0.0036	0.0009	-0.0034	0.0012	-0.0035
	(0.8914)	(-1.9692)	(1.6164)	(-1.9908)	(0.9482)	(-2.0157)	(1.2499)	(-1.9453)	(0.9090)	(-2.0054)	(1.0606)	(-1.9255)
book leverage	-0.0003	0.0017	-0.0007	0.0017	-0.0004	0.0017	-0.0006	0.0017	-0.0003	0.0017	-0.0006	0.0017
	(-0.7058)	(1.9430)	(-1.7184)	(1.6424)	(-0.7787)	(1.9698)	(-1.2425)	(1.6355)	(-0.7427)	(1.9634)	(-1.0303)	(1.6307)
dividend yield	0.0047	-0.0299	-0.0029	-0.0003	0.0052	-0.0298	-0.0028	0.0006	0.0049	-0.0298	-0.0027	0.0008
	(1.0548)	(-2.2281)	(-1.2352)	(-0.0125)	(1.0851)	(-2.1966)	(-1.0875)	(0.0204)	(1.0303)	(-2.1917)	(-1.0481)	(0.0290)
payout ratio	-0.0041	0.0321	-0.0030	0.0218	-0.0047	0.0322	-0.0020	0.0216	-0.0044	0.0322	-0.0018	0.0216
	(-0.6808)	(5.2823)	(-0.7803)	(1.9197)	(-0.8351)	(5.2546)	(-0.3809)	(1.9051)	(-0.7802)	(5.2551)	(-0.2833)	(1.9074)
sales growth	-0.0013	0.0033	-0.0008	0.0040	-0.0013	0.0033	-0.0006	0.0039	-0.0013	0.0033	-0.0005	0.0039
	(-2.0013)	(6.9573)	(-1.6448)	(12.937)	(-2.1783)	(6.9592)	(-0.7634)	(12.580)	(-2.1356)	(6.9560)	(-0.5705)	(12.375)
amihud liquidity measure	0.1779	-1.2717	0.0779	-1.4117	0.2043	-1.2716	0.0068	-1.4111	0.1928	-1.2705	-0.0066	-1.4108
	(0.7319)	(-8.2525)	(0.4990)	(-11.096)	(0.9129)	(-8.2576)	(0.0263)	(-11.093)	(0.8637)	(-8.2282)	(-0.0201)	(-11.089)
flag for multiple filings	0.0888	0.5817	0.1592	0.5442	0.0785	0.5798	0.1858	0.5352	0.0839	0.5786	0.1907	0.5347
	(0.6981)	(3.2368)	(1.6187)	(1.9512)	(0.6524)	(3.2361)	(1.5071)	(1.9541)	(0.6992)	(3.2342)	(1.3162)	(1.9620)
multiple occurrence (1st)	-0.0035	-0.4412	-0.0523	-0.3705	0.0060	-0.4414	-0.0708	-0.3704	0.0020	-0.4410	-0.0745	-0.3701
	(-0.0409)	(-6.1213)	(-1.4876)	(-5.8719)	(0.0761)	(-6.1460)	(-1.2379)	(-5.8896)	(0.0255)	(-6.1579)	(-0.9909)	(-5.8385)
multiple occurrence (2nd within 6MO)	0.0579	-0.4952	0.0377	-0.6882	0.0684	-0.4958	0.0021	-0.6916	0.0639	-0.4951	-0.0044	-0.6899
	(0.7223)	(-3.3420)	(0.4360)	(-3.8933)	(0.9452)	(-3.3406)	(0.0156)	(-3.9053)	(0.8665)	(-3.3446)	(-0.0256)	(-3.8932)
multiple occurrence (2nd after 6MO)	0.0931	-0.4796	-0.0010	-0.5380	0.1027	-0.4788	-0.0284	-0.5390	0.0983	-0.4784	-0.0334	-0.5387
	(0.8760)	(-5.1263)	(-0.0202)	(-6.1804)	(1.0468)	(-5.1097)	(-0.3127)	(-6.1821)	(1.0054)	(-5.1126)	(-0.2810)	(-6.2002)
flag for notice of delisting	0.1429	-0.3556	-0.0296	-0.5267	0.1503	-0.3587	-0.0560	-0.5292	0.1472	-0.3579	-0.0614	-0.5296
	(0.8288)	(-2.0889)	(-0.3837)	(-2.5630)	(0.8919)	(-2.1006)	(-0.5014)	(-2.5584)	(0.8824)	(-2.1026)	(-0.4533)	(-2.5531)
R-squared	-0.0868	0.409	-0.0299	0.4401	-0.1280	0.4086	0.0427	0.4388	-0.1092	0.4083	0.0455	0.4383
Adj. R-squared	-0.1070		-0.0586		-0.1490		0.0159		-0.1298		0.0188	
Partial R-squared		0.0032		0.0022		0.0029		0.0013		0.0024		0.0009
F-statistic	1.76e+15	13.963	-6.399e+14	11.022	7.735e+14	20.487	-1.134e+15	5.1653	-1.66e+15	16.312	-8.409e+14	2.8482
P-value (F-stat)	0.0000	0.0002	1.0000	0.0009	0.0000	6.004e-06	1.0000	0.023	1.0000	5.373e-05	1.0000	0.0915
number of observations	2362	2362	1288	1288	2362	2362	1288	1288	2362	2362	1288	1288

This table shows IV regression results for abnormal returns regressed over ownership stake (natural logarithm of dollar amounts). Values in parenthesis correspond to coefficients' t-statistics.
In all 5 panels, the dependent variable is *abnormal return* but with reference to a distinct pricing models. For example, *Panel A* shows regressions for which the dependent variable is abnormal return using the *market model*, while *Panel E* the dependent variable is abnormal return using the *Fama-French 5 factors model*.
Each panel contains 3 groups of results, each corresponding market returns calculated using different methodologies (*). From left to right, they are: *CRSP equal weighted*, *CRSP value weighted*, and *S&P500*. We use market returns to compute lagging 10-trading days market trend*(used as control variable and presented here for completeness) as well as deviations from the trend*(the instrumental variable). For each of the 3 market return categories, we present results for the 2 periods studied (2006-2022 and 2010-2019). For each period, we provide results for both the second stage and the first stage of two-stage least squares regressions (2SLS). The endogenous variable is *ownership*, and the instrument is the *deviation from market trend*. Regressions are controlled for market trend, industry and time fixed effects, firm-level variables (pre-event), and variables related to characteristics of the event (to control for biases in ownership data extraction).
Loadings for the pricing models were computed using the 100 trading days window ($t-121$ to $t-21$) that precedes the evaluation window. The evaluation window used to compute abnormal returns is the 41 days trading window from $t-20$ to $t+20$.

Table A.23: IV Regression full table: abnormal returns on ownership stake *(natural logarithm of dollar amounts) - continued*

Panel B: *Dependent variable: abnormal return (Fama-French 3 factors)*, ± *20 days*, t_0=*event date*

	*equal weighted**				*value weighted**				*S&P500**			
	2006-2022		*2010-2019*		*2006-2022*		*2010-2019*		*2006-2022*		*2010-2019*	
	2^{nd} stage	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*	*2^{nd} stage*	*1^{st} stage*
intercept	-1.7618	9.5759	-0.0359	8.9837	-1.9207	9.5610	0.6608	8.9647	-1.8126	9.5413	0.8977	8.9563
	(-0.9672)	(67.260)	(-0.0318)	(51.261)	(-1.1398)	(69.032)	(0.3559)	(51.409)	(-1.0832)	(69.630)	(0.3787)	(51.573)
ownership stake (log US$)	0.1589		0.0073		0.1762		-0.0708		0.1647		-0.0976	
	(0.8444)		(0.0572)		(1.0116)		(-0.3362)		(0.9513)		(-0.3637)	
deviation from trend*		-6.1175		-6.6633		-5.6910		-4.7863		-5.2386		-4.0406
		(-3.7366)		(-3.3200)		(-4.5263)		(-2.2727)		(-4.0388)		(-1.6877)
market trend*	-0.0523	1.3136	-0.6392	-2.4056	-0.2170	1.2639	-0.7448	-1.8793	-0.2016	1.2955	-0.6933	-1.6592
	(-0.1406)	(1.4884)	(-1.1783)	(-2.1747)	(-0.6192)	(1.4113)	(-1.1487)	(-1.7827)	(-0.5831)	(1.3977)	(-0.9897)	(-1.6827)
book-to-market	0.0394	-0.1463	0.0123	-0.2424	0.0417	-0.1464	-0.0060	-0.2397	0.0401	-0.1469	-0.0121	-0.2388
	(0.8190)	(-1.7730)	(0.4199)	(-5.5383)	(0.8677)	(-1.7625)	(-0.1302)	(-5.5131)	(0.8579)	(-1.7670)	(-0.2048)	(-5.4774)
cash-to-assets	0.0032	-0.0093	0.0011	-0.0071	0.0034	-0.0094	0.0006	-0.0071	0.0033	-0.0094	0.0004	-0.0071
	(1.4464)	(-5.6927)	(1.5565)	(-4.1546)	(1.5946)	(-5.6118)	(0.4857)	(-4.1537)	(1.5655)	(-5.5877)	(0.2475)	(-4.1634)
return-on-assets	-0.2481	0.7533	-0.1664	-0.0730	-0.2605	0.7474	-0.1702	-0.0749	-0.2516	0.7483	-0.1731	-0.0763
	(-0.8829)	(1.9417)	(-1.5476)	(-0.1489)	(-0.9387)	(1.9316)	(-1.7284)	(-0.1531)	(-0.8955)	(1.9251)	(-1.7887)	(-0.1564)
tobins' q	-0.0466	0.2468	0.0139	0.1857	-0.0508	0.2466	0.0291	0.1876	-0.0479	0.2467	0.0342	0.1882
	(-0.9531)	(6.1764)	(0.5748)	(5.8128)	(-1.1118)	(6.1458)	(0.7022)	(5.8361)	(-1.0511)	(6.1378)	(0.6505)	(5.8531)
profit margin	0.0000	-7.875e-05	-0.0000	-4.908e-05	0.0000	-8.074e-05	-0.0000	-4.847e-05	0.0000	-8.114e-05	-0.0000	-4.85e-05
	(0.5651)	(-1.7577)	(-0.1376)	(-1.0127)	(0.6010)	(-1.8357)	(-0.3909)	(-0.9833)	(0.5369)	(-1.8471)	(-0.4599)	(-0.9799)
cashflow	-0.0812	0.5236	0.2203	1.2296	-0.0921	0.5316	0.3145	1.2327	-0.0863	0.5321	0.3489	1.2359
	(-0.3912)	(1.9378)	(1.2450)	(3.4770)	(-0.4383)	(1.9832)	(1.2133)	(3.4944)	(-0.4127)	(1.9735)	(1.0750)	(3.5149)
market leverage	0.0010	-0.0034	0.0012	-0.0037	0.0011	-0.0035	0.0009	-0.0036	0.0011	-0.0034	0.0009	-0.0035
	(0.9671)	(-1.9692)	(1.2261)	(-1.9908)	(1.0210)	(-2.0157)	(0.8469)	(-1.9453)	(0.9849)	(-2.0054)	(0.6723)	(-1.9255)
book leverage	-0.0004	0.0017	-0.0005	0.0017	-0.0004	0.0017	-0.0004	0.0017	-0.0004	0.0017	-0.0004	0.0017
	(-0.7723)	(1.9430)	(-1.2455)	(1.6424)	(-0.8288)	(1.9698)	(-0.7873)	(1.6355)	(-0.7896)	(1.9634)	(-0.5983)	(1.6307)
dividend yield	0.0048	-0.0299	-0.0029	-0.0003	0.0053	-0.0298	-0.0028	0.0006	0.0050	-0.0298	-0.0027	0.0008
	(1.0471)	(-2.2281)	(-1.2096)	(-0.0125)	(1.0533)	(-2.1966)	(-0.7647)	(0.0204)	(0.9949)	(-2.1917)	(-0.6480)	(0.0290)
payout ratio	-0.0040	0.0321	-0.0015	0.0218	-0.0045	0.0322	0.0001	0.0216	-0.0041	0.0322	0.0007	0.0216
	(-0.7145)	(5.2823)	(-0.3915)	(1.9197)	(-0.8634)	(5.2546)	(0.0241)	(1.9051)	(-0.7923)	(5.2551)	(0.1087)	(1.9074)
sales growth	-0.0013	0.0033	-0.0005	0.0040	-0.0013	0.0033	-0.0002	0.0039	-0.0013	0.0033	-0.0001	0.0039
	(-1.9765)	(6.9573)	(-1.0527)	(12.937)	(-2.1514)	(6.9592)	(-0.2733)	(12.580)	(-2.1109)	(6.9560)	(-0.1214)	(12.375)
amihud liquidity measure	0.2112	-1.2717	0.0152	-1.4117	0.2332	-1.2716	-0.0949	-1.4111	0.2187	-1.2705	-0.1327	-1.4108
	(0.8459)	(-8.2525)	(0.0878)	(-11.096)	(1.0151)	(-8.2576)	(-0.3286)	(-11.093)	(0.9615)	(-8.2282)	(-0.3582)	(-11.089)
flag for multiple filings	0.0543	0.5817	0.1775	0.5442	0.0450	0.5798	0.2187	0.5352	0.0516	0.5786	0.2328	0.5347
	(0.4342)	(3.2368)	(1.8747)	(1.9512)	(0.3779)	(3.2361)	(1.6349)	(1.9541)	(0.4375)	(3.2342)	(1.4366)	(1.9620)
multiple occurrence (1^{st})	0.0069	-0.4412	-0.0657	-0.3705	0.0146	-0.4414	-0.0946	-0.3704	0.0095	-0.4410	-0.1046	-0.3701
	(0.0786)	(-6.1213)	(-1.5405)	(-5.8719)	(0.1785)	(-6.1460)	(-1.3601)	(-5.8896)	(0.1173)	(-6.1579)	(-1.1563)	(-5.8385)
multiple occurrence (2^{nd} within 6MO)	0.0718	-0.4952	0.0092	-0.6882	0.0801	-0.4958	-0.0457	-0.6916	0.0745	-0.4951	-0.0640	-0.6899
	(0.8754)	(-3.3420)	(0.1007)	(-3.8933)	(1.0716)	(-3.3406)	(-0.3130)	(-3.9053)	(0.9845)	(-3.3446)	(-0.3455)	(-3.8932)
multiple occurrence (2^{nd} after 6MO)	0.1034	-0.4796	-0.0287	-0.5380	0.1115	-0.4788	-0.0710	-0.5390	0.1060	-0.4784	-0.0852	-0.5387
	(0.9415)	(-5.1263)	(-0.4816)	(-6.1804)	(1.0901)	(-5.1097)	(-0.6814)	(-6.1821)	(1.0440)	(-5.1126)	(-0.6312)	(-6.2002)
flag for notice of delisting	0.1199	-0.3556	-0.0718	-0.5267	0.1263	-0.3587	-0.1132	-0.5292	0.1222	-0.3579	-0.1278	-0.5296
	(0.7212)	(-2.0889)	(-0.7736)	(-2.5630)	(0.7779)	(-2.1006)	(-0.8300)	(-2.5584)	(0.7626)	(-2.1026)	(-0.7716)	(-2.5531)
R-squared	-0.1685	0.409	0.0392	0.4401	-0.2141	0.4086	-0.0211	0.4388	-0.1833	0.4083	-0.0927	0.4383
Adj. R-squared	-0.1902		0.0124		-0.2366		-0.0497		-0.2053		-0.1233	
Partial R-squared		0.0032		0.0022		0.0029		0.0013		0.0024		0.0009
F-statistic	6.784e+14	13.963	-2.618e+14	11.022	-4.244e+14	20.487	6.484e+14	5.1653	3.61e+14	16.312	3.189e+14	2.8482
P-value (F-stat)	0.0000	0.0002	1.0000	0.0009	1.0000	6.004e-06	0.0000	0.023	0.0000	5.373e-05	0.0000	0.0915
number of observations	2362	2362	1288	1288	2362	2362	1288	1288	2362	2362	1288	1288

This table shows IV regression results for abnormal returns regressed over ownership stake (natural logarithm of dollar amounts). Values in parenthesis correspond to coefficients' t-statistics.
In all 5 panels, the dependent variable is *abnormal return* but with reference to a distinct pricing models. For example, *Panel A* shows regressions for which the dependent variable is abnormal return using the *market model*, while *Panel E* the dependent variable is abnormal return using the *Fama-French 5 factors model*.
Each panel contains 3 groups of results, each corresponding market returns calculated using different methodologies (*). From left to right, they are: *CRSP equal weighted*, *CRSP value weighted*, and *S&P500*. We use market returns to compute lagging 10-trading days market trend*(used as control variable and presented here for completeness) as well as deviations from the trend*(the instrumental variable). For each of the 3 market return categories, we present results for the 2 periods studied (2006-2022 and 2010-2019). For each period, we provide results for both the second stage and the first stage of two-stage least squares regressions (2SLS). The endogenous variable is *ownership*, and the instrument is the *deviation from market trend*. Regressions are controlled for market trend, industry and time fixed effects, firm-level variables (pre-event), and variables related to characteristics of the event (to control for biases in ownership data extraction).
Loadings for the pricing models were computed using the 100 trading days window ($t-121$ to $t-21$) that precedes the evaluation window. The evaluation window used to compute abnormal returns is the 41 days trading window from $t-20$ to $t+20$.

Table A.23: IV Regression full table: abnormal returns on ownership stake *(natural logarithm of dollar amounts) - continued*

Panel C: *Dependent variable: abnormal return (Fama-French 3 factors + momentum)*, ± *20 days*, t_0=*event date*

	*equal weighted**				*value weighted**				*S&P500**			
	2006-2022		*2010-2019*		*2006-2022*		*2010-2019*		*2006-2022*		*2010-2019*	
	2nd stage	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*
intercept	-1.4812	9.5759	-0.0629	8.9837	-1.5631	9.5610	0.5543	8.9647	-1.4254	9.5413	0.7668	8.9563
	(-0.7396)	(67.260)	(-0.0518)	(51.261)	(-0.8935)	(69.032)	(0.2815)	(51.409)	(-0.8121)	(69.630)	(0.3063)	(51.573)
ownership stake (log US$)	0.1344		0.0116		0.1434		-0.0577		0.1289		-0.0817	
	(0.6460)		(0.0847)		(0.7895)		(-0.2584)		(0.7061)		(-0.2887)	
deviation from trend*		-6.1175		-6.6633		-5.6910		-4.7863		-5.2386		-4.0406
		(-3.7366)		(-3.3200)		(-4.5263)		(-2.2727)		(-4.0388)		(-1.6877)
market trend*	-0.0738	1.3136	-0.7106	-2.4056	-0.2053	1.2639	-0.7771	-1.8793	-0.1758	1.2955	-0.7199	-1.6592
	(-0.2252)	(1.4884)	(-1.2598)	(-2.1747)	(-0.6561)	(1.4113)	(-1.1664)	(-1.7827)	(-0.5705)	(1.3977)	(-1.0026)	(-1.6827)
book-to-market	0.0366	-0.1463	0.0142	-0.2424	0.0378	-0.1464	-0.0019	-0.2397	0.0356	-0.1469	-0.0074	-0.2388
	(0.8495)	(-1.7730)	(0.4418)	(-5.5383)	(0.9218)	(-1.7625)	(-0.0383)	(-5.5131)	(0.9015)	(-1.7670)	(-0.1164)	(-5.4774)
cash-to-assets	0.0028	-0.0093	0.0012	-0.0071	0.0029	-0.0094	0.0007	-0.0071	0.0027	-0.0094	0.0006	-0.0071
	(1.2120)	(-5.6927)	(1.6524)	(-4.1546)	(1.3938)	(-5.6118)	(0.5893)	(-4.1537)	(1.3348)	(-5.5877)	(0.3453)	(-4.1634)
return-on-assets	-0.2372	0.7533	-0.1514	-0.0730	-0.2433	0.7474	-0.1548	-0.0749	-0.2321	0.7483	-0.1576	-0.0763
	(-0.7939)	(1.9417)	(-1.3749)	(-0.1489)	(-0.8491)	(1.9316)	(-1.5754)	(-0.1531)	(-0.7971)	(1.9251)	(-1.6605)	(-0.1564)
tobins' q	-0.0373	0.2468	0.0142	0.1857	-0.0394	0.2466	0.0277	0.1876	-0.0358	0.2467	0.0323	0.1882
	(-0.6990)	(6.1764)	(0.5449)	(5.8128)	(-0.8368)	(6.1458)	(0.6281)	(5.8361)	(-0.7529)	(6.1378)	(0.5793)	(5.8531)
profit margin	0.0000	-7.875e-05	-0.0000	-4.908e-05	0.0000	-8.074e-05	-0.0000	-4.847e-05	0.0000	-8.114e-05	-0.0000	-4.85e-05
	(0.0446)	(-1.7577)	(-0.1873)	(-1.0127)	(0.0893)	(-1.8357)	(-0.4079)	(-0.9833)	(0.0281)	(-1.8471)	(-0.4665)	(-0.9799)
cashflow	-0.0511	0.5236	0.2047	1.2296	-0.0576	0.5316	0.2883	1.2327	-0.0500	0.5321	0.3194	1.2359
	(-0.2674)	(1.9378)	(1.0381)	(3.4770)	(-0.2994)	(1.9832)	(1.0087)	(3.4944)	(-0.2611)	(1.9735)	(0.9018)	(3.5149)
market leverage	0.0010	-0.0034	0.0011	-0.0037	0.0010	-0.0035	0.0009	-0.0036	0.0010	-0.0034	0.0009	-0.0035
	(0.9509)	(-1.9692)	(1.1560)	(-1.9908)	(1.0176)	(-2.0157)	(0.8197)	(-1.9453)	(0.9677)	(-2.0054)	(0.6589)	(-1.9255)
book leverage	-0.0003	0.0017	-0.0005	0.0017	-0.0004	0.0017	-0.0004	0.0017	-0.0003	0.0017	-0.0004	0.0017
	(-0.7171)	(1.9430)	(-1.1331)	(1.6424)	(-0.7638)	(1.9698)	(-0.7331)	(1.6355)	(-0.7124)	(1.9634)	(-0.5658)	(1.6307)
dividend yield	0.0040	-0.0299	-0.0028	-0.0003	0.0043	-0.0298	-0.0026	0.0006	0.0039	-0.0298	-0.0026	0.0008
	(0.9133)	(-2.2281)	(-1.1251)	(-0.0125)	(0.9376)	(-2.1966)	(-0.7553)	(0.0204)	(0.8477)	(-2.1917)	(-0.6509)	(0.0290)
payout ratio	-0.0024	0.0321	-0.0016	0.0218	-0.0026	0.0322	-0.0001	0.0216	-0.0022	0.0322	0.0004	0.0216
	(-0.3901)	(5.2823)	(-0.3793)	(1.9197)	(-0.4911)	(5.2546)	(-0.0229)	(1.9051)	(-0.3997)	(5.2551)	(0.0550)	(1.9074)
sales growth	-0.0012	0.0033	-0.0005	0.0040	-0.0012	0.0033	-0.0003	0.0039	-0.0011	0.0033	-0.0002	0.0039
	(-1.6617)	(6.9573)	(-0.9699)	(12.937)	(-1.8831)	(6.9592)	(-0.2939)	(12.580)	(-1.8073)	(6.9560)	(-0.1535)	(12.375)
amihud liquidity measure	0.1824	-1.2717	0.0227	-1.4117	0.1940	-1.2716	-0.0750	-1.4111	0.1756	-1.2705	-0.1089	-1.4108
	(0.6611)	(-8.2525)	(0.1221)	(-11.096)	(0.8061)	(-8.2576)	(-0.2453)	(-11.093)	(0.7299)	(-8.2282)	(-0.2791)	(-11.089)
flag for multiple filings	0.0655	0.5817	0.1762	0.5442	0.0609	0.5798	0.2125	0.5352	0.0693	0.5786	0.2251	0.5347
	(0.4902)	(3.2368)	(1.7729)	(1.9512)	(0.5036)	(3.2361)	(1.5215)	(1.9541)	(0.5747)	(3.2342)	(1.3307)	(1.9620)
multiple occurrence (1st)	-0.0020	-0.4412	-0.0590	-0.3705	0.0020	-0.4414	-0.0847	-0.3704	-0.0045	-0.4410	-0.0937	-0.3701
	(-0.0209)	(-6.1213)	(-1.3133)	(-5.8719)	(0.0247)	(-6.1460)	(-1.1460)	(-5.8896)	(-0.0541)	(-6.1579)	(-0.9776)	(-5.8385)
multiple occurrence (2nd within 6MO)	0.0493	-0.4952	0.0095	-0.6882	0.0536	-0.4958	-0.0395	-0.6916	0.0465	-0.4951	-0.0559	-0.6899
	(0.5393)	(-3.3420)	(0.1012)	(-3.8933)	(0.6837)	(-3.3406)	(-0.2596)	(-3.9053)	(0.5780)	(-3.3446)	(-0.2892)	(-3.8932)
multiple occurrence (2nd after 6MO)	0.0870	-0.4796	-0.0255	-0.5380	0.0911	-0.4788	-0.0630	-0.5390	0.0842	-0.4784	-0.0758	-0.5387
	(0.7442)	(-5.1263)	(-0.4002)	(-6.1804)	(0.8791)	(-5.1097)	(-0.5699)	(-6.1821)	(0.8089)	(-5.1126)	(-0.5305)	(-6.2002)
flag for notice of delisting	0.1335	-0.3556	-0.0652	-0.5267	0.1370	-0.3587	-0.1021	-0.5292	0.1318	-0.3579	-0.1153	-0.5296
	(0.7916)	(-2.0889)	(-0.6853)	(-2.5630)	(0.8512)	(-2.1006)	(-0.7370)	(-2.5584)	(0.8275)	(-2.1026)	(-0.6840)	(-2.5531)
R-squared	-0.1154	0.409	0.0346	0.4401	-0.1355	0.4086	0.0015	0.4388	-0.1036	0.4083	-0.0501	0.4383
Adj. R-squared	-0.1361		0.0076		-0.1566		-0.0264		-0.1240		-0.0795	
Partial R-squared		0.0032		0.0022		0.0029		0.0013		0.0024		0.0009
F-statistic	7.711e+14	13.963	-3.808e+15	11.022	2.186e+15	20.487	-2.357e+15	5.1653	6.297e+15	16.312	-1.12e+15	2.8482
P-value (F-stat)	0.0000	0.0002	1.0000	0.0009	0.0000	6.004e-06	1.0000	0.023	0.0000	5.373e-05	1.0000	0.0915
number of observations	2362	2362	1288	1288	2362	2362	1288	1288	2362	2362	1288	1288

This table shows IV regression results for abnormal returns regressed over ownership stake (natural logarithm of dollar amounts). Values in parenthesis correspond to coefficients' t-statistics.
In all 5 panels, the dependent variable is *abnormal return* but with reference to a distinct pricing models. For example, *Panel A* shows regressions for which the dependent variable is abnormal return using the *market model*, while *Panel E* the dependent variable is abnormal return using the *Fama-French 5 factors model*.
Each panel contains 3 groups of results, each corresponding market returns calculated using different methodologies (*). From left to right, they are: *CRSP equal weighted*, *CRSP value weighted*, and *S&P500*. We use market returns to compute lagging 10-trading days market trend*(used as control variable and presented here for completeness) as well as deviations from the trend*(the instrumental variable). For each of the 3 market return categories, we present results for the 2 periods studied (2006-2022 and 2010-2019). For each period, we provide results for both the second stage and the first stage of two-stage least squares regressions (2SLS). The endogenous variable is *ownership*, and the instrument is the *deviation from market trend*. Regressions are controlled for market trend, industry and time fixed effects, firm-level variables (pre-event), and variables related to characteristics of the event (to control for biases in ownership data extraction).
Loadings for the pricing models were computed using the 100 trading days window ($t-121$ to $t-21$) that precedes the evaluation window. The evaluation window used to compute abnormal returns is the 41 days trading window from $t-20$ to $t+20$.

Table A.23: IV Regression full table: abnormal returns on ownership stake *(natural logarithm of dollar amounts) - continued*

Panel D: *Dependent variable: abnormal return (Fama-French 5 factors),* ± *20 days,* t_0*=event date*

	*equal weighted**				*value weighted**				*S&P500**			
	2006-2022		*2010-2019*		*2006-2022*		*2010-2019*		*2006-2022*		*2010-2019*	
	2nd stage	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*	*2nd stage*	*1st stage*
intercept	-1.8260	9.5759	-0.2363	8.9837	-1.9165	9.5610	0.4945	8.9647	-1.7866	9.5413	0.7418	8.9563
	(-1.0394)	(67.260)	(-0.2186)	(51.261)	(-1.1419)	(69.032)	(0.2727)	(51.409)	(-1.0654)	(69.630)	(0.3188)	(51.573)
ownership stake (log US$)	0.1600		0.0260		0.1700		-0.0559		0.1563		-0.0838	
	(0.8820)		(0.2132)		(0.9800)		(-0.2719)		(0.9002)		(-0.3183)	
deviation from trend		-6.1175		-6.6633		-5.6910		-4.7863		-5.2386		-4.0406
		(-3.7366)		(-3.3200)		(-4.5263)		(-2.2727)		(-4.0388)		(-1.6877)
market trend	-0.1012	1.3136	-0.5994	-2.4056	-0.2483	1.2639	-0.7140	-1.8793	-0.2224	1.2955	-0.6696	-1.6592
	(-0.2465)	(1.4884)	(-1.1084)	(-2.1747)	(-0.6506)	(1.4113)	(-1.0956)	(-1.7827)	(-0.5853)	(1.3977)	(-0.9518)	(-1.6827)
book-to-market	0.0426	-0.1463	0.0238	-0.2424	0.0439	-0.1464	0.0046	-0.2397	0.0419	-0.1469	-0.0018	-0.2388
	(0.8952)	(-1.7730)	(0.8196)	(-5.5383)	(0.9307)	(-1.7625)	(0.1040)	(-5.5131)	(0.9172)	(-1.7670)	(-0.0309)	(-5.4774)
cash-to-assets	0.0033	-0.0093	0.0014	-0.0071	0.0034	-0.0094	0.0008	-0.0071	0.0033	-0.0094	0.0006	-0.0071
	(1.4957)	(-5.6927)	(1.8766)	(-4.1546)	(1.5799)	(-5.6118)	(0.6734)	(-4.1537)	(1.5349)	(-5.5877)	(0.3846)	(-4.1634)
return-on-assets	-0.2570	0.7533	-0.1772	-0.0730	-0.2636	0.7474	-0.1812	-0.0749	-0.2530	0.7483	-0.1841	-0.0763
	(-0.9290)	(1.9417)	(-1.6431)	(-0.1489)	(-0.9506)	(1.9316)	(-1.8989)	(-0.1531)	(-0.8983)	(1.9251)	(-1.9844)	(-0.1564)
tobins' q	-0.0480	0.2468	0.0111	0.1857	-0.0503	0.2466	0.0270	0.1876	-0.0469	0.2467	0.0324	0.1882
	(-1.0100)	(6.1764)	(0.4663)	(5.8128)	(-1.0962)	(6.1458)	(0.6559)	(5.8361)	(-1.0170)	(6.1378)	(0.6186)	(5.8531)
profit margin	0.0000	-7.875e-05	0.0000	-4.908e-05	0.0000	-8.074e-05	-0.0000	-4.847e-05	0.0000	-8.114e-05	-0.0000	-4.85e-05
	(0.4118)	(-1.7577)	(0.1412)	(-1.0127)	(0.4395)	(-1.8357)	(-0.0696)	(-0.9833)	(0.3764)	(-1.8471)	(-0.1601)	(-0.9799)
cashflow	-0.0645	0.5236	0.2176	1.2296	-0.0718	0.5316	0.3165	1.2327	-0.0647	0.5321	0.3523	1.2359
	(-0.3020)	(1.9378)	(1.2378)	(3.4770)	(-0.3333)	(1.9832)	(1.2383)	(3.4944)	(-0.3022)	(1.9735)	(1.0995)	(3.5149)
market leverage	0.0009	-0.0034	0.0012	-0.0037	0.0009	-0.0035	0.0010	-0.0036	0.0009	-0.0034	0.0009	-0.0035
	(0.8227)	(-1.9692)	(1.2199)	(-1.9908)	(0.8439)	(-2.0157)	(0.8450)	(-1.9453)	(0.8011)	(-2.0054)	(0.6749)	(-1.9255)
book leverage	-0.0002	0.0017	-0.0005	0.0017	-0.0003	0.0017	-0.0003	0.0017	-0.0002	0.0017	-0.0003	0.0017
	(-0.4938)	(1.9430)	(-1.0803)	(1.6424)	(-0.5229)	(1.9698)	(-0.6514)	(1.6355)	(-0.4782)	(1.9634)	(-0.4837)	(1.6307)
dividend yield	0.0047	-0.0299	-0.0026	-0.0003	0.0050	-0.0298	-0.0025	0.0006	0.0046	-0.0298	-0.0024	0.0008
	(1.1356)	(-2.2281)	(-1.0873)	(-0.0125)	(1.0747)	(-2.1966)	(-0.7279)	(0.0204)	(0.9957)	(-2.1917)	(-0.6125)	(0.0290)
payout	-0.0045	0.0321	-0.0020	0.0218	-0.0048	0.0322	-0.0003	0.0216	-0.0043	0.0322	0.0003	0.0216
	(-0.8158)	(5.2823)	(-0.5059)	(1.9197)	(-0.9087)	(5.2546)	(-0.0515)	(1.9051)	(-0.8231)	(5.2551)	(0.0502)	(1.9074)
sales growth	-0.0013	0.0033	-0.0007	0.0040	-0.0014	0.0033	-0.0003	0.0039	-0.0013	0.0033	-0.0002	0.0039
	(-2.1025)	(6.9573)	(-1.3580)	(12.937)	(-2.1891)	(6.9592)	(-0.4253)	(12.580)	(-2.1295)	(6.9560)	(-0.2342)	(12.375)
amihud	0.2091	-1.2717	0.0399	-1.4117	0.2219	-1.2716	-0.0756	-1.4111	0.2045	-1.2705	-0.1150	-1.4108
	(0.8521)	(-8.2525)	(0.2422)	(-11.096)	(0.9551)	(-8.2576)	(-0.2678)	(-11.093)	(0.8846)	(-8.2282)	(-0.3160)	(-11.089)
flag for multiple filings	0.0424	0.5817	0.1694	0.5442	0.0372	0.5798	0.2128	0.5352	0.0451	0.5786	0.2275	0.5347
	(0.3616)	(3.2368)	(1.7612)	(1.9512)	(0.3252)	(3.2361)	(1.5945)	(1.9541)	(0.3970)	(3.2342)	(1.4103)	(1.9620)
multiple occurrence (1st)	0.0063	-0.4412	-0.0524	-0.3705	0.0107	-0.4414	-0.0827	-0.3704	0.0046	-0.4410	-0.0932	-0.3701
	(0.0743)	(-6.1213)	(-1.2065)	(-5.8719)	(0.1322)	(-6.1460)	(-1.2208)	(-5.8896)	(0.0569)	(-6.1579)	(-1.0569)	(-5.8385)
multiple occurrence (2nd within 6MO)	0.0704	-0.4952	0.0306	-0.6882	0.0751	-0.4958	-0.0268	-0.6916	0.0684	-0.4951	-0.0459	-0.6899
	(0.8598)	(-3.3420)	(0.3353)	(-3.8933)	(0.9681)	(-3.3406)	(-0.1861)	(-3.9053)	(0.8660)	(-3.3446)	(-0.2509)	(-3.8932)
multiple occurrence (2nd after 6MO)	0.0991	-0.4796	-0.0163	-0.5380	0.1037	-0.4788	-0.0606	-0.5390	0.0971	-0.4784	-0.0755	-0.5387
	(0.9494)	(-5.1263)	(-0.2867)	(-6.1804)	(1.0368)	(-5.1097)	(-0.6021)	(-6.1821)	(0.9710)	(-5.1126)	(-0.5743)	(-6.2002)
flag for notice of delisting	0.1299	-0.3556	-0.0577	-0.5267	0.1339	-0.3587	-0.1011	-0.5292	0.1290	-0.3579	-0.1163	-0.5296
	(0.8067)	(-2.0889)	(-0.6843)	(-2.5630)	(0.8382)	(-2.1006)	(-0.7853)	(-2.5584)	(0.8163)	(-2.1026)	(-0.7366)	(-2.5531)
R-squared	-0.1699	0.409	0.0220	0.4401	-0.1953	0.4086	0.0105	0.4388	-0.1605	0.4083	-0.0477	0.4383
Adj. R-squared	-0.1916		-0.0054		-0.2174		-0.0171		-0.1820		-0.0770	
Partial R-squared		0.0032		0.0022		0.0029		0.0013		0.0024		0.0009
F-statistic	-3.231e+14	13.963	1.65e+15	11.022	2.079e+14	20.487	9.098e+14	5.1653	8.609e+14	16.312	-4.562e+15	2.8482
P-value (F-stat)	1.0000	0.0002	0.0000	0.0009	0.0000	6.004e-06	0.0000	0.023	0.0000	5.373e-05	1.0000	0.0915
number of observations	2362	2362	1288	1288	2362	2362	1288	1288	2362	2362	1288	1288

This table shows IV regression results for abnormal returns regressed over ownership stake (natural logarithm of dollar amounts). Values in parenthesis correspond to coefficients' t-statistics.
In all 5 panels, the dependent variable is *abnormal return* but with reference to a distinct pricing models. For example, *Panel A* shows regressions for which the dependent variable is abnormal return using the *market model*, while *Panel E* the dependent variable is abnormal return using the *Fama-French 5 factors model*.
Each panel contains 3 groups of results, each corresponding market returns calculated using different methodologies (*). From left to right, they are: *CRSP equal weighted*, *CRSP value weighted*, and *S&P500*. We use market returns to compute lagging 10-trading days market trend*(used as control variable and presented here for completeness) as well as deviations from the trend*(the instrumental variable). For each of the 3 market return categories, we present results for the 2 periods studied (2006-2022 and 2010-2019). For each period, we provide results for both the second stage and the first stage of two-stage least squares regressions (2SLS). The endogenous variable is *ownership*, and the instrument is the *deviation from market trend*. Regressions are controlled for market trend, industry and time fixed effects, firm-level variables (pre-event), and variables related to characteristics of the event (to control for biases in ownership data extraction).
Loadings for the pricing models were computed using the 100 trading days window ($t-121$ to $t-21$) that precedes the evaluation window. The evaluation window used to compute abnormal returns is the 41 days trading window from $t-20$ to $t+20$.

Chapter 2

Is there a premium for sustainable development goals? The case of activist investors

Abstract

This paper investigates the relationship between company valuations after being targeted by blockholder activists when their stated investment objectives exhibit some degree of similarity to Sustainable Development Goals (SDGs) as developped by the UN. Our empirical approach employs Natural Language Processing (NLP) to establish similarity scores from the textual content of activist regulatory filings (investment objectives) and the SDGs. We find a robust positive relationship that is both economically and statistically significant. Switching from no similarity to the average similarity increases the abnormal returns around the filing date of an activist investor by approximately 2%, about one fifth of the average abnormal return. These findings imply that activist investors may play a crucial rule in the transition towards a more sustainable economy.

JEL Classification: G14, G23, G30, M14

Keywords: large shareholders, blockholders, activist investors, activism, corporate governance, sustainable finance, SDGs, textual analysis, NLP

2.1 Introduction

The literature on sustainable investments has grown substantially in recent years. One question that is of great interest is whether sustainable investments are detrimental to returns. While there is lot of grey literature and anecdotal evidence, in special during the late 2010's and early 2020's bull market, claiming evidence for the *sustainable finance business case*, scholars in Finance and Economics are reticent to assume conclusions on such ill and vaguely framed problem. We refer the reader to Edmans (2023) for a scholar perspective to many of the misconceptions that are currently taken for granted outside academia, as *"shareholder value is short-termism"*, *"more Environmental, Social and Governance (ESG) is always better"*, *"sustainable stocks earn higher returns"* and *"more investor engagement is always better"*, among others.

A close question, but conceptually different, is whether changes in company strategy that focus on sustainability are valued by market participants. If the previous question tries to evaluate if investing in companies with high level of sustainability is better, indifferent or worse than other investment strategies on a fundamental level, the latter question aims to evaluate how market participants anticipates the gains of moving companies strategies towards sustainable investment practices, independently if their beliefs are correct or not. The difference between these two questions, as seen in Pástor et al. (2021) and 2022, both theoretically and empirically, refers to the difference in expected returns, that given investor preferences for green assets should be lower, in contrast to observed/realized returns. As investor beliefs and preferences changes towards greener companies, there is an accommodation period with higher observed returns - which is obviously not to be confused as evidence of higher future expected returns. The change in preferences mechanism was also suggested by Choi et al. (2020): as retail investors revised their beliefs with respect to climate change, they sold carbon intensive companies and returns responded to the selling pressure without any connection to fundamentals.

In this paper we study how market participants react to changes into company strategies when these are connected to sustainability principles, as measured by textual similarity scores, using activist blockholders events as a laboratory. Blockholder activists are those with sizable ownership stakes that have stated intention to influence corporate decision-making and strategic direction. The literature on blockholders activism has produced considerable amount of evidence of the beneficial role these players have on targeted companies, by exerting close monitoring.[1] We evaluate if there is a premium post-intervention if a company is targeted by an activist that discloses investment objectives that can be to some extent linked to the Sustainable Development Goals (SDG)s. As activist investors drive change, this setup provides a way to measure a premium linked to sustainability due to a reorientation in the company strategy that is not driven from within.

As social or environmental concerns, once seen as investment niche, gained importance and are even becoming mandatory for institutional investors in many jurisdictions, investment decision processes necessarily incorporate risks and opportunities derived by those dimensions along traditional financial analysis. Conceptually, the formal investment analysis conducted by diligent investors have not changed. After all, *proper* investment decision processes have always taken into account governance, environmental and social aspects whenever they were deemed material to company businesses, independently of investment style or ideologies. However in practice, some important changes have occurred on the wake of intensification of the debate around corporate responsibility and the indisputable scientific evidence of antropogenic global warming.

First, as *regulations became tighter* for social and, most notably, environmental problems connected to companies activities, those dimensions gained economic importance on the investment analysis process. Some companies or sectors might be well positioned and benefit from tighter regulation, while the prospect for others might be dreadful (i.e. reclassify some assets as stranded). Moreover, there is much more information available to investors to take those aspects into consideration. Although far from ideal, non-financial disclosures and information derived from an array of alternative sources (usually provided by specialized vendors, including sustainability rating agencies) expand considerably the resources a diligent investor has at its disposal to perform analysis that take into account social and environmental aspects. To complete the scenario, there was an *increase in the salience* of sustainability related issues - in

[1] For a comprehensive review of the literature on blockholder's activism we refer the reader to Brav et al. (2022).

the age of social media, it does not take long for bad press on working conditions, human rights, animal rights and so on to affect consumers perception and consequently the revenues of a given brand. Hence as the investment landscape has changed, it is reasonable to expect that activist blockholders also start to target companies with salient sustainability-related issues. These new opportunities are considerably different from earlier blockholder activist interventions that were mainly concerned to governance aspects only.

Our paper makes multiple contributions to the Sustainable Finance literature. First, we contribute to the vast strand that discusses stock performance of sustainable investments. Most results in this literature focuses on single and very specific aspects of sustainability (i.e. carbon intensity, employment satisfaction). Bolton and Kacperczyk (2021) argue that emission intensity is a driver of higher stock returns as investors want to be rewarded by stocks with higher climatic risk. The effect is intensified as many institutional investors adopt exclusionary policies for those firms. Pástor et al. (2021) develops a theoretical model where investors adjust expectations of company cash-flows because of changes in preferences and regulations (i.e. due to environmental concerns) and extract utility by holding green assets. While green investors accept lower expected returns, the model predicts that green assets outperform brown assets when there is a shock to climate concerns. In an accompanying empirical paper, Pastor et al. (2022) contrast observed returns of those assets to their model prediction. They conclude that although realized returns were abnormally high for the period studied, expected returns for green assets should be below to ones for brown assets, consistent to the theory. We refer the reader to Giglio et al. (2021) for a literature review of climate finance, including pricing and hedging of climate risks, and awareness and attitudes of investors towards those risks. Still on that same paper, the authors observe, as climate change has gained importance among investors only recently, time series to estimate climate risk premium are rather short. This observation has important implications for our study, as we should expect to see larger market reaction only on the more recent part segment of our data series.

Our paper complements the results above in what concerns observed/realized returns, in various aspects. First it does not rely on a narrowly defined concept of sustainability. As, instead, we use a similarity measure that refers to all 17 SDGs, we have the potential to capture market reactions to themes as broad as biodiversity, water conservation or gender diversity, that will not be represented, for example in studies that uses greenhouse gas (GHG) emissions datasets. Second, by using our measure within the particular setting of blockholder activism, we can compare how events with SDG-aligned strategies drives a premium over events that do not have such objectives.

Of particular interest for us are the studies that connects prices, risks or behaviour to investor attention to climate risk. It is common for these studies to use textual data from news outlets to extract indices that can capture shocks to climate change concerns. Engle et al. (2020) creates a procedure to hedge climate change risk. Their approach uses textual similarity indices to measure attention (word count) and negative attention (sentiment analysis) on WSJ news to climate change vocabulary extracted from authoritative documents. Choi et al. (2020) documents increased attention to climate change (as proxied by Google searches) when temperature are higher in a geographic area and link these observations to retail investors selling high carbon intensive firms. Ardia et al. (2022) test empirically the model developed by Pástor et al. (2021) where the shocks to investors concerns is proxied by a index constructed using US major news outlets.

Notice that on all the papers cited above, natural language processing (NLP) is used to measure a climate change concern, while the proxy for firm-specific "greenness" is either ESG rating, its "E" component, or GHG emissions. Our paper differs in the way we apply NLP - we do not use it to extract generic - broad changes in climate concerns. Our usage is a specific, firm-level measure, for companies targeted by blockholder activists. This approach replaces traditionally used measures of sustainability, be it ESG ratings or ESG -related measures extracted from self-reported firm content (e.g. sustainability reports, MD&As). While both classes of sustainability measures are currently used in research, they are either prone to inconsistencies (ratings) or manipulation (self-reporting). On the other hand, the measure we use, similarity of blockholders investment objectives with the SDGs, is unlikely to suffer from those shortcomings. While we do not prove that activist blockholders do not intentionally apply vocabulary related to sustainability in regulatory fillings that could potentially induce other market participants to overreact to their intervention - and hence reap large profits faster, this is not likely to be the case. Activist blockholders literature has shown, on the contrary, that they take at least some months holding their position, and they actively intervene on investee companies, instead of immediately realizing profits after disclosing their positions. They do walk their talk.

We also contribute to the literature of sustainability-related investor activism. In recent years, there was an important shift on how shareholders' activism is perceived. As investors are called to be drivers of change, and to adhere to principles of responsible investment,[2] exclusionary policies or divestment ("exit") gave space to more active participation ("voice"). For example, Gormley et al. (2023) documents how the largest mutual funds managers have acted directly to increase board diversity, and how such measures have created externalities (i.e. have influenced change in policies of shareholders proxy services advisory). In addition to documenting success rates of such interventions (as measured in this early example by demographic diversity), most papers measure the impact of sustainability-related investor activism on company performance. The driver of change in performance in such studies is either shareholders proposals, or private engagements of specific groups or of a single institutional investor (Flammer (2015); Gillan and Starks (2007); Hoepner et al. (2022)). Our paper contributes to this literature without narrowing down on initiatives that need to be voted in order to be implemented (as in shareholder proposals) or being limited to outcomes that derive from activity of a single investor.

A positive impact in market performance that can be attributed to investor activism is of particular importance for institutional investors. As discussed on the very start of this introduction, there is a shortage of academic evidence (but plenty of agenda-driven gray literature) that backs-up the *general ESG investment case.* This gap in the literature is problematic for institutional investors, as they need ultimately to act in the best interest of their final beneficiaries - and some of their sustainability related beliefs might clash with those of their beneficiaries' that have very concrete investment objectives (e.g fund their children studies, retirement purposes). While basic tenets of responsible investment are indisputable (e.g human rights, zero tolerance of child labor, demographic discrimination or slavery), the term "sustainable investment" or " ESG -aligned investments" today incorporate more nuanced

[2] Increase in the number of signatories of the UN PRI exemplifies this trend.

topics, and attempts of institutional investors to influence companies on those can often times lead to heated debates of whether they (the investors) are breaching their fiduciary duties. Our paper comes to complement the literature in that regard, as it documents a positive sustainability-aligned premium that could be related to market participants pricing-in activist investor interventions, notably the ones related to sustainability. If those interventions drive larger market returns, such shareholder engagement is not in opposition to their fiduciary responsibilities.

Finally, we contribute to the literature on blockholders activism. Brav et al. (2008) inaugurated a long strand of literature of hedge fund activism. Particularly important for our paper is that the authors shows that not all abnormal excess returns after blockholders interventions are created equally. They depend on activist objective (e.g. sale of company, business strategy or capital restructuring). We complement their findings by identifying a subset of business strategy interventions, the ones related to sustainability, that became relevant in recent years.

In particular, while researchers have extensively studied the firm-level consequences of blockholders activism (operational, financial and market reaction) and whether activists returns are obtained in detriment of other stakeholders, there are some recent papers that used the blockholder activist setting to study sustainability related questions. Both Akey and Appel (2020) and Chu and Zhao (2019) use firm-level pollutants data and find that targeted firms do reduce toxic emissions after blockholder intervention, by closing high polluting plants and investing in green technologies. They also find a positive correlation of reduction of toxic releases and higher buy-and-hold returns. These results suggest that activist interventions are positive in terms of environmental impact, though they do not prove that activists were acting with the avert purpose to achieve better environmental indicators. After all, blockholder activists research has shown that they seek operational efficiency, through sale of less profitable plants among other strategic initiatives. It is likely that plants that are less profitable are probably older and more polluting - hence the ones that are closed. As for the increase in patents of green technologies is also expected as after being targeted, firms innovate more - in all areas. As patents in green technologies are a subset of patents in general - the positive relationship is maintained. Our paper contributes to that strand as we focus on the activist declared investment objective, instead on concrete pollution related outcomes. To our knowledge, our paper is the first one to explore the rich setting of acquisition of initial blocks to evaluate how market responds to shocks that changes company strategy that are to some extent aligned to sustainability goals.

As a final important practical implication of our results, the firm-wise evidence of anticipation of positive outcomes in market prices of costly measures that would be perceived to only materialize in the long run, can be a game changer for either some firms to act earlier or even to act at all, on sustainability related matters.

This paper is organized as follows: section 2.2 discusses traditional measures of sustainability and their limitations; section 2.3 outlines the construction of the similarity scores; in section 2.5 we present and analyse our results, with emphasis on how results for social aspects differs from the the ones related to environmental aspects. In section 2.6 we summarize our findings, discuss some limitations and give suggestions for future research.

Blockholder activism - non-causal relationship

At this point it is useful to underscore that findings of block activists studies refer to the following question: *"What is the effect of blockholder engagement, conditional on a company being targeted?"*, as opposed to the unconditional *"What is the effect of blockholder engagement?"*. The later question cannot be answered with real world data because targeted companies, are not randomly selected. Overall, have excess cash/low leverage, lower market capitalization, return on assets, profit margins and Tobin's Q.

The ideal experiment (Angrist and Pischke (2008)) to identify the effect block activist influence is to take pairs of identical companies and submit, for each pair, one to treatment (be targeted by a blockholder) and leave the other untreated. As there is not a second parallel universe for each firm, the empirical researcher needs to find alternatives for identification. If companies were randomly selected the average treatment effects (ATE) would be unbiased and be congruent to the results of the ideal experiment. However, companies are not randomly chosen because block activists are more likely to intervene on firms in which their engagement will provide potentially higher returns. Thus positive results would be due not only from activist intervention, but to a combination of it with stock picking. However researchers have found credible justifications to minimize the importance of stock picking in their findings (Albuquerque et al. (2022), Brav et al. (2022)).

It is important to highlight that by choosing the blockholder activist investor setting, the companies we study tend to have some characteristics specific to this group. There are no large firms, as for a investor to reach 5% participation of a large company it would represent an extremely large dollar amount. Second, they tend to have more cash and low Tobins' Q (value companies).

The block acquisition date and subsequent disclosure date (the filing date) marks two periods where we can study different regimes. This setting also lets us establish how market players perceive and thus price, corporate changes that lean on sustainability. Market participants drive stock prices upwards, once an activist blockholder discloses its presence, in anticipation of positive results that will be driven by the investor intervention. If additional abnormal returns are observed when companies are targeted by SDG-aligned activists, it would be used as a proxy for a sustainability premium.

A central assumption of our design is that *not* all blockholders have sustainability related objectives and that this understanding is shared by market participants. We assume only blockholders that express sustainability related concerns on the regulatory filings, the documents that disclose their presence, are perceived by the market as exerting influence on that direction. Later we develop an argument to support our belief that most blockholders until recently only focused on a rather narrow approach to create value as sustainability-related concerns became salient only recently.

2.2 The UN SDG similarity measure

In this section, we present topics concerning our selection of similarity measure. We start with a concise exploration of NLP applications in finance that leads to our choice of frequency-based vectorization over more sophisticated solutions like embeddings. Then we introduce the SDG as the foundation upon which sustainable finance has emerged, thereby justifying our utilization of SDG' goals and targets as the cornerstone for our similarity scores.

2.2.1 Natural language processing (NLP)

Tetlock (2007) seminal work marked a milestone in finance research by pioneering sentiment analysis[3] from unstructured data - financial news forecasts - to link it with market movements. Subsequent research explored various textual contexts, including social media, central bank communications, and press releases. Techniques employed have expanded beyond sentiment analysis, to include the large NLP toolkit (e.g. event extraction, summarization, topic modeling).

The fast-paced progress of textual analysis is clearly reflected in the subjects explored and methodologies adopted in financial literature. For example, Loughran and Mcdonald (2011) criticized the application of the psychology lexicon in Tetlock (2007), which prompted them to create a specialized lexicon for sentiment analysis in finance/accounting. Their approach involved categorizing terms based on their own industry knowledge. A decade later, machine learning might be used to tackle that problem from a completely different angle, with results that are arguably more efficient[4] and less susceptible to the subjectivity inherent in expert classifications.

NLP became a leading field within AI. Simultaneously, techniques have grown more potent, while also becoming more accessible and applicable. Open-source initiatives, cost-effective and scalable cloud computing services, along with transfer learning,[5] have collectively propelled NLP's adoption in both commercial applications and research endeavors within the finance domain.[6]

Particularly noteworthy is the pivotal role of NLP for sustainable finance. Distinct from conventional accounting metrics, the assessment of ESG factors predominantly relies on unstructured textual data sourced from various outlets such as news articles, social media, reports, and corporate communications. A crucial consideration for both researchers and practitioners

[3] The author used General Enquirer (GI) to extract sentiment scores. Developed in the 1960s, GI utilized the Harvard psychology lexicon to categorize words into 77 sentiment groups.

[4] Although the proposed solution might involve the fundamental challenge of labeling sentiments for a sample of texts in finance, an endeavor that should not be overlooked.

[5] In transfer learning, a pre-trained model is initially trained on a large dataset to learn general language features. Then, the model - that incorporates knowledge captured using resources intensively - can be fine-tuned or further trained on a smaller, domain-specific dataset to adapt it to a specific task or domain. Hugging Face's Transformers library is commonly used to access and fine-tune pre-trained models like those based on BERT, GPT, and other architectures for various NLP tasks.

[6] We refer the reader to Gentzkow et al. (2019) and Loughran and McDonald (2020) for a literature review in finance and accounting.

lies in discerning the origin of textual content. It can either stem from the company itself, potentially susceptible to greenwashing and biased viewpoints, or derive from a diverse audience where content manipulation is less likely. For instance, CSR reports, often showcasing company initiatives, might reflect marketing influences, skillfully tailoring a positive corporate social responsibility (CSR) impression. This underscores the limited utility of metrics extracted solely from CSR reports, or other self-reported sources as MD&As section of financial reports.

This discussion is particularly relevant for our analysis. Given that the textual data in this study originates from activist shareholders' regulatory filings, the potential manipulation of the text by them is a noteworthy consideration. Such manipulation could pose a predicament for our research design if investors deliberately introduced sustainability-related concerns as investment objectives with the intention to drive superior abnormal returns, such as profiting rapidly from price upswings. However, a more plausible scenario leans toward the contrary: that activists are unlikely to be overly concerned about furnishing information that accurately reflects investment objectives for specific investee company cases. Notably, investment purpose text is chiefly composed of boilerplate content, contrasting with the scarcity of unique, informative material. Frequently, only boilerplate content is present, particularly among investors who consistently hold blockholder activist positions and opt not to modify investment objectives in legal documentation on a case-by-case basis.

These characteristics — prevalence of boilerplate content, a scarcity of significant information, and crucially, whether informative content exists at all — limit the approaches we can use successfully to extract a similarity score. For example - word embeddings derived from potent generative models, such as ChatGPT - are no better than simple word frequency vectorization for our specific objective. While the former is powerful enough to detect writing style, its extra complexity (1,536 features) becomes detrimental to discerning whether there is a single reference to a term related to the SDGs. We provide some illustrative examples of embedding applied to our documents on Appendix C.

Changes in the related vocabulary

Before we proceed, there are two relevant observations to highlight here. First, simple NLP approaches do not preclude diligent work. Is one is to use word count of industry related "terms" it important both to be include all relevant terms in a target list and to understand that preference into using some specific terms do change over time. Let's illustrate it with a simple example. An expression that was used at large in the past,

UN SDGs and sustainability

The SDGs are a series of 17 goals, further broken down into 169 targets that in 2015 the UN member countries have pledged to implement by 2030.[7] The SDGs framework connects development objectives, as usually measured by economic growth, to social and environmental aspects. Although goals and objectives are adopted by nations and also at supra-national levels, it became clear that these goals could only be achieved with engagement of civil society and the private sector, whose actors are invited to act on micro-level initiatives (e.g. firm-level, communities).

SDGs are composed of three pillars: economic growth, social inclusion and environmental protection. The "economic" pillar is not accessory - but as relevant as the other two. The bottom line of SDGs its that economic progress is the vector of better living conditions and well-being, although it cannot be considered alone. To be clear, job creation and affordability (e.g. energy/housing/transportation) cannot be regarded as secondary to other social or environmental concerns. An example of the intertwining of these 3 pillars and how policy makers addressed them is the European Just Transition Mechanism (JTM). The transition to a low carbon economy, necessary and urgent for environmental reasons, affects regions (e.g. whole communities that depended on coal mining) and impacts other vulnerable population (e.g. increase energy prices). The term *just* in JTM, qualifies *transition*: JTM does set the foundations for any transition, but one done with justice, implementing mechanisms to support those mostly affected.

Sustainable finance has shared objectives with the SDGs and is a relevant element to help

[7]Other supranational initiatives preceded the SDGs and were later refined and incorporated into it, notably the Millennium Development Goals (MDGs).

achieve them. It emerges naturally as an aid for the capital allocation problem, as it (potentially) helps to identify investment options that are aligned with SDGs objectives. Besides re-orienting capital flows, sustainable finance can help reshape firms and businesses as these are led to disclose present social and environmental impacts, plans to improve in those dimensions and how associated risks are being managed. While sustainability is intimately linked to the Sustainable Development Goals (SDGs) we might lose sight of the big picture represented by the SDGs to privilege conventional measures as sustainability ratings. In part this is due to practical concerns - it seems useful to have measures to help redirect capital flows, gauge improvements and evaluate risks. In this paper we deviate from using ratings or other usual measures of sustainability. We created our sustainability scores based on text similarity of activist stockholders investment objectives, as filled in regulatory fillings and the SDGs.

Alternative measures of sustainability

Before we get into the details on how we construct our sustainability measures, we will discuss the shortcomings of two traditional approaches: ESG ratings (aggregate or single components) and data extracted from firms self-reported documents. Our objective with this discussion is to highlight that our results are not and should not be compared to research that use those metrics.

ESG ratings are widely used in the investment industry and also by academic scholars as a proxy of firm sustainability, despite being well-known and documented that rating agencies' ESG scores are so distinct that companies rankings, even in the same industry, are discrepant (Berg et al. (2022), Billio et al. (2021)). A crucial point for this divergence is how the three distinct ESG dimensions are aggregated to form a single score. But even when setting the weighting problem aside, when considering each dimension separately (E, S and G), the disparity persists because agencies use distinct approaches, methodologies, data sources and interpretations..[8]

The variability among ESG ratings, even the ones widely used in the industry as Sus-

[8]For example, one agency might use emission levels for environmental rating, while others might use emission per revenue. Some agencies might consider plans to curb future emission, others may score plans if there is evidence they are being implemented. Moving to a example for social thematic ratings, some might use NLP to get sentiment scores of employee reviews on Glassdoor, while others might not rely on any NLP tool, but rather use employee turnover, or number of labour disputes

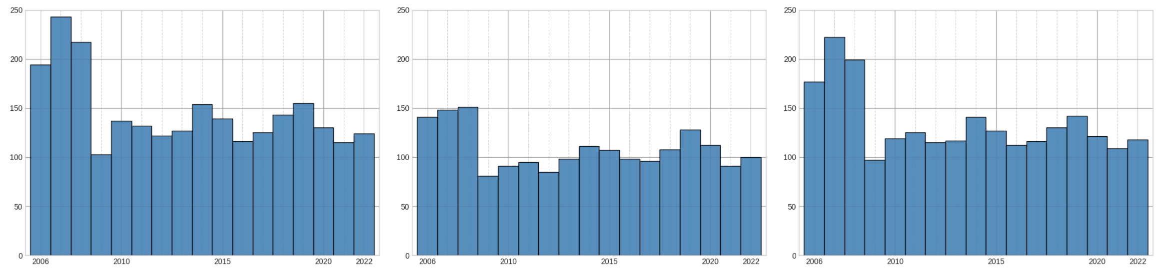
tainalytics, MSCI or Bloomberg evidences how difficult it is to rank firms objectively and in an unequivocal way. Research findings are often not robust to datasets from different rating providers. In other words, results are not generalizable across different rating agencies. Moreover, ESG rating disagreement, as proxied by the standard deviations of rating providers, seems to be an element that drives market premium and lowers demand for stocks (Rajna Gibson and Schmidt (2021), Avramov et al. (2022)), what further complicates their usage in studies such as the ones related to the so called *ESG business case*.

Other NLP-based metrics

These anecdotes exemplify an important concern for an analyst using NLP to obtain scores for non-structured data. Just as in non-financial reports, that are based on non-standardized and non-verified self reporting - speeches will often be crafted to reflect how one wants to be perceived. Some "ESG scoring" based on NLP might be less prone to manipulation, as for example, the ones that captures controversies based on thousands of news outlets.

2.3 Data

The main data source is the Electronic Data Gathering, Analysis, and Retrieval (SEC Edgar), from which we extract textual data from 13D beneficial ownership reports. We follow the methodology described in Cruz (2023) to extract only core blockholders events, meaning other non-blockholders filings (e.g. forms 4, 8Ks, DEFM14s, PREM14s) were also used. As defined in that paper, *core* events are the ones initiated by investors that have carefully selected target companies for which they have identified opportunities to enhance value with their intervention, which excludes among others, bankruptcies and reorganization, insider traders, derivatives and cases related to mergers. As result of using that methodology, our dataset starts on 2006 (the first year for which the identification of non-core events is easily attainable) and we also incorporate flags to control for multiple filings. It comprises 2,476 events initiated by 1,188 unique filers, targeting 1,762 unique targeted companies. Figure 2.1 shows the annual distribution of events, filers and targets.



(a) Events (b) Filers (c) Subjects

Figure 2.1: Unique Events, Filers and Subjects

This figure is based on **unique events** corresponding to 13D filings with unique *permno/date*, aggregated by year. Panel A shows the number of unique events, panel B shows the number of filers (unique CIKs) and Panel C shows the number of subject companies. The sample has been cleaned for duplicated records, companies outside the contiguous continental US and securities that are not common stocks. We dropped observations in the following cases: CUSIP, ownership stake or event date missing (or not parsed by our algorithm), no match CUSIP/PERMNOs, delta days between filing date and event date that is either negative or superior to 20 days.

Firm-related control variables, stock prices and volumes, as well as securities classes where obtained using Center for Research in Security Prices (CRSP)/Compustat, accessed via Wharton Research Data Services (WRDS) are reported in table 2.1

2.4 Similarity measure

2.4.1 Textual content

In this section we outline the procedure used to create the similarity measure. While we provide basic explanation for technical words from the field of NLP, we refer the reader to Jurafsky and Martin (2014) for further clarification of terms, concepts and techniques. In what follows we will use the terms *document* to refer to the text extracted from the section labeled *"item 4 - purpose of acquisition"*, for a single activist event and the term *corpus* to refer to the collection of all documents.

We perform a trivial documents pre-processing, consisting of exclusion of punctuation, stop words and boilerplate[9] text and then apply a stemming[10] algorithm. We refer to each semantic

[9] *"Boilerplate"* is a term to designate standardized text that is repeated across multiple documents that doesn't provide unique information

[10] *"Stemming"* means stripping suffixes from the end of words. Although this frequently leads to recovering

Table 2.1: Descriptive statistics: target fundamentals

	2006-2022				*2010-2019*			
	count	*mean*	*50%*	*std*	*count*	*mean*	*50%*	*std*
market capitalization	2476	1623.50	341.22	5665.37	1350	1725.21	341.22	5912.54
book-to-market	2476	0.60	0.49	0.66	1350	0.61	0.51	0.66
tobin's Q	2476	1.87	1.44	1.43	1350	1.83	1.41	1.38
sales growth	2287	0.56	0.05	15.18	1259	0.86	0.04	20.36
ROA	2344	0.02	0.08	0.27	1282	0.03	0.08	0.25
cashflow	2340	-0.02	0.05	0.30	1278	-0.01	0.04	0.31
market leverage	2475	22.67	13.83	24.92	1350	22.66	13.53	24.94
book leverage	2475	37.17	25.23	50.21	1350	37.85	25.31	53.66
cash-to-assets	2476	24.10	14.52	25.09	1350	23.64	15.04	24.03
dividend yield	2475	0.73	0.00	2.99	1350	0.64	0.00	2.14
payout ratio	2476	2.25	0.05	4.94	1350	2.10	0.09	4.32
profit margin	2404	-97.18	7.78	766.41	1318	-54.19	7.97	558.66
amihud liquidity measure	2476	0.31	0.11	0.58	1350	0.30	0.11	0.59

This table shows summary statistics for the targeted companies (in firm-months) in our sample for events initiated within two periods: from January, 2006 to December, 2022 and from January, 2010 to December, 2020. Market capitalization is in millions of dollars (May, 2023 dollar values); book-to-market is (book value of equity/market value of equity); tobin's Q is (book value of debt + market value of equity)/(book value of debt + book value of equity); ROA is EBITDA/lagged assets; cashflow is (net income + depreciation and amortization)/lagged assets; market leverage is total debt/(total debt + market value of equity); book leverage is total debt/(total debt + book value of equity); cash is (cash + cash equivalents) scaled by assets; dividend yield is common dividend/market value of equity; Payout ratio is (common dividend + share repurchases)/market value of equity. Amihud illiquidity measure is the yearly average (using daily data) of $10000\sqrt{\frac{|\text{Return}|}{\text{Dollar Trading Volume}}}$.

unit (the words after pre-processing) as *tokens*.

We get the textual content of the 17 SDGs, and apply to it the same pre-processing pipeline.

In our analysis we run all procedures for two versions of SDGs textual information: a *short version* based solely on the single statement that synthesizes the goal and a *long version*, that includes the SDG targets. We refer to the previous as *header* and the later as *long*. To illustrate, for the SDG 13, the **header** (before pre-processing) is:

> *"Take urgent action to combat climate change and its impacts",*

and the **long** (before pre-processing), includes the *goal* as well as its *targets*:

> *"Take urgent action to combat climate change and its impacts*
>
> *Strengthen resilience and adaptive capacity to climate-related hazards and natural disasters in all countries.*
>
> *Integrate climate change measures into national policies, strategies and planning.*
>
> *Improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning.*
>
> *Implement the commitment undertaken by developed-country parties to the United Nations Framework Convention on Climate Change to a goal of mobilizing jointly $100 billion annually by 2020 from all sources to address the needs of developing countries in the context of meaningful mitigation actions and transparency on implementation and fully operationalize the Green Climate Fund through its capitalization as soon as possible.*
>
> *Promote mechanisms for raising capacity for effective climate change-related planning and management in least developed countries and small island developing States, including focusing on women, youth and local and marginalized communities."*

words *roots*, this is not always the case, as *stemming algorithms* are rather simple.

2.4.2 Vectorization

Fundamental methods are *word count* and similarity. For instance, Flammer (2015) looked for the frequency of terms such as "profits", "performance" and "productivity" among others and their variations, as sign that shareholder proposal's were linked to a perception of value-enhancing. Conversely, Giglio et al. (2021) leverages authoritative texts to extract climate-related terms, offering a more robust approach compared to researcher-defined term selections.

Similarity scores are measures of distance between two vectors. These vectors can be any representation of textual documents, such as frequency-based vectors, where each feature is a different token (a word or pre-processed n-grams) or embeddings containing tokens. In our case we are measuring distance of vectors for each 13D event investment objective with respect to a reference vectors (SDGs).

We start by *training a vectorizer* using the SDG corpus. The trained vectorizer is an algorithm that receives a corpus as input and produces a matrix as output, where each row corresponds to a document and the columns contains features that corresponds to the tokens. In the simplest setup the features are comprised by 1-grams only (single tokens) - this is the base case for which we show results. In the appendix we show results using 2-grams (groups of 2 consecutive tokens). As the vectorizer was trained in the SDG pre-processed corpus, the number of features is limited by its number of tokens. For example, when training for 1-grams, the number of features is 110 and 761, for the header and the long versions respectively. We then feed the trained vectorizer with the investment objective corpus as the input. The output is a matrix where each row corresponds the 13D events. Finally the scores are computed, using cosine similarity.

2.5 Results

We investigate the importance of sustainable development goals for abnormal returns surrounding activist filings using modern machine learning methods for language similarities. One downside of using language based similarity measure is the lack of interpretability of results. To overcome this, we state results relative to the mean of the variables. Starting with the

Table 2.2: Descriptive statistics: SDG similarity scores

	2006-2022				*2010-2019*			
	count	*mean*	*50%*	*std*	*count*	*mean*	*50%*	*std*
goals	2477	0.059547	0.148372	0.16166	1349	0.058006	0.150694	0.164738
goals + targets	2477	0.049118	0.123741	0.11909	1349	0.047350	0.126484	0.121992

This table shows summary statistics for two types of similarity measures, of a composite SDGs textual content (concatenated) vs. activist investment objective, for two periods: from January, 2006 to December, 2022 and from January, 2010 to December, 2020. Years correspond to those where events were initiated (*event date*). There are two categories of similarity measures: *short* corresponds only to the SDG goals, and the *long* incorporates also textual content that refers to the targets. Similarity measures are obtained after applying same pre-processing for both the textual content of the reference (SDGs) and the investment objectives.

similarity measures, table 2.2 presents descriptive statistics for the similarity measures used in the regressions. More detailed tables, with additional years and more percentiles breakdowns are provided in the Appendix A.

Next, we present summary statistics for the abnormal returns. Table 2.3 presents descriptive statistics for abnormal returns, for different intervals centered around *filing date*, using as reference the market model. A more comprehensive table, with statistics for those same intervals, computed used other pricing models is presented on the Appendix, both for reference day on the *filing date* and on *event date* (Tables A.1 and A.2). Each panel contains abnormal returns calculated with respect to different pricing models.

Keeping the summary statistics in mind allows us to gauge the relevance of our main results as presented in the table (A.3). There, we see that the point estimate for the variable*SDG similarity score* is typically around *30*. Given that the sample mean of *SDG similarity score* is around *0.06* as displayed in Table 2.2. This implies that going from completely non-SDG related filing to an average filing implies an increase in abnormal returns of about 1.8 additional abnormal returns. An alternative economic interpretation relates to standard deviations. A one-standard deviation increase in *SDG similarity score* (a *0.16* increase) leads to a $30 * 0.16 = 4.8$ increase in abnormal returns. This is a meaningful increase given it is half the average abnormal return (or about 1/6 standard deviations of abnormal returns).

The results are fairly stable across specifications. They have a meaningful economic magnitude and a high statistical significance throughout, typically significant at the level of 5 percent level.

Similar to the other chapters, we run a large number of robustness checks relating to the way

Table 2.3: Descriptive statistics: abnormal returns *(market model)* centered around *filing date*

Panel A: *2006-2022*

interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	2360	6.36	21.38	-9.34	-2.99	3.21	11.84	23.15
[-10, 10]	2360	9.74	29.89	-15.50	-4.71	5.30	18.41	37.47
[-15, 15]	2360	10.79	34.04	-19.44	-6.64	5.28	22.28	42.87
[-20, 20]	2360	10.83	39.11	-25.51	-9.21	4.99	24.30	48.49
[0, 5]	2360	3.66	13.46	-6.34	-2.07	1.80	7.34	14.39
[0, 10]	2360	4.39	17.83	-9.46	-3.30	2.13	9.18	18.72
[0, 15]	2360	4.90	20.82	-12.26	-4.51	2.13	11.18	22.77
[0, 20]	2360	4.97	22.65	-15.10	-5.41	2.25	12.28	24.99

Panel B: *2010-2019*

interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	1282	5.78	16.89	-7.82	-2.50	3.05	10.93	21.01
[-10, 10]	1282	9.52	25.49	-12.88	-3.55	5.60	18.25	35.88
[-15, 15]	1282	10.44	30.00	-16.71	-4.89	5.80	21.47	39.65
[-20, 20]	1282	10.67	33.70	-21.08	-7.34	5.47	23.96	45.80
[0, 5]	1282	3.42	11.26	-5.20	-1.70	1.75	7.14	13.68
[0, 10]	1282	4.10	13.81	-7.32	-2.70	2.12	8.95	17.02
[0, 15]	1282	4.54	15.86	-9.52	-3.69	1.99	10.92	20.99
[0, 20]	1282	4.70	17.87	-12.12	-4.51	2.40	11.93	22.34

This table presents the buy and hold abnormal return, using the market model as reference, over various intervals. The reference period is the interval $[t_{\text{event}} - 120, t_{\text{event}} - 21]$. The number of days indicated in the column *interval* corresponds to trading days, with reference to the *filing date*. The abnormal returns are cumulative from the first day of the interval to the last day, inclusive. Note that, for each panel, the first four rows correspond to intervals where the limits are equidistant from the *filing date* (reference date), while the last four rows indicate abnormal returns without considering returns on lagging days. Statistics for the same intervals, using other 4 pricing models are presented in the Appendix, in Table A.1 (for returns centered around filing date) and Table A.2.

abnormal returns are calculated. We allow for simple market return corrections or corrections estimating betas in various benchmark models. Also, we check for different event windows, for example $\pm 5, \pm 10, \pm 15$ or ± 20 windows. Our results are robust to the exact specification. Tables with the robustness checks are presented in the Appendix ().

We also investigate results centered around filing date rather than trigger dates. Table A.2 presents descriptive statistics for abnormal returns, for different intervals centered around *event date*. Each panel contains abnormal returns calculated with respect to different pricing models. The reference intervals to compute loadings for the pricing models is $[-120, -21]$ with respect to *event date*.

2.6 Conclusion

Our paper is centered in the principle that some activist blockholders will acquire the investment block with the objectives to influence how companies treat sustainability related matters.

We ask a relatively simple question: do abnormal returns around activist filings exhibit higher abnormal returns if the language of the filing bears high resemblance to the language of the UN's sustainable development goals. We use modern natural language processing techniques embedded in a machine learning setting to construct a similarity score between the stated investment objectives of activist investors' 13–D filings with the UN's SDG goals. We find that across a broad range of specifications, a one standard deviation increase in the similarity score implies a 1/6 standard deviation increase in abnormal returns around the activist filing.

While we have investigated the overall reaction of market participants on the stated goals, further work needs to be done concerning the strategy and actual implementation of these proposed strategies. This is left for further research.

Table 2.4: Regression: Abnormal return over SDGs similarity *(market model)*

Dependent variable: *abnormal returns (market model) ± 15 days, t_0=filing date*

Panel A: SDG reference: *goals*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9987***	4.4178	4.2265	-0.5508	0.1645	10.3485***	-1.1559	-1.8219	0.1755	-0.3213
	(0.7154)	(4.7823)	(4.5078)	(7.1070)	(7.4495)	(0.8049)	(4.5762)	(4.1158)	(9.3725)	(9.2500)
SDG similarity *(ref: goals)*		20.3621	20.3233	25.9643**	26.0134**		37.8222**	36.9951**	36.4603**	35.1903**
		(13.3475)	(12.6429)	(11.7392)	(11.1782)		(16.3671)	(16.9561)	(15.9212)	(16.2533)

Panel B: SDG reference: *goals + targets*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9987***	3.6476	3.3146	-1.0950	-0.6013	10.3485***	-0.7569	-1.5269	1.0509	0.3747
	(0.7154)	(3.8125)	(3.6653)	(6.5563)	(7.0010)	(0.8049)	(3.7832)	(3.3631)	(8.8116)	(8.6388)
SDG similarity		30.8619**	31.0035**	37.5006***	37.9836***		40.9184**	40.1487**	39.4785*	38.6681*
		(12.3978)	(12.2954)	(11.7455)	(11.6850)		(19.2071)	(19.7876)	(20.2612)	(20.5638)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns (market model) over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Panel A and Panel B differs in terms of the reference textual content from SDG used to measure similarity scores. Panel A uses only the textual content of goals, and Panel B uses both goals and targets. Cumulative abnormal returns are measure over *(± 15 days*, centered around *t_0=filing date)*. Pricing model use as references observed returns over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in Cruz (2023), for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

References

Akey, P., & Appel, I. (2020). Environmental Externalities of Activism. *Working paper*. https://papers.ssrn.com/sol3/papers.cfm?abstractid=3508808

Albuquerque, R., Fos, V., & Schroth, E. (2022). Value Creation in Shareholder Activism. *Journal of Financial Economics*, *145*, 153–178. https://doi.org/10.1016/j.jfineco.2021.09.007

Angrist, J., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. www.degruyter.com/document/doi/10.1515/9781400829828/html

Ardia, D., Bluteau, K., Boudt, K., & Inghelbrecht, K. (2022). Climate Change Concerns and the Performance of Green Versus Brown Stocks. *Working paper*. https://ssrn.com/abstract=3717722

Avramov, D., Cheng, S., Lioui, A., & Tarelli, A. (2022). Sustainable Investing with ESG Rating Uncertainty. *Journal of Financial Economics*, *145*, 642–664. https://www.sciencedirect.com/science/article/abs/pii/S0304405X21003974

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance*, *26*, 1315–1344. https://academic.oup.com/rof/article/26/6/1315/6590670

Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, *28*, 1426–1445. https://doi.org/10.1002/csr.2177

Bolton, P., & Kacperczyk, M. (2021). Do Investors Care About Carbon Risk? *Journal of Financial Economics*, *142*, 517–549. https://doi.org/10.1016/j.jfineco.2021.05.008

Brav, A., Jiang, W., & Li, R. (2022). Governance by Persuasion: Hedge Fund Activism and Market-based Shareholder Influence. *ECGI working paper*. https://papers.ssrn.com/sol3/papers.cfm?abstractid=3955116

Brav, A., Jiang, W., Partnoy, F., & Thomas, R. (2008). Hedge Fund Activism, Corporate Governance, and Firm Performance. *Journal of Finance*, *63*, 1729–1775. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.2008.01373.x

Choi, D., Gao, Z., & Jiang, W. (2020). Attention to Global Warming. *Review of Financial Studies*, *33*, 1112–1145. https://doi.org/10.1093/rfs/hhz086

Chu, Y., & Zhao, D. (2019). Green Hedge Fund Activists. *Working paper*. https://ssrn.com/abstract=3499373

Cruz, P. P. G. (2023). Unveiling Non-Core Activist Blockholder Events. *Working Paper*.

Edmans, A. (2023). Applying Economics – Not Gut Feel – To ESG. *Financial Analysts Journal - forthcoming*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4346646

Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging Climate Change News. *Review of Financial Studies*, *33*, 1184–1216. https://doi.org/10.1093/rfs/hhz072

Flammer, C. (2015). Does Corporate Social Responsibility Lead to Superior Financial Performance? A Regression Discontinuity Approach. *Management Science*, *61*, 2549–2568. https://doi.org/10.1287/mnsc.2014.2038

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, *57*, 535–574. https://doi.org/10.1257/jel.20181020

Giglio, S., Kelly, B., & Stroebel, J. (2021). Climate Finance. *Annual Review of Financial Economics*. https://doi.org/10.1146/annurev-financial-102620

Gillan, S., & Starks, L. (2007). The Evolution of Shareholder Activism in the United States. *Journal of Applied Corporate Finance*, *19*, 55–73. https://onlinelibrary.wiley.com/doi/10.1111/j.1745-6622.2007.00125.x

Gormley, T. A., Gupta, V. K., Matsa, D. A., Mortal, S. C., & Yang, L. (2023). The Big Three and Board Gender Diversity: The Effectiveness of Shareholder Voice. *Journal of Financial Economics*, *149*, 323–348. https://doi.org/10.1016/j.jfineco.2023.04.001

Hoepner, A., Oikonomou, I., Sautner, Z., Starks, L. T., & Zhou, X. Y. (2022). ESG Shareholder Engagement and Downside Risk. *ECGI working paper*. https://papers.ssrn.com/sol3/papers.cfm?abstractid=2874252

Jurafsky, D., & Martin, J. H. (2014). *Speech and Language Processing*. Pearson.

Kedia, S., Starks, L. T., & Wang, X. (2021). Institutional Investors and Hedge Fund Activism. *Review of Corporate Finance Studies*, *10*, 1–43. https://doi.org/10.1093/rcfs/cfaa027

Loughran, T., & McDonald, B. (2020). Textual Analysis in Finance. *Annual Review of Financial Economics*, *12*(1), 357–375. https://doi.org/10.1146/annurev-financial-012820-032249

Loughran, T., & Mcdonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, *66*, 35–65. https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2010.01625.x

Pástor, Ľ., Stambaugh, R. F., & Taylor, L. A. (2021). Sustainable Investing in Equilibrium. *Journal of Financial Economics*, *142*, 550–571. https://doi.org/10.1016/j.jfineco.2020.12.011

Pástor, Ľ., Stambaugh, R. F., & Taylor, L. A. (2022). Dissecting Green Returns. *Journal of Financial Economics*, *146*, 403–424. https://doi.org/10.1016/j.jfineco.2022.07.007

Rajna Gibson, P. K., & Schmidt, P. (2021). ESG Rating Disagreement and Stock Returns. *Financial Analysts Journal*, *77*, 104–127. https://www.tandfonline.com/doi/abs/10.1080/0015198X.2021.1963186?journalCode=ufaj20

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance*, *62*, 1139–1168. https://www.jstor.org/stable/4622297

Appendix A

Additional tables and figures

Table A.1: Descriptive statistics: abnormal returns centered around *filing date*

Panel A: *abnormal return (CAPM), ± 15 days, t_0=filing date*

	2006-2022								2010-2019							
interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	2360	6.33	21.18	-9.58	-2.92	3.20	11.96	24.19	1282	5.76	16.80	-8.03	-2.51	3.20	11.05	21.10
[-10, 10]	2360	9.60	29.27	-15.92	-5.10	5.42	18.17	36.76	1282	9.44	25.35	-13.24	-3.67	5.53	18.13	34.94
[-15, 15]	2360	10.74	34.16	-20.03	-6.41	5.28	22.53	43.47	1282	10.41	29.82	-15.85	-4.59	5.76	21.62	39.41
[-20, 20]	2360	10.83	39.25	-26.36	-9.20	5.25	24.06	48.48	1282	10.71	33.77	-21.00	-6.96	5.96	23.97	45.95
[0, 5]	2360	3.65	13.36	-6.43	-2.07	1.69	7.40	14.56	1282	3.45	11.34	-5.18	-1.81	1.74	7.14	14.12
[0, 10]	2360	4.32	17.54	-9.91	-3.34	2.04	9.28	19.00	1282	4.10	13.91	-7.69	-2.90	2.12	8.86	17.16
[0, 15]	2360	4.79	20.61	-12.54	-4.69	2.12	11.29	22.64	1282	4.53	15.89	-9.69	-3.72	2.14	11.28	21.29
[0, 20]	2360	4.86	22.22	-14.99	-5.55	2.17	12.57	25.38	1282	4.70	17.87	-11.97	-4.39	2.58	11.86	22.58

Panel B: *abnormal return (Fama-French 3 factors model), ± 15 days, t_0=filing date*

	2006-2022								2010-2019							
interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	2360	6.35	21.37	-9.83	-2.81	3.46	11.75	24.49	1282	5.74	16.96	-8.36	-2.49	3.44	10.93	21.18
[-10, 10]	2360	9.66	29.54	-16.33	-4.70	5.36	18.00	36.66	1282	9.44	25.55	-13.38	-3.74	5.58	17.51	35.84
[-15, 15]	2360	10.72	34.47	-19.82	-6.62	5.60	22.70	42.73	1282	10.37	30.17	-16.18	-4.85	6.03	21.60	40.16
[-20, 20]	2360	10.70	39.26	-25.24	-9.36	5.11	23.66	48.33	1282	10.69	34.24	-20.38	-7.31	5.78	23.36	46.37
[0, 5]	2360	3.61	13.49	-6.63	-2.20	1.64	7.34	14.66	1282	3.41	11.37	-5.56	-1.80	1.74	7.06	13.58
[0, 10]	2360	4.35	17.76	-10.11	-3.44	2.00	9.43	19.08	1282	4.10	13.94	-7.78	-2.87	2.16	8.93	17.70
[0, 15]	2360	4.76	20.64	-12.48	-4.76	2.28	11.14	22.79	1282	4.52	16.02	-9.78	-3.91	2.54	11.15	20.72
[0, 20]	2360	4.81	22.29	-15.26	-5.71	2.24	12.00	25.20	1282	4.71	17.93	-12.28	-4.64	2.62	12.00	22.37

Panel C: *abnormal return (Fama-French 3 factors + momentum), ± 15 days, t_0=filing date*

	2006-2022								2010-2019							
interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	2360	6.20	21.48	-10.15	-3.05	3.38	11.75	24.45	1282	5.77	17.09	-8.19	-2.43	3.43	11.10	21.48
[-10, 10]	2360	9.41	29.71	-16.43	-4.91	5.21	18.14	36.97	1282	9.49	25.41	-13.22	-3.49	5.48	17.45	36.25
[-15, 15]	2360	10.40	34.49	-21.07	-7.03	5.64	22.70	41.85	1282	10.47	30.25	-17.15	-4.95	6.35	22.05	39.18
[-20, 20]	2360	10.33	39.63	-25.52	-10.06	5.26	24.01	47.76	1282	10.79	34.41	-20.17	-7.54	6.20	24.15	46.37
[0, 5]	2360	3.59	13.55	-6.79	-2.28	1.65	7.42	15.17	1282	3.46	11.49	-5.40	-1.90	1.82	7.27	14.05
[0, 10]	2360	4.31	17.98	-10.64	-3.51	1.98	9.51	19.29	1282	4.17	14.04	-8.23	-2.73	2.38	9.18	17.92
[0, 15]	2360	4.69	20.82	-12.72	-4.61	2.13	11.47	22.52	1282	4.60	16.14	-10.06	-3.73	2.42	11.54	21.25
[0, 20]	2360	4.66	22.51	-15.57	-6.06	2.14	12.01	25.66	1282	4.78	18.02	-12.61	-4.89	2.64	12.10	22.65

Panel D: *abnormal return (Fama-French 5 factors), ± 15 days, t_0=filing date*

	2006-2022								2010-2019							
interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	2360	6.40	21.63	-10.04	-2.84	3.56	11.66	24.14	1282	5.75	17.14	-8.60	-2.31	3.36	10.91	21.60
[-10, 10]	2360	9.64	29.89	-16.57	-4.81	5.26	18.61	36.01	1282	9.46	25.70	-13.42	-3.70	5.50	18.08	35.94
[-15, 15]	2360	10.66	34.90	-20.29	-7.05	5.24	22.20	42.51	1282	10.34	30.23	-16.16	-5.26	5.69	21.15	39.93
[-20, 20]	2360	10.57	40.12	-25.33	-9.55	5.06	23.40	48.83	1282	10.66	34.30	-20.32	-7.21	5.50	22.66	46.90
[0, 5]	2360	3.61	13.60	-6.76	-2.25	1.60	7.34	14.90	1282	3.44	11.53	-5.54	-1.88	1.62	7.08	14.01
[0, 10]	2360	4.30	17.66	-10.31	-3.61	1.96	9.70	18.86	1282	4.11	14.00	-7.79	-3.01	2.21	8.91	17.93
[0, 15]	2360	4.77	20.97	-12.32	-4.72	2.15	11.21	22.81	1282	4.59	16.25	-10.19	-3.88	2.34	11.01	21.23
[0, 20]	2360	4.73	22.58	-15.59	-6.08	2.10	12.14	25.17	1282	4.76	18.00	-12.09	-4.64	2.67	11.62	22.98

Each panel corresponds to the buy and hold abnormal return for different pricing models over various intervals. Abnormal returns are always calculated using loadings based on stock prices in the interval $[t_{\text{event}} - 120, t_{\text{event}} - 21]$. The number of days indicated in the column *interval* corresponds to trading days, with reference to the *filing date*. The abnormal returns are cumulative from the first day of the interval to the last day, inclusive. Note that, for each panel, the first four rows correspond to intervals where the limits are equidistant from the *filing date* (reference date), while the last four rows indicate abnormal returns without considering returns on lagging days.

Table A.2: Descriptive statistics: abnormal returns centered around *event date*

Panel A: *abnormal return (CAPM), ± 15 days, t_0=event date*

	2006-2022								*2010-2019*							
interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	2360	7.00	24.63	-11.32	-3.42	3.08	12.94	29.44	1282	6.91	22.03	-9.61	-2.54	3.18	12.36	28.77
[-10, 10]	2360	9.22	29.58	-16.41	-5.31	5.12	17.99	37.42	1282	9.19	26.97	-14.53	-3.85	5.20	17.72	36.11
[-15, 15]	2360	10.31	35.55	-21.25	-7.33	5.41	22.27	44.04	1282	10.13	30.27	-18.23	-5.06	5.44	22.55	42.30
[-20, 20]	2360	11.32	41.10	-24.95	-9.05	4.92	24.46	50.06	1282	10.81	33.86	-21.27	-6.90	5.44	24.64	48.10
[0, 5]	2360	6.03	19.91	-7.32	-1.99	2.55	9.46	22.05	1282	5.93	17.40	-5.97	-1.58	2.82	9.04	20.10
[0, 10]	2360	8.26	23.46	-8.89	-2.44	4.22	13.65	28.44	1282	8.10	19.78	-7.37	-1.76	4.49	13.21	27.82
[0, 15]	2360	8.89	25.62	-11.53	-3.71	4.68	15.80	32.27	1282	8.59	21.39	-9.38	-2.52	5.00	15.57	29.67
[0, 20]	2360	9.54	29.60	-14.35	-4.49	4.54	17.65	34.62	1282	8.93	23.18	-11.39	-3.38	4.68	17.62	33.01

Panel B: *abnormal return (Fama-French 3 factors model), ± 15 days, t_0=event date*

	2006-2022								*2010-2019*							
interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	2360	6.97	24.67	-11.46	-3.54	3.22	12.86	29.02	1282	6.83	22.25	-9.88	-2.61	3.29	12.08	28.00
[-10, 10]	2360	9.18	29.83	-16.81	-4.71	5.26	17.93	37.29	1282	9.09	27.24	-14.44	-3.64	5.26	17.30	35.42
[-15, 15]	2360	10.26	35.36	-21.37	-7.24	5.26	22.37	43.03	1282	10.09	30.61	-17.56	-5.08	5.62	21.60	41.65
[-20, 20]	2360	11.26	40.95	-24.32	-9.13	4.94	23.73	49.70	1282	10.90	34.51	-21.57	-7.21	5.82	23.60	48.40
[0, 5]	2360	6.02	19.93	-7.14	-2.06	2.63	9.12	21.43	1282	5.86	17.55	-5.90	-1.63	2.82	8.72	20.43
[0, 10]	2360	8.23	23.55	-9.16	-2.57	4.30	13.54	28.06	1282	7.98	19.83	-7.62	-1.88	4.49	13.12	27.42
[0, 15]	2360	8.89	25.75	-11.35	-3.39	4.68	15.66	32.24	1282	8.51	21.53	-9.32	-2.68	4.78	15.60	29.02
[0, 20]	2360	9.52	29.31	-13.98	-4.23	4.90	17.27	35.08	1282	8.91	23.42	-11.93	-3.17	5.15	16.72	33.19

Panel C: *abnormal return (Fama-French 3 factors + momentum), ± 15 days, t_0=event date*

	2006-2022								*2010-2019*							
interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	2360	6.80	24.54	-12.12	-3.90	3.36	12.77	28.39	1282	6.87	22.16	-9.92	-3.06	3.52	12.16	27.60
[-10, 10]	2360	8.99	29.99	-17.30	-5.03	5.22	17.99	35.82	1282	9.15	27.05	-14.14	-3.61	5.40	17.42	34.91
[-15, 15]	2360	10.02	35.67	-21.60	-7.49	5.46	22.08	42.65	1282	10.19	30.51	-17.92	-5.02	6.14	21.92	41.06
[-20, 20]	2360	11.00	41.17	-25.73	-9.61	4.96	23.96	49.37	1282	10.98	34.53	-21.37	-6.93	5.90	23.98	48.01
[0, 5]	2360	5.92	19.95	-7.41	-2.12	2.67	8.93	21.61	1282	5.88	17.52	-5.93	-1.69	2.89	8.58	20.83
[0, 10]	2360	8.11	23.72	-9.55	-2.63	4.14	13.64	28.32	1282	8.01	19.80	-7.61	-1.96	4.27	13.47	26.72
[0, 15]	2360	8.74	26.06	-11.82	-3.45	4.70	15.89	32.03	1282	8.58	21.59	-9.44	-2.41	4.87	15.89	29.91
[0, 20]	2360	9.32	29.61	-14.15	-4.44	4.77	17.75	34.92	1282	8.98	23.52	-11.17	-3.25	5.24	17.44	33.53

Panel D: *abnormal return (Fama-French 5 factors), ± 15 days, t_0=event date*

	2006-2022								*2010-2019*							
interval	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*	*count*	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
[-5, 5]	2360	7.03	25.20	-11.72	-3.69	3.08	13.11	29.06	1282	6.83	22.37	-9.99	-2.48	3.41	12.74	27.38
[-10, 10]	2360	9.18	30.47	-16.84	-4.96	5.24	18.04	36.34	1282	9.03	27.30	-14.27	-3.65	5.38	17.32	34.54
[-15, 15]	2360	10.17	36.02	-21.32	-7.33	5.22	21.91	42.73	1282	10.05	30.66	-17.96	-5.59	5.58	21.98	41.67
[-20, 20]	2360	11.15	41.58	-25.06	-9.82	5.44	23.50	49.60	1282	10.80	34.54	-21.19	-7.18	5.92	23.04	48.30
[0, 5]	2360	6.10	20.23	-7.36	-2.09	2.72	9.34	21.65	1282	5.91	17.60	-5.96	-1.66	2.80	8.78	20.87
[0, 10]	2360	8.26	23.81	-9.36	-2.65	4.28	13.90	28.18	1282	8.01	19.79	-7.23	-1.84	4.16	13.48	27.69
[0, 15]	2360	8.90	25.98	-11.84	-3.41	4.70	16.08	31.82	1282	8.60	21.64	-9.50	-2.58	4.90	15.86	29.73
[0, 20]	2360	9.57	30.03	-14.16	-4.57	4.90	17.35	35.99	1282	8.97	23.58	-11.47	-3.50	5.04	16.99	33.58

Each panel corresponds to the buy and hold abnormal return for different pricing models over various intervals. Abnormal returns are always calculated using loadings based on stock prices in the interval $[t_{\text{event}} - 120, t_{\text{event}} - 21]$. The number of days indicated in the column *interval* corresponds to trading days, with reference to the *event date*. The abnormal returns are cumulative from the first day of the interval to the last day, inclusive. Note that, for each panel, the first four rows correspond to intervals where the limits are equidistant from the *event date* (reference date), while the last four rows indicate abnormal returns without considering returns on lagging days.

Table A.3: Regression: Abnormal return over SDGs similarity
(reference textual content: SDG goals + targets)

Dependent variable: *abnormal returns* ± *15 days,* t_0*=filing date* **SDG reference:** *goals + targets*

Panel A: *pricing model:* CAPM

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9638***	3.8545	3.6366	-1.9989	-1.3017	10.3515***	-0.6193	-1.4503	1.5146	0.6116
	(0.7253)	(4.1112)	(3.8509)	(6.1721)	(6.4497)	(0.8028)	(3.5589)	(2.9857)	(8.8099)	(8.4602)
SDG similarity		30.3878**	30.3572**	36.7937***	37.1046***		40.8760**	40.0437*	38.8548*	37.9683*
		(13.3808)	(13.0486)	(12.2957)	(11.9972)		(20.2142)	(20.8115)	(20.9424)	(21.2730)

Panel B: *pricing model: Fama-French 3 factors*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9453***	5.0235	4.0636	-1.0311	-1.3901	10.3010***	-0.5495	-1.9098	0.4674	-1.0770
	(0.7274)	(4.9696)	(4.7141)	(6.5518)	(6.8453)	(0.8134)	(3.3656)	(2.8027)	(9.7798)	(9.4182)
SDG similarity		26.9851*	27.6727**	31.7447**	32.9156**		42.2571**	42.0798**	39.4117*	39.0796*
		(14.4686)	(14.0995)	(13.3606)	(13.0561)		(20.6278)	(21.1766)	(21.3699)	(21.6994)

Panel C: *pricing model: Fama-French 3 factors + momentum*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.6113***	7.4643	6.9342	1.8799	2.0976	10.4429***	1.2485	0.0848	2.2613	0.8973
	(0.7243)	(5.5173)	(5.2936)	(6.5715)	(6.8666)	(0.8205)	(3.2549)	(2.6183)	(10.0743)	(9.7179)
SDG similarity		25.8285*	26.1893*	30.1351**	30.9577**		41.0779**	40.5289**	38.4668*	37.7632*
		(14.5167)	(13.9710)	(13.3000)	(12.9230)		(20.1526)	(20.6198)	(20.8890)	(21.2060)

Panel D: *pricing model: Fama-French 5 factors*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9075***	3.3421	2.3375	-1.9863	-2.7853	10.3207***	-1.7074	-3.0793	-0.3999	-2.3417
	(0.7377)	(5.0701)	(4.8309)	(6.4675)	(6.6771)	(0.8183)	(3.4560)	(2.8596)	(10.4889)	(10.1621)
SDG similarity		29.5802**	30.3680**	34.5920***	36.0107***		43.9741**	43.9586**	42.1335**	42.0834**
		(14.4329)	(14.1221)	(13.0252)	(12.7807)		(20.1814)	(20.7343)	(20.9099)	(21.2076)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *(± 15 days*, centered around t_0*=filing date).* Pricing model references are measuared over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Table A.4: Regression: Abnormal return over SDGs similarity
(reference textual content: SDG goals)

Dependent variable: *abnormal returns* $\pm$ *15 days,* t_0*=filing date* — **SDG reference:** *goals*

Panel A: *pricing model:* CAPM

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9638***	4.4229	4.3323	-1.6481	-0.7502	10.3515***	-1.5221	-2.2642	0.1431	-0.5914
	(0.7253)	(5.1786)	(4.8470)	(6.7887)	(7.0151)	(0.8028)	(4.2811)	(3.6966)	(9.3077)	(9.0173)
SDG similarity		21.4008	21.2649	26.7375**	26.7160**		41.3319**	40.5172**	39.2794**	37.9786**
		(15.1632)	(14.3565)	(13.4064)	(12.7187)		(16.9608)	(17.5819)	(16.0978)	(16.4359)

Panel B: *pricing model: Fama-French 3 factors*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9453***	5.9715	5.1849	-0.2699	-0.3924	10.3010***	-1.1286	-2.3830	-0.5905	-1.9638
	(0.7274)	(6.3552)	(6.0701)	(7.4682)	(7.7694)	(0.8134)	(3.7666)	(3.1195)	(10.0667)	(9.7266)
SDG similarity		15.8521	16.0126	19.9079	20.3250		40.2360**	39.9163**	37.6238**	36.7743**
		(18.4921)	(17.6632)	(16.8689)	(16.1018)		(16.7264)	(17.1401)	(16.1561)	(16.3179)

Panel C: *pricing model: Fama-French 3 factors + momentum*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.6113***	8.7140	8.3827	2.9331	3.4095	10.4429***	0.9151	-0.1579	1.4772	0.2699
	(0.7243)	(6.8891)	(6.6140)	(7.6036)	(7.8791)	(0.8205)	(3.6725)	(2.9462)	(10.4232)	(10.0708)
SDG similarity		12.7370	12.4734	16.6196	16.6371		37.4978**	36.9622**	35.0674**	34.0218**
		(19.1007)	(18.2240)	(17.3164)	(16.5364)		(16.3683)	(16.7628)	(15.8150)	(15.9792)

Panel D: *pricing model: Fama-French 5 factors*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9075***	4.0558	3.2519	-1.4947	-2.0196	10.3207***	-2.3127	-3.5763	-1.4353	-3.1942
	(0.7377)	(6.3731)	(6.1067)	(7.2555)	(7.4766)	(0.8183)	(3.6860)	(2.9945)	(10.6407)	(10.3482)
SDG similarity		19.6911	19.7601	24.0221	24.3985		41.8894**	41.7179**	39.5862**	38.9248**
		(18.2369)	(17.5214)	(16.7831)	(16.1161)		(16.4178)	(16.8324)	(15.7054)	(15.9104)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *($\pm$ 15 days*, centered around t_0*=filing date).* Pricing model references are measured over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Table A.5: Regression: Abnormal return (CAPM) over SDGs similarity (targets) - full table

Dependent variable: *abnormal returns (pricing model: CAPM),* ± *15 days,* t_0*=filing date* **SDG reference:** *goals + targets*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9638***	3.8545	3.6366	-1.9989	-1.3017	10.3515***	-0.6193	-1.4503	1.5146	0.6116
	(0.7253)	(4.1112)	(3.8509)	(6.1721)	(6.4497)	(0.8028)	(3.5589)	(2.9857)	(8.8099)	(8.4602)
SDG similarity		30.3878**	30.3572**	36.7937***	37.1046***		40.8760**	40.0437*	38.8548*	37.9683*
		(13.3808)	(13.0486)	(12.2957)	(11.9972)		(20.2142)	(20.8115)	(20.9424)	(21.2730)
book-to-market				2.6684	2.4155				1.2308	1.1791
				(2.7729)	(2.7139)				(1.9972)	(2.0774)
cash-to-assets				0.0625	0.0632				0.0695	0.0743
				(0.0512)	(0.0530)				(0.0554)	(0.0554)
ROA				-1.2419	0.1283				-16.6570	-14.4224
				(12.6161)	(12.6464)				(11.3251)	(11.5469)
ln market capitalization				0.1369	-0.0936				-0.7607	-0.9977
				(0.6031)	(0.5776)				(0.6442)	(0.6626)
tobin's Q				0.3292	0.3926				1.2257	1.3401
				(0.9147)	(0.9032)				(1.2637)	(1.3259)
profit margin				-0.0007	-0.0007				-0.0008	-0.0005
				(0.0018)	(0.0019)				(0.0013)	(0.0013)
cashflow				-2.1099	-2.5927				21.2672***	20.6807***
				(13.6296)	(13.3722)				(6.4780)	(6.9215)
market leverage				0.0237	0.0207				0.1003***	0.0983***
				(0.0572)	(0.0590)				(0.0352)	(0.0339)
book leverage				-0.0047	-0.0066				-0.0391**	-0.0413**
				(0.0233)	(0.0233)				(0.0179)	(0.0175)
dividend yield				0.0030	-0.0132				-0.1885	-0.2194
				(0.5015)	(0.5194)				(0.2218)	(0.2269)
payout ratio				0.1514	0.0938				-0.2391	-0.2752*
				(0.1726)	(0.1678)				(0.1538)	(0.1455)
sales growth				-0.0641***	-0.0604***				-0.0424***	-0.0402***
				(0.0143)	(0.0135)				(0.0119)	(0.0126)
amihud liquidity measure				3.1856	2.7173				0.1213	0.1840
				(2.1328)	(1.9178)				(1.6585)	(1.6559)
flag for multiple filings				20.2666***	19.8652***				19.3329**	18.3623**
				(6.9500)	(7.0084)				(9.0782)	(9.1281)
multiple *(1^{st} occurrence)*				-4.6029**	-4.1584**				-3.9545*	-3.3337
				(1.9351)	(1.8041)				(2.3686)	(2.2522)
multiple *(2^{nd} within 6MO)*				-0.1271	-0.0107				1.2941	1.5638
				(2.7406)	(2.7619)				(3.6986)	(3.7669)
multiple *(2^{nd} after 6MO)*				0.9826	0.9722				-2.5177*	-2.5127*
				(2.1542)	(2.1724)				(1.4387)	(1.4578)
notice of delisting flag				-0.4190	-0.4458				-13.2691***	-13.3186***
				(5.7757)	(5.7308)				(3.8239)	(4.0114)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0052	0.0183	0.0260	0.0363	0.0000	0.0088	0.0195	0.0486	0.0594
R-squared adj.	-0.0000	0.0016	0.0074	0.0143	0.0175	0.0000	0.0021	0.0054	0.0276	0.0311
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *(± 15 days*, centered around t_0*=filing date)*. Pricing model references are measuared over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Table A.6: Regression: Abnormal return (Fama-French 3 factors model) over SDGs similarity (targets) - full table

Dependent variable: *abnormal returns (pricing model: FF3),* ± *15 days, t_0=filing date* **SDG reference:** *goals + targets*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9453***	5.0235	4.0636	-1.0311	-1.3901	10.3010***	-0.5495	-1.9098	0.4674	-1.0770
	(0.7274)	(4.9696)	(4.7141)	(6.5518)	(6.8453)	(0.8134)	(3.3656)	(2.8027)	(9.7798)	(9.4182)
SDG similarity		26.9851*	27.6727**	31.7447**	32.9156**		42.2571**	42.0798**	39.4117*	39.0796*
		(14.4686)	(14.0995)	(13.3606)	(13.0561)		(20.6278)	(21.1766)	(21.3699)	(21.6994)
book-to-market				2.8115	2.6915				1.6023	1.6017
				(2.6150)	(2.5775)				(2.0353)	(2.0993)
cash-to-assets				0.0576	0.0612				0.0750	0.0801
				(0.0562)	(0.0573)				(0.0620)	(0.0625)
ROA				-2.2517	-0.6694				-16.7132	-14.4038
				(12.4315)	(12.6037)				(11.5386)	(11.9281)
ln market capitalization				0.1522	-0.0395				-0.7437	-0.9594
				(0.5496)	(0.5374)				(0.7126)	(0.7276)
tobin's Q				0.6559	0.6854				1.6360	1.7278
				(0.8981)	(0.8958)				(1.3663)	(1.4211)
profit margin				-0.0016	-0.0016				-0.0008	-0.0005
				(0.0026)	(0.0026)				(0.0014)	(0.0014)
cashflow				-0.6414	-1.3642				21.4317***	20.6994***
				(12.9035)	(12.8681)				(6.0677)	(6.7394)
market leverage				0.0294	0.0275				0.1093***	0.1061***
				(0.0592)	(0.0613)				(0.0408)	(0.0406)
book leverage				-0.0098	-0.0112				-0.0396**	-0.0414**
				(0.0236)	(0.0238)				(0.0196)	(0.0190)
dividend yield				-0.1034	-0.1207				-0.1369	-0.1835
				(0.5150)	(0.5286)				(0.2231)	(0.2322)
payout ratio				0.2009	0.1646				-0.2319	-0.2700*
				(0.2092)	(0.2071)				(0.1577)	(0.1521)
sales growth				-0.0583***	-0.0549***				-0.0386***	-0.0370***
				(0.0145)	(0.0143)				(0.0132)	(0.0142)
amihud liquidity measure				2.9687	2.6882				0.5453	0.5513
				(2.0834)	(1.9237)				(1.9600)	(1.9409)
flag for multiple filings				17.9385***	17.5596***				17.8711**	16.9738*
				(6.7139)	(6.7852)				(8.9684)	(9.0426)
multiple *(1^{st} occurrence)*				-4.9630***	-4.6189***				-4.0555*	-3.5645*
				(1.5585)	(1.5110)				(2.1542)	(2.0538)
multiple *(2^{nd} within 6MO)*				-0.2101	-0.0935				1.2478	1.5266
				(2.5439)	(2.5849)				(3.7322)	(3.7919)
multiple *(2^{nd} after 6MO)*				0.8593	0.8775				-2.9303**	-2.9746**
				(2.2418)	(2.2574)				(1.4480)	(1.4651)
notice of delisting flag				-1.7701	-1.9193				-14.6861***	-14.7531***
				(5.6330)	(5.6417)				(3.9603)	(4.1315)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0044	0.0130	0.0250	0.0316	0.0000	0.0086	0.0169	0.0504	0.0591
R-squared Adj.	-0.0000	0.0008	0.0021	0.0133	0.0127	0.0000	0.0020	0.0028	0.0294	0.0308
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *(± 15 days*, centered around *t_0=filing date)*. Pricing model references are measuared over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Table A.7: Regression: Abnormal return (Fama-French 3 factors + momentum) over SDGs similarity (targets) - full table

Dependent variable: *abnormal returns (model: FF3 + momentum)*, ± *15 days*, t_0=*filing date* **SDG reference:** *goals + targets*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.6113***	7.4643	6.9342	1.8799	2.0976	10.4429***	1.2485	0.0848	2.2613	0.8973
	(0.7243)	(5.5173)	(5.2936)	(6.5715)	(6.8666)	(0.8205)	(3.2549)	(2.6183)	(10.0743)	(9.7179)
SDG similarity		25.8285*	26.1893*	30.1351**	30.9577**		41.0779**	40.5289**	38.4668*	37.7632*
		(14.5167)	(13.9710)	(13.3000)	(12.9230)		(20.1526)	(20.6198)	(20.8890)	(21.2060)
book-to-market				2.8364	2.7795				1.5335	1.5157
				(2.3095)	(2.3328)				(2.1123)	(2.1796)
cash-to-assets				0.0462	0.0498				0.0871	0.0922
				(0.0578)	(0.0578)				(0.0591)	(0.0592)
ROA				-3.4653	-1.8010				-15.8208	-13.8033
				(12.1465)	(12.4101)				(11.7834)	(12.1832)
ln market capitalization				0.0784	-0.1348				-0.7452	-0.9604
				(0.5707)	(0.5578)				(0.7126)	(0.7243)
tobin's Q				0.8829	0.9404				1.5549	1.6633
				(0.8703)	(0.8816)				(1.3468)	(1.4065)
profit margin				-0.0020	-0.0021				-0.0011	-0.0008
				(0.0027)	(0.0028)				(0.0014)	(0.0014)
cashflow				1.5058	0.8018				20.9233***	20.4200***
				(12.2290)	(12.2618)				(6.8540)	(7.4006)
market leverage				0.0232	0.0249				0.1012**	0.0990**
				(0.0546)	(0.0561)				(0.0455)	(0.0455)
book leverage				-0.0059	-0.0086				-0.0342	-0.0364
				(0.0229)	(0.0229)				(0.0227)	(0.0222)
dividend yield				-0.1516	-0.1540				-0.1436	-0.1802
				(0.5533)	(0.5661)				(0.2371)	(0.2495)
payout ratio				0.2572	0.2355				-0.2093	-0.2460
				(0.2169)	(0.2150)				(0.1685)	(0.1643)
sales growth				-0.0554***	-0.0519***				-0.0391***	-0.0369***
				(0.0129)	(0.0129)				(0.0127)	(0.0137)
amihud liquidity measure				2.8523	2.6147				0.6200	0.6104
				(2.2957)	(2.1410)				(2.0015)	(1.9951)
flag for multiple filings				17.7749***	17.1559***				18.1505**	17.3250*
				(6.4892)	(6.5887)				(8.9744)	(9.0543)
multiple *(1^{st} occurrence)*				-4.8694***	-4.4838***				-3.7521*	-3.2428
				(1.5727)	(1.5704)				(2.1742)	(2.0738)
multiple *(2^{nd} within 6MO)*				-1.0078	-0.8926				0.9459	1.2440
				(2.5141)	(2.6142)				(3.6644)	(3.7188)
multiple *(2^{nd} after 6MO)*				0.5386	0.5556				-2.9370*	-2.9487*
				(2.0383)	(2.0339)				(1.5412)	(1.5633)
notice of delisting flag				-1.0019	-0.8813				-14.4966***	-14.5204***
				(5.7148)	(5.7067)				(3.9066)	(4.0541)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0037	0.0118	0.0243	0.0307	0.0000	0.0066	0.0142	0.0470	0.0550
R-squared Adj.	-0.0000	0.0001	0.0009	0.0126	0.0118	0.0000	-0.0001	0.0001	0.0260	0.0266
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *(± 15 days*, centered around t_0=*filing date)*. Pricing model references are measuared over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Table A.8: Regression: Abnormal return (Fama-French 5 factors model) over SDGs similarity (targets) - full table

Dependent variable: *abnormal returns (pricing model: FF5),* *± 15 days, t_0=filing date* **SDG reference:** *goals + targets*

	2006-2022					2010-2019				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9075***	3.3421	2.3375	-1.9863	-2.7853	10.3207***	-1.7074	-3.0793	-0.3999	-2.3417
	(0.7377)	(5.0701)	(4.8309)	(6.4675)	(6.6771)	(0.8183)	(3.4560)	(2.8596)	(10.4889)	(10.1621)
SDG similarity		29.5802**	30.3680**	34.5920***	36.0107***		43.9741**	43.9586**	42.1335**	42.0834**
		(14.4329)	(14.1221)	(13.0252)	(12.7807)		(20.1814)	(20.7343)	(20.9099)	(21.2076)
book-to-market				3.4693	3.3846				1.9304	1.9517
				(2.5660)	(2.5392)				(2.1121)	(2.1681)
cash-to-assets				0.0689	0.0732				0.0909	0.0964*
				(0.0587)	(0.0592)				(0.0577)	(0.0580)
ROA				0.1678	1.9310				-17.3216	-14.8707
				(12.5267)	(12.7830)				(11.5220)	(11.9675)
ln market capitalization				-0.0466	-0.2279				-0.9704	-1.1753
				(0.5426)	(0.5337)				(0.7476)	(0.7615)
tobin's Q				0.7198	0.7207				1.6366	1.7148
				(0.8784)	(0.8809)				(1.3883)	(1.4418)
profit margin				-0.0017	-0.0016				-0.0005	-0.0002
				(0.0026)	(0.0027)				(0.0015)	(0.0015)
cashflow				-1.6098	-2.4867				23.0186***	22.0679***
				(13.0627)	(13.0813)				(6.0595)	(6.7543)
market leverage				0.0074	0.0048				0.1078**	0.1051**
				(0.0668)	(0.0688)				(0.0463)	(0.0462)
book leverage				0.0077	0.0067				-0.0302	-0.0318
				(0.0274)	(0.0277)				(0.0219)	(0.0211)
dividend yield				-0.1115	-0.1252				-0.1698	-0.2197
				(0.5290)	(0.5415)				(0.2516)	(0.2611)
payout ratio				0.1516	0.1251				-0.2091	-0.2438*
				(0.1967)	(0.1956)				(0.1509)	(0.1456)
sales growth				-0.0642***	-0.0607***				-0.0409***	-0.0397***
				(0.0149)	(0.0147)				(0.0124)	(0.0133)
amihud liquidity measure				2.6417	2.4750				0.1933	0.2157
				(2.0196)	(1.8577)				(1.8859)	(1.8765)
flag for multiple filings				16.1172**	15.7743**				17.8002*	16.9470*
				(6.5921)	(6.6628)				(9.1769)	(9.2675)
multiple *(1^{st} occurrence)*				-5.2088***	-4.8312***				-3.1835	-2.7122
				(1.6045)	(1.6101)				(2.2100)	(2.1046)
multiple *(2^{nd} within 6MO)*				-0.5525	-0.4074				1.8812	2.1652
				(2.6311)	(2.6906)				(3.6780)	(3.7267)
multiple *(2^{nd} after 6MO)*				0.1082	0.1505				-2.9036*	-2.9575*
				(2.1405)	(2.1524)				(1.5230)	(1.5500)
notice of delisting flag				-0.6647	-0.8201				-14.2832***	-14.3692***
				(5.7128)	(5.7253)				(3.8181)	(4.0078)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0051	0.0119	0.0243	0.0295	0.0000	0.0105	0.0175	0.0528	0.0605
R-squared Adj.	0.0000	0.0015	0.0010	0.0126	0.0106	0.0000	0.0038	0.0034	0.0318	0.0323
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *(± 15 days*, centered around *t_0=filing date).* Pricing model references are measuared over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Table A.9: Regression: Abnormal return (CAPM) over SDGs similarity (goals) - full table

Dependent variable: *abnormal returns (pricing model: CAPM),* ± *15 days,* t_0=*filing date* **SDG reference:** *goals*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9638***	4.4229	4.3323	-1.6481	-0.7502	10.3515***	-1.5221	-2.2642	0.1431	-0.5914
	(0.7253)	(5.1786)	(4.8470)	(6.7887)	(7.0151)	(0.8028)	(4.2811)	(3.6966)	(9.3077)	(9.0173)
SDG similarity		21.4008	21.2649	26.7375**	26.7160**		41.3319**	40.5172**	39.2794**	37.9786**
		(15.1632)	(14.3565)	(13.4064)	(12.7187)		(16.9608)	(17.5819)	(16.0978)	(16.4359)
book-to-market				2.6661	2.4064				1.3587	1.2992
				(2.7711)	(2.7139)				(1.9452)	(2.0264)
cash-to-assets				0.0608	0.0612				0.0685	0.0728
				(0.0514)	(0.0534)				(0.0563)	(0.0565)
ROA				-0.8194	0.4210				-16.4593	-14.3063
				(12.6009)	(12.6432)				(11.1664)	(11.3697)
ln market capitalization				0.1779	-0.0483				-0.6964	-0.9299
				(0.6105)	(0.5861)				(0.6453)	(0.6655)
tobin's Q				0.3598	0.4316				1.2937	1.4092
				(0.9174)	(0.9043)				(1.2384)	(1.2988)
profit margin				-0.0008	-0.0008				-0.0009	-0.0006
				(0.0018)	(0.0019)				(0.0012)	(0.0012)
cashflow				-2.3838	-2.7806				21.1338***	20.5912***
				(13.6514)	(13.4103)				(6.4254)	(6.8623)
market leverage				0.0213	0.0187				0.0970***	0.0951***
				(0.0576)	(0.0594)				(0.0349)	(0.0338)
book leverage				-0.0044	-0.0065				-0.0388**	-0.0412**
				(0.0234)	(0.0235)				(0.0182)	(0.0178)
dividend yield				0.0276	0.0112				-0.1780	-0.2071
				(0.5046)	(0.5214)				(0.2218)	(0.2284)
payout ratio				0.1455	0.0874				-0.2420	-0.2765*
				(0.1741)	(0.1691)				(0.1564)	(0.1486)
sales growth				-0.0643***	-0.0607***				-0.0441***	-0.0417***
				(0.0141)	(0.0133)				(0.0122)	(0.0131)
amihud liquidity measure				3.2351	2.7500				0.2580	0.3160
				(2.1197)	(1.9115)				(1.6767)	(1.6741)
flag for multiple filings				20.1974***	19.7885***				19.4470**	18.4727**
				(6.9569)	(7.0105)				(9.2621)	(9.3134)
multiple *(1st occurrence)*				-4.6997**	-4.2679**				-4.0412*	-3.4222
				(1.9239)	(1.7844)				(2.3036)	(2.1847)
multiple *(2nd within 6MO)*				-0.1685	-0.0600				1.3272	1.6075
				(2.7565)	(2.7735)				(3.6634)	(3.7343)
multiple *(2nd after 6MO)*				0.9759	0.9665				-2.4890*	-2.4792*
				(2.1251)	(2.1448)				(1.4450)	(1.4633)
notice of delisting flag				-0.2707	-0.2844				-12.7188***	-12.7795***
				(5.6734)	(5.6318)				(3.6748)	(3.8538)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0048	0.0178	0.0254	0.0357	0.0000	0.0113	0.0219	0.0509	0.0614
R-squared Adj.	-0.0000	0.0011	0.0069	0.0137	0.0169	0.0000	0.0047	0.0079	0.0299	0.0332
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *(± 15 days*, centered around t_0=*filing date)*. Pricing model references are measuared over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Table A.10: Regression: Abnormal return (Fama-French 3 factors model) over SDGs similarity (goals) - full table

Dependent variable: *abnormal returns (pricing model: FF3),* *± 15 days, t_0=filing date* **SDG reference:** *goals*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9453***	5.9715	5.1849	-0.2699	-0.3924	10.3010***	-1.1286	-2.3830	-0.5905	-1.9638
	(0.7274)	(6.3552)	(6.0701)	(7.4682)	(7.7694)	(0.8134)	(3.7666)	(3.1195)	(10.0667)	(9.7266)
SDG similarity		15.8521	16.0126	19.9079	20.3250		40.2360**	39.9163**	37.6238**	36.7743**
		(18.4921)	(17.6632)	(16.8689)	(16.1018)		(16.7264)	(17.1401)	(16.1561)	(16.3179)
book-to-market				2.8084	2.6793				1.7276	1.7198
				(2.6150)	(2.5801)				(1.9898)	(2.0528)
cash-to-assets				0.0563	0.0595				0.0740	0.0786
				(0.0564)	(0.0576)				(0.0628)	(0.0634)
ROA				-1.9010	-0.4238				-16.4813	-14.2532
				(12.4536)	(12.6320)				(11.4600)	(11.8261)
ln market capitalization				0.1878	0.0006				-0.6811	-0.8925
				(0.5509)	(0.5395)				(0.7136)	(0.7291)
tobin's Q				0.6792	0.7169				1.7073	1.8017
				(0.9002)	(0.8964)				(1.3328)	(1.3853)
profit margin				-0.0017	-0.0016				-0.0009	-0.0006
				(0.0026)	(0.0026)				(0.0013)	(0.0013)
cashflow				-0.8369	-1.4839				21.3033***	20.6180***
				(12.9913)	(12.9708)				(6.0676)	(6.7217)
market leverage				0.0274	0.0258				0.1061***	0.1031**
				(0.0598)	(0.0619)				(0.0409)	(0.0409)
book leverage				-0.0097	-0.0113				-0.0395**	-0.0413**
				(0.0237)	(0.0239)				(0.0200)	(0.0194)
dividend yield				-0.0851	-0.1020				-0.1270	-0.1714
				(0.5197)	(0.5320)				(0.2289)	(0.2388)
payout ratio				0.1966	0.1598				-0.2352	-0.2719*
				(0.2116)	(0.2095)				(0.1599)	(0.1547)
sales growth				-0.0581***	-0.0548***				-0.0401***	-0.0383***
				(0.0146)	(0.0144)				(0.0134)	(0.0145)
amihud liquidity measure				2.9922	2.6983				0.6677	0.6704
				(2.0493)	(1.9002)				(1.9835)	(1.9651)
flag for multiple filings				17.8194***	17.4281***				17.9305**	17.0294*
				(6.6874)	(6.7569)				(9.1420)	(9.2244)
multiple *(1^{st} occurrence)*				-5.0445***	-4.7085***				-4.1500**	-3.6614*
				(1.5483)	(1.4930)				(2.0893)	(1.9856)
multiple *(2^{nd} within 6MO)*				-0.2475	-0.1364				1.2868	1.5770
				(2.5539)	(2.5941)				(3.6972)	(3.7593)
multiple *(2^{nd} after 6MO)*				0.8484	0.8665				-2.9009**	-2.9397**
				(2.2148)	(2.2299)				(1.4533)	(1.4697)
notice of delisting flag				-1.6706	-1.8025				-14.1763***	-14.2495***
				(5.5117)	(5.5271)				(3.8656)	(4.0284)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0037	0.0123	0.0242	0.0307	0.0000	0.0104	0.0187	0.0520	0.0605
R-squared Adj.	-0.0000	0.0001	0.0014	0.0125	0.0118	0.0000	0.0038	0.0046	0.0310	0.0323
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *(± 15 days*, centered around *t_0=filing date)*. Pricing model references are meausured over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Table A.11: Regression: Abnormal return (Fama-French 3 factors + momentum) over SDGs similarity (goals) - full table

Dependent variable: *abnormal returns (model: FF3 + momentum),* ± *15 days, t_0=filing date* **SDG reference:** *goals*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.6113***	8.7140	8.3827	2.9331	3.4095	10.4429***	0.9151	-0.1579	1.4772	0.2699
	(0.7243)	(6.8891)	(6.6140)	(7.6036)	(7.8791)	(0.8205)	(3.6725)	(2.9462)	(10.4232)	(10.0708)
SDG similarity		12.7370	12.4734	16.6196	16.6371		37.4978**	36.9622**	35.0674**	34.0218**
		(19.1007)	(18.2240)	(17.3164)	(16.5364)		(16.3683)	(16.7628)	(15.8150)	(15.9792)
book-to-market				2.8326	2.7650				1.6525	1.6263
				(2.3104)	(2.3352)				(2.0755)	(2.1418)
cash-to-assets				0.0451	0.0484				0.0862	0.0909
				(0.0579)	(0.0580)				(0.0599)	(0.0601)
ROA				-3.1424	-1.5804				-15.5711	-13.6375
				(12.2070)	(12.4707)				(11.6959)	(12.0748)
ln market capitalization				0.1122	-0.0973				-0.6862	-0.8977
				(0.5741)	(0.5623)				(0.7155)	(0.7277)
tobin's Q				0.9028	0.9678				1.6262	1.7365
				(0.8713)	(0.8811)				(1.3170)	(1.3745)
profit margin				-0.0021	-0.0022				-0.0012	-0.0009
				(0.0027)	(0.0028)				(0.0014)	(0.0014)
cashflow				1.3497	0.7237				20.8031***	20.3483***
				(12.3361)	(12.3785)				(6.8405)	(7.3761)
market leverage				0.0214	0.0233				0.0982**	0.0962**
				(0.0552)	(0.0567)				(0.0457)	(0.0458)
book leverage				-0.0060	-0.0088				-0.0342	-0.0364
				(0.0230)	(0.0230)				(0.0231)	(0.0225)
dividend yield				-0.1364	-0.1386				-0.1345	-0.1690
				(0.5571)	(0.5687)				(0.2446)	(0.2574)
payout ratio				0.2536	0.2316				-0.2128	-0.2482
				(0.2191)	(0.2173)				(0.1708)	(0.1670)
sales growth				-0.0550***	-0.0516***				-0.0404***	-0.0380***
				(0.0130)	(0.0129)				(0.0129)	(0.0139)
amihud liquidity measure				2.8608	2.6102				0.7275	0.7145
				(2.2588)	(2.1152)				(2.0310)	(2.0247)
flag for multiple filings				17.6191***	16.9856***				18.1665**	17.3409*
				(6.4565)	(6.5529)				(9.1326)	(9.2188)
multiple *(1^{st} occurrence)*				-4.9453***	-4.5625***				-3.8494*	-3.3402*
				(1.5589)	(1.5481)				(2.1197)	(2.0169)
multiple *(2^{nd} within 6MO)*				-1.0445	-0.9322				0.9881	1.2962
				(2.5230)	(2.6221)				(3.6255)	(3.6821)
multiple *(2^{nd} after 6MO)*				0.5246	0.5410				-2.9082*	-2.9146*
				(2.0132)	(2.0083)				(1.5470)	(1.5688)
notice of delisting flag				-0.9280	-0.7908				-14.0351***	-14.0672***
				(5.5940)	(5.5948)				(3.8390)	(3.9765)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0029	0.0109	0.0233	0.0296	0.0000	0.0077	0.0153	0.0480	0.0558
R-squared Adj.	-0.0000	-0.0008	-0.0000	0.0116	0.0107	0.0000	0.0011	0.0012	0.0270	0.0275
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *(± 15 days*, centered around *t_0=filing date).* Pricing model references are measuared over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Table A.12: Regression: Abnormal return (Fama-French 5 factors model) over SDGs similarity (goals) - full table

Dependent variable: *abnormal returns (pricing model: FF5),* ± *15 days, t_0=filing date* **SDG reference:** *goals*

	2006-2022					*2010-2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	10.9075***	4.0558	3.2519	-1.4947	-2.0196	10.3207***	-2.3127	-3.5763	-1.4353	-3.1942
	(0.7377)	(6.3731)	(6.1067)	(7.2555)	(7.4766)	(0.8183)	(3.6860)	(2.9945)	(10.6407)	(10.3482)
SDG similarity		19.6911	19.7601	24.0221	24.3985		41.8894**	41.7179**	39.5862**	38.9248**
		(18.2369)	(17.5214)	(16.7831)	(16.1161)		(16.4178)	(16.8324)	(15.7054)	(15.9104)
book-to-market				3.4668	3.3739				2.0630	2.0774
				(2.5596)	(2.5338)				(2.0713)	(2.1251)
cash-to-assets				0.0674	0.0713				0.0899	0.0949
				(0.0588)	(0.0595)				(0.0586)	(0.0592)
ROA				0.5602	2.2087				-17.0647	-14.6994
				(12.5414)	(12.8063)				(11.4262)	(11.8435)
ln market capitalization				-0.0081	-0.1840				-0.9043	-1.1042
				(0.5471)	(0.5392)				(0.7496)	(0.7639)
tobin's Q				0.7475	0.7572				1.7136	1.7952
				(0.8813)	(0.8832)				(1.3633)	(1.4148)
profit margin				-0.0018	-0.0017				-0.0006	-0.0004
				(0.0026)	(0.0027)				(0.0014)	(0.0014)
cashflow				-1.8529	-2.6478				22.8833***	21.9833***
				(13.1437)	(13.1797)				(6.0998)	(6.7714)
market leverage				0.0052	0.0029				0.1044**	0.1018**
				(0.0674)	(0.0694)				(0.0468)	(0.0469)
book leverage				0.0079	0.0067				-0.0301	-0.0318
				(0.0275)	(0.0279)				(0.0224)	(0.0217)
dividend yield				-0.0894	-0.1028				-0.1594	-0.2069
				(0.5350)	(0.5462)				(0.2592)	(0.2695)
payout ratio				0.1463	0.1193				-0.2127	-0.2460*
				(0.1996)	(0.1984)				(0.1533)	(0.1483)
sales growth				-0.0643***	-0.0608***				-0.0425***	-0.0410***
				(0.0151)	(0.0149)				(0.0126)	(0.0136)
amihud liquidity measure				2.6815	2.4982				0.3196	0.3391
				(1.9977)	(1.8451)				(1.9140)	(1.9042)
flag for multiple filings				16.0312**	15.6711**				17.8476*	16.9900*
				(6.5786)	(6.6465)				(9.3300)	(9.4305)
multiple *(1^{st} occurrence)*				-5.2991***	-4.9341***				-3.2865	-2.8183
				(1.5942)	(1.5918)				(2.1407)	(2.0330)
multiple *(2^{nd} within 6MO)*				-0.5921	-0.4549				1.9245	2.2211
				(2.6446)	(2.7022)				(3.6242)	(3.6754)
multiple *(2^{nd} after 6MO)*				0.1002	0.1422				-2.8722*	-2.9198*
				(2.1172)	(2.1283)				(1.5271)	(1.5530)
notice of delisting flag				-0.5353	-0.6754				-13.7520***	-13.8418***
				(5.5876)	(5.6058)				(3.7322)	(3.9144)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
industry fx effects	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0046	0.0112	0.0236	0.0287	0.0000	0.0124	0.0194	0.0544	0.0618
R-squared Adj.	0.0000	0.0009	0.0003	0.0119	0.0098	0.0000	0.0058	0.0053	0.0334	0.0337
number of observations	2199	2199	2199	2199	2199	1202	1202	1202	1202	1202

This table shows the coefficients and standard errors (in parenthesis) for regression of abnormal returns over a measure of similarity between the stated investment objective of activist investors and UN SDGs. Cumulative abnormal returns are measure over *(± 15 days*, centered around *t_0=filing date)*. Pricing model references are measuared over the period t_{-120} to t_{-20}. Columns 1 to 5 refers to the period 2006 to 2022 and Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. Standard errors are clustered at year level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy). Firm-specific controls are predetermined: the latest available value before the evaluation window. For summary statistics of these controls as well as their formulas, please refer to Table 2.1. The *stated investment objective* is extracted from the informational element *Item 4*, of SC 13D filings. Filings have been pre-processed, using the methodology proposed in 2023, for keeping only core-events. The last 5 regression controls are also defined on the aforementioned paper.

Appendix B

UN Sustainable Development Goals

1. End poverty in all its forms everywhere
2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture
3. Ensure healthy lives and promote well-being for all at all ages
4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
5. Achieve gender equality and empower all women and girls
6. Ensure availability and sustainable management of water and sanitation for all
7. Ensure access to affordable, reliable, sustainable and modern energy for all
8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all
9. Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation
10. Reduce inequality within and among countries progressively achieve and sustain income growth of the bottom per cent of the population at a rate higher than the national average
11. Make cities and human settlements inclusive, safe, resilient and sustainable
12. Ensure sustainable consumption and production patterns
13. Take urgent action to combat climate change and its impacts*.
14. Conserve and sustainably use the oceans, seas and marine resources for sustainable development
15. Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss ensure the conservation, restoration and sustainable use of terrestrial and inland freshwater ecosystems and their services, in particular forests, wetlands, mountains and drylands, in line with obligations under international agreements
16. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels
17. Strengthen the means of implementation and revitalize the global partnership for sustainable development

Appendix C

Embeddings

This chapter presents plots generated from embedding our documents using OpenAI. We have two primary aims here. Firstly, is to provide an overall characterization of the textual content our dataset, and contrast them to the various SDGs used as references in our study. Secondly, it helps illustrate why a more sophisticated approach, such as embedding, is not better suited as the foundation for our similarity measure than simple frequency-based vectorization of pre-processed text.

Embedding vectors are generated using complex models that consider the context of surrounding words to create multidimensional representations. The values along each dimension are derived from fitting the textual data to a trained machine learning model and, except for explicit examination within the realm of explainable machine learning, one should not expect to establish a straightforward relationship of embedding dimensions with specific attributes of the original text. This stands in stark contrast to frequency-based vectors, where each element contains the count of occurrences of individual words (or other pre-processed semantic units/tokens).

We used OpenAI's GPT-3 *text-embedding-ada-002* model to create embeddings for both the event documents and SDG goals in our corpus. In Figure C.1, we showcase the embeddings' high dimensionality by plotting a heatmap of the its first 100 dimensions across a sample of 100 activist events. Each column in the heatmap corresponds to one embedding dimension, while the rows represent individual events.

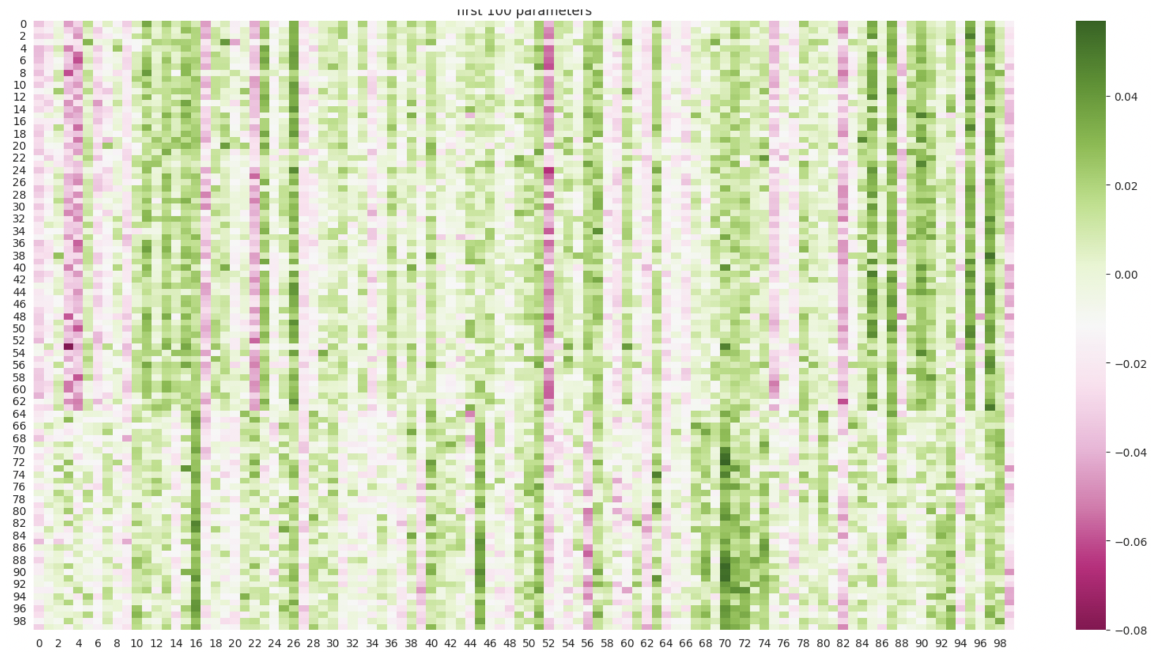


Figure C.1: Embedding: Parameters' Heatmap

This figure shows a heatmap for the embeddings for a subsample for documents containing the investment objectives (Item 4 of SC 13D filings). Each column corresponds to embedding parameters. The plot only includes the first 100 parameters (out of 1536). Each row corresponds to a single document.

In figure C.2 we employed Stochastic Neighbor Embedding (SNE)[1] to reduce the embedding vector's dimensionality from 1,536 to 2. This reduction allows us to visualize the data in a 2-dimensional plot. Blue dots on the plot corresponds to activist events, while other points clustered in the upper right corner represent SDGs. Those marked with a red X symbolize the goals, while the green dots represent goals along with their associated objectives. Remarkably, the inclusion of objectives doesn't substantially alter the positioning of SDG-related points, despite the latter encompassing larger texts with additional unique words.

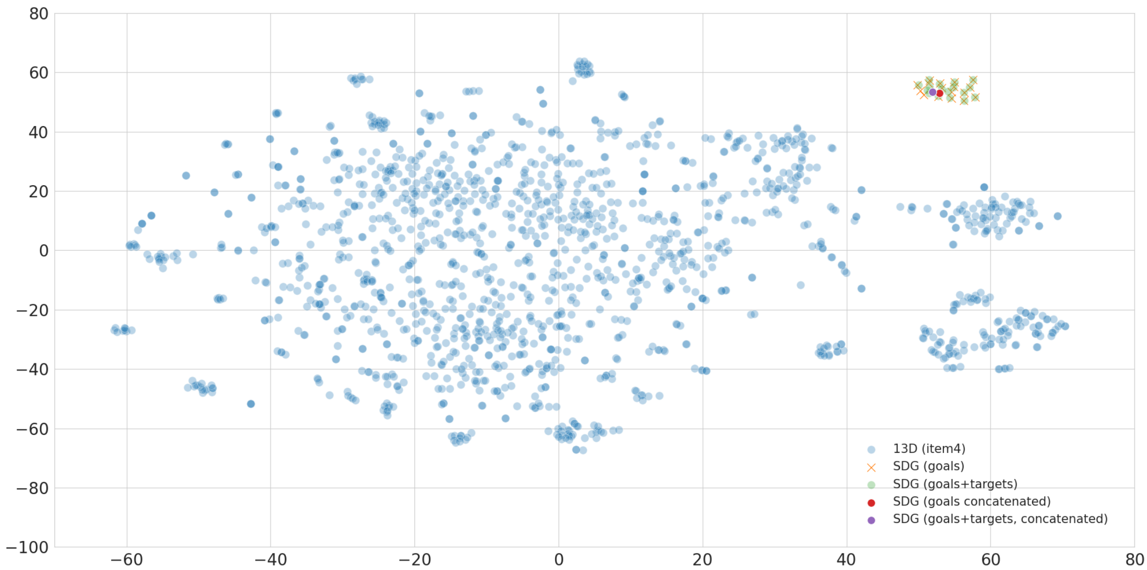


Figure C.2: Embeddings: investment objective and SDGs

This picture shows the plot of embeddings obtained using OpenAI GPT-3 *text-embedding-ada-002*. The embedding vector dimension was reduced from 1,536 to 2 using Stochastic Neighbor Embedding (SNE).

These embeddings capture not only semantic content but also encompass style, tone, and contextual nuances present in the text snippets. These aspects explain two characteristics of the blue dots' 2D distribution: a larger degree of variability compared to the 17 SDGs documents and the emergence of clusters. This variability arises from diverse authors and writing styles, with some utilizing more boilerplate text and offering limited informative content, while others provide detailed references to individuals and specific situations.

[1]SNE is used to visualize high-dimensional data in a lower-dimensional space while preserving the pairwise similarities between data points.

When distinct authors (filers) contribute to the variability, it leads to the formation of clusters around the central cloud of dots. Filings from the same filer often employ a consistent template, with minor adjustments for coherence with specific events. As a result, the authorship footprint significantly influences the embeddings, what is extremely problematic for our application. For a concrete example, even for a filer that adjusts his template to provide informative content about his objectives, the effect that a reference to to topics like pollution, climate, or discrimination will have on the embeddings will be overshadowed by authorship and style, rendering the resulting similarity metrics ineffective.

There are evidently more sophisticated ways to devise embedding models that can be trained to capture those specific traits on our texts, but this would represent another research endeavor. As we are interested in the economic application, we just want to find a method that can be easily applied and that works in our context. For our purposes word count seems to be more effective and suitable for our needs.

When distinct authors (filers) contribute to the variability, it leads to the formation of clusters around the central cloud of dots. Filings from the same filer often employ a consistent template, with minor adjustments for coherence with specific events. As a result, the authorship footprint significantly influences the embeddings, which is extremely problematic for our application. For a concrete example, reference to topics like pollution, climate, or discrimination will have a minimal impact on the embeddings of filers initiated by the same investor due to the overshadowing effect of authorship and writing style. This renders the resulting similarity metrics based on the embeddings ineffective.

Given our focus on economic application, we do not investigate how to devise alternative models that could potentially capture these specific traits in our texts - this would entail a separate research endeavor and be more on the real of Computer Science / Applied NLP than Economics. For our purposes, word count appears to be a suitable and effective choice - a method that can be easily applied and yields effective results within our context.

Chapter 3

Unveiling Non-Core Activist Blockholder Events

Abstract

Blockholder activism is characterized by large shareholders with a declared intention to exert influence in corporate decision-making. Researchers have frequently used blockholder activist events as a means to investigate a variety of topics in Corporate Finance. While regulatory filings, the primary source of data, are publicly available, the process of extracting a suitable events dataset, one that exclusively includes instances in which an external investor targets a company with the intent to exert influence on its business, lacks thorough documentation, making it unfeasible to be reliably reproduced. In this paper, we discuss key considerations for using public data to compile an activist dataset that identifies *non-core* activist blockholder events. Furthermore, we investigate the consequences of failing to exclude these *non-core events* on two common research outcomes: *short-term abnormal returns* and *ownership stakes*. This work represents an initial step toward establishing a fully reproducible process for dataset assembly, potentially reducing barriers to future research in this field.

JEL Classification: G14, G23, G30, G32, C81

Keywords: activist investors, blockholder activism, corporate governance, data collection, data extraction, reproducibility

3.1 Introduction

Blockholder activism is a well-established strand of literature in Corporate Finance. Although the term *activist* can be used elsewhere in a fairly broad sense—mainly referring to engagement with respect to a specific topic – it is usual in the literature to refer to *activists* as the group of large shareholders (blockholders) that have acquired at least a 5% share of securities issued by a publicly traded company (target) with the declared intention to influence the target's business. The 5% threshold is somewhat arbitrary, as highlighted by Edmans, 2014, because it is not grounded in economic or financial principles. The widespread adoption of this number as a reference owes to the fact that, in the United States,[1] regulations requires investors who exceed this threshold – and intend to exert influence on a company's business – to disclose various details, including their *ownership stake*, the *date of the block acquisition* and their *investment objective*. Therefore, the 5% threshold is the *de facto* reference among researchers primarily because of data availability.

A common factor in empirical papers on blockholder activism is the use of data obtained, directly or indirectly, from regulatory filings, specifically *Schedule 13D*.[2,3] Over the years,

[1]This paper's focus is limited to examining activist events that occur within the US market, specifically involving companies incorporated in the United States with common shares traded on US stock exchanges.

[2]A *Schedule 13D* is a disclosure document mandated by the US Securities and Exchange Commission (SEC). The number *"13"* refers to the Section of the law (Section 13 of the Securities Exchange Act of 1934) outlining reporting requirements for *beneficial ownership* disclosures. Some subsections specify precise investor qualifications along with their respective filing requirements. Subsection *"D"*, for instance, pertains to those who intend to *actively* influence a company's strategic decisions, usually referred to as *activist blockholders*. These filings are titled *Schedule 13D*, also known as *SC 13D* or simply *13D*, deriving their name from the specific Section and Subsection in which they are referenced in the legal framework. Other filings outlined in Section 13 include *Schedule 13E* (related to going-private transactions) and *Schedule 13G* (primarily for passive investors).

[3]We distinguish the term *activist dataset* as employed in this study (*"13D-based datasets"*) from a common alternative interpretation (*"campaign datasets"*). In this paper, the term *activist dataset* refers to datasets primarily sourced from *13D filings*, and then subjected to a process of cleaning, consolidation, and exclusion. While researchers occasionally gather, from alternative sources, events related to stakes below the legal threshold to incorporate them into *13D-based activist datasets*, these additions have a minimal impact on the overall dataset size. In contrast, there are datasets, also commonly referred to as *activist datasets*, that compile what are known as *campaigns*, regardless of investor stake size. In this context, *campaigns* refer to instances in which investors publicly express their views, that may include criticism or advocacy for change within an investee company, through various media channels like social media, white papers, or interviews, with the primary aim of shaping public perception. While regulatory filings are also used to source information, these are supplementary. Defining what qualifies as a *campaign* can be nuanced, as it spans from casually mentioning a company in an interview to launching a full-fledged attack on the company's management decisions. While *campaign datasets* have their own merits and uses, we emphasize that they are not the *datasets* we refer to when we mention

researchers have employed various data collection methods, progressing from manual collection to using commercial databases once they became available, and eventually moving towards creating custom algorithms tailored to their specific needs. This evolution reflects the progress of technology and the researchers' growing familiarity with programming languages for data manipulation. The data collection methods available at each phase also influenced the kind of questions that could be addressed, shaping the substance and design of those studies across the years. However, despite considerable effort invested in collecting and categorizing data, this has not been translated into clear reproducible procedures.

Rather than being justified by the simplicity of the process,[4] the limited attention given to data extraction is a byproduct of its secondary role in the articles where it is introduced: it supports the core objective of investigating well-defined research questions in Corporate Finance. Hence, while a fundamental component for good research, data extraction has not been properly documented. In this paper, we address this information gap by exploring the core aspects of compiling an activist dataset. Occasionally, we take a detour to introduce field-specific concepts and terms, up to the minimum necessary to discuss our main topic. This work represents an initial effort to document such a methodology, laying the groundwork for collaboration and the establishment of an authoritative dataset.

The remainder of this introductory section is organized into three parts, where we lay the foundations for contextualizing our work. First, in Section 3.1.1, we provide a succinct overview of the manual process employed in collecting blockholder activism data up to the mid-2000s, contending that it constitutes the main cause for the limited scope, by today's standards, of the studies carried out at that time. Following this, we argue that the subsequent introduction of commercial datasets, though seemingly simplifying data collection for research, was not a panacea. Thus, we present a couple of shortcomings that lead some scholars to opt not to use them altogether.

Second, in Section 3.1.2, we take a somewhat technical stance to characterize two distinct yet essential phases in activist dataset handling: *parsing*, which involves extracting information from

activist datasets in this study.

[4]As this paper demonstrates, far from being simple, creating research-quality activist datasets requires significant work, involving tasks such as data extraction and industry knowledge.

raw sources and organizing it in a structured way; and *event identification*, which consists of consolidating filings into single events and subsequently assigning labels to distinguish between those that fall under the category of *core* and *non-core* events (we explore *core/non-core events* categorization in Part 3.1.2.8). To ensure accessibility even for non-experts, we revisit fundamental concepts related to data extraction, such as *core parsing mechanism* and *interactive refinement*. As we elaborate on the challenges of extracting information, our analysis naturally progresses into exploring the consequences of deficient documentation and the absence of an authoritative public dataset, that motivates this study.

We conclude this introduction in Section 3.1.3 by outlining our approach, contributions, and providing a high-level overview of the topics explored in detail throughout the rest of this chapter. This includes the design and implementation of a methodology to compile activist datasets, along with the presentation of some empirical findings.

3.1.1 Data extraction shifts across time

In earlier research on blockholder activism, as exemplified in the works reviewed by Gillan and Starks, 2007, covering articles published from 1990 to 2006, scholars predominantly examined events initiated by a sole large investor or related to a single investor category, often within relatively short time horizons. The scope was limited because data collection was manual and time-consuming, compelling researchers to narrow down the focus of their studies to make them compatible with the available resources. Although the introduction of Electronic Data Gathering, Analysis, and Retrieval (Edgar)[5] in 1994 provided convenient centralized access to filings submitted to the SEC from that date onward, facilitating collection to some extent, the process remained predominantly manual for decades. We defer the discussion of operational aspects related to extraction to the next section (3.1.2), as the current first part is dedicated to presenting a broader overview of the qualitative changes in the data acquisition process.

To offer a more practical perspective, most studies mentioned in Gillan and Starks, 2007

[5] Edgar is a comprehensive electronic system managed by the US Securities and Exchange Commission (US SEC). It serves as a central repository for various regulatory disclosures, including insider trading (*Form 4*), institutional ownership (*13F*), and material disclosures (*8K*), among other documents. Today, Edgar has evolved into a web service that can be conveniently accessed through an API, streamlining the process of accessing and analyzing regulatory information.

examines typically a horizon of 1-2 years, with only a few extending up to 5 years. A prevalent topic is the assessment of the effectiveness of *proxy fights*[6,7] initiated by activist investors, inherently limiting their scope to *activists* and respective *targeted companies* directly involved in such battles. Regarding the sponsors of activism, certain investigations have predominantly featured public pension funds, with a significant number of articles examining the activism led by California Public Employees' Retirement System (CalPERS).[8] Most of the remaining articles, which purportedly covers all types of activists, also fall into the category of proxy fight/shareholder proposals.

As data providers expanded their offerings to include access to datasets related to activist investors, the prospect of freeing up researchers' time from data collection sounded promising. With comprehensive datasets at their disposal, researchers could potentially widen the scope of their inquiries, incorporate a greater variety of activist categories, analyze a more extensive array of target companies, and extend their analysis over longer time horizons. Nevertheless, its practical outcomes were somewhat minor because commercial data has limitations, including a restricted range of features, and often lack transparency about the methodologies employed.[9]

An illustrative example that exposes one aspect of such limitations is *percentage ownership*, a value that is typically computed based on the *ownership stakes* of various beneficiaries associated

[6]The *proxy battles* we refer to corresponds to those within the literature cited in Gillan and Starks, 2007. Since those papers were published from 1990 to 2006, and datasets typically have a considerable lag from the last data point of the study to the publication date, the datasets in those studies cover short periods located somewhere in the interval from the 1980s to the early 2000s. Within that timeframe, *proxy battles* typically revolved around either governance enforcement, addressing agency issues, or scenarios that culminated in takeovers.

[7]Interest in the topic *proxy fights*, and more specifically, *proxy voting* (not necessarily involving a conflict), has experienced a resurgence over the last 15 years, with a notable uptick in the recent 5 years. This resurgence is a natural outcome of both the growing influence of investors, who actively shape the policies and practices of the companies they invest in, and a heightened interest in subjects that have entered the investors' agenda, that extends beyond traditional governance-related concerns to encompass social and environmental factors, aligning with the principles of responsible investment. While these studies can be broadly categorized as related to *investment activism*, we emphasize once more that when we refer to *investment activism* within the context of this paper, we are specifically addressing ***blockholders'*** *activism*. So, this recent wave of *proxy voting* studies falls outside the scope of our interest.

[8]CalPERS, the largest public pension fund in the US, is well known for its active engagement with the companies in which it invests.

[9]Data vendors are responsible for conducting internal quality assessments and validation procedures to ensure the accuracy and reliability of their data. They may provide additional information, such as metadata or documentation, to specify their data sources, collection methods, and any steps taken to clean or validate the data. However, it's not uncommon for data vendors to offer limited details about their methods and to avoid openly discussing the challenges and potential limitations associated with their data extraction processes, especially in highly competitive industries where such disclosures are rare.

with a single filing. Intuitively, *total ownership* is simply the aggregation of individual *ownership stakes*. However, it is often the case that when multiple beneficiaries are listed within a sole filing, the list includes instances of *indirect ownership*,[10] hence simple summation will lead to overstated figures. This problem was, for example, detected in Dlugosz et al., 2006 for a then much-used commercial dataset of activist events. The authors assessed the data quality of activist *ownership stakes* in that dataset by contrasting them with figures they obtained from proxy filings. Their evaluation revealed inaccuracies, along with significant biases due to double counting. In the aftermath of this episode, that service was discontinued, leading to a subsequent period of several years during which there were no commercially available alternatives to activist datasets that included ***aggregate***[11] ownership stakes among their features.

The aforementioned example pinpoints two drawbacks of *commercial data dependency*.[12] In addition to exemplifying that commercial datasets are not immune to errors, a fact often overlooked or underestimated by users, it also illustrates that these products may not always offer all the features a researcher might want or need. While this specific case concerns a service suspended due to uncovered biases, various other considerations, such as cost-effectiveness under specific quality standards and specialized research interests, factor into whether a feature is offered in commercial datasets or not. Some features are so specific and have such limited demand that they do not justify the investment required, making them commercially unviable choices.

3.1.2 Activist datasets: parsing and event identification

Our primary goal is to provide comprehensive guidance to a wide-ranging audience of those working with activist investors' datasets. In particular, we aim to make this material accessible to Corporate Finance researchers who may possess more limited expertise in data extraction. To enhance clarity, we address fundamental concepts and common challenges, which might

[10] Examples that illustrates both *indirect* and *direct* ownership include listings that feature *feeder funds* alongside *master funds*, as well as *final beneficiaries* alongside either the *companies they control* or the *funds in which they hold quotas*.

[11] The identification of ownership stakes for individual beneficiaries does not pose problems; eventual challenges emerge when consolidating these stakes under a single entity.

[12] *Commercial data dependency* is the practice of relying heavily on commercially available datasets, without critical evaluation or consideration of potential limitations and inaccuracies in the data.

appear elementary to those well-versed in working with unstructured or semi-structured data. Nonetheless, data extraction is inherently an interdisciplinary task, requiring not only technical data manipulation skills, but also field-specific knowledge. Thus, the upcoming discussion on regulatory and investment industry considerations is informative not only for newcomers to the field, but also for individuals with substantial experience in data extraction.

This second (middle) part of the introductory section is rather technical and somewhat lengthy. We characterize the acquisition of *13D events* dataset from information sourced from Edgar, comprising two distinct phases: *parsing* (3.1.2.6) and *event identification* (3.1.2.8). *Parsing* issues arise from the non-standardized nature of the *13D documents*. On the other hand, *event identification* is not directly linked to extracting objective information from a document; instead, it consists in interpreting the data for consolidation and categorization. Before we explore each of these phases, we provide an overview of the *data processing*[13] that converts the filer's input into the structured data stored in Edgar. While we steer clear of excessive details, we do offer concise explanations for specific tasks, such as *data submission*, *storage*, and *indexing* to the necessary extent so to have elements to contrast various Edgar's data retrieval tools. This will be instrumental in assisting the selection of the most suitable extraction method, contingent on the *targeted informational element*, while exposing common sources of errors, thereby helping to avoid them.

3.1.2.1 Data bundles

When a company submits a filing electronically through Edgar, it indicates the *filing company*, which is limited by the user's access credentials; selects the *filing type*; and, for most cases,[14] they are required to upload the *main filing document*, a file containing the *"filing itself"*.[15] With each submission, the system associates metadata, such as the *filing date* based on the

[13] *Data processing* here refers to the process of taking raw data submissions and transforming them into structured, organized, and usable information.

[14] While the majority of filing types, including *8K*, *10K*, *10Q*, *13D*, and *13F*, require uploading the *main filing document* onto the Edgar system, for some others, like insider filings (*Form 3*, *Form 4*, and *Form 5*), data is input directly into electronic form fields, with no accompanying document uploads.

[15] In addition to the *main filing document*, *filers* have the discretion to upload supplementary files they deem necessary to provide comprehensive information, such as contracts, purchase agreements, correspondence, or excerpts from news published by media outlets.

submission timestamp; and an uniform resource locator (URL), a web address that points to the *main filing document*. While the *filing type* determines distinct data schemas,[16] there are certain *fundamental elements*, including *URL*, *filing date*, *filer company*[17] and, naturally, the *filing type* itself, that are present in all schemas. For those *types* with schemas containing additional *elements* alongside the basic ones, *filers* are required to input that information into the system. For example, *13D filers* are tasked with entering the *targeted company*, while *8K filers* select the *items* their filing refers to, from a preset list.[18]

Upon filing, the combination of user input and metadata forms a structured data bundle, serving as an information unit. There are some alternatives for efficiently accessing large batches of historical archives containing the *elements* of these bundles through Edgar application programming interface (API),[19] but in this study, we single out only two of them. A popular option is to use *ASCII index files*, which are structured 5-field fixed-format text files, each covering all *filing types* for a given quarter.[20,21] Alternatively, individual JSON files, packaged into a zip archive titled *submissions.zip*, contain entries corresponding to all filings related to a single company. Both choices, *index files* or *submissions.zip*, provide access to the main *elements* of the structured data bundle, allowing for efficient retrieval of *data elements* like the pair of *filing dates/companies* for all entries related to a specific *filing type*.[22]

[16] A data schema refers to the structure and format in which data is organized and represented for a specific category or type of information. Different data schemas are used to standardize and categorize information to ensure consistency and ease of retrieval.

[17] In reality there are two *elements* that refer to *filer company*: *filer company name* and *filer company CIK*. We will keep references as *filer company* to help exposition as, at this point, the distinction between these two *elements* is not important. The same logic will be applied later when we refer to *target company*.

[18] In the context of *8K filings*, *items* refer to the *events* or *topics* that trigger an *8K filing* requirement. Since late 2005, users select the relevant *items* from a predefined list when submitting *8K filings through* Edgar. This protocol encodes *items* as categorical data within the *8K data bundles*, streamlining retrieval of these *elements*. Prior to this update, the sole method for fetching *8K items* was parsing of *8K main filing document*.

[19] While querying Edgar through a graphical user interface graphical user interface (GUI) or manually downloading files is an option, these methods are neither efficient for handling larger datasets nor aligned with best practices for reproducibility.

[20] Index files are structures used to store and organize information about other data, acting as a catalog reference system that makes information searchable and readily accessible.

[21] *American Standard Code for Information Interchange (ASCII) index files* are popular among the public of Edgar users for two main reasons. Firstly, they've been around since the early days of Edgar, predating more convenient alternatives. Secondly, their tabular presentation appeals to users who may not be familiar with other formats, such as javascript object notation (JSON), Extensible Markup Language (XML), or hypertext markup language (HTML) format. In the past, users had to manually download these files, but now data acquisition can be seamlessly integrated into an automated data retrieval pipeline.

[22] Note that for certain filing types, those that involves a *subject company* (e.g. *13D*, *PX14A6G*), both the elements *company name* and *company central index key (CIK)* in the *index files* will be ambiguous, as they might refer either to the *filer company* or to *subject company*. Resolving ambiguity requires fetching the *.htm* file

While the pieces that constitute the bundle are informative on their own, it is usually the case that researchers depend on additional *data elements* that are neither explicitly provided by the *filer* as system inputs nor system-generated metadata. Gathering these *extra* elements requires parsing the uploaded *main filing document*, which entails two additional steps: retrieving the document and then processing its contents. The retrieval process is straightforward; the document is requested using the URL included in the data bundle. However, once obtained, parsing its contents can be laborious, with the extent of this effort varying according to the specific *filing type*. Some filings are available in a machine-readable format, requiring minimal parsing work. In contrast, others, such as *13D filings*, may require extensive data cleaning, as well as extraction procedures that involve customization to address unique cases.

In summary, compiling an activist dataset involves supplementing information obtained from structured data sources with *elements* parsed from *main filing documents*; a frequently underestimated task that often results in inaccuracies. In the upcoming section we show that, beyond the challenges posed by non-standardized content, technological advancements have led to the emergence of various *cohorts* of raw documents, each demanding tailored adjustments for parsing (e.g. pre-cleaning, parsing rules). While this may seem peculiar given their consistent visual layout, the apparent contradiction arises because the distinction is not present in the rendered documents but rather in the *underlying raw text*, which serve as the *raw material* for most parsers.

3.1.2.2 Parsing filings: overview

In the early days of Edgar, while the identification of relevant filings and the retrieval of associated documents became significantly more efficient, the process of parsing data from *main filings* still closely resembled the methods employed in pre-digitalization era. Parsing consisted of visually identifying relevant information in scanned/printed documents and manually entering the data elements[23] into a spreadsheet, a process that is clearly labor-intensive and error-prone. Despite the potential for algorithmic parsing, the presentation of documents in

that corresponds to that entry using the URL indicated in the *index file*.

[23]Manual data entry, also called data tabulation, is the process of entering data into a system by typing it or copying and pasting from an electronic document.

a non-strict standard semi-structured[24] format across all *filing types* meant that automation efforts, particularly with the technologies available at the time, were laborious, inefficient, and susceptible to inaccuracies. Moreover, those attempts did not eliminate manual intervention completely; on the contrary, the process remained hybrid, demanding a significant amount of manual *verification* and, to some extent, still relied on manual *data entry*. In many cases, automation not only failed to improve efficiency but was, in fact, counterproductive. It resulted in additional work and introduced errors that wouldn't have arisen if *only* manual data extraction methods were employed. Given these drawbacks, manual extraction remained as the practical choice.

This scenario underwent significant transformation with the introduction of extensible business reporting language (XBRL).[25,26] Once *filings*, such as *10Q*, *8K*, *10K*, and *13F*, adopted the XBRL standard, parsing those documents could be efficiently accomplished using an algorithmic approach. In fact, one of the primary objectives of the XBRL standard, much like any extensible markup language, is to facilitate the unambiguous recognition of data elements by tagging them according to specific taxonomies.

However, *13D filings*, despite being well-suited for precise tagging, field allocation, and concept association – qualities that render them highly compatible with machine readability – have not undergone an equivalent level of standardization as other filing types. Although there were universal changes since the early days of Edgar, such as mandating HTML format for uploaded documents in the 2000s, those changes are of secondary importance in the context of rule-based parsing.[27,28] Non-standardization issues persist because HTML does not equate to a

[24]In this context, *"semi-structured"* refers to documents that emulate the visual layout created by a typewriter, where characters are positioned in predefined columns on the page. This format contrasts with modern machine-readable formats that adhere to strict standardization. Although algorithms can be employed to parse *semi-structured* content, this process is not as straightforward as parsing machine-readable documents.

[25]XBRL is a global standardized language for the electronic communication of business and financial data. Various extensions are tailored to specific reporting needs within the broader XBRL framework. For instance, the **US GAAP!** Taxonomy Extension for Credit Losses is employed to report credit losses in accordance with the Current Expected Credit Loss (CECL) accounting standard.

[26]Taxonomies vary by jurisdiction; for instance, within the European Union (EU), the European Securities and Markets Authority (ESMA) has implemented the European Single Electronic Format (ESEF) taxonomy for XBRL-based reporting of financial statements, according to the International Financial Reporting Standards (IFRS).

[27]A different rule-based paradigm have the potential to be less affected by these inconsistencies: rules that take into account the positioning of characters in web rendering, but this implementation was not considered in this study.

[28]In this paper, while our primary objective is to provide a conceptual understanding, we also present an implementation that favors a rule-based paradigm over alternatives relying on machine learning (ML). This choice

fully machine-readable standard that establishes connections between information elements and tags with an associated taxonomy.

Therefore, compiling a dataset for blockholder activist events that includes elements beyond *company* and *filing date* still involves typical challenges of processing non-standardized texts. These issues results from variations introduced by *filers*, such as layout changes, omitted fields, typographical errors, modifications in phrasing, and the use of unconventional date formats, among others. Compounding these problems are certain business practices, such as poorly implemented text-to-HTML conversion,[29] which produces anomalous texts that require additional effort for parsing.

While demanding, complete scripting the parsing of these documents addresses many of the problems caused by the absence of a clear, replicable, and upgradable methodology, as discussed later, in Part 3.1.3. Before exploring the specifics of working with these documents, we first describe, next, in Part 3.1.2.3, the mechanics of the parsing process and some of its operational aspects. In particular, we present the different potential outcomes for those cases where either elements or entire entries cannot be parsed, which are central to the *conceptual problems* discussed later, in Part 3.1.2.6.

is rooted in three key reasons. Firstly, it aligns with the skillset and resource profile of a specific subset within our potentially diverse audience—Corporate Finance researchers. Secondly, it effectively addresses common issues encountered with legacy datasets and algorithms that our audience may have access to, guiding them in assessing the integrity and comprehensiveness of those. Thirdly, while it may not achieve the same level of accuracy as a well-implemented ML solutions, it effectively serves as a guiding thread for our discussion of fundamental problems, which is helpful, to some extent, in the development and validation of scripts that use either of these two paradigms.

[29]When the SEC mandated the use of HTML format for filings, many *filers* adopted text-to-HTML services to comply with the new directive. However, in those early days, converters were rudimentary, focusing on rendering each character in web browsers while neglecting semantic meaning. To further complicate, this transition took place during the early stages of HTML, when inline styling, rather than the cascade style sheets (CSS) used today, led to documents filled with excessive and confusing tags. It took several years before both the format and conversion services became more sophisticated. Consequently, non-semantic, poorly structured HTML content is prevalent in filings submitted throughout the 2000s.

3.1.2.3 Parsing mechanism

Among the many techniques used in parsing documents,[30] regular expression (regex) is typically a fundamental component for recognizing raw text patterns.[31] The regex rules are in fact, usually, the primary means by which the process of *locating*, the identification and isolation of *targeted elements* within the document, is executed. Generally, for each entry, multiple *elements* are targeted, and for each of those *elements*, there are specific rules customized to its unique structure and formatting. As a result, parsing rules will consist of not one, but several sets, each tailored for extracting a distinct *targeted informational element*, and these sets will be used for processing individual entries.[32]

In general, raw text patterns used to locate elements may relate to text content, styling markups, or characters/strings positioning within the document. Patterns that appear to match certain pieces of information without ambiguity are identified and then, for rule-based parsers like the one studied here, translated, most likely, into regex to be integrated into the parsing script. Hence, the initial stages of a document parser's design requires understanding the distribution of information within the specific document, as a first step to devise those rules.

The starting version of this set of expressions is formulated with reference to a *base case*, an initial representation of the structural data layout. The *base case* is inferred from one or more sources, the number of sources depending on the level of document standardization. To illustrate this point, next we provide examples for two categories of Edgar's documents used as data sources for this paper: *index.htm* (standardized documents) and *main filings* (non-standardized documents).

Devising rules: standardized documents

[30] As a reminder, note that here we are referring to documents that follow some type of structure, which might or might not be strictly standardized.

[31] Some programming languages, like PHP, provide built-in parsers for XML and HTML, while others use libraries like BeautifulSoup in Python for document manipulation. These parsers interact with the Document Object Model (DOM), representing the document's structure in a hierarchy. Note that, *under the hood*Note that, under the hood, the aforementioned instances, either built-in or external libraries, fundamentally perform text parsing to achieve this functionality.

[32] As mentioned previously, *main filing documents* in *13D archives* have evolved in format, from plain ASCII text used for typesetting to earlier HTML versions with suboptimal text-to-HTML conversions, and now, they are available in well-structured HTML. Consequently, parsing, cleaning, and preprocessing methods are tailored to each specific document type, resulting in the potential for multiple sets of rules for each element, depending on the filing cohort.

When dealing with standardized documents, developing parsing rules is usually uncomplicated, as each instance follows a consistent layout with minimal deviations from the established standard, if any. The *base case* is often derived from a single document instance used as reference, and the initially devised parser rules are likely to require only modest further adjustments.

An illustrative example of such documents are the Edgar *index.htm* files. These documents are automatically generated from data entered by *filers* into electronic forms within designated fields for each individual *informational element*. Because these documents are formatted in HTML and share a consistent DOM structure, parsing is a simple task. The consistent structure, characterized by a clear hierarchy, means the parsing rules are effective across the full dataset.

Devising rules: non-standardized documents

In contrast, when dealing with non-standardized documents, such as Edgar's *SC 13D main filings*, deriving parser rules becomes significantly more complex. Neither SEC regulatory guidelines that specify the filing layout, nor a single filed instance serves as an adequate reference, when considered in isolation. Relying on a single source will lead to parsing rules that cannot accommodate the many variations introduced by *filers*, resulting in a high number of unprocessed entries, upon execution. Consequently, in practical terms, the underlying references used to establish a *base case* for *13D main filings* consist of the SEC's prescribed model combined with, not one, but multiple concrete examples extracted from filed documents instances.

Therefore, the initial parser rules, though *"initial"* will incorporate some adjustments to cater for deviations from the mandated model, particularly those that can be readily identified. Nonetheless, even though *"enhanced"*, these rules will still be too narrow to contemplate the entirety of user-induced changes,[33] as well as variations present in each distinct technology-driven document group.[34] Thus, in addition to drawing from various sources to establish the *base case*, the initial rules must undergo extensive refinement, entailing numerous modifications. Upon

[33] So it is more efficient to keep inspection effort to enhance the initial rules to a very limited effort.

[34] Alongside changes initiated by users, there are modifications arising from shifts in system functionality, as well as from business practices. As discussed, later, in Part 3.1.2.6, these alterations create, as byproduct, distinct document cohorts: initial ones adopting layouts generated via typewriter settings, which were subsequently succeeded by various stages of HTML development, encompassing non-semantic designs with inline styling, and culminating in modern HTML5 with a unified and semantic structure. Moreover, there are *"unusual cases"*, documents that result from poorly executed text-to-HTML conversions.

concluding this iterative process, the final set of parser rules will be distinct and, notably, more complex than the initial one.

Irrespective of whether the *base case* is derived from one or multiple sources (i.e., referring to standardized or non-standardized documents), the subsequent step after formulating the *initial parsing rules* is their implementation within a prototype.

Next, we outline how parsers convert raw text into structured output and discuss various approaches to handle instances where information cannot be located. While we discuss the topic in a general context, these concepts directly apply to the specific case of the mentioned parser prototype.

From input to output: parsed, unlocated, and unprocessable

We now turn to practical considerations about the outcome of executing the rule-based prototype. For each filing, the input data is comprised by the raw text-based document, which is read and then (hopefully) successfully processed,[35] so that the *targeted elements* are located and isolated. Occasionally, an element cannot be located, whether due to its nonexistence or the inability of the devised rules to capture it. Under these circumstances, considering the numerous *elements* to be parsed within a single input entry, an appropriate approach is to designate a *sentinel value*[36] as the output.

In addition to element-wise failures in locating data, sometimes the *entire* entry is unprocessable. In such cases, it is reasonable either to retain the entry and assign custom *sentinel values*[37] for *all* of its *targeted elements* or, alternatively, to discard it (meaning the entry will not have a corresponding record in the parsed output). Whenever choosing the latter option, errors should be logged, allowing the entry to be tracked for subsequent verification.[38] Both

[35]This phase includes pre-processing the document (e.g. cleaning) to optimize it for information extraction.

[36]A *sentinel value*, such as *NA* (Not Available), serves as a special marker used to represent missing or undefined data. In the context of parsing, a *sentinel* indicates the parser's inability to locate the intended informational element, which can result from the element's absence or other factors impeding the parser's recognition. In cases where a labeled dataset ground truth (GT) is available for a given sample, these scenarios can be differentiated. However, when labeled data is unavailable, distinguishing whether the *sentinel* indicates the absence of the element or the parser's incapability to parse it is not possible without referencing the original documents. Subsequent elaboration on characterizing these scenarios is provided in Part 3.1.2.4.

[37]The *sentinel values* assigned to all the *elements* of unprocessable entries should ideally be distinct from those assigned to a single element that could not be located.

[38]While tracking non-parsed entries is the recommended approach, it's not always followed in some real implementations, as we will explore in Part 3.1.2.6.

of these non-parsing scenarios, whether they concern the non-processing of a single element or an entire entry, are central to a topic we will soon explore: *conceptual evaluation*, which is discussed later in Part 3.1.2.6.

Each individual entry yields multiple parsed *elements*, that are aggregated into a singular output record. As we process numerous entries,[39] whether they are from a selected subsample or from an entire *static dataset*, the resulting output consists of a structured collection of those *"single"* records. This output is subsequently employed to assess parsing rules' performance against the ground truth (GT),[40] with the objective to identify potential opportunities for refinement, which is the task we address next.

3.1.2.4 Parser refinement

In the preceding sections, we mentioned that, in the initial phases of developing a rule-based document parser, we *attempt* to identify unambiguous patterns that are then translated into rules to be implemented in a prototype. These parsing rules are, in fact, *preliminary*, in particular for non-standardized documents[41]; not only they fail to parse numerous entries/*elements* but, additionally, modest changes can easily improve them. The refinement of these rules typically consists of parser execution, followed by critical evaluation of the algorithm's performance. Comparing the obtained output against the GT, helps reveal underlying latent issues, assisting in subsequent adjustments and corrections. Following the adaptation of the rules, the refined version undergoes the same (repetitive) process: the parser is executed, and every new output is consistently assessed against the GT. The whole procedure (execute parser, evaluation of the output performance, rules refinement) is reiterated, producing after each cycle a progressively refined version of the rules. As we will see next, the number of iterations necessary to achieve satisfactory parser performance can range anywhere, from just a few to several.

[39]For those unfamiliar, parsing a set of documents might be initially associated with a sequential process (iterating over entries). However, in reality, its execution can be implemented using any processing paradigm (e.g., parallel).

[40]In this paper, we use the term *ground truth (GT)* to generically refer to the *known correct output*, whether it is pre-labeled or inferred post-running by contrasting the output to actual results. *Observed parsed outputs* are compared against the GT to assess the performance of parsing algorithms, as discussed further ahead, in Part 3.1.2.7, on *empirical evaluation*.

[41]That is precisely why we referred to the pattern identification as an *attempt*: it is highly improbable that those initially devised patterns matches consistently the *targeted elements* of all document's instances.

Refining: standardized and non-standardized documents

Much like our earlier discussion in the context of establishing the *base case* (Part 3.1.2.3), the *degree of standardization* also influences the *number of iterations* necessary to attain a refined version of the rules that will ultimately be accepted as the final one. On one extreme of the spectrum are those documents that stick to a consistent format, such as standardized web content with instances that follow the same DOM structure. The parsing rules initially implemented for such cases are not significantly different from those that will ultimately constitute the final set, meaning their refinement is usually concluded within just a couple of iterations.

To illustrate how the procedure for *standardized documents* unfolds, we revisit the example of *index.htm* files, introduced in Part 3.1.2.3. However, we now approach it from a *concrete* perspective, recounting how the observed output led to adjustments in the actual parser implementation conducted for this study. When we tested the initially devised rules for those documents on a sample of entries, only those originated from filings created up to a certain date could be accurately parsed. Failures of this kind, causing the output to be partitioned into two distinct, time-consecutive groups – one with correctly *parsed elements* and another without (or vice versa) – are indicative of modifications in the document structure made at either the system or the regulation level.[42]

Detecting the root cause of problems of this nature and fixing it is usually straightforward. It often suffices to randomly select a single entry from the corresponding period for which parsing fails, and examine how it differs from the *base case*. In this particular example, the parsed failed because there was a change in how *targeted informational element "subject company"* was identified in the dataset.[43] Once the cause was detected and subsequently corrected, the refined version yielded an appropriately accurate outcome and no other further changes were necessary. This example illustrates the typical refinement process for standardized documents -

[42]These errors, consistently presented sequentially, stand in contrast to user-induced errors, which are more likely to occur randomly.

[43]When user-entered data is input into Edgar, *subject (target) companies* and *filer companies* are entered into clearly distinct designated fields. However, the storage of this data in *index.htm* files lacks such clarity as both *elements* are housed in sibling containers with identical structures. Differentiation, instead, is achieved through labels that the system appends to the company name strings. Initially, the label used was **(target)**, but at a certain date, it was replaced by **(subject)**. This is a good illustration of some challenges found in real-world applications, where systems, at times poorly designed or patched (e.g. to incorporate new information), may lead to suboptimal structures.

it is concise, as eventual issues are usually easy to recognize and rectify, being resolved after minimal interventions.

In contrast to the scenario just outlined, refining non-standardized documents requires multiple iterations, subjecting the *initial rules* to significant changes. This occurs not solely due to the unbounded nature of user-induced modifications but also because many problems only become apparent after addressing other concerns beforehand.

As the algorithm undergoes substantial fine-tuning, with the majority of issues addressed, we eventually reach a stage where refining no longer significantly enhances the overall parser performance. At this point, although there are still returned outcomes that are incorrect, rectifying them is, in practice, instance-specific. Is it justified to engage in such granular refinement, considering the implications of increased resource usage, heightened complexity, and potential compromise on *generalization*[44]? To answer this question, we shall distinguish between the goals of a generic parser and the specific application discussed in this paper, as discussed next.

Documenting the data transformation of a *static* dataset

In the preceding discussion, while we provided occasional examples, we primarily explored the general workings of parsing techniques. We now turn our attention to our specific application, which consists of *generating a dependable, verifiable, and reproducible activist investor dataset* to be used in research. There are two distinguishing features that will set apart our application from a generic document parser[45]: the static nature of the input and the explicit goal of using the parser as a tool for documenting the data transformation process.

Though the source dataset is dynamic, with new 13D entries consistently being added to Edgar, we work with *13D main filing documents* within a given observed time interval. Hence, in practice, our input is an *static dataset*.[46] The static nature of the input means that despite the

[44]In the context of a parser, *generalization* is the ability to effectively process diverse inputs, other than those used for developing it. Excessive detail in refinement may limit the parser's adaptability, leading to a form of *overfitting* — similar to what is observed in the ML domain. Overfitting in ML occurs when a model learns the training data too well, including its noise and random fluctuations, and subsequently fails to generalize effectively to new data.

[45]While these two share similarities, they are essentially distinct devices; some features required here are fundamentally incompatible with those sought after in generic parser applications.

[46]A *static dataset*, also known as an *immutable dataset*, refers to a collection of data that remains fixed and

potential for an infinite array of user-induced variations, there is only a finite number of observed deviations from the *base case*.[47] Of course, the parser should ideally be also instrumental for updating activists' datasets – able to parse entries that do not currently exist. However, this should not be a primary concern. Not only will the updating marginal effort be minimal, as work on the (bulky) historical portion had been addressed,[48,49] but more importantly, at some point in the near future, new entries should be machine-readable.

The second distinctive feature of our application is its central objective of documenting the data transformation. This calls for a fully automated protocol, captured in an open-source script, where raw data is processed without any additional manual intervention. While the majority of the transformation is deemed to be achieved through rules for matching predefined patterns, there are instances that prove challenging to address solely with those rules. Even worse, some cases are so specific, that devising a rule is not so different from outright output assignment. Although this approach would be inadequate in the context of a generic parser, in our application, it is not only justifiable but also entirely suitable. Assigning output on a case-by-case basis for those instances not captured by parsing rules is consistent with the objective of maintaining a clear record of the data transformation, as thoroughly documented in the script.

A corollary of complementing parsing rules with the *brute force* approach is that obtaining 100% accuracy is something very concrete. While validating all outputs against the GT and occasionally resorting to a *brute force* may resemble the manual processes employed in earlier activist studies (refer to Part 3.1.2.6), it fundamentally differs from them for at least two reasons. Firstly, most of the location still results from executing parsing rules, making *brute force* a residual effort for entries that could not be correctly parsed. Secondly, whenever using *brute force* this is explicitly documented in the *script*. As such, future contributors can easily spot it and eventually develop rules or pre-cleaning methods to replace the direct assignment.

unaltered, with no new entries or updates being added after its creation.

[47]We introduced the concept of *base case* in Part 3.1.2.3

[48]Besides, our application is expected to manifest a certain level of what we could (*loosely*) term *"generalization"*. However, as explained further in this section, this is neither an objective of our application nor a concept hardly extensible to our context.

[49]As our main goal is to parse correctly a static set, any potential updates that aim to append more recent data, should be accompanied by verification of the correctness of the additional transformed entries.

At this point, we have most elements to help address the question left unanswered in the preceding section: Is it justified to engage in such granular refinement, considering the implications of increased resource usage, heightened complexity, and potential compromise on *generalization*?

The answer is yes; it is justifiable and we do work on that direction. As clarified by the objective of our specific application, which aims to document the data transformation, *generalization* is not a target (perhaps not even a concept that can be considered applicable). Moreover, we deliberately employ *brute force* to address cases that rules couldn't capture. Note that while we can draw an analogy of this application with a generic parser — it resembles an "overfitted model" — prioritizing high accuracy on our static data — one should be cautious not to confound those. The specificity of our application is deliberate and does not carry the negative connotation it would have if it were a generic parser.

Despite achievable, going granular to the extreme to achieve 100% accuracy, is a bit ambitious. In practice, for this paper, we do use broad rules that capture the majority, but not the totality of *informational elements*. As just discussed, we then complement them with specific corrections for particular cases. As, we do not perform validation for all results — leaving some special cases not addressed, there is room for future work.[50]

As a result of stopping short of complete verified accuracy, we should lean on some sort of criteria to orient the refinement processes (what to prioritize and when to stop). That is what we discuss in the following section, where we describe an approach suitable to dealing with extensive *static datasets* when resources are limited.

Steering the refinement process

Stopping short of achieving confirmed 100% accuracy[51] has practical implications; it calls for guidelines to steer the refinement process, rather than pursuing improvements aimlessly. In the following, we will introduce a framework to assist in that objective. To provide a better understanding, our strategy is to first outline an intuitive, uninformed approach to parsing

[50] *Future work* we refer to here is likely to be the result of collective effort, and might eventually build upon the initial approximation we devise in this paper.

[51] Notice that the reference to complete accuracy, which can both be achieved and verified, is only warranted for the specific case we are reconsidering here – of a *static dataset*.

refinement and then contrast it against a proposed technical alternative.

A naive approach to algorithm refinement is to set a goal of *"maximizing the number of parsed entries"*. This objective is not only loosely defined, which could lead to misleading interpretations, but also unsuitable for our intended purposes. A high proportion of parsed entries is far from being a good proxy for indicating an efficient parser. It can, for example, result from many entries with *incorrectly extracted elements*.[52] This misconception is the same that underlies the use of the term *successfully parsed* to describe entries that yield an outcome, mistakenly implying a contrast with those that do not. To clarify, an entry should be considered *successfully parsed* only when the searched *elements* are both located, and their accuracy is confirmed.

In contrast to the simplistic approach, a well-informed technical strategy balances the goal of maximizing parsed entries with outcome precision. An effective way to address this optimization is to draw parallels to binary classifiers, which find applications in disciplines as diverse as Signal Detection Theory (SDT), Risk Management, and ML. As such, a robust and well-established framework exists for fine-tuning the parameters of these classifiers. The framework prioritizes enhancing performance metrics aligned with the specific usage of the classifier output. Next, we provide a concise summary of the original framework, opting to use the terms commonly employed in ML, given its widespread popularity, before translating it to the context of document parsing.

In brief, regarding binary classifiers, the trade-off between maximizing True Positive cases (*recall*) and optimizing the number of True Positives relative to False Positives (*precision*) is known as the *precision-recall tradeoff*. This concept is traditionally introduced through the framework of a *confusion matrix*,[53] a tabular structure comprising two dimensions: *ground-truth (GT)*, representing the correct expected results, and *model prediction*, denoting the observed model's output. Each of these dimensions can assume one of two mutually exclusive values (e.g.

[52]Shortly, we will introduce a framework for document parsing performance measurement. But in advance, note that *incorrectly extracted elements* will correspond to False Positives using that framework.

[53]The *confusion matrix* is known by various names, including *"contingent matrix"* and *"confusion table"*.

"Y" or *"N"*).[54,55] Several metrics have been devised, besides *precision* and *recall*, to assist in setting objectives for fine-tuning (i.e. *accuracy*, *F1-score*, *AUC-ROC*). The choice of prioritizing one metric over another depends on the specific application of the model. Similarly, in document parsing tasks, one must address the trade-off between parsing a larger number of entries and potentially introducing inaccuracies, as opposed to parsing a smaller number of entries with a higher degree of precision.[55]

We tailor the framework of the *confusion matrix*, traditionally used for binary classifiers, to the domain of document parsers, by first, for ease of exposition and labeling, assuming that the parser always returns an output: either an *informational element* or a *sentinel value*.[56] Next, we redefine its traditional dimensions: the *first dimension* now represents *output correctness*, with possible values of *"Y"* (for correct) or *"N"* (for incorrect)[57]; the *second dimension* distinguishes whether the output is an *informational element* or a *sentinel value*.[58] This redefined framework leads to four possible scenarios, akin to those encountered in binary classifiers. While not identical in meaning, we adopt the same nomenclature — the cartesian product of true/false with positives/negatives — meaning we can apply the same metric formulas for the document parser analogue.

Note that we can further incorporate the *analogous confusion matrix* within a two-stage *hierarchical framework*. At the initial stage, this framework distinguishes whether the returned output is an *informational element* or a *sentinel*, and subsequently categorizes it based on its

[54]Within the *confusion matrix*, various terms are used to quantify the classification results and are later used in formulas to define performance metrics. True Positives (*TP*) represent cases where the model correctly predicts a positive outcome. True Negatives (*TN*) denote instances where the model accurately predicts a negative outcome. False Positives (*FP*) occur when the model incorrectly predicts a positive outcome when the GT is negative. Lastly, False Negatives (*FN*) are cases in which the model erroneously predicts a negative outcome when the actual result is positive.

[55]Performance metrics are derived from the fundamental concepts in a *confusion matrix*: *TP*, *FP*, *FN*, and *TN*. *Precision* measures the model's ability to correctly identify positive cases, and it is calculated as $P = \frac{TP}{TP+FP}$. *Recall* quantifies the model's ability to capture all positive cases and is calculated as $R = \frac{TP}{TP+FN}$.

[56]In this characterization, when the parser can't find pertinent information, it returns a *sentinel value*, which is a robust implementation choice. However, more importantly, this choice facilitates the translation of the confusion matrix concept to the document parsing analogue. However, in practical scenarios, we encounter parsers that either completely skip processing certain entries or generate empty outputs, as we will discuss in Section 3.1.2.6.

[57]The original *confusion matrix* dimension *"ground truth"* is replaced by *"whether the observed output conforms to the ground truth"*.

[58]This dimension associates the labels *negative/positive* to whether the parser returns *sentinel values* or not.

correctness. The *second stage* operates contingent on the observed *first-stage* scenario, which can be any of the four cases (*TP*, *TN*, *FP*, *FN*). In three of these scenarios (*TP*, *TN*, *FN*), the second stage does not alter the outcome that would have been obtained with the *confusion matrix*; both the *hierarchical framework* and the *confusion matrix* yield exactly the same classification in these instances.

In the case of the fourth remaining scenario, *FP* (returned output is an *informational element* and does not conform to the *GT*), the second stage branches into two distinct subtypes. The first subtype occurs when an *informational element* is returned, but the identification is somewhat incorrect, possibly incomplete, or even wholly equivocal; we will call this subtype *False Positive - Incorrect (FP-I)*. The second subtype, on the other hand, arises when an *informational element* is returned, whereas the correct output should have been a *sentinel value*, indicating the absence of the *element*; we will refer to it as *False Positive - False Detection (FP-FD)*. While we aim to avoid both scenarios (*FP-I* and *FP-FD*), in certain instances, the errors that lead to the second subtype might be more detrimental than the ones that causes the first.[59]

Next, we concisely present each possible outcome of the *analogous confusion matrix*, along with their respective interpretations. For convenience, we include the split of False Positive (*FP*) that results from the *2-stage hierarchical framework*.

- *TP*: an *informational element* is returned, and it is correct;
- *FP*: an *informational element* is returned, but it is incorrect. It can be subdivided into:
  - *FP-I* (incorrect): GT contains an *informational element*[60];
  - *FP-FD* (false-detection): GT has no *informational element*[61];
- *TN*: a *sentinel value* is returned correctly (*informational element* absent);
- *FN*: a *sentinel value* is returned incorrectly, as there is an *informational element*.[62]

Mapping parser output: prioritize precision

[59]For instance, when *FP-I* arises due to minor deviations from the GT.

[60]However, the returned output does not match the GT.

[61]The accurate output should be a *sentinel*, but *false detection* returned spurious information instead.

[62]Note that in the *confusion matrix analogue*, *FN* does not hold a perfect parallel with the reference case of the binary classifier. In the parser context, *FN* is the scenario in which the *sentinel value* is attributed due to the impossibility of parsing an element. Non-parsing, while not what we want, holds a meaning that is somewhat different to get a *"wrong"* value. The impossibility to parse is somewhat expected to occur, not too often, but it is slightly different than an error.

Up to this point, we have invested significant effort in characterizing a *analogous confusion matrix framework*, within the context of document parsers. Now, we demonstrate how mapping the parser output into this framework aids in guiding the refinement process, which was our primary goal when we first introduced it. To start with, it is clear that, without constraints, we would ideally maximize both *"TP"* and *"TN"*. However, akin to the ML classifier application, once a certain level of refinement is attained, an improvement in one aspect often leads to a decline in the other; they exhibit an inverse relationship.[63]

It is at this point that the framework proves helpful in assisting the selection of corrections to prioritize. Given that the *intended use* of the parser output in our application is to *serve as input in regression analysis*, our priority should be on maximizing *precision*,[64] because errors originating in the *parsing process* can impact regression coefficients and their statistical validity. Focusing on precision means, alongside other aspects, avoiding incorporating potentially biased or flawed data points, even though it may lead to a reduction in the overall number of parsed output that contains *informational elements*.

In the pursuit of precision, a potential caveat is to create overly strict rules, which lead to the exclusion of valid *informational elements*. Criteria that are excessively stringent can introduce bias if the resulting *FN* scenarios (failing to parse *existing elements*) are not randomly distributed; for instance, if rules consistently fail to capture *elements* from a specific cohort. Nevertheless, the risk of attributing incorrect values poses a more significant threat, because determining whether the parsed output conform to the truth, in unverified data, is considerably more challenging than identifying biases in non-parsed entries. While the former issue might only be detected through labor-intensive individual inspection of parsed results against the original document,[65] the later problem is likely to be easily uncovered through simple *empirical evaluation*, as further explored in Part 3.1.2.7. Similarly, we should exercise caution not only when considering stringent rules but also when dealing with more complex ones that aim to

[63]Note that when considering a *static dataset*, it's theoretically possible to validate 100% of the parsing outputs. In this scenario, we'd have only *TP* and *TN*, meaning there are no incorrect classifications left. However, this would require labor-intensive, manual correction effort to address each user-induced variation for every document.

[64]*Precision* is a measure indicating the proportion of True Positive predictions among all the positive predictions: $P = \frac{TP}{TP+FP}$. Higher precision implies fewer False Positives.

[65]This is the case when considering non-labeled datasets, meaning GT is done by inspection. Evidently, this is not the case when dealing smaller labeled sub-samples, where verification can be done automatically.

capture a broader range of parsing instances. Increased complexity, while expanding the overall count of parsed entries, may raise significantly the number of *FP* (i.e. increase the number of *"apparently"* successfully parsed intries).

If overly stringent or excessively complex rules are not advisable, opting for simpler algorithms is not necessarily the solution either, as they too can result in inaccuracies and biases. However, issues in simple algorithms are easier to identify and can often be pinpointed with a quick and cursory analysis. In contrast, identifying issues in more advanced algorithms demand a comprehensive examination, which may require multiple stages and diverse methodologies, before eventually uncovering them.[66]

This section concludes our examination of fundamental *parsing* concepts. We covered the stages of parser development, from creating a prototype based on preliminary rules to refining it, prioritizing precision over sheer volume. With this theoretical foundation and equipped with parsing terminology, we now turn to the analysis of concrete, real algorithms.

3.1.2.5 Legacy algorithms: lessons for data extraction

We adopt a pragmatic approach to compile the most common problems frequently overlooked in the data extraction of *activist events*: we draw from *scripts* developed by practitioners that were accessible online. Rather than representing the kind of work a professional developer would deliver today, these *scripts* are, in essence, historical artifacts — past attempts that remain accessible on the web.[67] We then reverse-engineer these *scripts* to translate then back into their fundamental *algorithms*,[68] to concentrate on their distinctive features.[69,70]

[66]The challenges of understanding more elaborated algorithms, can be mitigated with thorough comprehensive code documentation. The *scripts* we examined in this study, starting on Part 3.1.2.6, do include some level of documentation, mainly basic comments on essential tasks, which, considering their simplicity, were not completely absent. However, there is room for improvement, particularly in the parsing aspect. For instance, providing illustrative examples of *strings* that will be captured by specific regex patterns can significantly enhance the algorithm's comprehensibility.

[67]Legacy scripts, although not representative of what experienced developers would create, have value in pointing out common mistakes that are likely to be made by those lacking this specific expertise, even today.

[68]In this context, an algorithm is regarded as a code-agnostic set of instructions that can be articulated in plain language.

[69]As we extract *algorithms* from legacy scripts, we also refer to them as *legacy algorithms*.

[70]Our primary emphasis lies in the logic and methodology of problem-solving, without going into any specific programming syntax or implementation aspects. But while not central to our study, in this note, we provide a concise characterization of the *scripts* we have assessed. Our evaluation encompassed two categories of *scripts*: one implemented in *R*, employing procedural scripting and designed for batch processing of filings, with

While this section explores the failures in those legacy algorithms, criticism is not the final objective; this exercise serves simply as an accessory to better understand the process of extracting activist events. The main benefits of this approach are twofold. First, it provides insights for those developing or evaluating algorithms for effective *13D data* extraction. Second, it raises awareness among users of pre-compiled datasets, whether commercial or not, prompting them to be vigilant about potential issues in the datasets they rely on.[71]

The legacy algorithms we assess here lead to rudimentary implementations, similar to the early-phase *scripts* and prototypes we mentioned in Part 3.1.2.3. As we identify parsing issues, we employ the same investigative procedures as those used during the initial development phase of a brand new parser. Consequently, the unfolding description serves as an illustration for those developing their own parsers, should they choose to do so using fundamental principles and best practices.

We demonstrate that parsimonious approaches prove inadequate for extracting data from *13D filings*, resulting in datasets plagued by errors and biases. Furthermore, we outline the appropriate approach as an iterative process involving testing, fine-tuning, and validation to ensure accuracy, efficiency, and *dependability*.[72]

In this section, we briefly address issues that arise in either *conceptual* or *empirical evaluation* of those algorithms. Our aim is to provide sufficient context for the extraction of meaningful data, but we keep the discussion at a high level. A more comprehensive examination is deferred to later sections of this paper, specifically in Section 3.2 and Appendix A. Before we proceed, please note that the evaluation presented in the upcoming Parts, 3.1.2.6 and 3.1.2.7, pertains only to the first phase of the *events dataset* creation, which is *parsing*. The second phase, *event*

intermediary as well as final parsed results stored in *SQLite*. The other written in *Python* with more modular structure. The later were primarily designed for processing either individual entries or a very limited number of entries, based on user interaction, rather than a batch processing approach. Furthermore, these later *scripts* featured diverse functionalities, including searching tools and parsing of other filing types. Despite their broader scope, they were useful for our exploration, as our focus was solely on their parsing modules and the rules devised therein.

[71] An additional, though minor, benefit of these examples is to illustrate that readily available online solutions, although easily accessible, often fall short in delivering effective results. The mere availability of such solutions online can potentially mislead those lacking the required knowledge in two significant ways. First and most concerning, they might choose to use these scripts or the output generated by them. Second, even if they don't use them, they are led to mistakenly conclude that efforts to devise an algorithm for extracting an events dataset are akin to *"reinventing the wheel"*, that is far from being the case, as shown in the current Section.

[72] *Dependability* refers to the parser's reliability and trustworthiness in consistently delivering outputs as expected without unexpected failures or errors.

identification, will be addressed only later, in Part 3.1.2.8, as the *scripts* mentioned here do not implement it.

3.1.2.6 Parsing problems: conceptual evaluation

The conceptual evaluation consists in a comprehensive analysis of the algorithmic design and its theoretical functionality. We single out two conceptual issues found on the legacy algorithms: *mismanagement of duplicate records* and *skipping entire entries*. *Duplicates* act as additional weight to entries. As this is a relatively simple issue, we leave its discussion to the main text. On the other hand, *skipped records* lead to a set of problems. Hence, in the following discussion, we provide some background on it and present its main repercussions, referring back to Part 3.1.2.3 (explaining the core parsing mechanism) and Part 3.1.2.4 (discussing the performance framework).

A closer look into the conceptual error of *"skipping records"*

It is expected and acceptable that each and every entry *will not* be parsed. However, when running scripts based on *legacy algorithms*, the resulting output will contain notably fewer records than the total number of input documents. The reduced size of the resulting dataset is mainly caused by procedures that skip those entries for which the parser cannot locate *informational elements*. At first glance, one might be tempted to attribute high failure levels to the simplicity of parsing rules (e.g., single regex patterns that do not accommodate variations). In what follows, we show that, instead, this simplicity is more of a symptom than a cause. By skipping entries, these algorithms become incompatible with parser refinement. While not the sole reason, this subpar procedure is likely to lead to rules that resemble preliminary ones, as those mentioned in Part 3.1.2.4 and as seen in the legacy implementations, rather than more advanced versions.

In Part 3.1.2.3, we introduced the sole three scenarios where parsing does not return an *informational element*, and we now connect them to the framework detailed in Part 3.1.2.4. In the first scenario, the *element* is indeed absent in the original document; hence, *no element returned* is the expected correct output, corresponding to *TN*. The next one relates to *FN*:

the *element* exists in the original document, but the devised rules were unable to capture it. The third one results from *processing errors*, which can occur either when processing the entire entry (i.e. the whole entry is unreadable) or when attempting to parse a specific *informational element*. Note that for the third scenario, the corresponding *elements* might or might not exist in the original document; hence those could refer either to *FN* or *TN* scenarios.

Although all of these three scenarios ultimately refer to *not locating* an *informational element*, in the first two, the entries are indeed processed, but *elements* are not found. In the first case, *not locating* is in fact the correct outcome, while in the second case, it corresponds to a *parsing error*.[73] On the other hand, in the third scenario, which corresponds to *unprocessable* entries, the setting significantly differs from the previous two, as it requires intervention to prevent program halting.

For any of these aforementioned scenarios, the parser output should ideally return a *sentinel value*,[74] just as we have conveniently assumed when formulating the *analogous confusion matrix framework*, in Part 3.1.2.4. However, in stark contrast to this ideal approach, the analyzed algorithms not only dismiss unprocessable entries as if they don't exist[75,76] but also employ the same procedure (skipping records) for entries in which any given, single, *targeted information element* could not be located, even though these entries *were* processable. In sum, all three cases are treated in a simplistic and improper way: instead of assigning a *sentinel* to the non-located *elements*, the whole records are simply skipped. This practice leads to a dysfunctional development process, as we will elaborate on in further detail.

"Skipping records" **equates to dysfunctional development**

As explained in 3.1.2.4, developing a parser for non-strictly standardized content demands an iterative refinement procedure. However, the combination of simple parsing rules with the failure-

[73]The distinction can be verified when working with pre-labeled GT for validation purposes. However, for unverified data, discerning whether the absence of a returned element pertains to one of the two scenarios is only feasible by cross-referencing with the original document.

[74]Ideally distinct *sentinels* would provide clear differentiation between the situations of *unable to find an element* and *process error*. Acceptable but suboptimal alternatives include either assigning a single *sentinel code* for all cases (resulting in no differentiation) or skipping entries and logging the errors for later examination.

[75]Handling exceptions for unprocessable entries by completely bypassing them.

[76]An alternative for unprocessed entries is to log the errors—this would lead to a separate approach to verify these issues if they were not incorporated into the final output.

skip mechanism compromises the refinement process. Most entries resulting in non-parsing (*FN* and *TN*) cannot be verified, as they have been entirely eliminated. As only those entries that return *informational elements* have corresponding records in the output, the problems that are observable and thus amenable to identification are solely those corresponding to (*FP-FD*) or (*FP-I*). This not only hampers effective refinement but also foregoes the opportunity to address simple cases that could be easily spotted and fixed.

But what are exactly the consequences of outright *skipping entire entries*? First, this practice masks the real proportion of entries that couldn't be parsed, as these items are simply ignored: the true extent of parsing failures rests concealed. Thus, it is impossible to accurately calculate performance measures, rendering the technical framework shown in Part 3.1.2.4 unusable. Second, it hinders the identification of the root causes of non-parsing, leaving aside many simple cases that could be easily addressed. Finally, the combination of *employing simple parsing rules* and *outright skipping entries* is likely to result in non-representative samples. Therefore, excluding non-parsed records in its entirety not only distorts the parser's success rate but also, more critically, omits output referring to a significant number of data entries that, if handled differently, could have been successfully parsed with modest additional effort.

The failure-skip practice represents a *conceptual* flaw that can be identified through algorithm inspection. It also has direct implications for *empirical evaluation*. When entries are skipped, *empirical evaluation* of the non-parsed results is infeasible. Conversely, in the full refinement process, where all entries are retained and non-parsed elements return *sentinel values*, two steps which were absent in the flawed case, can then be performed. One involves exploring ways to develop parsing rules that capture additional elements.[77] The other involves examining non-parsed entries to identify patterns that indicate non-representativeness. These and other aspects of *empirical evaluation* will be explored in detail next, in Part 3.1.2.7.

[77]Reducing number of non-parsed entries is not inconsistent with our earlier emphasis on prioritizing *precision* over *recall*. Contrary to not paying sole attention to *precision*, dealing with all easily solvable issues in the early stages of parsing development should not be overlooked, as minor adjustments at this point can yield significant improvements. Essentially, both *precision* and *recall* could be improved without compromising one another in the early refinement stages. It is only as the refinement process progresses and simple enhancements are exhausted that we turn to the established framework to guide us in prioritizing certain refinements, because it at that point that the *precision-recall trade-off* strikes-in.

3.1.2.7 Parsing problems: empirical evaluation

After examining *conceptual* issues, we turn our attention to the practical assessment of the parsed output. Similar to the discussion in the previous part, although the examples presented next are framed as assessments of legacy algorithms, they share commonalities with tasks that are integral to developing a custom parser. In particular, this type of evaluation holds parallels with the procedures that should either precede or, equivalently, conclude each iteration of the refinement process. Hence, we can abstract from the task of *spotting errors on the legacy algorithms* to picture the discussion that follows as *steps taken within the refinement process*. In that regard, we borrow from these examples to explore, beyond error identification, potential solutions or fixes.

For brevity, we briefly cover only two aspects of *empirical evaluation* in this introductory part: *pattern analysis* and *inspection of skipped* filings. The exposition that follows is succinct, consisting in pointing a couple of issues and providing illustrative, logical solutions for the chosen examples. A more detailed examination is left for Appendix A.

Identification of output patterns: *event date* anomaly

The initial stage of *empirical evaluation* consists of examining the output through the lens of standard exploratory data analysis (EDA) to identify outliers or anomalies, which informs subsequent refinements. Our first illustrative example, detected when running legacy *scripts*, is an anomaly in the distribution of *event dates* (dates when blockholders reach the regulatory threshold), consisting of an atypical concentration on the first day of each month. These irregular patterns are common when handling dates with partial information. In instances where a specified *day*[78] is absent, most date parsers default to assigning it the first day of the month. Similarly, when both the *day* and *month* are missing, the parsers default to setting the date to the first day of the year.

Once the problem is detected, the next step is to explore potential solutions. An obvious candidate is to configure the parser parameters to only accept *complete* dates, meaning strings containing the *month* but missing a corresponding *day* will not be parsed. After implementing this

[78]In this case the *day* is missing, but *month* and *year* are present.

adjustment, the outcomes obtained upon execution exposed an additional problem. Inspection of entries that previously contributed to the date anomaly, and therefore were not parsed post-adjustment, revealed that some indeed contained complete dates in their original version. Therefore, the date anomaly was, in part, a symptom of an issue *other than* incomplete dates, as first assumed. In fact, some of the strings fed to the date parser were truncated, inadvertently leaving out part of the original date characters.

Information loss is not uncommon when working with non-standardized documents and demands additional attention, especially during the seemingly harmless data cleaning/preprocessing. This is particularly likely to be a problem for non-standardized datasets, especially when compounded by factors such as the ones discussed in Part 3.1.2.6 (e.g., sub-par text-to-HTML conversion).

Next, we revisit the topic of non-parsed data introduced during our exploration of *conceptual evaluation* (see part 3.1.2.6), but now we examine it from an *empirical* perspective, seeking patterns that indicate what might be preventing the location of *informational elements*.

Identification of output patterns: non-parsed entries

As discussed in Part 3.1.2.4, *refinement*, consisting in the execution of the prototype and evaluation of its performance against GT data, is a fundamental component in the development of any document parser, especially for non-strictly standardized documents. Though we might instinctively relate refinements to corrections of returned outputs that are obviously wrong,[79] looking for the causes for those cases leading to non-located elements and fixing them are as important, if not more so, than the former.

However, as explained in Part 3.1.2.6, though non-located elements should ideally return *sentinels*, in practice this is not always observed. Conceptual flaws, such as outright skipping of non-parsed records, leads to faulty *empirical evaluation*, implying that refinement cannot be employed for remediating *false negatives (FN)*. This is evident in the naive algorithms discussed here - had them preserved instances that did not return parsed outcomes, these would have been analysed and some of those, corrected, resulting in a significant increase in the number

[79] Either for cases spotted by chance in a random inspection or those found by examining data patterns, as in the example just discussed about for *date anomaly*.

of elements successfully located. Subsequently, we present a concrete example illustrating how these algorithms could be substantially enhanced with minor adjustments.

A concrete example

Empirical evaluation is the tool that enables refinements: it reveals problems that could not be previously envisioned. In most cases, these are easy to address - they neither demand an approach that is too complex nor difficult to implement. Once implemented, the new, adjusted version of the parser should then be tested to ensure the changes were effective and that they do not compromise previously correct results.

To illustrate this point, we examine the parsing of the *informational element* CUSIP. In the template provided by SEC, the CUSIP,[80] is located just above a line containing the string **(CUSIP Number)**.[81,82] Hence, the intuitive strategy to get this *element*, based on the SEC mandated model, is to find that *indicative string* and capture the line just above it. However, as aforementioned elsewhere in this introduction, these SEC templates are not strictly enforced. The non-strict standardized nature of these documents leads to all sorts of variations, bringing about the need for evaluation of the parser's output for adjustments aimed at capturing entries deviating from the model.

Returning to the specific example, while *filers* often follows the established template, instances where CUSIP information appears below the designated string (instead of above it) or even on the same line (either preceding or following it) are not uncommon. Furthermore, the label **CUSIP Number** is frequently changed (e.g. **CUSIP No.** or simply **CUSIP**).

Hence, a more pragmatic approach, one with higher odds of capturing more entries, consists in using less limiting search rules. First, the search should aim to detect the string **CUSIP**, disregarding the more complete **CUSIP Number** and other variations. Second, the search should span multiple lines around the one containing the indicative string. Third, the search needs to be supplemented with rules for discerning CUSIPs alphanumeric structure, characterized by a

[80] Committee on Uniform Security Identification Procedures (CUSIP) is a unique standardized identifier assigned to financial instruments, including stocks.

[81] Although the template typically encloses the string within parentheses, this varies in actual documents.

[82] The illustration presented here is evidently very simplified. It contains just the sufficient elements to highlight shortcomings and to motivate abetter approach to parsing - one that is both computationally efficient but that also captures a high number of special cases.

fixed number of characters occasionally presented with separators.[83]

The *legacy algorithms* embody these traits. Their design indeed consider variations in CUSIP representation and they implicitly acknowledge that the positioning varies. The algorithms look for matches of alphanumeric patterns within the two lines above and two lines below the string **CUSIP**.

Now, though these rules seem to be good in theory, at least at first glance, the picture is different when running those scripts. Once this is done, multiple cases fail to be captured.[84] Among instances that do not yield correct results are those where the alphanumeric code and the string **CUSIP** are placed more than two lines away, and those where the string **CUSIP** is simply absent.[85],[86]

Given that neither the original indicative string **(CUSIP Number)** nor the sub-string **CUSIP** might be effective strategies for finding this *informational element*, would it be better to just search for its alphanumeric pattern? This is an easy-to-refute choice - completely ignoring the indicative text is not only computationally inefficient,[87] but it will also fail in multiple cases. *Filers* incorporate various changes - for example, they exclude verification digits, combine multiple characters to serve as single separator in between the elements of the code and position those separators on the most inappropriate way. Hence, in practice, the multitude of patterns needed to capture a reasonable number of CUSIPs are likely to return spurious results, things other than the CUSIP codes.

The solution we use in our application, is to employ a targeted search, that concentrates on the most probable location of the *informational element*, supplemented with fallback mechanisms. If the initial match is unsuccessful, the first fallback mechanism is activated. If there is still

[83] The representation will eventually contain spaces, dashes, and slashes as separators or a combination of those.

[84] As a reminder, in these examples, the *skipping entries* procedure - when the rules do not match the information, the entry is skipped altogether - interfere in this analysis. Hence, for the empirical evaluation conducted here, we were diligent to go after those skipped non-parsed entries.

[85] Likewise, there are examples of filings where analogous omissions are made for other *informational elements* such as address, filing company, target company, and others

[86] Though a nuisance for the development of a rule-based parser, the outright omission of the indicative string is not a problem for those who occasionally consult this document; information can be understood from the context - based on the alphanumeric structure and its approximate relative position to the rest of the text.

[87] Neglecting the indicative text leads to searching for a variety of possible alphanumeric patterns across large portions of the document. Instead, it is more efficient to search for a simple, unique string, and once it is found, confine search to a small text segment.

no returned output, a second fallback is triggered, and so on. Depending on the *informational element* there will be multiple fallbacks organized in a sequence: from those that capture more cases leading towards those that are more resource-intensive and capture marginal cases.

Back to the legacy algorithms, we can borrow their examples to illustrate how the refinement approach would be developed. Let's assume, for simplicity's sake, that those initial rules presented are a good first shot. We can keep them but need to complement them with fallbacks. For example, if the somewhat arbitrary search within a two-line distance[88] away from the indicative text is unsuccessful, the first fallback could be searching within an expanded segment. If there are still no returned results, the search might be conducted between two other indicative strings, delimiting a larger interval where the information should be contained. If this approach is not successful, contingent on the amount of unparsed results, additional fallbacks should be implemented. This cascade means that with each iteration, the devised rules end up targeting a smaller number of entries. As discussed in Part 3.1.2.4, our refining efforts will exhibit decreasing marginal effects, and at some point, we might resort to *brute force*.

Notice, that the rules in the cascade might incorporate strategies that in isolation would be completely inappropriate. This is the case, for example, of seeking the alphanumeric patterns anywhere in the raw text, as just discussed. Used as *first* parsing rule, it will be inadequate, but as a *late* fallback mechanism, it can be very effective.[89] We explore the idea of sequential searches in more detail in the next, in Part 3.1.2.7.

Solutions: focused search with fallbacks

As illustrated in the previous *empirical evaluation* examples, problems might, at first glance, seem simple when considered in isolation. But the reality is different: once problems are tackled, other hidden issues become apparent. Moreover, the resolution of the latter brings forth another wave of emerging issues, and so on. This was the case when we configured the date parser to accept only complete date input to prevent the *event date anomaly*. Then, we observed that

[88] This does not preclude changing the number of lines in the initial rules to another choice, though too many lines will lose the benefits from the confined search + fallbacks.

[89] Even when used as a last resort in a search for CUSIP, this strategy should be used cautiously to prevent undue patterns from making it into the dataset. A possibility is to use additional rules to verify the extracted results, such as checking the (found) CUSIP against a preset CUSIP database.

among those instances impacted by the change,[90] there were some that corresponded to entries where date was complete in the raw text, but the string was truncated before reaching the parser, compromising part of the date information. This example shows that once a problem – the parser default to assigning the first day of the month (or year) to incomplete dates – was addressed, another one was revealed, prompting a completely different investigation (the root cause of the truncation).

Moving on, as we discussed *CUSIP identification* we saw that rather than having a single, complex rule, there should ideally be a cascade of simple rules.[91] Each rule is activated only if the preceding rule fails to produce any output.[92] Hence, the cascade continues up until an output is successfully obtained. Those rules that require more computational overhead are reserved for later stages. Not only they will take longer to be processed, but they are inadequate to parse simple cases that would be easily processed with simple procedures. We have covered such an example: if the first rule used to extract CUSIP relies on alphanumeric patterns *alone*, disregarding any relative positioning,[93] it is likely to yield inaccurate output. The appropriate approach is to begin with a basic rule that effectively identifies the *element* in the vast majority of entries. The unbounded search for alphanumeric patterns is deferred in the cascade and is activated only for the smaller subset of entries, those that did not produce an output.

An additional feature that is very effective in complementing fallbacks is *focused search*. *Focused search* means that we confine the search effort within a specific segment, rather than scanning the entire document. The simplest implementation of focused search, in the context of a 13D parser, is to split the document into two parts, e.g., *cover page*[94] and remaining content. For example, as CUSIP is expected to be on the cover page, the first set of rules to search for the string **CUSIP** is confined to that substantially smaller text segment. As the refinement process progresses, and if needed, subsequent fallbacks may expand the search to broader sections and,

[90] The entries affected were those that contributed to the anomaly (returning an initial day of the month or initial day of the year inappropriately) that, after the change, did not return any date.

[91] For that matter, that conclusion applies for every *targeted informational element*.

[92] This scenario also includes those that returns a *sentinel values*.

[93] Here we refer to searching for alphanumeric patterns on the full text. As Cusip representation is non-uniform across entries, requiring the use of diverse patterns, these often inadvertently match with other elements in the corpus.

[94] *Cover page* is a term used informally to describe the initial section of a SEC filing. In our paper we use the term to refer to the portion of the raw text that includes *Item 1* (about the security and issuer) and *Item 2* (about identity and background) of the 13D *main text filing*.

eventually, scan *nearly*[95] the entire document. As this broader search would occur only in rare cases – those that were not captured in early phases of the cascade–, it should not compromise the parser's efficiency.

As in any project involving software development, there is a trade-off between promptly implementing the *13D parser* and thoroughly addressing all unique special cases. The rudimentary algorithms mentioned earlier serve as extreme examples of hasty implementations that compromise accuracy and limit the dataset size. On the other hand, it is not feasible within the scope of this study to account for every conceivable special case.[96,97] The optimal approach lies in middle ground: while we refrain from validating the entire output, we assess for biases or other systematic issues. In case they are identified, we only proceed after remediating them.

Though we have made extensive contributions, documenting considerations[98] that are not available elsewhere, we acknowledge that their reach is rather limited. In this context, our effort can be regarded primarily as an initial phase for a collaborative project, which has the potential to evolve into a publicly maintained, open-source curated repository of activist datasets.

To summarize, our parser application consists in *focused search*, supplemented by a sequence of *fallback mechanisms*. For some *informational elements*, such as CUSIP, there are later fallbacks that, when activated, search for patterns within a broader text body. Moreover, as we progress in the refinement process, we resort to outright assignment. Finally, as we do not validate output integrally, there is work yet to be done. We hope that our implementation and results will benefit from collective and collaborative efforts.

3.1.2.8 Event identification

In the preceding discussion, we touched upon common mistakes and shortcomings in *13D filing parsing*. However, *parsing* alone is not enough; it is merely the initial, albeit fundamental,

[95]Though more rarely, the search is done on the complete document.

[96]Here we refer do addressing special cases either by using *brute force* or by devising rules.

[97]To address each unique special case, we need to perform manual validation of every parsed instance against the original data source. This enlarged effort could be the aim of collective effort, given activists dataset can benefit multiple researchers. Numerous examples of such collective efforts exist, particularly on the Computer Science domain, and the technology needed for these initiatives is well-established, such as version-controlled shared repositories.

[98]Here we refer to the combination of this introductory section, the main text, or in the Appendix.

step in obtaining an activist event dataset. For it to materialize into a useful resource for research, *parsing* needs to be complemented by *event identification*. *Event identification* consists in consolidating filings into *single events* and categorizing them into either *core* or *non-core events*. Contrary to *parsing*, which received extensive coverage in the introduction, we defer most of the exploration for *event identification* to the main text. In this introductory section, we provide only a brief overview of its components: *event consolidation* and *categorization*.

Beginning with *event consolidation*, though each entry usually corresponds to one event, there are instances where we need to combine multiple entries. While some cases involve mere aggregation, this is not always adequate, and the decision to maintain the entry as a separate entity, aggregate it with another entry (or many others), or discard it is not unequivocal. We revisit this topic in more depth in the main text and provide many details in Appendix A.

Moving on to *event categorization*, *core events* are those that align with the *"idealized activist blockholder setting"* assumed in academic research. They consist of events initiated by investors who select target companies wherein they perceive opportunities to enhance value through their intervention in relatively standard business conditions. Everything else that doesn't fit this definition is labeled as *non-core event*, including arbitrage. While the primary notion of a *13D filing* is that it represents an activist event, there are numerous events that would be categorized as *non-core*, yet they trigger a *13D* submission.

To illustrate the concept of a *non-core event*, consider a debtholder initially with *no* active interest in influencing company matters. In a hypothetical case of a debt-to-equity conversion during bankruptcy relief, if they end up holding more than 5% of the company's shares, they will have switched from a passive to an active role, even if such active involvement was not their primary intent. This event prompts the filing of a *13D*, despite active beneficial ownership in this episode being circumstantial.

In a second example, we explore a somewhat similar scenario, but now an investor intentionally acquires distressed debt at heavily discounted prices. While both examples refer to distressed company debt restructuring, the circumstances that lead to active ownership are completely distinct. Only in the second case there is an overt intention of influence upon first investing in the company. However, *vulture investing* does not contribute to answering the questions usually addressed in activist investor studies. The return profile of such investments is more dependent on the circumstances that allow the company to overcome its adverse situation. Moreover, within narrow windows around either the *event date* or *filing date*, the price movements will reflect little of the debtholder influence on company matters. Hence, this is an interesting, but a separate, area of investigation.

Activist investor studies, as Brav et al., 2008,[99] and Lilienfeld-Toal and Schnitzler, 2020, explicitly refer to excluding events related to *bankruptcies*. Besides distressed companies, they also remove cases they could identify for *M&A arbitrage* and *insider trading*, that we defer for the main text.

In summary, after parsing, filings are consolidated into events and categorized as either *core* or *non-core* ones. Though the consolidation and categorization are fundamental for the proper handling of *non-core events*,[100] they are often overlooked. Neglecting *event identification* has potential undesirable consequences, that can vary from increased noise, signal attenuation, or a substantially higher probability of obtaining inconclusive findings that would have been otherwise clear and meaningful if *non-core events* were properly addressed.

In this section, we presented a concise overview of event identification, offering only the essential elements to support the discussion of literature gaps and the motivation for this study, as detailed in the next section. Given the centrality of this topic in our study, we defer more in-depth considerations to the main text and Appendix A.

3.1.2.9 The missing links in data collection and documentation

As we have just characterized, the extraction of a suitable activist investor database consists in two distinct and essential phases that cannot be overlooked: *parsing data* and *events identification*.[101] Notably, the aforementioned rudimentary algorithms, as well as earlier commercial databases, solely address the first phase (parsing). Moreover, while currently reputable data providers consolidate filings into events, this only partially addresses the *event identification*

[99] This also applies to all subsequent papers using the original dataset and its updated versions.

[100] *Handling non-core events* means either excluding them from the dataset, or neutralizing them using dummies in regression exercises.

[101] *Event identification* consists of conversion of filings into events and identification of *non-core* ones, that can be later excluded or signed by dummies on regression exercises, as described in 3.1.2.8.

process. The challenge remains in identifying *non-core events* that, if retained, may introduce noise into empirical results. Additionally, the presentation of data in commercial datasets can further complicate this identification process, as we will extensively discuss later in this paper.

The significant effort required for *event identification*, even when one has access to parsed content from *13D filings*, along with the other potential limitations of commercially available datasets (such as lack of transparency, uncertain quality, and limited features), serves as a motivating factor for researchers to invest in their own customized solutions. In fact, many of the main results published in this field post-2005 rely on datasets directly extracted from primary sources, which are extensively reused across different studies.

Among such papers, Brav et al., 2008 inaugurated a new strand of literature that focuses on hedge fund activism, linking it to investee companies' post-intervention improvement in financial and operational performance that persist over the long run. Subsequent papers have explored other direct and indirect outcomes of activism, such as improved capital allocation, enhanced innovation, and increased productivity (Brav, Jiang and Kim (2010, 2015), Brav, Jiang, Ma and Tian (2018), Barry et al., 2020, Gantchev et al., 2019). Authors have also examined the characteristics of targeted firms (Brav, Jiang, and Li, 2022), the costs of activism (Gantchev, 2013), and the symbiotic role of blockholder activists with other investors. While the presence of an activist attracts other investors hoping to benefit from their oversight, since activists do not hold majority stakes, they depend on persuading those investors as well to support their proposed changes (Brav, Dasgupta, and Mathews, 2022, Brav, Jiang, and Li, 2022). For a comprehensive literature review of blockholder activism in the context of hedge funds, we refer the reader to Brav et al., 2022.

Datasets employed in these papers serve as foundational resources for numerous subsequent research studies. Authors of the follow-up studies often refer readers back to the original research papers that initially introduced these datasets for details about the dataset acquisition. However, the information provided in these references is usually inadequate for reproducing the dataset, particularly concerning *event identification*. While there are mentions to event exclusions, such as those related to *mergers*, *bankruptcy reorganizations*, or *insider trading*, specific procedures for identifying these events remain largely undisclosed. Key details, such as which data sources were used (including whether they were integrated into any commercial databases or merged with freely available public information), and specifics about the chosen time span for characterizing exclusion cases (e.g., weeks, a month, six months) are often omitted. Furthermore, references to the manual handling of such cases tend to lack clarity and sufficient detail necessary for reproducibility.

In conclusion, researchers have invested significant time and resources in extracting their own blockholder activist datasets, a process that, as we have seen, demands careful handling to create an appropriate dataset for research applications. While prior efforts have been instrumental in advancing the current body of knowledge in this area, these datasets often suffer from a lack of comprehensive documentation, posing a challenge in formulating replicable procedures. The absence of thorough documentation, particularly concerning the isolation of *core events*, results in two primary consequences. First, findings can only be replicated from pre-refined, non-verifiable datasets, rendering full reproducibility unattainable. Second, this approach is inefficient overall, as researchers frequently rely on outdated datasets, leading to substantial temporal gaps of up to five years or more between their research and the latest available data point.

3.1.3 Enhancing data extraction

Researchers can only build upon earlier datasets either by borrowing them, relying on the collection process in their best faith, or by using comprehensive documentation on data acquisition to reproduce them. Evidently, in the past, where manual collection was prevalent (see 3.1.1), choices were limited essentially to borrowing them. However, in the current state of technology, where all processes to obtain information can be encoded, from data ingestion to parsing and identification, as extensively discussed in this introduction, that needs no longer to be the case.

However, while many reputable journals mandate authors to provide scripts for replicating primary publication results, this requirement does not extend to data collection. Consequently, the current practice in contemporary research papers is to ommit such documentation. As a result, the specific elements and procedures necessary to replicate datasets from primary sources remain unknown. This is quite intriguing, as the advantages of documenting data collection procedures are evident. It eliminates the dilemma of either initiating the process from the beginning, resulting in duplicated efforts; updating existing work with potentially different criteria; or relying on outdated data. Additionally, it exposes the collection methodology to overall scrutiny, which will either attest its soundness or raise concerns that otherwise would not be addressed. Moreover, as other researchers use the public methodology for their own specific needs, they can contribute by reviewing, refining, and improving collection methods, as well as updating the datasets with more recent entries. In sum, documentation not only prevents unnecessary duplication of work but also accelerates advancements in the field, leading to genuinely fully reproducible work.

This paper represents the initial step toward addressing the lack of data collection documentation in the context of activist investors. With this stated objective, we have explored, in this introduction, how *13D filings information* is first input and then, stored into Edgar. We used this background to distinguish those *informational elements* that can be sourced from structured sets from the ones that can only be obtained by searching semi-structured textual content.

In the following sections, we present the basic steps for extracting and cleaning data, as well as for identifying *non-core events* in a procedural manner. Our procedure can be fully encoded, leading to many advantages over the *manually/non-documented* data extraction that is prevalent in the literature. As our approach uses free public data[102] and is fully reproducible with modest effort, it lowers the barrier to entry for research in this topic and fosters collaboration - fellow researchers, or any interested person for that matter, can further enhance it. In addition, as updating[103] the dataset with more recent data points can be done without much difficulty, the

[102]The only data we use that is not publicly available *for free* is market and fundamental data from CRSP/Compustat. However, those resources are included in the WRDS basic package, to which typical research institutions already have a subscription. Additional packages on WRDS provide information on activist events, but they are additional to the basic package. We do not use this data source in this paper.

[103]U.S. SEC, 2023 mandates the transition to a machine-readable standard, but this is set to come into force only from 2025. Hence the parsing step as discussed in this paper will be much more simpler and efficient. Those changes will not, however, eliminate the need for event identification. Nevertheless, with machine-readable filings, some of the challenges currently found for identification will be mitigated.

use of recycled datasets that conclude many years in arrears[104] would become less common. After all, information on Electronic Data Gathering, Analysis, and Retrieval (SEC Edgar) can be accessed as close as real-time.[105]

Our paper contributes to the literature on blockholders activism in several ways. Firstly, we provide the building blocks for a methodology to extract and categorize information from SEC filings[106] procedurally, that serves as practical guidance for designing and implementing this process. While we acknowledge that our methodology cannot entirely replace the commendable and painstaking efforts of widely recognized scholars in this area – who promoted the field's advancement by compiling such datasets for their studies – our approach can be particularly valuable when used alongside theirs. We devise supplementary procedures that can help identify cases that may escape manual collection efforts, such as instances where *item 4 - Investment objective*, is either absent or non-informative (e.g. field contains *"boilerplate"* text[107]).

Secondly, we present empirical results that emerged as a direct outcome of the tests we conducted to gauge the effectiveness of the *non-core events* identification approach. In our study we use a combination of automatic procedures either to flag or to exclude those events related to mergers, notice of delisting, insider trading, among others. Under specific circumstances, we employ these indicators for *non-core events* as dummies on regressions where the dependent variable is either *abnormal returns* or *ownership stakes*. By analysing of the resulting dummies' coefficients we gather evidence that retaining *non-core events* in the sample might distort results of common research outcomes.

Finally, much of the literature on blockholder's activism restricts data to a single investor type, such as pension funds or hedge funds, with an important emphasis on the later since the seminal work of Brav et al., 2008. Our study, however, serves as a complement to those less

[104]Across academic fields, especially in the Social Sciences, using datasets with substantial time gaps (from the last data point to the research date) is a common practice. While once justifiable, this should gradually stop being the case, particularly when relying on public data sources, once researchers incorporate reproducibility principles for all research steps, including data collection. Unfortunately, this is yet to come, as researchers rarely provide documentation and coding scripts that cover data collection. Implementing such practices would promote result replication and simplify extension of existing datasets, reducing redundant efforts.

[105]The SEC Edgar website provides two ways to stay promptly informed about filing updates: *RSS* and *Latest Filing Search*.

[106]While our primary objective is to document the dataset extraction process for activist investors, a considerable portion of the methodology and practical guidance presented here can be extended to other types of filing types.

[107]*Boilerplate* refers to standardized, often generic text used in various similar documents, lacking specific information unique to a particular instance.

frequently investigations that covers a broader spectrum of shareholder activists, as Lilienfeld-Toal and Schnitzler, 2020. In the later, the authors compile a dataset spanning the period 2001-2016, and observed differences on stakes and types of targeted firm, conditional on investor type. However they did not find any statistically significant variation on abnormal returns. Notice though that while our study does include all investor types, we do not discriminate among them, hence we do not use taxonomies as dependent variable.

Nevertheless, after implementing our methodology to exclude *non-core events*, an examination of the top 100 blockholder activists, as detailed in the tables added in Appendix B, suggests, in line with most to the literature, the significant prevalence of hedge funds. This observation contradicts the notion proposed by Bebchuk et al., 2013 that the *vast majority* of activists are non-hedge funds benefiting from hedge fund activism on targeted companies. While this holds true in terms of raw data, the vast majority of *non-hedge-fund events* seems to be composed of *non-core events*, which should not be included in the datasets. However, a more robust conclusion can be only reached if we undergo a classification of investors as conducted in Lilienfeld-Toal and Schnitzler, 2020.

This paper is organized as follows: Section 3.2 presents the main points on the data gathering, which is integrated with event identification (consolidation and detection of *non-core events*). For the later, we rely on regulatory filings unrelated to blockholder activism to identify those events that should be excluded. We postpone a more complete description of the procedure to Appendix A. Section 3.3 presents main empirical results, where we regress usual outcomes against the dummies created for *non-core events*. These results serve not only to ratify our methodology approach to identify *non-core events*, but holds also empirical importance by themselves. They bring evidence on the effects of failing to exclude *non-core events* on two traditionally studied outcomes: *short-term abnormal returns* and *ownership stakes*. Section 3.4 summarizes our findings and concludes with a discussion on how access to up-to-date reliable *13D datasets* can be beneficial for research on Blockholder Activism.

3.2 Data and Methodology

This section covers the steps for extracting activist data and identifying *core/non-core events* from regulatory filings using an algorithmic approach. One of the main advantages of this method is its potential to reduce the workload for updating the dataset.[108] Although running the script on new (unseen) entries calls for results' validation that will likely lead to script adjustments (refer to Part 3.1.2.4), sticking to an algorithm contributes to maintaining consistency between the historical dataset and the newly incorporated entries. Furthermore, this approach retains *informational elements* that aid in *event identification* (see 3.1.2.8), which are often discarded from commercial datasets (see 3.1.1). In addition, we keep the URLs corresponding to each extracted event (see *data bundle* in Section 3.1.2.1) to efficiently check, in just one click, the results obtained against their original entry source. In summary, the methodology described here aims to efficiently compile an activist dataset that can be easily updated and conveniently verified for consistency. It also allows for reproducible results and it can serve as a platform for collaboration and refinements within the research community.

In this study we use datasets sourced from two data providers: SEC Edgar and Wharton Research Data Services (WRDS).[109] The core data elements are acquired through sequential API calls,[110][111] followed by dataset joins, that lead either to more cleaning or aids in events consolidation. We perform four *series* of requests to Edgar to retrieve files containing *regulatory filing information*: *crawler files*, *index.htm files*, *13D main filing documents*[112] , and

[108] As already noted, the part of this algorithm that relates to *parsing* will be simplified for processing events filed from 2025 on, when the changes mandated by U.S. SEC, 2023 to report machine-readable 13D filings will come into force. The algorithm for *event identification*, though, should not be considerably altered.

[109] The databases accessed though WRDS are CRSP, Compustat and Fama-french factors.

[110] In line with the purpose of this work and to ensure reproducibility, the data ingestion is entirely algorithmic, including the connection to Edgar web services and WRDS server.

[111] For each batch of files, we open a single connection to the web service and then submit multiple requests through that link, avoiding connection overhead.

[112] The archived *13D main filing documents* are presented in either ASCII or HTML, for which the information is not machine-readable. The new regulation mandating beneficial owners to file them as machine-readable documents (2023), will only come into force in 2025.

the *Edgar bulk download zip file*.[113,114] As for *market* and *company-level data*, these are obtained through the WRDS web service using SQL queries that combine one of more of the following datasets: CRSP/Compustat merged (*ccm*), CRSP security daily (*dsi*), Fama-French factors (*ff_all.factors_daily* and *ff_all.fivefactors_daily*), and Compustat company fundamentals (*compa*).[115]

Each new query aims to either detect/isolate elements of interest (e.g., CUSIP extracted from *main filing documents*) or gather information that helps identify entries related to *non-core events*[116] (e.g., indicators derived from filings archived in the *bulk download zip file*). In between each series of data retrievals, there are numerous integrity checks, including validation against raw data (see 3.1.2.4) and empirical evaluation (see 3.1.2.7). This means that before joining new data to the compiled dataset, a substantial amount of workload takes place: multiple iterations for refinements, similar to those described in 3.1.2.7; outright assignments documented in the script; or simple processing to extract data (e.g., filtering to retain only 13D filings from *crawler files*, or parsing *targeted elements* from *13D main filing documents*).

In the following section, however, we abstract from these intermediary processing steps. Instead, we contunue on a high-level overview, but this time we prioritize two aspects: the role of each source dataset in the final compiled set and the extent to which each step reduces the number of observations in the sample. Hence, we focus solely on the alternation of requests to Edgar API and WRDS API, leaving aside parsing and refinements, which are detailed in Appendix A.

[113] For simplicity, we fetch complete Edgar *index files* and *bulk download zip files* in full, and process them locally (e.g., filter for *filing type*).

[114] The first three files require multiple requests, done in batches. We retrieve all *crawler files* over the studied period, making nearly 250 requests as each file covers one quarter. From those we keep only the records corresponding to *13Ds*. We edit the URLs therein so these will point to *index.htm files*. We make about 200,000 requests to fetch all of them. After cleaning, we acquire all *13D main filing documents* remaining in the dataset (one per entry), totaling around 60,000 requests. The *Edgar bulk download zip* is a single file, hence it is fetched with a single request.

[115] Due to the sheer size of the datasets in CRSP and Compustat, as opposed to the Edgar files, data acquisition via WRDS is restricted to the daily observations around the study window and is obtained using SQL queries. This window assumes at the most 140 trading days, comprised of reference period and evaluation period, corresponding to -120 to +20 days around *event date*

[116] The concept of *core/non-core events* was introduced in Part 3.1.2.8.

3.2.1 Activist data

To derive the activist events dataset, we started by fetching the SEC Edgar *crawler files*. Our data collection spans from the series' inception in 1994 until August 2023.[117] Within the information contained in *crawler files*, we obtained URLs that pointed to *index.htm* files.[118] We filtered the entries to keep only those that refered to *SC 13D initial*[119] *filings*. The URLs therein were then employed in approximately 120,000 requests, each retrieving an *index.htm* file. We read each file, to extract the *filer company*, *target company*, and their respective *postal addresses.*,[120] as detailed in Section A.3.3 of the Appendix A. By combining the pair *filer/target* with *event date*, we spotted and subsequently removed duplicated records.[121][122]

The information in the first two series of retrieved documents (*crawler files* and *index.htm files*) is structured. As mentioned in Part 3.1.2.3, parsing information in these cases is straightforward. However, the same cannot be said for the next set of documents to be retrieved: *13D main filing documents*. Since these are semi-structured (non-standardized), parsing them requires a significant amount of intermediate processing, particularly in terms of refining. To avoid unnecessary work in empirical evaluation and refinement, as well as to reduce the number of requests (each *13D main filing document* requires one request), we drop those records that are either duplicated, incomplete or that refer to non-US companies. The resulting set, for the referred interval, consisted of 59,904 unique records.

At this initial phase, the dataset is very limited. It contains only those features available

[117]Later in the paper, we narrow our focus to the period from 2006 to 2022. This restriction is based on the availability of auxiliary data to enhance the sample's cleanliness.

[118]The URLs presented in the *crawler files* originally point to the *main filing documents*. However, by removing the *name* of the file from the web address, and replacing it with the string *"index.htm"*, the assembled address points to *index.htm* files instead.

[119]To be clear: (1) we do not incorporate amendments *(SC 13D/A)*; (2) *crawler files* are not the same as *filing documents*. The former contains only structured data (5 elements) from the *data bundle*.

[120]Each *index.htm* file is generated by Edgar, so it has a consistently uniform structure, where *filer* and *target* are clearly labeled (see 3.1.2.4).

[121]For SEC filings that inherently feature two companies, as in the case of *13Ds* where a *filer company* invests in a *target company*, there are typically (at least) two entries in the *crawler.idx* file for the same event: one for the *filer* and another for the *targeted company*. Identifying and removing these duplicates requires additional information, as the only coinciding elements among these *crawler files* records representing the same filings are the event date. All other elements, such as company, CIK, and URL, are distinct. For a comprehensive understanding of the detailed procedure to identify and remove duplicates, please refer to Appendix A.

[122]Duplicated records, as described here, is unlikely to be an concern in datasets like the one presented in Brav et al. (2008), where only events initiated by specific investor types, e.g. hedge funds, are included. In such cases, duplicated records are naturally discarded when the *company field* is filled with the *targeted company*, as the later is unlikely to be a hedge fund.

in the *data bundle* that are obtained exclusively from structured sources: the *filing date* and *filer/targeted company* elements (*name*, *CIK*[123] and *postal address*). Despite the limited number of informational elements at this stage, *postal address* can be used as the basis for two cleaning steps. First, we dropped entries for which *filer's* and *target's* postal addresses matched, assuming if those coincide they refer to an insider event.[124] Second, after geoprocessing *filer/target* postal addresses, we retained only those entries with corresponding addresses located within the contiguous continental US. Following this step, our sample comprised 36,764 records.

In sequence, using batch requests, we retrieved the *main filing documents* for those filings that remained in our sample.[125] We processed the *raw text* of each of these documents to parse *CUSIP* (identify the specific security acquired), *event date* (aka *trigger date*), *ownership stake*[126] (in percentage) and the *activist investment objective* (*Item 4* of *13D filing*).

With this data in hand, we proceeded with additional cleaning, removing duplicated submissions, amendments incorrectly submitted as initial filings, and cases for which we could not parse CUSIP[127] – resulting in a sample of 31,465 observations. Further, we excluded 950 entries that either lacked an *event date* in the original records (e.g., the field was populated with *N/A* or anything at all) or for which we could not parse the *event date*. Most of these incomplete filings corresponds to years pre-dating 2001, a time interval usually excluded in empirical papers due to its known incompleteness.[128] After this removal, the sample consisted of 30,515 entries and incorporated elements extracted from the non-structured *main filing documents*.

[123] CIK stands for Central Index Key. It is a unique identifier assigned by the US SEC to entities (e.g. firms, investment funds, organizations, individuals) that are required to file reports with the SEC.

[124] A *company executive*, as *filer*, would use the same address as the *target company* for investments qualified as insider trading in two situations. Either it is the companies commercial address (which happens to be the same as the one for its' executive), or whenever the filing is submitted on behalf of the executive by the companies' legal representative, that also happens to file as such for the company itself. This process is clearly not foolproof, as the address are also manually input.

[125] For every *main filing document* obtained, a web request was made. Hence, the initial cleaning, which removed entries related to identical filings and non-US companies, prevents unnecessary requests and saves computational resources. In our case, this process resulted in a roughly 40% reduction in the number of filings to be retrieved.

[126] Note that extracting *ownership stake* is a rather complex procedure, mainly due to group filings, as mentioned in Part 3.1.1. Although done with care, including data validation, we acknowledge that the our figures for *ownership* variable are likely to be biased upwards . Later we alleviate this bias using dummies based on patterns of multiple filings for a variety of cases (see subsection 3.2.2).

[127] Consistent with the explanation in Subsection 3.1.2, while our algorithm efficiently seeks a regular expression match on a limited segment of the raw *filing document* where information is most probably located, if the match is unsuccessful, fallback mechanisms search the full text.

[128] Similarly, in your empirical study (Section 3.3), earlier dates are not included as well as we miss information to classify them into *core/non-core*.

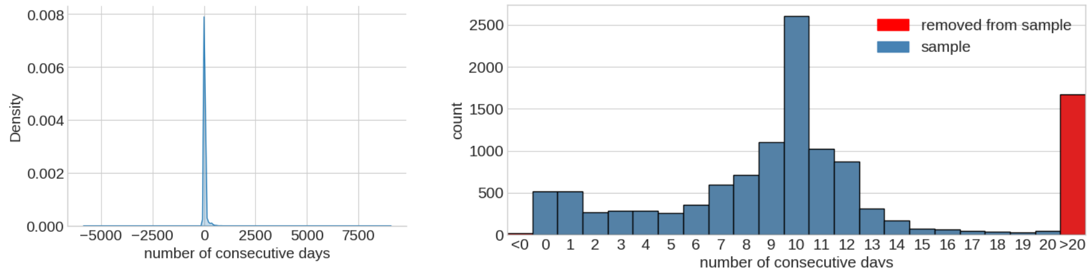
At this point we were ready for the first query in WRDS, using the pair *CUSIP/event date*, to get CRSP/Compustat data. Numerous companies had multiple filings. Specifically, there were 16,719 unique CUSIP,[129] of which 16,617 were matched to CRSP permanent number (PERMNO).[130] Among those filings related to the same securities, some were filed within a short time span, either within a few days or a couple of months, while for others the interval between filings were longer, exceeding six months and often years apart. We set aside, temporarily, the pursuit of multiple filings, as we will dedicate next part (3.2.2) for it.

At this point, we performed a cleaning procedure that is commonly described in the literature on blockholder's activism.[131] We excluded *non-core events* based on *targeted company type* and *security type*, using information from CRSP (*siccd* and *shrcd*). We removed utilities (siccd 49), financial companies (siccd 60-67), and *non-US incorporated securities* (shrcd 10-11), resulting in the elimination of 266, 2,859, and 385 PERMNOs respectively. This step leads to a substantial reduction in the number of observations, with only 11,934 PERMNOs retained out of the 16,617 present in the previous step. Note that this is not information "loss"; rather, it results from the deliberate exclusion of securities that should not be part of our sample.

Equipped with *security unique codes* and *event dates*, we went through a first stage of *event consolidation*.[132] We collapsed multiple filings (same PERMNO/date) into single events, aggregating *ownership stakes* according to the rules outlined in Appendix A. This aggregation resulted in a reduction of total events to 11,861 observations.

As widely known and documented, the time deltas between the *event date* and the date a *13D* is filed often do not satisfy the regulatory bound of 0 to 10 calendar days. Figure 3.1 shows the time deltas for our sample up to this point. Panel A shows the full distribution: while there is a clear concentration around the regulatory time interval, the distribution contains outliers, including negative numbers. Panel B shows the histogram: one bin for negative timedeltas, then 21 daily bins, from 0 to 20 consecutive trading days deltas and a final bin collecting all time deltas above 20 trading days. We kept only those events that refers to the blue bars in the plot (between 0-20 days) and dropped the other filings, shrinking the sample size to 10,169

[129] CUSIP is a unique alphanumeric identifier assigned to financial securities.
[130] PERMNO is a unique six-digit numerical identifier assigned to individual securities in the CRSP database.
[131] We mention this step here for completeness.
[132] Technically, we performed some event consolidation by eliminating duplicated events.



(a) all events

(b) events clipped (0-20 days)

Figure 3.1: Time deltas of 13D Events

This figure shows the distribution of time deltas between filing date and trigger date, for 13D Events from 1994 up to May, 2023. The events sample was cleaned for non-US incorporated companies, duplicates, missing CUSIPs, event dates and utilities and financial companies. Panel A shows the distributions for all observations. Panel B shows counts binned by 1-day time deltas, from 0 to 20, plus the counts for negative time deltas on the first column and the count for time deltas superior to 20 days (last column). There extremes (marked in red) were removed from the sample.

observations.[133]

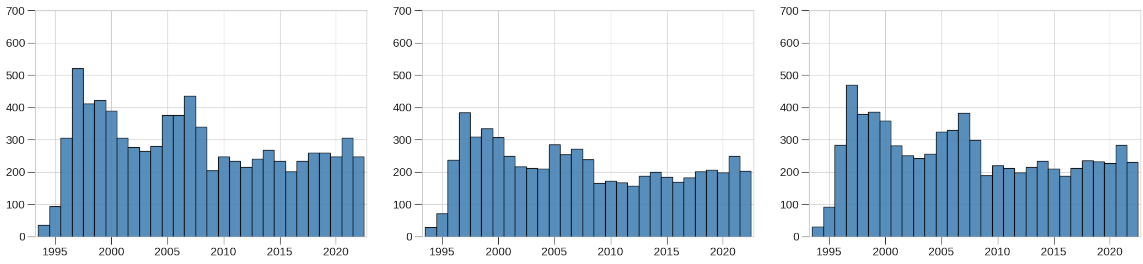
We hold a particular interest in *ownership stakes* as it indicates whether an investor holds a *minority stake* or not. Investors with majority stakes are not within our scope of interest, as they do not contribute to addressing the typical questions related to *investor activism*. Hence, for our purposes, events corresponding to majority stakes were excluded. In addition, we dropped 343 events due to missing ownership information (either nonexistent or we could not parse them).

Once again, following the literature, we excluded filings whenever the *filer* is a non-financial corporation. At this point from 5,424 unique *filers*, 1,033 (19%) were corporates.[134] After removing those events, the sample size decreased by 15%, resulting in 8,342 events.

While we had carefully cleaned our sample up to this point, we have done so based solely on info from *SC 13D* filings' raw text or PERMNOs classifications (sector or address, security type, place of incorporation). Before we move ahead, we passed a fine toothed comb to remove remaining foreign companies that were misclassified in CRSP as US-incorporated (remember we have removed most of them when we geoprocessed the *target company* addresses taken from the filings). Our strategy was to check, for the CIK of *targeted companies*, the existence of filings that are specific for foreign companies (6-K, 40-F, 20-F filings) *This can be done by either*

[133]We discuss cases related to negative time deltas and those above 20 days in Appendix A.

[134]Among the corporates with multiple *13D filings* the most active are Pharmaceuticals (e.g. Abbott, Eli & Lilly) and Technology companies (e.g. AT&T, HP, Microsoft).



(a) events (b) filers (c) targets

Figure 3.2: Unique events, filers and targeted companies

This figure is based on **unique events** corresponding to 13D filings with unique *permno/date*, aggregated by year. Panel A shows the number of unique events, panel B shows the number of filers (unique CIKs) and Panel C shows the number of subject companies. The sample has been cleaned for duplicated records, companies outside the contiguous continental US and securities that are not common stocks. We dropped observations in the following cases: CUSIP, ownership stake or event date missing (or not parsed by our algorithm), no match CUSIP/PERMNOS, delta days between filing date and event date that is either negative or superior to 20 days.

searching the crawler files or each JSON file in the bulk download for each corresponding CIK.. We identified 23 of such cases, related to 30 events. Hence, our final cleaned event sample size is 8,312, that relates to 5,170 unique *target companies*. Figure 3.2 depicts the counts of unique events, unique *filers*, and unique *targets* in the event sample after all cleaning procedures outlined thus far.

3.2.2 Multiple fillings

Up to this point we cleaned the data for a variety of cases and formed events by aggregating filings with coinciding *targeted securities* and *date* (either *filing* or *trigger date*). But there still remain securities for which there are multiple filings – if the dates do not coincide they were kept as separate events. We flag those filings so we can add correspondent dummies as covariates in our regressions.

We use the following flags for multiple filings for the same CRSP security identifier (PERMNO):

- **flag related to consolidation**
  - **has multiple filings**: applies to all events with PERMNOs that had multiple events associated to it in the raw dataset but for which only one event was kept (i.e. more

than one event for with same PERMNO and *filing date*, but different *filers*). The flagged events should include fillings that are filed by groups, but potentially includes also private placements and derivatives. There were 5,059 such flags (60% of the sample).

- **flags related to sequential filings**
  - **first filing**: flags the first event of a PERMNO that have multiple events associated to it, but neither *filing date* nor *event date* is coincident. There were 1,954 such flags (23% of the sample).
  - **sequential filing within 6 months from a previous one**: flags sequential filings (skip the first one) if the interval between it and the previous one is less than six months apart. There were 553 such flags (6.6% of the sample).
  - **sequential filing above 6 months from previous one**: flags sequential filings (skip the first one) if the interval between it and the previous one is more than six months apart. There were 2,552 such flags (31% of the sample).

Note that entries flagged as *has multiple filings* can also assume any of the three alternatives for *sequential filings*. Hence, the *reference case*, where all four flags above equals zero, corresponds to a filing that has not been obtained through consolidation (corresponds to a single *13D entry*) and for which the targeted company has never been targeted more than once (neither before nor after it).

Notice that being targeted more than once might allude to events that should not be present in our activist database; that is why we added the flags related either to *consolidation* as well as for *sequential filings*. A case in point is when multiple filings refer to targeting the same security within a short interval apart (<6 months). One hypothesis for such sequential filings is that they refer to selling opportunities (e.g., including but not restricted to private placements/distributions). Such cases might be rather distinct from those concerning buying efforts initiated by an activist. For the former, the dummy coefficient observed should, at least partially, neutralize the effects on regression of abnormal returns, hence assuming a negative sign.

However, these flagged events would eventually refer to *wolf packs*,[135] instead. If this is the case, various scenarios are possible. Among them, the dummy coefficient would have a negative sign if the initial buyer benefited from acting undercover, and the ones in the *pack* bought at later dates, already incurring newly marked prices. In that scenario, the *pack* would get smaller abnormal returns when compared to the initial buyer (the reference case). However, there are alternative scenarios where the dummy coefficient could, instead, assume a positive sign. This would happen if the price dynamics induced by the *pack* led to subsequent upward movements in prices, for example, by triggering algorithmic trading buying orders.

In contrast to filings presented in quick succession, for sequential instances occurring after a more extended interval our initial assumption is that the firm is being targeted for the second time by a different investor. If this assumption is correct, the instances marked with *sequential filing above 6 months* would indeed configure an activist event. However, we identified peculiar cases that do not conform to that hypothesis. On those, an institutional investor filed an ***initial*** *13D* upon reallocating stocks to different funds within their portfolio. In our view, these filings are not indicative of a *"new"* activist blockholder activity; the *additional* (second) "initial" filing is more of a bureaucratic artifact.[136] Given these possibilities, it seems that the dummy for sequential filings above 6 months, can assume either positive or negative sign. Furthermore it is possible that the coefficients do not exhibit statistical significance at all. We postpone the discussion about coefficient interpretation of this dummy, along with those related to the other sequential flags, until we present the results from regressions in Section 3.3.

In summary, there are four flags related to sequential filings targeting the same security. All of them are derived solely from information already integrated into the dataset up to this point. They can refer to consolidation of filings into a single event or to the identification of sequential filings corresponding to different events. In the latter case, we distinguish between the first occurrence to subsequent ones, either within short- or long-term intervals. Next, we will discuss the flags related to *non-core events*, for which the information is obtained from *other* regulatory

[135]For empirical model on *wolf packs* see Brav, Dasgupta, and Mathews (2022) and for Law Scholar perspective see Coffee and Palia (2015). However we consider that this is unlikely to be the case, because most *wolf packs* are formed by shareholders with stakes around 1%-2% (not captured by *SC 13D filings*).

[136]Though we did not explore the motivations for such cases, we suspect, however, that the new filing was deemed necessary in those instances in which none of the final beneficiaries mentioned in the "first" initial filing was currently holding the position.

filings.

3.2.3 Other SEC Edgar sources for *event categorization*

A critical aspect of dataset construction consists in identifying and keeping only those entries related to *"idealized activist blockholders' events"*. These events are initiated by investors who select *target companies*, for which they see opportunities to enhance value through their intervention. In Part 3.1.2.8, we introduced the term *event categorization* to specifically refer to the task of distinguishing between *core* and *non-core* events.

In the literature on blockholders' activism, authors mention certain *non-core events* that they exclude from their datasets. These exclusions comprises *insider trading*, *risk-arbitrage* (due to mergers), and *company reorganizations*, especially those resulting from bankruptcies. However, as discussed in Part 3.1.2.9, authors are either silent about the process they use to identify those cases[137] or refer to painstaking manual collection (i.e., examining, one-by-one, the text in *Item 4. investment objective*). The later, given the sheer amount of work involved, equates to not being reproducible.[138]

To address this gap, we document in the current Section[139] a cost-effective approach, that relies on SEC Edgar filings to identify such cases. Specifically, we use *Forms 8K*, merger-related proxy filings, notice of delisting and insider trading filings, as detailed below.

- **Mergers**: Before companies proceed with mergers, they must obtain shareholders' approval. For such cases, US regulations mandate the filing of proxy statements bearing special codes that include the letter *"M"* (*PREM14*, *DEFM14* or related filings). While around 38% of *targeted companies* have merged at some point (before or after the *13D event*), mergers that hold significance for our study are only those announced *before* the *activist event*, and within a rather tight window. We create dummies that assumes the value 1 (one) if there is a merger-related proxy filing for the same *targeted company*, within a given *pre-13D*

[137] One possibility is that flags might have been extracted from WRDS datasets that are available for an extra fee (not included in the basic subscription). However, this is just a conjecture, as there is no explicit mention in the papers.

[138] In addition, we have demonstrated that *legacy scripts* fall short of implementing those exclusions, and *commercial datasets* also need to go though event consolidation (see Part 3.1.2.8).

[139] Please refer to Appendix A for additional explanations, including operational aspects.

interval (the window before the activist *trigger date*). This exercise was conducted for the following intervals: 0-3 months, 3 months-6 months, and 6 months-1 year (for which 506, 139, and 100 dummies assumed the value 1 (one), respectively).

- **Bankruptcy**: Bankruptcies and reorganizations are *material events*. As all *material events*, they must be promptly disclosed by filing an *8K*.[140] Filings referring to these events contain the reference *Item 1.03 — Bankruptcy or receivership*.[141] We create dummies that assumes value 1 (one) if there is a *8K* filing that refers to *Item 1.03*, within the following pre-13D intervals: 0-3 months, 3 months-6 months, and 6 months-1 year (for which 51, 17, 23 dummies assumed the value 1 (one), respectively).

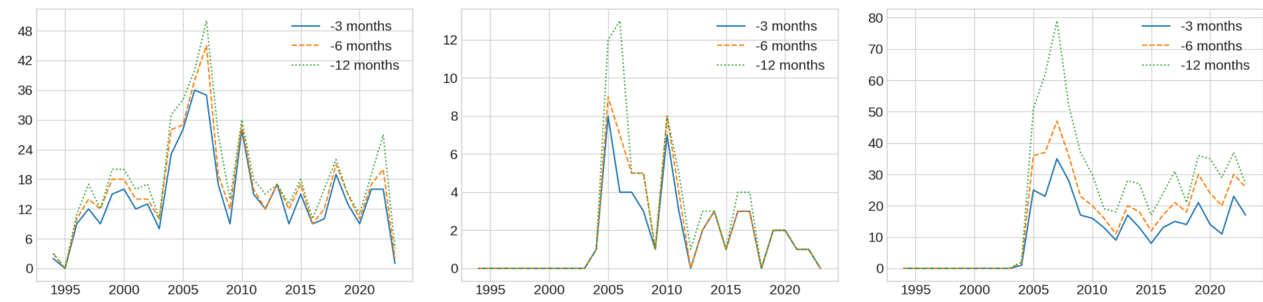
- **Notices of delisting**: We also flagged *targeted companies* for which there were *8K* filings referring to *Item 3.01, Notice of Delisting or Failure to Satisfy a Continued Listing Rule or Standard; Transfer of Listing;* within the pre-13D filing intervals defined between 0, 3, 6 and 12 months.[142] Companies need to submit these filings for example, if they receive a notice for non-compliance (e.g. failure to file *10K/10Q*, or stock price falls below the standards set by the exchange where it is listed) from the exchange where their securities are listed. There were 371, 232, 372 flags for each of the windows defined in between 0, 3, 6 and 12 months.

Figure 3.3 shows the number of flags aggregated by year for the events sample. Panel A displays mergers that preceded *13D filings*, while Panel B illustrates *bankruptcies*, and Panel C represents *notice of delistings*. Data on Panel B and Panel C starts in mid-2005 because that is when information about the *8K's items* began to be conveniently incorporated into the SEC Edgar JSON files packed in the *bulk download zip file*.

[140] Although a company can disclose *material events* on a *10Q* or *10K*, it is extremely unlikely that the bankruptcy event will coincide with the date of the quarterly or annual reports, so the examination of *8K* filings sounds as a reasonable choice for identifying bankruptcy and reorganizations.

[141] The SEC regulation lists all *material events* that should be disclosed via an *8K* and assigns a reference number for each. A single *8K* may refer to one or multiple *material events*.

[142] This is analogous to the procedure done for mergers and bankruptcies.



(a) mergers (b) bankruptcies (c) delistings

Figure 3.3: Flags on pre-13D windows (-3/-6/-12mos)

This figure shows the number of flags for windows prior to 13D events, aggregated by year (reference year of the 13D filing date. Panel A shows mergers that preceded 13D filings, Panel B shows bankruptcies and Panel C, delistings. All the panels share the same legend line styles: the dotted line covers a window of 12 months before 13D filing, the dashed line a period of 6 months and the solid line, 3 months. Data on Panel B and Panel C starts on mid-2005, because that is the date when the information about the 8Ks *items* started to be published in SEC EDGAR index records. The base dataset with 13D events used as source for matching the flags had 27.295 unique pairs *target company/13D filing date*.

- **Insider trading**: In the first part of this section (Part 3.2.1), we removed events for which the postal addresses for *filers/targets* were the same (the procedure is detailed in the Appendix, Section A.3.4). With that, most of the filings that may potentially involve insider trading were dropped. However, this approach has limitations, so we complemented it with information from SEC *Form 4* (Insider Trading).[143,144] We flagged events for which, within the 20-days and 365-days windows *pre-13D filing date*, there is a *Form 4* for the same pair of *filer/target company* (in practice, we match *filer CIK* and *target CIK*). There are 394 and 552 cases, respectively. Notice that it is not sufficient just to match SEC *Form 4* for *target companies* only: it is important to match the *filer* as well.

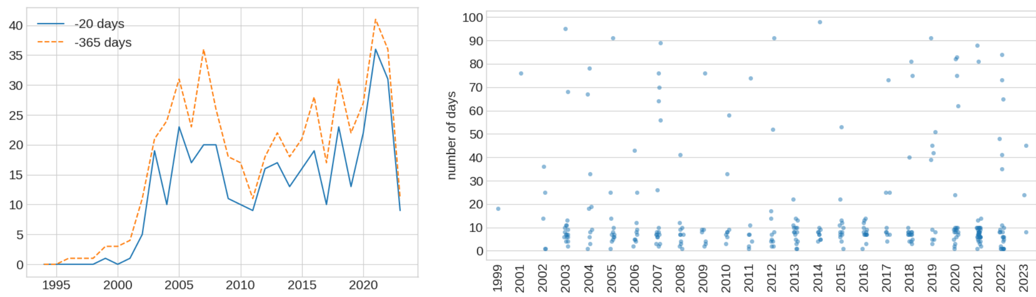
  Instead of dropping the flagged filings outright, as we have done with most of the insider trading data (those with coincident addresses), we retained these residual filings and introduced dummy variables referring to them in the regressions. This approach allows us to observe how the inclusion of insider *13Ds* influences the results of blockholder activism variables that are typically presented in the literature.

Figure 3.4 refers to flags for insiders. Panel A shows the amount of insider flags aggregated

[143] If we had checked the Insider Trading forms before dropping filings with coincident postal addresses for filer/target, the percentage of filings marked as insider would be much larger, around 30%.

[144] The trigger for filing a *Form 4* is any transaction with the company's securities by an insider, and it must be filed within two business days following the triggering event.

by year for the 20 days intervals. We add for reference also a 365 days interval. Panel B shows, for every observation flagged, binned by 13D-year, the time difference between the *Form 4 filing date* and the 13D *filing date*. This later plot helps in determining the ideal window (pre-13D filing) for flagging insider trading: 10 days is too narrow, so we chose to use 20 days, instead. In this interval, when there is a match it is likely that both *13D* and *Form 4* refers to the same event that happens to be covered by the two distinct regulations. This is the case whenever an insider has reached, for the first time, 5% participation on the *targeted company*.



(a) all events (b) days between Form 4 / 13D filings

Figure 3.4: Insider flags: interval between insider filing and 13D filing

This figure refers to 13D events that were flagged as insider events. The flag is attributed whenever there is a form 4 filing (insider trading) with the same pair filer/targeted company in the time window that precedes the 13D event. Panel A shows flag counts (solid line for 20 days interval and dashed line for 365). Panel B shows the number of days between the 13D and form 4 filings, binned by year in the for the interval 0 to 100 consecutive days.

3.2.4 Market and fundamental data

We merge data from SEC and CRSP/Compustat using the combination of CUSIPs and *filing date* (extracted from SEC *13D filings*) to find the correspondence to Compustat global company key identifier (GVKEY)s and PERMNOs.[145]

We collect market information, including stock prices, trading volume, and the number of shares, from CRSP. Following the literature on blockholder activism, we select firm-specific control variables that figure as items in Table 3.1 or 3.2, and their respective definitions can be found in those tables' captions. The financial accounting data required to compute these variables is obtained from Compustat. In addition, when computing *abnormal returns* with

[145]The correspondence is extracted from the CRSP/Compustat merged dataset (*ccm*).

respect to CAPM, FF3, FFM, and FF5, we rely on common factors reported on Kenneth French's website to derive loadings (betas). All data from CRSP, Compustat, and the common factors were retrieved from WRDS using a basic subscription.

3.2.5 Descriptive statistics

Over the period covered by our study, there are 15,497 common stocks in the CRSP/Compustat universe, for a total of 142,027 firm-year observations. On average, over one third of the companies was targeted at least once by a blockholder activist, at some point within that time span.

3.2.5.1 Firm fundamentals

Table 3.1 presents firm-year summary statistics (number of observations, mean, and median) for two groups of companies: those targeted by blockholder activists at some point (columns 1-3) and the entire CRSP/Compustat universe (columns 4-6). The last column shows the proportion of (targeted) firms in our sample with respect to all companies.[146] Notice that in the first three columns of Table 3.1, there is no distinction between the periods referring to pre- or to post-blockholder intervention. Hence, as these statistics combine figures for the pre-target period with eventual outcomes from activities or initiatives taken by the activist along the investee firm, they serve only for a rather broad overview and are not suitable for deriving robust conclusions about the characteristics of companies targeted.

After performing cleaning and consolidation, as outlined in Part 3.2.1 and 3.2.2, our activist dataset contains 8,312 events. However, incorporating fundamentals, which are used as controls in regressions, leads in practice to a reduction in the dataset size, as some company/date pairs lack corresponding data in the Compustat database.[147] For instance, we obtain data for as much as 86% of the events for *cash-to-assets* and *payout*, but for as little as 69% for *sales growth*.

Table 3.2 offers summary statistics for the fundamentals of *targeted companies* in the single-

[146] Column *(7)* equates to column *(1)* over column *(4)*.

[147] The loss of entries could have been avoided through data imputation. However, as we opted not to impute data, entries with missing values for the covariates were consequently removed. This approach is consistent with the literature on Investor Activism.

Table 3.1: Target firms fundamentals: firm-months for all years *(1994-2023)*

	event sample			*CRSP/Compustat*			
	firms-year	*mean*	*median*	*firms-year*	*mean*	*median*	*ev/Comp*[1]
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
market capitalization	50493	2835.87	389.94	140960	4902.54	392.86	35.8%
book-to-market	50755	0.58	0.45	141642	0.63	0.49	35.8%
tobin's Q	50755	2.21	1.53	141640	2.21	1.39	35.8%
sales growth	46655	0.79	0.08	128725	0.77	0.08	36.2%
ROA	47538	0.06	0.11	130888	0.04	0.08	36.3%
cashflow	47516	0.01	0.07	129900	0.01	0.05	36.6%
market leverage	50690	21.60	13.22	141280	24.60	16.93	35.9%
book leverage	50786	35.70	25.40	141723	39.00	30.41	35.8%
cash-to-assets	50844	22.20	11.81	142027	20.37	9.39	35.8%
dividend yield	50711	0.98	0.00	141390	2.98	0.00	35.9%
payout	50844	2.19	0.02	142027	2.23	0.03	35.8%
profit margin	49416	-73.71	9.25	136868	-75.78	11.65	36.1%

[1] Ratio of firm-years of event firms over firm-years of Compustat firms.

This table shows summary statistics for the targeted companies (in firm-months) in our sample (3 first columns) and reference Compustat universe (next 3 columns). The last column shows the percentage of firms-months that covers all targeted firms per total Compustat firms-months for the period from January, 1994 to May, 2023. Variables are winsorized at 1% and 99% levels. Market capitalization is in millions of dollars (May, 2023 dollar values); book-to-market is (book value of equity/market value of equity); tobin's Q is (book value of debt + market value of equity)/(book value of debt + book value of equity); ROA is EBITDA/lagged assets; cashflow is (net income + depreciation and amortization)/lagged assets; market leverage is total debt/(total debt + market value of equity); book leverage is total debt/(total debt + book value of equity); cash is (cash + cash equivalents) scaled by assets; dividend yield is common dividend/market value of equity; Payout ratio is (common dividend + share repurchases)/market value of equity.

Table 3.2: Targeted firms fundamentals before trigger date

	count	*std*	*mean*	*10%*	*25%*	*50%*	*75%*	*90%*
market capitalization	6408	8639.84	1571.85	29.89	76.99	241.67	822.55	2580.91
book-to-market	6410	0.78	0.65	0.09	0.26	0.50	0.88	1.40
tobin's Q	6410	2.39	1.95	0.87	1.05	1.40	2.07	3.31
sales growth	5742	157.07	2.48	-0.21	-0.05	0.06	0.20	0.52
ROA	5896	0.26	0.03	-0.26	-0.02	0.09	0.15	0.23
cashflow	5891	0.29	-0.02	-0.32	-0.07	0.05	0.11	0.18
market leverage	6399	25.73	24.10	0.00	0.91	15.26	40.07	64.57
book leverage	6412	57.43	38.16	0.00	1.79	27.50	55.72	81.02
cash-to-assets	6432	25.59	22.55	0.86	3.04	11.49	34.29	64.60
dividend yield	6403	16.40	1.09	0.00	0.00	0.00	0.00	1.60
payout	6432	4.73	2.00	0.00	0.00	0.00	1.67	6.37
profit margin	6183	597.93	-76.84	-55.84	-1.34	7.33	14.93	25.03

This table shows summary statistics for the targeted companies fundamentals before the 13D trigger event (variables taken from the window -13 months to -1 month). Variables are winsorized at 1% and 99% levels. Market capitalization is in millions of dollars (May, 2023 dollar values); book-to-market is (book value of equity/market value of equity); tobin's Q is (book value of debt + market value of equity)/(book value of debt + book value of equity); ROA is EBITDA/lagged assets; cashflow is (net income + depreciation and amortization)/lagged assets; market leverage is total debt/(total debt + market value of equity); book leverage is total debt/(total debt + book value of equity); cash is (cash + cash equivalents) scaled by assets; dividend yield is common dividend/market value of equity; Payout ratio is (common dividend + share repurchases)/market value of equity.

year period preceding the *13D event*. As for each entry we take only figures corresponding to the last *pre-event* firm-year observation available, these figures portray the average value for targeted companies' fundamentals in the period leading up to the activist intervention. The aggregate raw data in this table is in line with those patterns suggested in the literature: *targeted companies* are smaller (have below-average sizes as measured by *market capitalization* in constant prices), are less leveraged, have diminished *payout* and *dividend yield*, along with lower *Tobin's Q* and *book-to-market* ratios.

In Table 3.3, we break the summary statistics for fundamentals by year. As the data therein is raw, there are substantial variations among observations. This is in great part due to the distinct prevailing macroeconomic factors across the years, as well as the various different stages of the business cycle. As a comparable table for the CRSP/Compustat universe (not included) would also exhibit variations, the figures in Table 3.3 alone do not provide much information on the unique characteristics of targeted companies relative to others.

Thus far, the tables presented in this section, Table 3.1, 3.2, and 3.3, fall short to communicate how fundamentals of *targeted companies*[148] compare to those of the entire aggregate of firms, for the same corresponding point in time. To address this aspect, we provide context on pre-target fundamentals by collecting their frequency into corresponding Compustat universe decile bins, which are recomputed for each year. The results of this more precise approach are presented in Figure 3.5. In general, the relative status of targeted companies' fundamentals are aligned to those suggested in blockholder activists' literature. For example, activist investors target those firms with market *capitalization*, *tobin's Q*, *ROA*, and *profit margin* below market averages.

However, a close inspection reveals that some facts taken for granted in the literature are not that clear-cut. For instance, it is usually assumed that blockholder activists act like value investors "because they invest in low *book to market* (*btm*) companies". While aggregate averages do indicate a slightly higher *btm* for *targeted companies*, there are many years in which this observation does not hold (2011, 2013, 2017), while for other years, the difference is minimal.[149]

As for *leverage*, it can be observed that though targeted companies traditionally have very

[148]At the time immediately preceding the activist event.

[149]As there is a substantial number of targeted companies each year (on the hundreds); this variation is not driven by low representativity.

Table 3.3: Target firms fundamentals before 13D trigger date

	market capitalization			*book-to-market*			*tobin's Q*			*sales growth*			*ROA*			*cashflow*		
	count	*mean*	*50%*	*count*	*mean*	*50%*	*count*	*mean*	*50%*	*count*	*mean*	*50%*	*count*	*mean*	*50%*	*count*	*mean*	*50%*
1994	72	882.62	228.97	79	0.60	0.53	79	1.65	1.36	79	0.08	0.07	79	0.09	0.13	79	0.06	0.09
1995	211	1931.97	133.69	209	0.60	0.45	209	2.19	1.40	180	0.45	0.09	186	0.06	0.12	186	-0.00	0.07
1996	409	613.42	141.26	409	0.55	0.43	409	2.15	1.54	324	0.44	0.14	329	0.06	0.12	328	0.00	0.07
1997	332	1793.95	148.83	329	0.54	0.42	329	1.99	1.57	292	0.39	0.12	300	0.04	0.10	300	-0.02	0.05
1998	326	823.98	153.70	326	0.68	0.53	326	1.86	1.35	298	0.37	0.09	295	0.07	0.11	295	0.01	0.07
1999	301	2582.63	169.92	301	0.84	0.65	301	2.78	1.23	265	0.30	0.10	263	0.09	0.13	263	0.03	0.08
2000	281	588.48	156.46	281	1.03	0.69	281	1.99	1.12	249	1.07	0.12	252	-0.02	0.10	252	-0.09	0.05
2001	203	1518.35	122.66	203	1.04	0.65	203	1.66	1.17	193	0.12	0.02	194	0.01	0.07	194	-0.06	0.02
2002	179	392.26	106.30	179	0.94	0.82	179	1.51	1.10	169	0.05	-0.02	172	-0.01	0.04	172	-0.06	0.01
2003	207	1725.14	176.30	207	0.72	0.57	207	1.83	1.29	195	0.13	0.02	197	0.06	0.08	197	0.00	0.04
2004	268	1401.46	259.42	268	0.53	0.48	268	1.93	1.51	250	0.13	0.07	253	0.07	0.11	253	0.04	0.07
2005	292	1236.85	402.68	292	0.58	0.52	292	1.74	1.48	276	0.12	0.08	279	0.08	0.11	279	0.04	0.07
2006	346	2070.61	449.90	346	0.48	0.44	346	2.06	1.66	328	0.22	0.07	335	0.04	0.09	335	0.00	0.06
2007	290	1792.97	294.49	290	0.54	0.46	290	1.92	1.56	263	45.26	0.06	269	0.04	0.08	269	-0.00	0.05
2008	193	627.55	201.09	193	1.03	0.70	193	1.41	1.16	184	0.45	0.03	186	0.07	0.09	186	0.01	0.06
2009	179	764.79	180.60	179	0.71	0.58	179	1.61	1.32	173	0.08	-0.04	177	0.03	0.07	177	-0.03	0.03
2010	212	1676.51	291.31	212	0.60	0.51	212	1.66	1.41	200	0.14	0.05	201	0.10	0.11	201	0.06	0.09
2011	156	984.89	372.20	156	0.78	0.62	156	1.54	1.24	150	0.21	0.12	152	0.07	0.10	152	0.03	0.06
2012	169	1527.88	348.60	169	0.67	0.60	169	1.62	1.32	163	0.15	0.03	164	0.06	0.10	164	0.03	0.07
2013	197	1667.90	393.20	197	0.60	0.48	197	1.88	1.46	182	0.09	0.01	185	0.03	0.07	183	-0.02	0.03
2014	203	2992.11	491.12	203	0.52	0.46	203	1.96	1.44	176	0.08	0.02	181	0.05	0.08	181	0.01	0.05
2015	166	1710.83	312.14	166	0.51	0.42	166	1.94	1.43	148	0.07	0.05	157	-0.01	0.06	156	-0.06	0.03
2016	187	1616.20	314.07	187	0.47	0.40	187	1.88	1.45	163	0.15	0.04	172	-0.03	0.05	171	-0.08	0.01
2017	195	1510.57	287.40	195	0.40	0.36	195	2.39	1.58	170	5.46	0.02	178	-0.09	0.04	178	-0.11	0.00
2018	200	1634.41	371.98	200	0.57	0.52	200	1.79	1.27	175	0.28	0.06	185	-0.03	0.06	185	-0.07	0.03
2019	184	1882.45	330.84	184	0.64	0.59	184	1.88	1.28	160	0.07	0.02	174	-0.03	0.06	174	-0.08	0.01
2020	165	4413.03	346.27	165	0.65	0.53	165	2.52	1.43	119	0.16	-0.08	133	-0.10	0.03	133	-0.16	-0.04
2021	207	2394.38	419.96	207	0.52	0.38	207	2.34	1.75	151	0.27	0.16	173	-0.09	0.05	173	-0.12	0.02
2022	78	1225.95	198.38	78	0.72	0.63	78	1.80	1.19	67	0.20	0.11	75	-0.12	-0.03	75	-0.14	-0.11

s table shows summary statistics for the targeted companies fundamentals before the 13D trigger event (variables taken from the window -13 months to -1 month) broken by (event) year. Variables are winsorized at 1% and 99% levels. Market capitalization is in millions of dollars (May, 2023 dollar values); book-to-market is (book value of equity/market value of equity); tobin's Q is (book value of debt + market value of equity)/(book value of debt + book value of equity); ROA is EBITDA/lagged assets; cashflow is (net income + depreciation and amortization)/lagged assets; market leverage is total debt/(total debt + market value of equity); book leverage is total debt/(total debt + book value of equity); cash is (cash + cash equivalents) scaled by assets; dividend yield is common dividend/market value of equity; Payout ratio is (common dividend + share repurchases)/market value of equity.

Table 1.3: Target firms fundamentals before 13D trigger date *(continued)*

	market leverage			*book leverage*			*cash-to-assets*			*dividend yield*			*payout ratio*			*profit margin*		
	count	*mean*	*50%*	*count*	*mean*	*50%*	*count*	*mean*	*50%*	*count*	*mean*	*50%*	*count*	*mean*	*50%*	*count*	*mean*	*50%*
1994	79	30.23	27.22	82	81.39	39.66	82	9.76	5.37	79	0.34	0.0	82	0.97	0.00	81	9.39	10.71
1995	208	25.60	20.05	211	47.94	36.34	211	16.11	5.24	209	0.44	0.0	211	1.05	0.00	205	-69.97	7.31
1996	407	23.54	16.94	407	34.18	29.37	409	20.73	7.75	407	1.93	0.0	409	0.81	0.00	400	-36.61	7.28
1997	329	24.29	18.41	330	39.31	33.81	332	17.74	6.28	328	0.18	0.0	332	1.13	0.00	324	-33.07	7.29
1998	324	27.29	21.35	324	39.28	32.74	329	16.42	5.72	326	0.31	0.0	329	2.27	0.00	322	-39.08	8.99
1999	297	31.82	28.27	301	44.94	37.32	301	16.14	4.94	298	0.46	0.0	301	2.83	0.00	294	-29.61	8.66
2000	280	31.25	24.66	281	38.31	33.70	281	20.13	8.52	281	8.22	0.0	281	2.78	0.00	277	-77.66	5.81
2001	203	27.21	17.07	203	36.77	22.60	203	22.17	13.04	203	1.05	0.0	203	1.79	0.00	200	-53.18	5.40
2002	179	23.97	14.81	179	40.00	21.26	179	26.92	12.99	179	0.37	0.0	179	1.52	0.00	175	-56.02	4.68
2003	207	25.30	18.04	207	33.21	27.14	207	22.56	11.13	207	0.59	0.0	207	1.45	0.00	204	-52.72	6.48
2004	268	20.10	11.00	268	30.42	24.01	268	22.81	12.78	268	0.43	0.0	268	1.34	0.00	264	-59.61	9.03
2005	292	19.38	10.76	292	31.77	23.11	292	20.65	12.28	292	0.62	0.0	292	1.85	0.00	291	-25.70	8.81
2006	346	19.51	12.11	346	33.48	21.83	346	23.65	14.25	346	0.53	0.0	346	2.18	0.00	341	-118.44	7.49
2007	290	19.34	9.36	290	35.22	21.50	290	26.37	14.25	290	1.54	0.0	290	2.48	0.00	279	-107.03	7.31
2008	193	26.98	16.95	193	35.33	22.17	193	24.13	13.65	193	1.03	0.0	193	4.95	0.00	190	-53.51	6.95
2009	179	21.32	9.44	179	28.62	14.79	179	28.70	19.41	179	0.37	0.0	179	1.89	0.00	176	20.02	6.55
2010	212	18.11	8.10	212	28.98	15.63	212	25.11	18.31	212	0.53	0.0	212	1.81	0.01	211	6.08	10.47
2011	156	24.53	11.83	156	36.13	23.28	156	22.24	13.53	156	0.41	0.0	156	2.45	0.12	154	-22.64	9.52
2012	169	24.05	15.08	169	35.71	22.88	169	21.41	13.83	169	0.90	0.0	169	3.10	0.68	167	-27.29	10.24
2013	197	21.31	11.01	197	36.43	24.41	197	21.76	14.51	197	0.66	0.0	197	2.08	0.04	192	-30.68	7.92
2014	203	21.82	13.96	203	37.02	25.35	203	22.80	12.34	203	0.61	0.0	203	1.95	0.04	189	-7.87	7.77
2015	166	24.86	14.34	166	47.27	26.75	166	24.51	16.00	166	0.71	0.0	166	2.17	0.08	154	-139.38	6.41
2016	187	24.67	16.65	187	46.49	35.43	187	23.43	15.24	187	0.71	0.0	187	1.96	0.09	178	-191.83	5.56
2017	195	22.02	14.13	195	43.38	30.53	195	26.16	14.18	195	0.85	0.0	195	1.85	0.01	179	-161.15	5.13
2018	200	26.45	19.48	200	38.96	28.20	200	25.34	9.31	200	0.49	0.0	200	2.00	0.25	185	-189.54	5.92
2019	184	26.26	17.37	184	38.51	28.62	184	22.43	10.18	184	1.16	0.0	184	3.10	0.35	164	-209.99	6.13
2020	165	26.50	17.13	165	41.94	29.83	165	24.99	13.97	165	1.40	0.0	165	1.19	0.00	137	-416.94	3.19
2021	206	20.86	13.80	207	42.33	31.07	218	36.82	25.40	206	1.24	0.0	218	1.58	0.01	181	-160.05	5.22
2022	78	28.32	14.66	78	41.99	25.62	78	33.03	20.16	78	0.55	0.0	78	2.06	0.08	69	-113.66	1.86

This table shows summary statistics for the targeted companies fundamentals before the 13D trigger event (variables taken from the window -13 months to -1 month) broken by (event) year. Variables are winsorized at 1% and 99% levels. Market capitalization is in millions of dollars (May, 2023 dollar values); book-to-market is (book value of equity/market value of equity); tobin's Q is (book value of debt + market value of equity)/(book value of debt + book value of equity); ROA is EBITDA/lagged assets; cashflow is (net income + depreciation and amortization)/lagged assets; market leverage is total debt/(total debt + market value of equity); book leverage is total debt/(total debt + book value of equity); cash is (cash + cash equivalents) scaled by assets; dividend yield is common dividend/market value of equity; Payout ratio is (common dividend + share repurchases)/market value of equity.

(a) market capitalization

(b) book-to-market

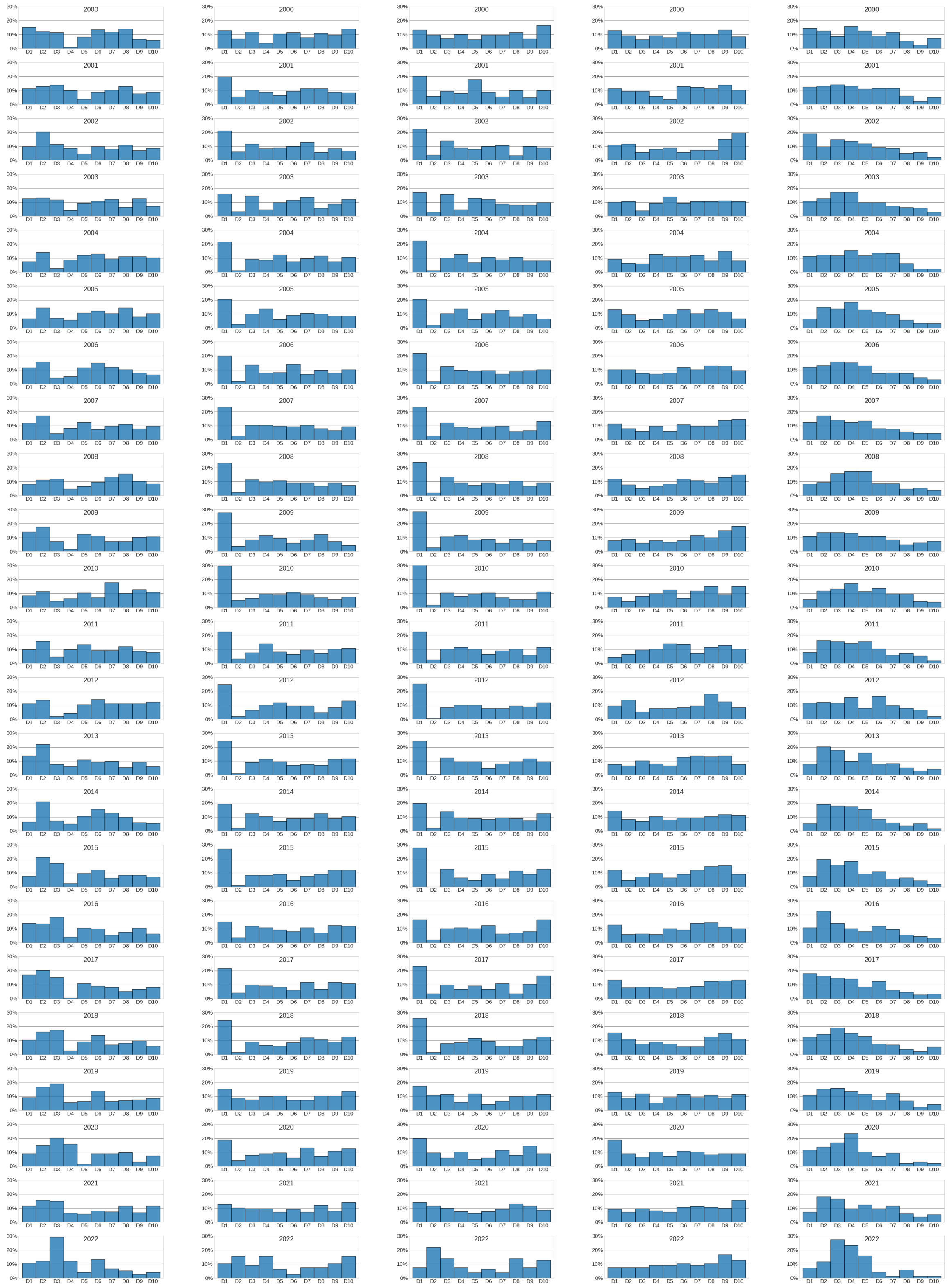
(c) tobin's Q

(d) sales growth

(e) return-on-assets

Figure 3.5: Target companies fundamental data by market decile

This figure presents the changes in the distribution of fundamental statistics for target companies across different years. Each year-plot displays the count of targeted companies, categorized into ten bins. These bins are defined based on the deciles of the same fundamental characteristics, computed using data from the Compustat universe. The decile breaks are recalculated annually to account for changes in the composition of the Compustat universe and the evolution of its fundamentals.



(f) cashflow | (g) market leverage | (h) book leverage | (i) cash-to-assets | (j) profit margin

Figure 3.5: Target companies fundamental data by market decile - continued

This figure presents the changes in the distribution of fundamental statistics for target companies across different years. Each year-plot displays the count of targeted companies, categorized into ten bins. These bins are defined based on the deciles of the same fundamental characteristics, computed using data from the Compustat universe. The decile breaks are recalculated annually to account for changes in the composition of the Compustat universe and the evolution of its fundamentals.

low *leverage* (around 20% of them had either *book* or *market leverage* in the lowest market decile), in recent years the distribution relative to the market became less skewed (see columns *g* and *h* of Figure 3.5).

Targeted companies are also known for holding high levels of *cash to assets*. Indeed, one emblematic strategy used by activists is to acquire blocks in companies with excess cash and, subsequently, influence them to distribute it to shareholders. On average, too much cash seems to be the rule for *targeted companies*, except for 2020, the year of the pandemics.[150]

Finally, when the analyzed statistic is *cash flow*, there is no clear pattern that repeats along all years, although in the last decade, low cash flow has been the norm.

3.2.5.2 Descriptive stats: abnormal returns and ownership stakes

The next (and last) four tables with descriptive statistics are dedicated to the regression outcomes that will be studied soon, in Section 3.3. All of those tables follow the same structure: statistics are presented first by year, and at the bottom of the table, they are aggregated for three different periods. The first period refers to the full dataset (1994-2022). The next one covers the interval for which we have data in a convenient way to flag *non-core events* (2006-2022). Finally, the last period (2010-2019) is a subset of the second one, excluding years of the financial crisis and the pandemics.

Abnormal returns and market returns

Table 3.4 presents descriptive statistics for buy and hold abnormal returns of targeted common stocks in the ± 20 trading days interval around the triggering date. We present the mean and standard deviation with respect to four different pricing models commonly used in the literature and in the industry: Capital Asset Pricing model (CAPM) (columns 1-2), Fama-French 3 factors model (FF3) (columns 3-4), Fama-French 3 factors + momentum model (FFM) (columns 5-6) and Fama-French 5 factors model (FF5) (columns 7-8). They convey pretty similar results, in terms of mean abnormal returns around 10%, but, above all, an extremely

[150] A credible explanation for this anomaly is that companies, overall, adopted a conservative strategy by maintaining larger cash reserves and postponing investments, given the substantial uncertainty prevailing at that point. This behavior influenced the decile breakpoints, distorting the resulting illustration.

high degree of variability. Though mean abnormal returns are positive, a one standard deviation is much larger than the mean, about 2.5 times, whether considered in single years or when aggregated in large intervals (bottom of the table). We will see later that when we use abnormal returns as dependent values in regressions, the covariates do capture much of that variability, such that their coefficients are statistically significant.

To help give context to the raw abnormal returns presented, we incorporate Table 3.5 with daily averages of *market returns* over those same days centered around the *trigger date* using three distinct references: value-weighted (*vwred*), equal-weighted (*ewred*), and SP500. Columns 1-3 present daily returns in daily rates, while columns 4-6 show those rates compounded yearly.

Ownership stakes: percentage and log dollar

Table 3.6 and 3.7 present summary statistics for *ownership stakes* in percentages and in dollars (log), respectively. In contrast to the observation for abnormal returns, the variation for ownership is much more contained: standard deviation is a bit smaller than average figures.

Now, comparing the two ownership stakes distributions, there is a marked difference. Percentage stakes are skewed to the right, with some high values influencing the average upwards. However, when we evaluate dollars log, the skewness disappears or is even slightly to the left for some years. Though there is some reducing effect from computing logs, this alone wouldn't be enough to change the direction of the skewness. The balancing effect comes from the inverse relationship between the size of the company and the percentage stake. Investors acquire larger stakes in smaller companies, and vice versa. As the dollar ownership is computed as the product of percentage ownership with company market capitalization, it counteracts the skewness observed solely in percentages.

The distribution of the *dollar log* holds some interesting properties when compared to that for *percentage ownership*. Percentage ownership is a variable bounded both on left and right. It can take values from a minimum of 0% to a maximum of 100%. Whereas the *log dollar ownership* is still bounded on the left, it ends being much less affected as company sizes are virtually infinite. The boundedness of the *percentage ownership* seems to take a toll on the coefficients obtained in regressions. As we will see in the next section, certain regression coefficients that do not show statistical significance in percentage ownership regressions become relevant when we switch to *dollar log*.

As briefly described in 3.1.2, extracting percentage stakes from *13D filings* involves more than simply aggregating values from multiple parsed *13D uploaded documents*. It requires supplementary information and additional steps for *event identification* (see Part 3.1.2.8), and even then, we were not able to completely remove the upward bias due to double counting. We do not expect our procedure to be flawless; on the contrary, we recognize that double-counting is a pervasive problem if one does not manually check each and every filing document. That being said, we take measures to address that problem and also rely on Dlugosz et al. (2006)'s findings, indicating that when *ownership stakes* are positioned on the left-hand side of regression equations (as in our case), the regression coefficients are not biased.

Table 3.4: Summary statistics: targeted stocks cumulative abnormal returns

		CAPM		*FF 3 factors*		*FF3 + momentum*		*FF 5 factors*	
	count	*mean*	*std*	*mean*	*std*	*mean*	*std*	*mean*	*std*
		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
1994	10	0.085795	0.170753	0.097680	0.162734	0.093101	0.167000	0.089905	0.157645
1995	68	0.068876	0.336010	0.068067	0.337543	0.066719	0.331667	0.068533	0.346089
1996	190	0.062210	0.370478	0.069864	0.378325	0.062412	0.378198	0.074573	0.385357
1997	339	0.032032	0.314044	0.033917	0.325617	0.034489	0.327673	0.034914	0.329537
1998	264	0.079853	0.428633	0.094307	0.474510	0.102874	0.540025	0.095484	0.474385
1999	259	0.129468	0.429647	0.110046	0.434304	0.109949	0.432457	0.091637	0.415201
2000	242	0.153113	0.590587	0.156842	0.659873	0.161071	0.711899	0.133067	0.588925
2001	190	0.131418	0.592526	0.100872	0.540157	0.119752	0.533743	0.091727	0.545185
2002	157	0.135684	0.804994	0.162274	0.869125	0.165328	0.859277	0.152313	0.911097
2003	136	0.110760	0.460954	0.107217	0.469371	0.118355	0.449948	0.116497	0.474130
2004	151	0.080820	0.324750	0.076533	0.327923	0.078134	0.326189	0.072561	0.331116
2005	209	0.081162	0.284102	0.078988	0.284330	0.079529	0.283282	0.083522	0.295550
2006	206	0.074824	0.287912	0.090382	0.296058	0.086841	0.289427	0.085161	0.303907
2007	263	0.081948	0.319993	0.082923	0.324911	0.074768	0.321685	0.090682	0.348744
2008	229	0.211034	1.122476	0.205130	1.008315	0.202299	0.976023	0.205271	1.008316
2009	116	0.191020	0.522331	0.161290	0.483546	0.136016	0.523304	0.147347	0.503071
2010	148	0.121093	0.332889	0.108227	0.334309	0.105950	0.333939	0.108819	0.339504
2011	145	0.100683	0.330976	0.107075	0.336821	0.105043	0.328850	0.112881	0.330994
2012	129	0.131939	0.350853	0.130284	0.348143	0.136846	0.349213	0.131322	0.341850
2013	133	0.094408	0.245914	0.099842	0.247329	0.099368	0.245884	0.097705	0.241310
2014	156	0.066777	0.315870	0.065573	0.316991	0.071543	0.330857	0.067649	0.317629
2015	145	0.095364	0.317774	0.090686	0.323175	0.091992	0.327533	0.088879	0.327673
2016	123	0.137452	0.279887	0.125594	0.291298	0.129012	0.304137	0.119130	0.296655
2017	133	0.069998	0.352835	0.087218	0.351161	0.085200	0.348611	0.084507	0.348192
2018	149	0.128576	0.369009	0.133455	0.378684	0.129892	0.378012	0.128523	0.395197
2019	157	0.145916	0.453762	0.146985	0.478614	0.145251	0.471047	0.148599	0.475450
2020	132	0.215523	0.706568	0.191028	0.703632	0.152957	0.652793	0.189598	0.718112
2021	124	0.133533	0.862595	0.151147	0.825390	0.161779	0.855097	0.162669	0.852867
2022	134	0.161190	0.620583	0.130584	0.523016	0.134406	0.525304	0.129316	0.530611
1994-2022	4837	0.111448	0.502449	0.109904	0.495551	0.109658	0.498431	0.107617	0.497386
2006-2022	2622	0.125356	0.531539	0.123030	0.503358	0.119587	0.498050	0.122898	0.510040
2010-2019	1418	0.109045	0.341319	0.109321	0.347775	0.109759	0.348230	0.108760	0.348792

This table shows summary statistics for average buy and hold cumulative abnormal returns for targeted companies for the period ± 20 trading days centered around trigger date (observation window). We show statistics using 4 distinct models: CAPM, FF3, FFM and FF5. The reference period used to compute loading consists in the 100 trading days that precedes the start of the observation window.

Table 3.5: Summary statistics: average market daily returns

	count	*daily rates*			*annualized rates*		
		value weighted	*equal weighted*	*S&P500*	*value weighted*	*equal weighted*	*S&P500*
		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
1994	110	0.000663	0.001602	0.000468	0.182	0.497	0.125
1995	858	0.001248	0.001907	0.001166	0.369	0.617	0.342
1996	2541	0.000671	0.001045	0.000685	0.184	0.301	0.189
1997	4191	0.000957	0.001277	0.000977	0.273	0.379	0.279
1998	3355	0.001171	0.000899	0.001276	0.343	0.254	0.379
1999	3124	0.000492	0.001211	0.000299	0.132	0.357	0.078
2000	3047	-0.000179	0.000517	-0.000145	-0.044	0.139	-0.036
2001	2508	-0.000262	0.001119	-0.000380	-0.064	0.326	-0.092
2002	1892	-0.000303	0.000130	-0.000352	-0.074	0.033	-0.085
2003	1771	0.001090	0.002226	0.000867	0.316	0.752	0.244
2004	1804	5.734694	0.000294	-3.727235	0.015	0.077	-0.009
2005	2475	0.000436	0.000434	0.000265	0.116	0.116	0.069
2006	2519	0.000215	0.000215	0.000187	0.056	0.056	0.048
2007	3190	0.000308	-7.514800	0.000150	0.081	-0.002	0.039
2008	2915	-0.001635	-0.001254	-0.001708	-0.338	-0.271	-0.350
2009	1463	0.001821	0.003158	0.001584	0.582	1.214	0.490
2010	1815	0.001091	0.001347	0.000929	0.316	0.404	0.264
2011	1771	-0.000155	-0.000445	-0.000124	-0.038	-0.106	-0.031
2012	1529	0.000475	0.000582	0.000385	0.127	0.158	0.102
2013	1562	0.000951	0.001109	0.000906	0.271	0.322	0.256
2014	1936	0.000351	1.992977	0.000418	0.093	0.005	0.111
2015	1793	-7.612688	-0.000295	-4.519199	-0.019	-0.072	-0.011
2016	1573	0.001082	0.001463	0.000907	0.314	0.446	0.257
2017	1694	0.000693	0.000651	0.000647	0.191	0.178	0.177
2018	1826	-0.000305	-0.000304	-0.000349	-0.074	-0.074	-0.084
2019	1892	0.000941	0.000772	0.000942	0.267	0.215	0.268
2020	1683	-0.000235	-1.352546	-0.000243	-0.058	-0.003	-0.060
2021	1573	0.000688	0.000264	0.000888	0.190	0.069	0.251
2022	1683	-0.000199	-0.000344	-0.000240	-0.049	-0.083	-0.059
1994-2022	60093	0.000366	0.000602	0.000314	0.097	0.164	0.083
2006-2022	32417	0.000257	0.000289	0.000207	0.067	0.076	0.054
2010-2019	17391	0.000492	0.000465	0.000452	0.132	0.124	0.121

This table shows summary statistics for average market daily returns during the 40 trading days centered around trigger date (observation window). We show statistics for 3 distinct references: value weighted (vwret), equal weighted (ewret) and S&P500. Columns 1-3 presents average daily abnormal returns in daily rates and columns 4-6 the rates were compounded to yearly rates.

Table 3.6: Summary statistics: ownership stakes *(% market capitalization)* by year

	count	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
1994	10	12.063	10.122	5.460	5.857	6.600	16.700	25.640
1995	69	14.016	10.383	5.196	6.030	8.700	20.700	27.120
1996	197	15.133	11.906	5.200	5.922	9.500	20.800	34.620
1997	329	15.227	11.704	5.300	6.000	10.600	19.400	35.164
1998	257	15.993	11.823	5.310	6.400	11.290	22.100	34.700
1999	244	15.972	12.098	5.292	6.285	10.650	22.222	34.970
2000	235	14.712	11.558	5.200	5.900	9.900	18.700	33.700
2001	187	15.744	11.778	5.272	6.005	11.480	21.850	33.320
2002	154	14.984	11.428	5.200	6.147	9.950	21.475	34.300
2003	136	14.536	11.087	5.225	6.087	8.450	22.225	30.150
2004	140	14.931	11.671	5.200	5.752	9.500	21.450	32.184
2005	205	16.028	12.129	5.294	6.400	10.500	23.300	35.434
2006	203	15.161	11.340	5.200	5.900	9.900	20.600	32.990
2007	262	16.192	12.295	5.200	6.200	10.700	23.567	37.970
2008	234	16.312	12.748	5.279	6.177	9.864	22.602	38.958
2009	107	15.861	11.695	5.209	5.775	11.610	21.390	34.760
2010	147	15.589	11.216	5.268	6.205	11.800	21.330	33.880
2011	141	14.334	11.084	5.100	5.700	8.700	21.300	30.000
2012	127	14.995	11.449	5.068	5.510	9.700	20.570	35.640
2013	132	15.320	11.270	5.191	5.770	9.945	21.225	31.470
2014	152	14.959	11.061	5.300	5.940	9.805	20.150	33.390
2015	142	15.293	10.445	5.300	6.000	11.500	21.550	30.760
2016	122	15.160	11.114	5.100	5.925	10.440	20.550	35.464
2017	131	14.373	11.489	5.300	5.800	8.700	20.045	30.300
2018	144	14.803	10.608	5.273	6.537	9.990	21.100	30.100
2019	153	16.681	12.188	5.340	6.380	11.600	24.400	36.320
2020	131	16.012	11.423	5.500	6.800	13.000	20.800	34.000
2021	119	15.308	11.740	5.400	6.295	9.600	21.550	32.260
2022	129	15.850	11.297	5.194	6.500	11.400	23.500	32.872
1994-2022	4739	15.414	11.600	5.208	6.030	10.000	21.400	34.240
2006-2022	2576	15.489	11.522	5.200	6.020	10.100	21.702	34.200
2010-2019	1391	15.170	11.186	5.200	5.900	9.990	21.400	32.880

This table shows summary statistics for percentage ownership for the companies in our final sample, after cleaning the dataset and keeping only events for which we can match Compustat data.

Table 3.7: Summary statistics: ownership stakes
(log dollars) by year

	count	*mean*	*std*	*10%*	*25%*	*50%*	*75%*	*90%*
1994	10	7.926	1.538	5.674	6.837	8.565	9.224	9.443
1995	68	7.032	1.812	5.235	5.702	6.760	8.567	9.260
1996	197	6.689	1.795	4.606	5.457	6.450	7.848	8.844
1997	329	6.891	1.534	5.086	5.856	6.686	7.756	8.863
1998	257	7.087	1.752	4.920	5.829	6.886	8.238	9.227
1999	243	7.234	1.773	4.982	6.012	7.132	7.998	9.399
2000	235	7.079	1.852	4.824	5.825	6.903	8.384	9.537
2001	187	7.045	1.883	4.509	5.638	6.962	8.543	9.422
2002	154	7.093	1.974	4.601	5.676	7.073	8.345	9.436
2003	136	6.625	1.708	4.348	5.481	6.567	7.773	8.862
2004	140	7.137	1.881	4.903	5.846	6.874	8.496	9.252
2005	205	7.764	1.686	5.739	6.580	7.682	8.813	9.806
2006	203	7.953	1.662	5.781	6.675	8.088	9.217	9.956
2007	262	8.143	1.800	5.857	6.766	8.137	9.182	10.480
2008	234	8.131	1.891	5.913	6.683	7.858	9.300	10.773
2009	107	7.211	1.652	5.127	6.284	7.093	8.235	9.282
2010	147	7.712	1.629	5.590	6.620	7.693	8.786	9.980
2011	141	7.744	1.819	5.391	6.499	7.523	8.988	10.129
2012	127	7.828	1.808	5.443	6.946	7.569	8.876	10.504
2013	132	8.037	2.019	5.698	6.663	7.997	9.639	10.513
2014	152	8.181	1.722	5.993	6.894	8.305	9.317	10.349
2015	142	8.575	1.757	6.470	7.501	8.442	9.858	10.531
2016	122	8.329	1.808	6.111	7.063	8.314	9.619	10.712
2017	131	7.894	1.886	5.381	6.703	7.949	9.305	10.334
2018	144	8.489	1.716	6.473	7.085	8.447	9.787	10.647
2019	153	8.379	1.694	6.094	7.105	8.406	9.533	10.652
2020	131	8.020	1.829	5.569	6.897	7.870	9.179	10.574
2021	119	8.292	1.858	5.866	6.932	8.412	9.607	10.408
2022	127	8.547	1.802	6.298	7.334	8.474	9.937	11.035
1994-2022	4739	15.414	11.600	5.208	6.030	10.000	21.400	34.240
2006-2022	2576	15.489	11.522	5.200	6.020	10.100	21.702	34.200
2010-2019	1391	15.170	11.186	5.200	5.900	9.990	21.400	32.880

This table shows summary statistics for dollar ownership (log) for the companies in our final sample, after cleaning the dataset and keeping only events for which we can match Compustat data.

3.3 Empirical results

In this section, we study how pre-trigger events, such as *mergers*, *bankruptcy*, *insiders*, and *notice of delisting*,[151] impact activists' *abnormal returns* around *trigger dates*, as well as *ownership stakes*. In addition we also distinguish how those outcomes are related to the *order*[152] and *delay*[153] between multiple filings targeting the same company. The objective here is twofold. First, we assess the effectiveness of the methodology proposed in this paper in capturing those specific episodes. By finding statistically robust coefficients, we are in good hopes of being on the right track into addressing the problems exhaustively discussed on Section 3.1.[154] In particular, as proposed in Part 3.1.3, our methodology does not rely solely on manual collection; instead it is fully documented and, thus, easily reproducible.[155] The second objective of this exercise is to evaluate the extent to which failing to remove these pre-trigger events affects usual research outcomes. We retain the flagged cases and add corresponding dummies,[156] so their coefficients provide the quantification we are looking for.

In what follows, we present tables with regression results, each displaying outcomes for one specific regressand among three: *abnormal return*, *ownership (%)*, and *ownership (log USD)*. All these tables follow the same structure, with columns covering two distinct periods: the first half corresponds to period 2006-2022 (columns 1-5), and the second, 2010-2019 (columns 6-10). We relate most of our discussion to the results obtained for the longer period (columns 1-5), since most of them are confirmed for the shorter period (columns 6-10) as well and, overall, the later hold less predictive power due to the fewer data points. However we do turn to the shorter period, whenever it provides additional insights.

[151] *Notice of delisting* includes failures to comply with Exchange rules that might have been resolved later.

[152] Whether the event is the first one targeting that company, out of at least two; or if it is a subsequent event.

[153] If a company was targeted before the event in question, did it happen within a short period (less than 6 months), or only after a longer interval?

[154] For a more targeted and concise coverage please refer to 3.1.2.9.

[155] As discussed earlier, we performed the *event identification* by consolidating multiple filings into single events (as seen in Part 3.2.2) and employing information from Edgar public data to flag pre-trigger events (3.2.3).

[156] As already noted, besides the dummies, we include control variables and/or time fixed effects, as indicated in the footer rows of each table (Y/N). Please refer to Appendix B for tables that presents all coefficients, including those of the controls.

To enhance readability, the tables presented here, in the main text, only present coefficients and standard errors (in parenthesis) for the studied dummies. Comprehensive tables, incorporating results for all controls, can be found in Appendix B. Errors in the tables are clustered by industry.

3.3.1 Abnormal returns: effect of mergers, bankruptcies, insiders

We start with Table 3.8, depicting the results for regressing *abnormal returns* as dependent variable.[157] Columns 1 to 5 covers the full period for which we have flag information, from 2006 to 2022 (see figure 3.3 for the counts of flagged events per year). Column 1 contains results from a simple regression over the unit. Hence, the coefficient therein represents a summary statistic: the raw average abnormal return for events within the interval 2006-2022 is 11.16%. When we include firm-specific controls the average return is slightly higher: 12.58% (column 2). In addition, when we also control for year fixed effects (column 3), the abnormal return for the base year 2010 is 13.56%. Columns 2 and 3 have the same controls as columns 4 and 5, respectively, with the difference being that the latter includes the dummies. The intercepts of column 2-3 do not change significantly from those on columns 4-5 . Overall, we observe that when the regression outcome is *abnormal return* (Table 3.8), those dummies that are unrelated to *duplicated records* exhibit negative coefficients that are both economically and statistically significant at the 1% level. Next we will briefly discuss the interpretation of each of these coefficients in separate.

3.3.1.1 Insider trading

As expected, the coefficients for *insider trading dummies* in Table 3.8 are negative (-9.60% and -10.53% for columns 4 and 5, respectively), indicating that these events have noticeably smaller *abnormal returns* than those initiated by non-insiders. On average, *13D events* associated to insiders yield abnormal return around 3%, over the ± 20 days around the event date.

Our main hypothesis for explaining the small abnormal returns associated with insider trading is quite intuitive: most of these transactions are likely to be linked to stock option

[157] For definitions of the outcome variables and descriptive statistics, please refer to Part 3.2.5.2.

Table 3.8: Regression: abnormal return over flags

Dependent variable: *abnormal return (CAPM), ± 20 days, t_0=event date*

	2006 to 2022					*2010 to 2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	0.1116***	0.1258***	0.1396***	0.1292**	0.1390**	0.1102***	0.1769***	0.1876***	0.2366***	0.2450***
	(0.0089)	(0.0463)	(0.0519)	(0.0607)	(0.0637)	(0.0103)	(0.0597)	(0.0615)	(0.0713)	(0.0717)
group filings flag				0.0494	0.0492				0.0370	0.0406
				(0.0423)	(0.0411)				(0.0539)	(0.0538)
multiple *(1^{st} occurrence)*				-0.0374*	-0.0338				-0.0753***	-0.0656**
				(0.0214)	(0.0214)				(0.0274)	(0.0273)
multiple *(2^{nd} within 6MO)*				-0.0182	-0.0134				-0.0547	-0.0459
				(0.0273)	(0.0269)				(0.0361)	(0.0357)
multiple *(2^{nd} after 6MO)*				0.0233	0.0245				-0.0152	-0.0152
				(0.0260)	(0.0262)				(0.0270)	(0.0268)
F4 flag				-0.0960***	-0.1053***				-0.0462	-0.0489
				(0.0369)	(0.0368)				(0.0492)	(0.0484)
merge flag				-0.1113***	-0.1025***				-0.1327***	-0.1237***
				(0.0261)	(0.0264)				(0.0340)	(0.0340)
notice of delisting flag				0.0611	0.0526				-0.0421	-0.0508
				(0.0586)	(0.0565)				(0.0469)	(0.0489)
bankruptcy flag				-0.7071***	-0.7178***				-0.5967***	-0.6032***
				(0.1658)	(0.1654)				(0.2023)	(0.1953)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
firm controls	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0144	0.0228	0.0263	0.0341	0.0000	0.0212	0.0324	0.0376	0.0474
R-squared adj.	0.0000	0.0098	0.0127	0.0190	0.0213	0.0000	0.0134	0.0191	0.0250	0.0296
number of observations	3176	2822	2822	2822	2822	1823	1633	1633	1633	1633

This table shows the coefficients and standard errors (in parenthesis) for the flags (dummies) when the dependent variable is abnormal returns with reference to CAPM. The estimation window covers 100 trading days (t-120, t-20) that precedes the observation window. The observation window spans over 40 trading days centered around the trigger date (t-20, t+20). Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table - full table is available in Appendix A). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

vesting operations.[158] We prefer the vesting hypothesis over interventions on the open market because insiders are closely monitored by regulatory bodies. Engaging in trades to capitalize on material non-public information constitutes a violation of regulations governing fair and transparent financial markets. This breach could result in various legal consequences, including lawsuits, financial penalties, and being legally barred from holding executive positions within a corporation. Additionally, there is the potential for significant damage to the insider's reputation.[159] These constraints means that insider trading disclosures are not likely to convey strong signals. However, stock option conversions might still be perceived as a sign of insiders' confidence in the company's future performance, explaining the small positive abnormal returns observed in our regressions.

We came with additional considerations that are aligned to our main hypothesis. First, many of these insider trades are well-anticipated by the market, as information about vesting schedules is public and companies often provide details about their stock option plans in public filings. Second, as insiders need to file an *SEC Form 4* within 2 days after the trigger event, by the time a *13D* is filed, the information on insider trading has already been made public.[160] Third, some of those instances might refer to marginal changes in insiders' positions, for example, from 4.8% to 5%. Although this additional investment triggers a *13D filing*, the change on the investor stake is not material.[161]

This analysis suggests that excluding entries related to insiders or, equivalently, neutralizing them by using regression dummies, is a reasonable choice. However, we acknowledge that we could draw more concrete conclusions with further investigation. While various avenues can be explored for this purpose, we would prioritize those relying solely on open-source data to align with the primary aim of our study.

[158]Where insiders convert their stock option positions upon reaching full ownership.

[159]Despite occasional insider scandals, we find it reasonable to consider that, on average, this is not the case for deals involving insiders in our sample.

[160]To address this mismatch of publication dates, one possible approach would be to adjust the intervals used for measuring returns.

[161]The position held by insiders is public information, available in materials such as proxy filings and *prior Form 4 filings*.

3.3.1.2 Mergers

Next, we analyze the marginal effects for the *mergers dummy*. The inclusion of this flag is motivated by a common practice in activist investment literature to remove cases related to a process dubbed *"M&A risk arbitrage"*.[162]

There are some limitations to identifying those cases manually. The most common manual method consists of searching, within the *13D Item 4* section, for textual references that might refer to mergers. However, it is customary for the textual content within that section to consist solely of *"boilerplate"* text: an extensive list of potential interventions a hypothetical activist might contemplate regarding an equally hypothetical targeted company. Hence, the *13D Item 4* text is often non-informative, as it lacks specifics for each particular investment case. For this reason, not only manual searches to identify *risk arbitrage* using the *13D Item 4* section will often fail, but so will automatic (as opposed to manual) text searches. In sum, searches for terms such as *M&A* or *merger* are likely to return a positive flag, even if there is no envisioning of mergers involving the targeted company.[163,164] Hence, the presence of these strings do not equate to concrete intentions of acting upon mergers.

Given the limitations of text search,[165] we choose an alternative approach that relies on merger-related proxy filings. These filings are mandatory for those companies undergoing a merger or reorganization, as shareholders need to approve merger-related matters (see 3.2.3 for further information about merger-related proxy filings).

We obtained coefficients for *merger dummies* that are both economically and statistically relevant at 1% level. The previous literature is silent on which direction mergers impacts abnormal returns, as those events are, at least, theoretically, removed. In our case, we observe *merger dummies' coefficients* that assumed negative signs: average abnormal returns, for firms

[162] In *"M&A risk arbitrage"*, investors acquire a block with the objective of influencing the company to vote favorably towards a merger, so they can profit from price convergence in case the merger is successful.

[163] *Boilerplate text*, spanning an exhaustive list of potential corporate actions that are unlikely to ever materialize in full, serves as a convenience artifact employed by legal advisors to avoid filing amendments in case the blockholder's objectives change in the future, with relation to their initial goals upon acquisition.

[164] Note that this problem is not constrained to *merger-related* instances. A useful indicator for the informativeness of *Item 4* textual content, is how broad/generic is its scope, as is often the case.

[165] In addition to *13D's Item 4*, some scholars also mention searching for news related to mergers in public or paid outlets. However, the results of these efforts are never explicitly documented; hence, they cannot be verified. While we do not consider such approach in this paper, it could potentially be included in future revised versions of the methodology presented here, provided it is adequately documented.

that have disclosed plans involving mergers, are significantly reduced, being almost neutralized under the period 2006-2022.

Although our initial motivation was to flag cases of *merger arbitrage*, there are some caveats in our strategy. Our approach does not differentiate between the latter and other alternative scenarios also associated with mergers. Aside form merger arbitrage,[166] the impact of mergers on stocks range from extremely positive in cases the perceived outcome is beneficial (e.g. synergies) to negative repercussions (e.g. financially distressed scenarios that lead to shareholder value dilution). Interestingly, it is worth noting the contrast in results for the two different intervals studied (columns 1-5 vs. columns 6-10): when we exclude periods with crises, merger-related targeted companies exhibit substantial abnormal returns. However, these returns represent approximately half of those achieved by reference targeted companies (as seen in columns 9 and 10).

Finally, there are other aspects that could be integrated into this analysis to aid in characterizing numerous types of mergers. Although addressing these considerations fall outside the scope of our work, it is worth briefly mentioning them to expose potential limitations of our findings and guide future refinements. Firstly, our *merger dummies* only indicate that merger-related activities were made public for shareholder approval. Consequently, the final outcome remains unknown in our context. It might be the case that by the time of the block acquisition, uncertainty might be resolved (i.e., either the merger has materialized or was aborted), or alternatively, it might still be unclear if it will come to a close. These considerations become more significant when we add to the fact that the dummies incorporated in our regressions refer to the 6-month interval before the activist event. This means that we might eventually capture, by chance, different alternative scenarios in our $\pm$ 20-day interval. Of course, the statistically relevant coefficients suggest that we are capturing a type of event that seems to be part of a distinct group.

[166]Though the anticipated coefficient sign in cases involving solely merger arbitrage is unclear, we find it reasonable that coefficients associated with them would not possess substantial explanatory power, given the uncertainty of the merger materializing.

3.3.1.3 Bankruptcies

Now we evaluate the marginal effect of the dummy *bankruptcy*. This *dummy* is associated to targeted companies that filed 8K's specifying liquidations, financial reorganizations and other distressed situations, under *Item 1.03. Bankruptcy of Receivership*. Indeed, as expected, the coefficients are negative with extremely large absolute values (around -70%) and statistically significant to the 1% level over the period 2006 to 2022. In conclusion, the methodology used for flagging those distressed cases seems to be very effective.

3.3.1.4 Notice of delisting and flags for multiple filings

Finally, we incorporated a group of flags related to *notice of delisting* and for *multiple filings*. Interestingly, these dummies are neither economically nor statistically significant on the 2006-2022 interval when regressing *abnormal returns*. However, the scenario changes when examining *ownership stakes*, which is the object of our study in Part 3.3.2.

3.3.1.5 Removing crisis: analysis of the 2010-2019 period

Columns 6-10 of the table 3.8, refer to a shorter period that excludes crisis (2008 financial crisis and the pandemics), narrowing the studied period to 10 years, from 2010 to 2019. Interestingly, excluding crisis years, the raw coefficient (column 6) is barely the same of the one observed for the full period (column 1). However, once we include firm-specific effects, the average *abnormal returns* around *trigger dates* increase considerably in the non-crisis sample, to around 17.5% (column 7) when compared to 12.58% (column 2). In addition, when we include *insider*, *merger*, and *bankruptcy* flags (equivalent to shutting down these channels), *abnormal returns* jump to the 24% area (columns 9-10), from a much less pronounced excess return, in the 13% area (columns 4-5).

Notice, though, that, for this short time span, the *insider flag* is less pronounced and not statistically significant. This comes as no surprise: as, we have previously mentioned, most cases for *insiders* were eliminated based on coincident addresses for *filer/target* and this fact compounds to the lower predictive power for this reduced time span.

3.3.2 Ownership stakes: effect of multiple filings

In this section, we analyse the regression results when the dependent variable is *ownership stake*. Our main interest is on understanding how biases in *ownership stakes* can be captured by dummies associated with multiple filings. As explained in Part 3.2.2, these dummies indicate either single events derived from consolidating various filings; or multiple events targeting the same company. In the latter case, specific dummies differentiate sequential events based on the *order* of appearance and the *time lag* between them.

We conducted regressions for two representations of ownership stakes: *percentage* and *log dollar*. Table 3.9 presents regression results for *percentage ownership*, while Table 3.10 displays results for *ownership in log dollars*.

As shown in 3.2.5.2, these two ownership representations, *percentage* and *log dollar*, exhibit distributions with distinct characteristics. Notably, *percentage ownership* is constrained between 0% and 100%.[167] In practice, for our sample, the bounded interval is likely to be even narrower, with stakes for non-controlling activist blockholders falling within the range of 5% to less than 50% ownership. On the other hand, ownership in *log dollars* is the result of multiplying the (bounded) percentage variable by market capitalization (a variable virtually unbounded to the right) and subsequently applying a log transformation. Hence, though *log dollars* do carry a component from a constrained interval, these transformations mitigate its impact, providing regression results with more robust coefficients. Hence, although the findings in both tables seem to be consistent, in the subsequent analysis, we reference our discussion to the robust results for the *log dollars* regression exercise (Table 3.10).

Next, we discuss the obtained results for each specific dummy. We follow the same practice adopted in Part 3.3.1, referencing our discussion to the period 2006-2022 (columns 1-5). Furthermore, considering the current analysis involves dependent variables not expressed in constant dollars, we specifically concentrate on the results of regressions with time fixed effects (columns 3 and 5). The omitted year-variable is 2010, so the dummies on columns 3 and 5 refer

[167] While in practice, short positions would typically represent negative values, and long leveraged positions might correspond to percentages above the unit, this does not appear to be the case for the filings studied here. However, with the update in the regulation enacted in 2023, it is advisable to review if future cases would require different treatment.

to that year reference. Starting from column 3, which includes company-level controls and time fixed effects, the intercept indicates that, on average, dollar stakes are \$3 million dollars, for the year 2010. With the same controls, but now adding the dummies, these average stakes raise to \$5.2 million (column 5).[168]

As it will become evident up ahead, all dummies that exhibited statistically significant regression coefficients when dependent variable was *abnormal returns* no longer hold significance when we switch to *ownership stakes*. On the flip side, the dummies' coefficients irrelevant for abnormal returns are now statistically significant at the 1% level.

3.3.2.1 Filings consolidation

The coefficient of the *group filings flag* (column 5 on Table 3.10) indicate that events derived from the consolidation of multiple filings are substantially larger, approximately 2.6 times greater,[169] than those obtained from a single filing. For a specific numerical example, in the year 2010, events obtained through the consolidation of filings had an average size of USD 13.7 million, while the reference case was USD 5.2 million.

We propose two plausible explanations, likely to coexist in our dataset, to justify these larger stakes. Firstly, it is credible that when various investors participate in deals as a group, the corresponding total stakes are higher. It is important to note that we are not implying causality here. Larger stakes can either exist because higher stakes demand involvement from various pockets, or because when multiple pockets are engaged, it is likely that they will target a larger total stake. Therefore, the positive coefficient captures this almost mechanic aspect of deals filled by groups in separate filings.

The second explanation relates to a potential methodological problem: we cannot dismiss the possibility of double counting or incorrectly aggregating multiple filings that should not be consolidated. This may well be the case, given the significant size of the coefficient, 2.5x. In this scenario, the coefficient captures biases incorporated during data collection. Once again, as with other conclusions drawn from our dummies, we refrain from going deeper into distinguishing

[168] Intercepts were converted from logarithms to their respective approximate dollar values.
[169] That is the result of the anti-log of the dummy coefficient.

Table 3.9: Regression: ownership stake (percentage) over flags

Dependent variable: *ownership stake (% market capitalization)*

	2006 to 2022					*2010 to 2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	21.9690***	20.5037***	20.3117***	18.5262***	18.1570***	21.2209***	19.1846***	19.3299***	17.3499***	17.2861***
	(0.3863)	(2.2466)	(2.5831)	(2.2380)	(2.4280)	(0.4939)	(3.0587)	(3.2860)	(3.3186)	(3.4507)
group filings flag				34.1661***	34.2527***				32.0382***	32.2170***
				(3.1643)	(3.1722)				(4.3744)	(4.3600)
multiple *(1^{st} occurrence)*				0.0597	-0.2878				-0.1240	-0.2588
				(1.0836)	(1.0855)				(1.3598)	(1.3702)
multiple *(2^{nd} within 6MO)*				-2.9853**	-3.2521***				-3.8797**	-4.0086**
				(1.2008)	(1.2149)				(1.6008)	(1.6312)
multiple *(2^{nd} after 6MO)*				-1.1136	-1.2865				-0.1620	-0.3737
				(0.8639)	(0.8709)				(1.1231)	(1.1453)
F4 flag				2.5026	2.5440				3.3019	3.2896
				(2.2914)	(2.3109)				(3.4569)	(3.4495)
merge flag				-0.9368	-0.9805				1.0344	1.0524
				(1.8866)	(1.8604)				(2.3548)	(2.3169)
notice of delisting flag				7.9896***	8.0144***				8.1523***	8.2161***
				(1.9585)	(1.9582)				(2.8994)	(2.8881)
bankruptcy flag				-4.7960	-5.1522				-16.7341***	-17.5655***
				(14.0871)	(13.8963)				(3.7919)	(3.8059)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
firm controls	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	-0.0000	0.0155	0.0212	0.1260	0.1317	-0.0000	0.0166	0.0203	0.1209	0.1254
R-squared adj.	-0.0000	0.0109	0.0110	0.1194	0.1202	-0.0000	0.0087	0.0069	0.1094	0.1090
number of observations	3176	2822	2822	2822	2822	1823	1633	1633	1633	1633

This table shows the coefficients and standard errors (in parenthesis) for the flags (dummies) when the dependent variable is the natural logarithm of ownership stake in dollars. Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table - full table is available in Appendix A). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

Table 3.10: Regression: logarithm of ownership stake (dollars) over flags

Dependent variable: *dollar ownership stake (natural logarithm)*

	2006 to 2022					*2010 to 2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	8.3043***	8.5032***	8.2838***	8.8751***	8.6100***	8.3444***	8.6720***	8.4122***	9.0790***	8.8020***
	(0.0345)	(0.1443)	(0.2037)	(0.1468)	(0.2047)	(0.0451)	(0.1455)	(0.2085)	(0.1641)	(0.2195)
group filings flag				0.9850***	0.9585***				0.9559***	0.9309***
				(0.1652)	(0.1620)				(0.2179)	(0.2125)
multiple *(1st occurrence)*				-0.4589***	-0.4189***				-0.4147***	-0.3946***
				(0.0882)	(0.0898)				(0.1137)	(0.1162)
multiple *(2nd within 6MO)*				-0.6118***	-0.5794***				-0.8525***	-0.8409***
				(0.1126)	(0.1138)				(0.1369)	(0.1402)
multiple *(2nd after 6MO)*				-0.4700***	-0.4720***				-0.4683***	-0.4819***
				(0.0686)	(0.0681)				(0.0879)	(0.0885)
F4 flag				-0.0777	-0.1269				-0.0668	-0.0897
				(0.2001)	(0.1980)				(0.2773)	(0.2752)
merge flag				-0.1153	-0.0623				-0.0410	-0.0183
				(0.1254)	(0.1243)				(0.1761)	(0.1707)
notice of delisting flag				-0.3452***	-0.3428***				-0.6504***	-0.6374***
				(0.1248)	(0.1227)				(0.1723)	(0.1717)
bankruptcy flag				-0.5757	-0.6053				-0.4987	-0.5849
				(0.7259)	(0.7083)				(0.7711)	(0.7811)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
firm controls	N	Y	Y	Y	Y	N	Y	Y	Y	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.3225	0.3348	0.3480	0.3586	0.0000	0.3274	0.3390	0.3623	0.3729
R-squared adj.	0.0000	0.3197	0.3281	0.3433	0.3503	0.0000	0.3224	0.3304	0.3544	0.3616
number of observations	3166	2822	2822	2822	2822	1823	1633	1633	1633	1633

This table shows the coefficients and standard errors (in parenthesis) for the flags (dummies) when the dependent variable is the natural logarithm of ownership stake in dollars. Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Columns 1 and column 6 are regressions over the constant only. The other columns include firm-specific controls (omitted from the table - full table is available in Appendix A). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).

between these cases, leaving this for future research. Regardless, in the absence of a clear conclusion, it seems reasonable to keep this flag in regressions, allowing us to analyze the obtained coefficient and decide whether to incorporate or ignore it.

3.3.2.2 Multiple sequential events

We collectively term as *sequential events* any two or more events that targets the same company, during the timeframe covered by our sample. Within *sequential events*, we distinguish three dummies, based on their order of appearance and the time lag between them. The coefficients for these dummies are all statistically significant at the 1% level when the dependent variable is *ownership stake in log dollars*. Specifically, activists spend less dollars on companies targeted more then once, when compared to those targeted only a single time: coefficients' negative signs are observed irrespective of the order in the sequence or the time interval between events. The approximate marginal effects for companies targeted multiple times during the 2006-2022 interval (see column 5 of Table 3.10), with time fixed effects are as follows:[170]

- First-time a company is targeted: stakes are 34.3% smaller than the reference case.[171]
- Subsequent events:
  - Within an interval $\leq$ 6 months: stakes are 45.2% smaller than the reference case.
  - After $>$ 6 months: stakes are 37.7% smaller than the reference case.

Overall, these numbers indicate that investing in companies targeted multiple times on average corresponds to smaller dollar stakes compared to those targeted only once, regardless of the order (first-time investing or subsequent) and the time span between interventions. Notably, the first investor commits larger dollar stakes than the subsequent, in particular when the second event occurs within a short interval.

We have some hypotheses that are consistent with these observations. First, not very

[170] The numbers on the list were calculated by subtracting from one (100%) the coefficients of column 5 converted to percentages ($e^{-0.42} = 0.657$, $e^{-0.6} = 0.548$, $e^{-0.47} = 0.624$). The reference case corresponds to events related to single-targeted companies. For the year 2010 (omitted year-dummy), the reference stake is USD 5.2 million. Consequently, for the same year, the corresponding average stakes are as follows for first-time targeted, targeted subsequently ($\leq$ 6 months), targeted subsequently (>6 months): USD 3.4 mm, USD 2.8 mm, USD 3.2 mm , respectively.

[171] To clarify, the *reference case* equates to events corresponding to single-targeted companies. The events in the list, on the contrary, relates to companies that figured more than once on our sample.

excitingly, we can once again resort, at least partially, to a mechanical explanation:[172] investors assuming an activist role are not likely to decrease their positions within a short period, thus limiting the volume accessible on the market that can be accumulated as a block. As a consequence, for multiple blocks to be available to (multiple) activist investors, particularly if the interval between events is small ($\leq$ 6 months), each individual event is likely to correspond to smaller stakes. For longer intervals, the mechanical impact tends to diminish. Over the extended term, some activists will have exited their positions, potentially making them available for acquisition by other activists.

Industry specific (non-mechanical) explanations sound more interesting. We hypothesise that *sequential events' coefficients* are capturing a combination of two factors: some *trend setting* and *block selling efforts*. *Trend setting* relates to a larger investor, a leader, that, upon buying a stock, is followed by smaller-scale investors who typically track their trades. This dynamic is observed in *wolf packs* but extends beyond that context. Note that though a significant portion of the trading associated with *wolf packs* may not be captured by *13D filings*, there should still be a handful of activist followers, though not many, that reach the 5% threshold.

The other factor that complements our hypothesis, *block selling efforts*, is compatible with the absence of relative differences in abnormal returns. These *efforts* consists of a substantial block being sold in the secondary market or, perhaps less likely, in the primary market.[173] In either scenario, the block would end up being dismembered, ultimately landing in the hands of more than one investor. This hypothesis is still compatible with *trend setters*, as block trades and/or private placements of less well-known companies often feature an anchor investor who is typically prominent, equipped with a reputable analysis team and an infrastructure both for evaluating such deals and later for acting to influence companies' business. Once these investor approves/acquires their stake, other satellite, smaller investors follow,[174] essentially free riding on the *trend setter* expertise.

Notice that we are making a subtle distinction between investors actively bidding for a

[172] Though this correspondence is not as direct for dollar stakes as it is for percentage stakes.

[173] Given the small dollar stakes, if in the primary market, this would likely be a private placement, exempt from distribution registration.

[174] It is reasonable for settlement dates to vary in such cases. The buyer and seller will reach an agreement in terms of the buyer's cash availability once the deal is approved, and constraints of the seller (who, depending on the demand, would not accept the risk of no settlement for an extended period) would come into play.

position and those evaluating the purchase of offered blocks. The former is typically how the scenario for *wolf packs* is presented in activist investor literature. However, *centralized selling efforts*, as just described, are also plausible scenarios for leaders to attract *wolf packs*. The difference that matters for our context is that *bids* drive prices upwards, leading to abnormal returns in the short run. On the other hand, when considering *centralized block offers* that end up being dismembered, the ask price is likely to remain relatively steady for both initial and subsequent buyers. Though our sample must include instances corresponding to both scenarios, we believe the later is a bit more consistent to the results we obtained. At least, this interpretation aligns with *not* finding any discernible relationship of *abnormal returns* with *sequential events dummies*.[175] Finally, even though *centralized selling efforts* may occur over extended periods, if events targeting the same company are more than one year apart, they are highly likely to be completely independent of each other.

We continue the discussion of *sequential events dummies* by addressing a potential methodological concern. Though these variables are tied to *sequential events*, our sample in columns 1-5 of Table 3.10 was not designed for capturing this type of data adequately. To illustrate this point, envision a hypothetical scenario in which a block is being actively promoted in the secondary market. Suppose an investor, after assessing the deal's conditions, approves the transaction and, following settlement, discloses activist intentions through a *13D filing*. If this example occurs towards the very end of 2022, we will miss subsequent deals related to the same block on later dates simply because they have not yet occurred.[176] Hence, we will fail to identify subsequent deals targeting the same company. This will result in all the *multiple events dummies* corresponding to this instance being assigned the value zero, whereas, under our research design, the appropriate value for the *first time targeted dummy* should be one.

This concern is particularly relevant for sequential events within larger intervals. However, even for shorter intervals, such as 6 months, it is prudent to cap the sample up to, in our case, 2021 at the most. This cap would provide an additional one year of observations to identify

[175] The lack of supporting evidence does not allow us to dismiss the presence of a relationship; our acknowledgment is limited to recognizing that, for the time being, we have been unable to establish one.

[176] Challenges of this nature are inherent in cohort analysis, a phenomenon notably observed in actuarial fields (e.g., mortality tables) and credit analysis (e.g., defaults). Take, for instance, credit risk analysis, where cohorts needs to be tracked over a substantial length of time to observe the defaults related to more seasoned contracts.

those filings that were targeted at least up to a second time. Consequently, we refer to the results for the shorter period (columns 6-10) for a robustness check, as its time span is even more conservatively capped (two years). Specifically, when observing columns 9-10 and comparing them to columns 4-5 for the longer and potentially methodologically flawed interval, the results for the capped sample are entirely consistent. The signs remain stable, the absolute size of the coefficients are similar, and the statistical significance observed for *sequential events dummies* in the long-period sample is maintained at the 1% level in the capped sample as well.

Finally, we conclude with some suggestions for further refining this methodology concerning *multiple events dummies*. The primary interventions would be to (1) incorporate additional dummies,[177] (2) introduce variables that interact dummies with other controls, (3) replace categorical dummies with the interaction of a numerical variable measuring the time between events targeting the same company and their order of appearance. Additionally, we could gather supplementary public information, such as registration (public distribution), exemptions from registration (private placements), and 13D amendments,[178] to identify selling efforts or blocks changing hands that would provide stronger support for our hypothesis.

3.3.2.3 Non-13D filings

Finally, there is a dummy we have incorporated from non-13D Edgar filings, *notice of delisting.*. Events with this flag, corresponds to targeted companies that have received a notice of non-compliance regarding listing requirements on stock exchange.[179] These dummy coefficients are statistically significant at 1% level both for *dollar stakes* as well as for *percentages*. However the coefficients' signs are diverge: While dollar stakes are 30% smaller (), the percentage ownership

reveals large percentage ownership of these non-exchange-compliant companies. There are

[177]For instance, these could include the effect of pairs of events targeting the same companies, subdividing the dummy corresponding to the first intervention to differentiate with respect to the later interval of the next event, or adding more dummies to represent alternative intervals longer than 6 months. Regarding the latter, if implemented, careful consideration should be given to capping the sample to ensure suitable 'future' observations characterizing companies targeted more than once.

[178]Especially those final amendments filed when the investor no longer holds any position with that specific stock

[179]For example, this may refer to missed filing a regulatory document or the stock price going bellow the exchange set threshold)

some possible alternatives to explain this coefficient. First, these non-compliant companies tend to have smaller market caps, which in turn is correlated to larger stakes. Furthermore, these might be companies going through difficulties that end up being targeted by activists envisioning to conduct a turnaround, that demands with larger stakes. We will return to that fact next when we analyze log dollar stakes.

The dummy for *dup permno first* flags the first event of a company that is followed by other subsequent events (no matter when). We have hypothesized that multiple events for the same company might be of two categories. The first category includes companies that are targeted multiple times, notably by different blockholders (either short time as in *wolf pack* activism mentioned in Brav, Dasgupta, and Mathews (2022), or after a long time span. This last case includes the exit of the first activist if he sells his position to another activist blockholder or, if at some point further in the future (in special for long-lived companies), it is an event without any relationship to the first activist event. The second category includes private placements, that could be started either by the issuer or by any distributor selling a large stake that ends up being placed among different investors. Private placements often are offered at discount to gather interest of investors, because investment companies have limited resources to evaluate all private offers they receive. They are most likely to include an offer in their busy analysis pipeline if there is some discount from the current price (in many cases this is a *sine qua non* condition). The financial settlement need not to be done at the same date for different investors. For large stakes, typically, the distributor will set up a date that is convenient to the buyer upon the approval of the investment within a reasonable time window. Hence, we should see some private placements that end up being distributed for different activist blockholders settled some days, or even months, apart in our dataset.

The negative sign on *dup permno first* dummy (statistically significant) suggests that we are capturing some events that are indeed private placements distributed for multiple investors. Returns becomes negative when the first block is sold, due to the discount being reflected on the settlement price. The discounts are around 6%-8% for the first settlement. The problem with this interpretation is that in order to confirm it, we should find evidence that subsequent block acquisitions of for the same *targeted company* should not present excess returns because the price has been already reset to a lower level to incorporate the discount of the private placement offer. However, this the dummy that could evidence this pattern, *dup permno 6MO*, is economically small (does not offset the regression constant) and it is not statistically significant. So at first blush we could interpret that *dup permno first* is capturing something else. Actually a careful analysis of how *abnormal returns* are computed in our dataset vis-a-vis the timing of private placement settlements for each investor shows that insignificant *dup permno 6MO* is not inconsistent with our hypothesis.

As seen in the analysis above, we gathered evidence that our procedure to flag *non-core events* seems to be doing their job, identifying events that should not be included in the sample and helping quantify the effect on shareholder activism results on non-core cases.

3.4 Conclusion

In this paper, we explore the process of acquiring a research-ready dataset for studying activist events using SC 13D filings. Our motivation comes from the fact that despite blockholder activism being a well-established field in Corporate Finance, there is a lack of clear documentation that enables the replication or updating of datasets. This limitation hinders *full* reproducibility of research results, leads to the reuse of outdated datasets, creates barriers for new researchers, and obscures opportunities to refine data collection methods.

Rather than solely presenting isolated practical solutions, we construct our approach on top of two foundational pillars: understanding of how the corresponding regulatory raw data is generated/stored and outlining the role of industry-specific elements that hold implications for the data extraction process. We start with a thorough examination of the data bundle generation process within Edgar, which combines *data input by the filer* and *system-generated metadata* upon filing submission. This examination lead us to clearly distinguish between structured data readily available within the bundle (i.e. *filing date*, *filing/target company*) and data that can only be obtained by parsing the raw text of the *submitted filing document* (i.e. *security identification*, *event date*, *ownership stake*).

We then characterize two distinct phases in dataset construction: *parsing* and *event identi-*

fication. In the context of *parsing*, we draw from legacy algorithms to provide examples that demonstrate how biases and errors can inadvertently be introduced into the resulting datasets. We then propose ways to mitigate these shortcomings. Concerning *event identification*, we propose a methodology grounded in Edgar information extracted from other filings (non-13Ds) and patterns discerned within multiple filings with coinciding targeted company.

At first glance, the availability of both legacy *parsing* scripts and activists datasets, some of which are even accessible online, might suggest that developing new scripts for extracting *13D events* is a redundant and potentially wasteful effort, especially when considering the sizable time and resources it demands when done comprehensively. However, as we demonstrated in our study, such efforts are warranted. Regarding legacy algorithms that one might eventually be tempted to use, we demonstrate that they fall short of being an alternative for serious work. Concerning existing datasets, while scholars may have addressed at least some of the challenges outlined in this paper, in their previous research, the entirety of their insights and operational procedures have largely remained undisclosed to the public. This lack of transparency makes it challenging to assess whether these datasets are reliable, or to adapt them for specific applications. Updating pre-existing datasets with new data points based on the same methodology is unfeasible, as authors often claim their datasets result from manual collection procedures. Though some references are provided for certain interventions or approaches, these are far from comprehensive in incorporating all the necessary steps to assemble the dataset. Moreover, at the time of writing this paper, there was only one activist dataset that became public in the wake of the publication of Dlugosz (2006), and it has not been updated since then. Concerning other potential data sources, they are either fee-based commercial alternatives or the property of scholars that are not currently in the public domain. Even in the latter cases, they can benefit from the methodology proposed in this paper.

Our work challenges the perception that commercial alternatives are *convenient*, purportedly freeing up researchers' time from data acquisition. In line with existing literature that uncovered inconsistencies in commercially available sets (Anderson and Lee (1997), Dlugosz et al. (2006)), we demonstrate that they are not only costly but also constrained from a research perspective. Before it can be appropriately employed as suitable, research-ready data, these commercial sets still need to undergo additional work, particularly with respect to *event identification* – a task that is unavoidable and at least as relevant as the *parsing* process itself. However, the simple approach to *event identification* outlined in this Appendix often becomes more laborious or even infeasible, when dealing with pre-manipulated sets, undermining the *convenience* narrative. In conclusion, what appears to be the more convenient solution, instead, is to eliminate dependency on commercial data altogether. Our work contributes to diminishing that dependency, ultimately fostering reproducibility, validation, and collaboration.

As regulations evolve, a portion of the discussion presented in this Appendix will become obsolete for those events filed under updated *rules* that mandate machine-readable *13D main filings* submission. The change in regulation will represent a significant advancement that will, unarguably, facilitate data *parsing*. However, the challenges in creating a *13D events* dataset relate not only to *parsing* content but also, and no less important, to identifying *core events*. So, while *parsing challenges*, which are in great part due to the non-standardized nature of the data, will cease to exist once these regulations are implemented, issues concerning the identification of *core events* are likely to persist. Hence, the insights provided here will continue to serve as a guide for identifying *core/non-core events* as well as for aggregating multiple filings into single events.

In summary, the transition to machine-readable filings submission will not diminish the significance of the work presented here for two main reasons. First, the permanence of historical data warrants comprehensive documentation of both its access and transformation. After all, though the legacy format of historical raw data archives is invariable, the underlying data that ultimately is the *raison d'être* for those records may not have been adequately decoded. Devising open-source, well-documented scripted rules unleashes the potential of collective efforts, promoting efficiency by avoiding redundant work. Refinements are built upon each other, resulting in an exponential effect, whether targeting mild updates or more relevant corrections. Secondly, the need for event identification persists, whether one decides to extract the dataset independently – irrespective of machine readability (historical archives or data shaped under new regulation mandates) – or opts for any pre-compiled dataset, whether commercial or not.

Our primary contribution is a methodology that covers the entire workflow to retrieve a research-ready database of activist events, encompassing the extraction of features from SC 13D filings, the consolidation of filings into *single events*, and the identification of *non-core events*. We implement both the *parser* and *event identification* steps, which are outlined in the main body of the paper. In Appendix A, we provide comprehensive discussion and documentation to facilitate replication, covering practical aspects, including the handling of large datasets. By using freely available Edgar public data and firm/market-specific datasets from Center for Research in Security Prices (CRSP) and Compustat hat requires only the basic subscription of WRDS, we not only reduce the cost of data acquisition but also enhance research reliability, minimize the need to rely on outdated datasets and increase conditions for potential collaboration.

Throughout our study, we have secondary outcomes, and in here we mention two of them. Firstly, we provide updated descriptive statistics, till end of 2022, including a comparison of the annual distribution of fundamental data for targeted companies with that of the CRSP universe. This analysis illustrates how certain fundamental aspects associated with targeted companies have evolved in more recent sub-periods. Secondly, we delve into the analysis of the impact of including *non-core* blockholder activist events, such as those involving *mergers*, *insider trading*, and *bankruptcies*, on *short-term abnormal returns* and *ownership stakes*. Our findings suggest that previous studies may have underestimated *abnormal returns*.

In conclusion, our paper presents a comprehensive examination of the challenges associated with obtaining an activist events dataset. It offers practical implementation considerations and underscores the impact of using inadequate datasets on conventional research outcomes. Our methodology serves as a practical guide for dataset extraction, relying predominantly on publicly available (either free or low cost) data sources that can be effortlessly replicated. This is particularly significant as it not only provides a replicable methodology for reproducing these datasets but also serves as a valuable resource for researchers who may need to assess and augment pre-compiled datasets with *event identification*.

Finally, our contribution extends beyond the immediate scope of our study, aligning with the broader realm of reproducibility in scientific research and providing insights into the transparency and suitability of commercial databases. While we are aware of the limitations inherent in our methodology, particularly the resource-intensive nature of obtaining a higher-quality dataset, we consider it as an initial stepping stone. We invite collaboration from fellow researchers interested in blockholder research, with the shared objective of refining and improving the procedure outlined here.

References

Anderson, R. C., & Lee, D. S. (1997). Ownership Studies: The Data Source Does Matter. *Journal of Financial and Quantitative Analysis*, *32*. https://www.jstor.org/stable/2331202

Barry, J., Brav, A., & Jiang, W. (2020). Hedge Fund Activism Updated Tables and Figures. *Working paper*. https://people.duke.edu/~brav/HFactivismMarch2019.pdf

Bebchuk, L. A., Brav, A., Jackson, R. J., & Jiang, W. (2013). Pre-Disclosure Accumulations by Activist Investors: Evidence and Policy. *Journal of Corporation Law*, *39*, 1–34. http://www.sec.gov/news/speech

Brav, A., Dasgupta, A., & Mathews, R. D. (2022). Wolf Pack Activism. *Management Science*, *68*, 5557–6354. https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2021.4131

Brav, A., Jiang, W., & Kim, H. (2010). Hedge Fund Activism: A Review. *Foundations and Trends® in Finance*, *4*, 185–246. https://www.nowpublishers.com/article/Details/FIN-026

Brav, A., Jiang, W., & Kim, H. (2015). The Real Effects of Hedge Fund Activism: Productivity, Asset Allocation, and Labor Outcomes. *Review of Financial Studies*, *28*, 2723–2769. https://doi.org/10.1093/rfs/hhv037

Brav, A., Jiang, W., & Li, R. (2022). Governance by Persuasion: Hedge Fund Activism and Market-based Shareholder Influence. *ECGI working paper*. https://papers.ssrn.com/sol3/papers.cfm?abstractid=3955116

Brav, A., Jiang, W., Ma, S., & Tian, X. (2018). How Does Hedge Fund Activism Reshape Corporate Innovation? *Journal of Financial Economics*, *130*, 237–264. https://doi.org/10.1016/j.jfineco.2018.06.012

Brav, A., Jiang, W., Partnoy, F., & Thomas, R. (2008). Hedge Fund Activism, Corporate Governance, and Firm Performance. *Journal of Finance*, *63*, 1729–1775. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.2008.01373.x

Coffee, J. C., & Palia, D. (2015). The Wolf at the Door: The Impact of Hedge Fund Activism on Corporate Governance. *Columbia Law and Economics Working Paper*. https://papers.ssrn.com/sol3/papers.cfm?abstractid=2656325

Dlugosz, J., Fahlenbrach, R., Gompers, P., & Metrick, A. (2006). Large Blocks of Stock: Prevalence, Size, and Measurement. *Journal of Corporate Finance*, *12*, 594–618. https://doi.org/10.1016/j.jcorpfin.2005.04.002

Edmans, A. (2014). Blockholders and Corporate Governance. *Annual Review of Financial Economics*, *6*, 23–50. https://www.annualreviews.org/doi/abs/10.1146/annurev-financial-110613-034455

Gantchev, N. (2013). The costs of shareholder activism: Evidence from a sequential decision model. *Journal of Financial Economics*, *107*, 610–631. https://doi.org/10.1016/j.jfineco.2012.09.007

Gantchev, N., Gredil, O. R., & Jotikasthira, C. (2019). Governance under the Gun: Spillover Effects of Hedge Fund Activism. *Review of Finance*, *23*, 1031–1068. https://academic.oup.com/rof/article/23/6/1031/5184904

Gillan, S., & Starks, L. (2007). The Evolution of Shareholder Activism in the United States. *Journal of Applied Corporate Finance*, *19*, 55–73. https://onlinelibrary.wiley.com/doi/10.1111/j.1745-6622.2007.00125.x

Klein, A., & Zur, E. (2009). Entrepreneurial Shareholder Activism: Hedge Funds and Other Private Investors. *Journal of Finance*, *64*, 187–229. https://doi.org/10.1111/j.1540-6261.2008.01432.x

Lilienfeld-Toal, U., & Schnitzler, J. (2020). The anatomy of block accumulations by activist shareholders. *Journal of Corporate Finance*, *62*. https://www.sciencedirect.com/science/article/abs/pii/S092911992030064X

U.S. SEC. (2009). Questions and Answers of General Applicability for Exchange Act Sections 13(d) and 13(g) and Regulation 13D-G Beneficial Ownership Reporting (2009, November 16). http://www.sec.gov/divisions/corpfin/guidance/reg13d-interp.htm

U.S. SEC. (2023). Modernization of Beneficial Ownership Reporting - Final Rule and guidance. https://www.sec.gov/files/rules/final/2023/33-11253.pdf

U.S. SEC. (n.d.). Rules and Regulations Under the Securities Exchange Act of 1934. https://www.ecfr.gov/current/title-17/chapter-II/part-240/subpart-A

Appendix A

Enhancing Transparency and Reproducibility in *13D filings* Data Extraction

The central theme of this Appendix is the characterization of the main steps and components required for implementing a script that assembles *activist events* data, primarily sourced from SEC Edgar. The detailed, and sometimes operational-oriented, content complements the insights given in the main paper, serving as a useful companion for the development of such scripts. The exposition presented here also explores practical subpar outcomes, illustrating common pitfalls and constraints that arise from solely relying on *13D filings documents*. We show that such simplistic approach fail to effectively inform either about events that should be consolidated or those that should be excluded. Therefore, this Appendix can also assist readers in recognizing the problems likely to be encountered, to some degree, in legacy studies and datasets.

Though we dedicate a substantial section to the implementation of a rule-based parser,[1] this fact is not overly restrictive on its scope of applicability. The parser is just, currently,[2] a necessary element of a larger analytical body. The material presented here has broad application, as it contains information that holds significance beyond any of its constituent parts. Moreover, while we build upon a concrete implementation to aid in our exposition, this is without loss

[1]Necessary for extracting specific data elements currently found exclusively in semi-structured, non-standardized content.

[2]It will soon, hopefully, be obsolete for new data entries, once new regulation is enacted.

of generality: the foundational concepts used to craft it still assist those seeking to develop solutions using other paradigms, such as those interested in alternative advanced document understanding (ADU) approaches. ADU has advanced significantly in recent years, capitalizing on progress in optical character recognition (OCR), machine learning, artificial intelligence (AI), and natural language processing (NLP). However, its effectiveness depends on contextual understanding and industry-specific knowledge – precisely the contributions offered in this document: industry insights that enable technical implementation.

A.1 Data

We exclusively rely on publicly available sources through the SEC Edgar API,[3] in combination with market and firm-specific data obtained from CRSP[4] and Compustat,[5] respectively, both accessible through a basic WRDS[6] subscription. Our strategy to assemble an *activist investor events* dataset consists of five main steps, as follows:

1. Get *crawler.idx* and/or *submissions.zip* files.
2. Filter data to narrow down for certain *filer* categories or filing types. Among parsed features, get the URL that points to *index.htm* files.
3. Request the *index.htm* files using the URLs obtained in *step 2* and then extract the contents of the retrieved object. These files contain information about both *filers* and *target* companies, which is necessary to help identify and subsequently eliminate duplicated records[7] (as explained in Section A.3.3). Furthermore, these files contain the path segments to construct the URLs that lead to the *filing documents*, which will be used in the next step.

[3]SEC Edgar API is provided by the SEC for accessing data and documents filed by companies with the SEC through its Edgar system.
[4]CRSP is a database offering historical stock price and market data for research and analysis, linked to the University of Chicago Booth School of Business.
[5]Compustat is a database, owned by S&P Global, that contains data from financial reports of publicly traded companies that have been standardized.
[6]WRDS is a web-based platform that provides access to financial, economic, and accounting datasets that are the standard for academic research in Finance and Economics, such as CRSP and Compustat.
[7]*Target company* is referred to as *subject company* in the *index.htm* files.

4. Request *filing documents*[8] using the URL obtained in the previous *step* and then parse the data to obtain *event date*, *investment stakes* and *investment objectives* from the *13D filings*.

5. Clean and filter data; convert filings into events and identify *non-core events*.

Although some basic *data elements* from filing records can be easily obtained in *steps 2* and *3*,[9] such as *company's name, CIK*,[10] *address*, or *filing date*; others, such as *ownership stakes* or *event dates*, require executing the entire crawling sequence, up to *step 4*: obtain URL to *index.htm*; from its response, obtain URL to the *filing document*; request the *filing document* and scrape contents from the last response object. Extra effort is required to obtain *informational elements* of the second category because this information is only available in the *main filing document*, which, in the case of *SC 13Ds*, is a semi-structured non-standardized text document. In contrast, those *elements* that are easy to retrieve are part of the structured *information bundle*, that is composed from data either entered by *filer* upon filing submission or generated as metadata.

There are two main challenges in obtaining an *activist events* database from raw data. The first challenge is *parsing* (the last part of *step 3*), as the data is presented in a non-standardized semi-structured format. For *filing types* that are submitted in XBRL[11] (e.g. *10K*), parsing is straightforward. However, for those that are not, as *Schedule 13D*, creating a effective parser can be time-consuming due to the need to address various special cases and adapt to changes introduced by the *filer*.

The second challenge involves transforming individual filings into coherent events and distinguishing *non-core events*, which we conveninently term concisely as *event identification*.

[8] In this study, where our aim is to construct a database for activist events, the only *filing document* (the text file containing the actual document) we retrieve is the one referring to *13D filings*. The complementary information used for identification of *non-core events* that is derived from other filing types (i.e. *8K*, *Form 4*) is obtained directly either from *index.htm* files or from the files contained in the *submissions.zip* files—an approach that is effective and easy to implement, as that data in it conveniently presented in structural, consistent, form.

[9] These steps are straightforward due to the structured nature of the data (fixed-width format files) or the standardized nature of Edgar-generated HTML documents. Routines to retrieve features, including the URL path segments for crawling, are relatively simple. They eventually require only minor adjustments to handle variations in the HTML structure introduced over time. This is in contrast to situations where **HTMLs!** (**HTMLs!**) are non-standard, as seen in *13D documents* submitted by *filers*. In these cases, each *filer* creates their own HTML, resulting in substantial discrepancies among raw content.

[10] CIK is a unique identifier assigned by the US SEC to individuals and entities that submit filings.

[11] XBRL is used to represent financial and business data in a format that is easily readable by both humans and computers. It is required format by many regulatory authorities for some disclosures.

This challenge emerges for various reasons, primarily due to the existence of multiple filings associated with a single event (group filings) and the fact that not all filings are relevant to the study of activist investors. For instance, insider traders also need to file *13Ds*. We turn single filings into events and identify *non-core events* by integrating supplementary information from other filings in combination with detecting patterns within the dataset.

We elaborate on these challenges and present our approach to mitigate potential issues in the subsequent sections, with a specific focus on the practical implementation of an algorithm, notably in relation to the information derived from processing raw data from the SEC Edgar.

A.2 EDGAR API access

Access to the SEC Edgar API is free and public. Establishing a connection to the Edgar API does not require any specific credentials; users only need to provide a header in their web request containing agent identification, such as an email address, and and observe the SEC Edgar web crawling guidelines, which currently[12] specifies a maximum request rate of 10 requests/second.

A.2.1 Starting point

In the content that follows, we outline how we programmatically extract data from the Edgar API through automated procedures. The process we describe is not unique; it's simply one of several potential methods for obtaining the data, all of which should yield the same information. Our main goal in explaining and sharing our data acquisition pipeline is to provide a clear path for others to replicate our research. Furthermore, we aim to foster contributions that enhance our methodology and offer a foundational resource for fellow researchers to construct their own datasets.

In our paper, we use data extracted from a variety of filings, including *13D*, *20F/40F*, *Form 4*, *PREM14*, *DEFM14* and *8K*. Depending on the filing type, we have used two distinct approaches as starting point to gather data: either the *crawler.idx* files or the *submissions.zip* files.

[12]At the time this text is being written, on the second semester of 2023.

The *crawler.idx* file contains only basic information (filing date, filing type, company name, and CIK) presented in a structured format (fixed-width format), along with the path segment to assemble URLs that points to *index.htm* files. For every quarter, since the inception of Edgar in 1994,[13] there is a separate *crawler.idx* file. Due to the limited number of features, individual files, as well as the concatenation of them is lightweight, and computationally efficient to manipulate. Filtering by file type, one can obtain the resulting URLs to request *filing documents* for subsequent extraction. Later in the text, we provide a detailed example of using this approach to extract *trigger date*, *investment objective*, and CUSIPs from *13D filings*.

Alternatively, another way to start involves bulk downloading references for all filings using the *submissions.zip* file. This file contains JSON[14]-format files for each corporation or individual person (i.e. separate files for each). These JSON files encompass additional information beyond what is found in the *crawler.idx* files, such as *previous company names*, *state of incorporation*, and reference to the specific *items* contained in forms *8K* (e.g., *Item 3.02, Unregistered Sales of Equity Securities* or *Item 5.01, Changes in Control of Registrant*). For corporations with a significant number of filings, there are supplementary JSON files with past records, slightly varying in structure.

The choice between these starting points, whether it's bulk downloading (*submission.zip*) or concatenating *crawler.idx* files, primarily depends on user preference and the convenience of accessing specific information. For instance, bulk downloading is advantageous when collecting *items* associated with *8K* filings, as it can be done in a single step. In the following sections, we will discuss potential challenges and pitfalls related to obtaining information from various filing types and provide practical recommendations for efficiently handling the substantial volume of data when working with Edgar filings using both starting points.

[13] Note, however, that electronic submission became mandatory only from January 1999 onwards, as mandated by Regulation S-T of April 1998.

[14] JSON is a lightweight data interchange format that is easy for humans to read and write, and easy for machines to parse and generate. It is commonly used for representing structured data and exchanging information between a server and a client, making it a popular format for web applications and APIs.

A.3 13D scraping

A.3.1 Get *crawler.idx*

For *13Ds* scraping, we begin by obtaining all the *crawler.idx* files, one per quarter, totaling 121 full quarters from 1994 to May 31, 2023.[15] We iterate over *year* and *quarter*, assemble each URL, send requests to the Edgar API, and add each retrieved file to a list. The records are then concatenated into a single dataframe. While there are 23,178,541 records in the final consolidated dataframe, the resulting object is not large since each record holds a small amount of characters, corresponding to 5 simple features (described soon). We then parse the fixed width format records into individual features and save it in a *parquet* file, partitioned by *filing type*, as we will work with one filing type at a time. Over the covered period, there are 721 different *filing types*, resulting in the same number of partitions. This process is swift, taking about 10 minutes to complete the entire loop for requesting, retrieving, concatenating, and saving the partitioned parquet file.

The resulting object contains the records that refers to all Edgar filings for that period. Each record has 5 attributes: **date**, **type**, **company**, **CIK**, **URL**.

- **date**: the *filing date*.
- **type**: the filing type (e.g. *SC 13D*, *8K*, *10Q*, *PX14A6G*)
- **company**: company[16] name. There is a single field for *company*. However, numerous *filing types* necessarily refer to two distinct entities. For example, PX14A6G is filed by an investor (filer) with considerations about the voting matters of a (subject) company, while *13D filings* are filed by an activist blockholder (filer) who has acquired a security block in a publicly traded company (*subject* or *target*). In such cases, there will be two entries in the *crawler.idx* files pertaining to a single filing. While the *filing date* and *type*

[15] Note, the 122nd quarter isn't complete.

[16] We use the term *company* in the context of SEC filings for any agent, not necessarily a firm. It could be a person, a Non-Governmental Organization (NGO), pension fund, or any other entity.

will match, all other fields will differ, including the URL.[17] Given that we cannot discern whether the company in a *crawler.idx* record is a *filer* or a *subject* company, we disregard company information in this phase. In a subsequent step (explained in Section A.3.3), we identify the *filer/subject* company pairs, each with its respective CIK, enabling us to eliminate duplicates.

- **CIK**: we follow same procedure adopted for the field *company*: ignore it in this step.
- **URL**: This refers to the text segment used to construct the URL that points to a file ending in *index.htm*. Parsing *index.htm* files is straightforward as they are internally generated by the US SEC.[18] Further information on how we scraped *index.htm* files is provided in Section A.3.3.

A.3.2 Single filing data

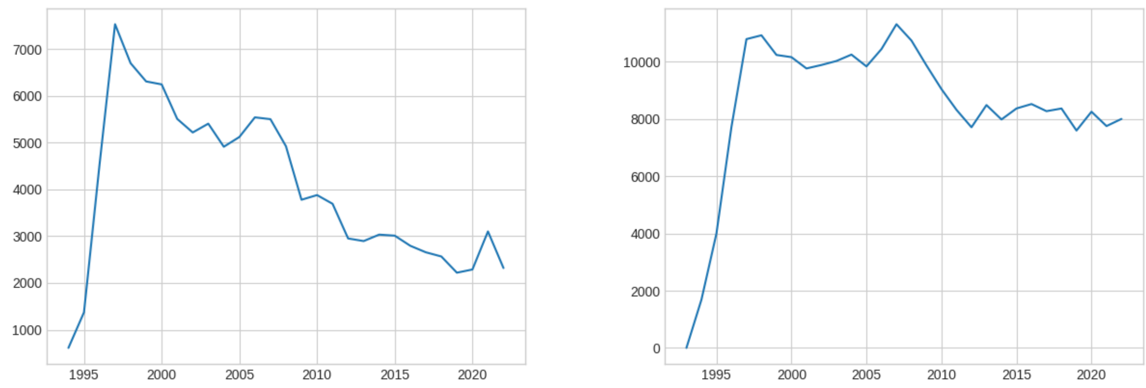
As we have saved all records partitioned by *file type*, we just need to read data for the *SC 13D* partition, a procedure that is much more efficient than reading the full dataset and then filtering. Figure A.1 shows the number of **records** referring to *13Ds* for each year. Panel A refers to initial filings only and Panel B, to amendments. Data from year 2023 was excluded from the plots as it does not refer to a full year.

For the period studied, there is a total of 117,486 records on *crawler.idx* that refer to *13D filings*. It is important to notice though that this number does not correspond to the number of unique filings. On the contrary these figures are roughly doubled, as each *13D filing* refers to, at least, two distinct entities, a *filer* and a *subject* company. So in most[19] cases there will be two records referring to the same filing: one with *company* and *CIK* denoting the *filer*, and another with *company* and *CIK* representing the *subject company*. We can only properly identify *filer* and *target* either by scraping it directly from the *filing document* textual content (which takes considerable time and somewhat error-prone) or, easier, by scraping it from *index.htm* files (our choice). Later we also identify and remove other forms of duplicates (e.g., submissions made to

[17]Although there will be two different URLs, both will point to identical *filing documents*.

[18]There are minimal changes in the HTML tree for different periods that can easily be addressed.

[19]Although in most cases there are 2 entries for each corresponding *13D filing*, this is not a rule. There are single entries as well as cases for more than 2 entries.



(a) initial filings (b) amendments

Figure A.1: Number of records - 13D filings

This figure show the number of 13D **records** aggregated per year, from 1994 to 2022 (last full year on our sample). Panel A shows records for initial filings and Panel B for amendments. As each 13D filing have on general 2 records on *crawler.idx*, one for the filer company and another for the target/subject company, these plots show numbers that roughly corresponds to double the amount of unique filings.

correct prior ones,[20] multiple submissions of identical filings but under different CIKs[21]), but these are not as significant in number as the duplicates eliminated in this step.

A.3.3 Scrape *index.htm* files

We describe now the process of identifying *filer* and *subject* companies in order to exclude duplicated references to the same filing. We use the path segment extracted from the *crawler.idx* files to assemble URLs that point to *index.htm* files. Given the large number of URLs (exceeding a hundred thousand), this process does take some time (approximately 5 hours); however, it is more efficient than the alternative of retrieving and then parsing the original documents. Conducting this process in batches is advisable to preserve intermediary results.[22]

For each retrieved file, we navigate through the HTML tags to parse data, occasionally employing text manipulation techniques (i.e. recognizing patterns using *regular expressions*)

[20] These are not "amendments" (*SC 13D/A*), but re-submissions with corrections.

[21] Such cases are rare. Generally, groups submit filings only once under a single CIK.

[22] This process can be adapted for various types of filings with minimal adjustments. In this example pertaining to *13D filings* corresponding to 121 quarters, we used 20 batches, with each cycle processing slightly fewer than 6,000 files. The complete cycle (request/retrieve, parse, and append results to a parquet file) for each batch typically took around 15-20 minutes to finalize, mainly due to adhering to the request rate specified in the SEC Edgar web crawling guidelines.

when necessary. This allows us to extract information regarding *target* and *filer companies* (names and CIKs). During this step, we also gather the *filing date* (although we could have used the one from the previous step) and addresses (city, state, and zip code) for both *filers* and *subject* companies. These addresses will be later used as a preliminary filter to exclude non-US-based companies and potential insiders. The parsing procedure for each element is as follows.

First, we locate the container with the *id=filerDiv*. Subsequently, we retrieve its child containers with *id=companyInfo* (one for the *filer* and another for the *subject*). While the *subject company* is typically the initial occurrence of *companyInfo*, a pragmatic approach to ensure accurate identification, instead of relying on positioning, is to locate the child container *class=companyName* containing the sub-string *(Subject)*. If this sub-string is present, the data in the parent container pertains to the *targeted company*; otherwise, it pertains to the *filer*. We proceed to clean the company names by removing the parenthetical identification text and trimming the resultant string. For each company, we navigate through the sibling tags of *companyName* to collect the CIK, city, state, and zip code. Finally, as our study requires extracting information exclusively present within the *filing document* and not available elsewhere, we need to access its contents. Here, we have two choices, each accompanied by corresponding URL paths in the *index.htm* file. The *filing document*'s content can be accessed either directly, through a *.txt* file for earlier filings or a *.htm* file for more recent ones, or through the *complete submission file*. This file is a text document that includes a couple of XML tags,[23] and the *filing document* is presented within a designated section. We discuss our choice of using the *complete submission file* in detail in Section A.3.5.

Once *filers* and *target* companies are properly labeled, we can identify the records that correspond to the same filing: the ones with matching *filing date*, *filer*, *targeted* company, and *file name*.[24] Out of the 117,486 records, nearly half are duplicates (one record for the *filer* and one for the *target company*), as expected. The preliminary count of unique *13D filings* over the covered period amounts to 59,904. This count is preliminary since we later eliminate

[23]While XML tags are present, their primary purpose is to distinguish the filing type and individual documents submitted at the time of the filings. They do not tag specific information of interest; instead, this information is embedded within the extensive text content of the *filing document*.

[24]The *filename* is derived from the last segment of the URL, which points to the *complete submission file*.

additional instances of duplicates and amendments that were mistakenly submitted as initial filings. However, these adjustments are minor, and as a result, the overall picture remains largely unchanged: the count of *13D* initial filings per year declines over time, dropping from approximately 3,500 filings at the end of the 1990s to around 1,000-1,500 in recent years.

Despite this cleaning process, there are still 1,025 records that exhibit identical *filer*, *target*, and *filing date*, yet possess different *file names*. We identified the following among these cases: records correspond to different stock classes, resulting in identical *target* **CIKs!** (**CIKs!**) but distinct CUSIPs (it refers to another security); multiple submissions in succession, often to split lengthy tables into individual files (in special for filings before 2005); double submission, either due to oversight or to include missing data - which do not qualify as amendments under the regulation (these are not *13D/A filings*). We should exclude records in cases where securities are identical, giving precedence to the last submission (often corrections to previous filings). Additionally, when securities differ, we should eliminate those referencing non-common stocks. As we do not have security identification yet (up to this stage), these records must be retained in the sample temporarily. We address these issues at a later stage, once we obtain securities' CUSIPs by scraping *13D filing documents*.

We also investigate cases where *targets* and *filers* coincided, but *filing dates* were different. There are 5,908 cases that are duplicates of *target* and *filer* CIK but do not share the same *filing date*. Causes for these submissions include situations where a *filer* acquired different classes of shares, passing the trigger thresholds for each class on different dates, or unfortunate cases where SC 13D/As (amendments) were mistakenly submitted as initial filings in the SEC Edgar system. While not encountered in our sample to this point, we cannot outright dismiss the possibility that an activist investor initially filed a *13D*, divested, and later reinvested in the same company to submit a *new* initial *13D filing* (investment/divestment/reinvestment). We will address this discussion once we have collected CUSIPs (Section A.3.7), enabling us to identify and subsequently eliminate any remaining duplicates.

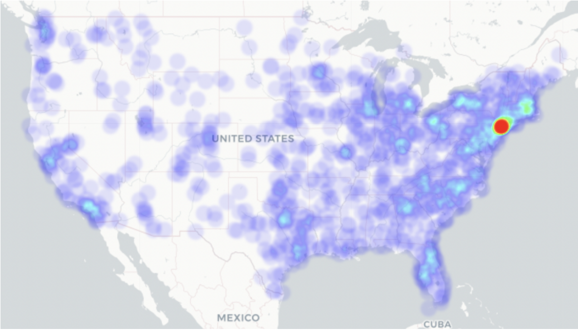
A.3.4 Geocoding: drop non-US companies and some insiders

Up to this point, our main focus has been on removing duplicate entries, while the challenges mentioned at the beginning of this appendix, involving parsing *filing documents* and eliminating *non-core events*, still need to be addressed. Before initiating the parsing process, which is time consuming, we can eliminate certain *non-core events* based on the names and addresses for both *filers* (including city, state, and zip code) and *subject* companies. This approach saves both time and resources, as we already possess the necessary data to execute this step, which was acquired during the scraping of *index.htm* files. By excluding these *non-core events* at this stage, we conserve resources, avoiding unnecessary requests and document parsing for filings that will ultimately be discarded.

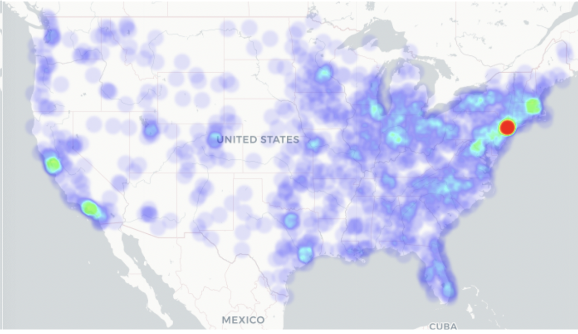
In our study, we focus on companies located within the contiguous continental US. Hence, we first filtered out records with addresses clearly outside the contiguous US, reducing the sample to 47,265 observations. We then geocoded the remaining addresses to obtain their latitude and longitude coordinates, after correcting over 600 unprocessable addresses, often due to typos or incorrect state assignments, such as “NY” instead of “NJ”. Following successfully geocoding the corrected batch of addresses, we proceeded with two sequential exclusion steps. Firstly, we removed cases where both *filer* and *target* shared the same name (1,942 cases) and/or address (9,635 cases), assuming these instances involved insiders (such as company executives, related entities, or shared legal representative for both *filer* and *target*). All instances where the company names for *filer* and *target* coincided also shared the same address, thereby reducing the sample size to 37,630 records. Secondly, we retained only records associated with companies within the contiguous continental US by clipping the records (geoprocessing operation). Following this clipping process, our final dataset consisted of 36,764 records. Fig A.2 shows the spatial distribution of *filers* and *targeted* companies, after the cleanings we have described until now.

The cleaning process thus far has primarily relied on data obtained from the scraping of *index.htm* files. While substantial progress has been made in terms of reducing the sample size by eliminating duplicates, non-US entities, and potential insider filings, more cleaning efforts lie ahead. This requires information from other sources like *8K* and *Form 4* filings.

Figure A.3 illustrates the count of *13D filings* aggregated by year for each cleaning step



(a) filers



(b) targeted companies

Figure A.2: Spatial Distribution of Filers and Targeted Companies

This figure shows the spatial distribution of filers and target companies of 13D filings, after cleaning for filers' and targets' coinciding addresses, dropping amendments and keeping only US companies. Panel A shows spatial distribution of Filers and Panel B the one for Targeted Companies.

described so far. The top-most line (dotted) represents the number of *13D records* from the *crawler.idx* files. The dashed line indicates the count after removing duplicates based on coinciding *filing date*, *filer*, *targeted* company. Finally, the solid line shows the series after excluding filings where the *filer/subject* had the same address and considering only companies located in the contiguous continental US. This represents a pre-cleaned sample, as further cleaning steps are required to identify and eliminate any remaining duplicates. The main upcoming steps involve gathering specific security information, which can only be acquired after scraping *filing documents* and matching security identification with the CRSP database.

A.3.5 Access *13D filing documents*

When a filing is submitted via Edgar, it automatically generates a *complete submission file*, which is a text file containing XML tags. Within the "BODY" section, in addition to the *main filing document*, there is content sourced from supplementary files such as contracts, reports, and correspondence. Each distinct document within this section is enveloped by either "TEXT" or "DOCUMENT" tags. Therefore, we can choose to extract the *filing document* content interchangeably, either from these "complete submission files" or directly from the *main filing document*, as the latter can be extracted from the former.

We, therefore, extract the text content from the *filing document* for information scraping.

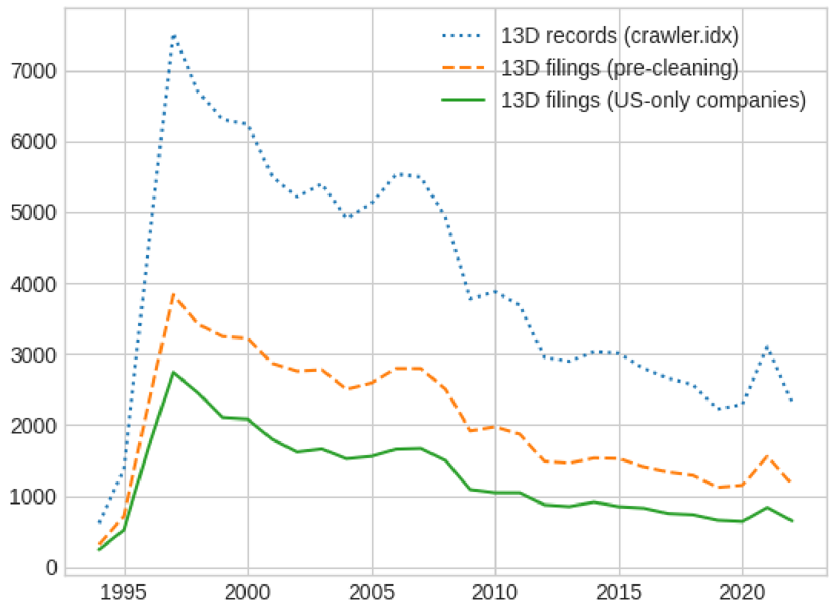


Figure A.3: 13D initial filings counts - before/after preliminary cleaning

This figure shows the number of 13D filings, aggregated by year. The dotted line, shown for reference, is the number of **records** that refers to 13D filings as in *crawler.idx* list. This number roughly duplicates unique filings, as there is, in general, one record for the filer and one for the targeted company. The dashed line shows records with unique combination of filer, target company and filing date were removed, after using information scraped from *index.htm* files. The solid line excludes companies outside the US and a single address for companies/filers. This last line is not yet the final sample used on our study, as we still perform various other cleaning steps.

However, the approach for handling this content varies depending on how the file was created. Over time, these files have evolved from typewritten style to non-semantic HTML, eventually transitioning to semantic HTML, albeit non-standard.

During the early stages of Edgar services, electronically submitted filings were typically presented in plain text format that used visual cues, similar to typewritten documents. For example, elements like tables were created by the physical positioning of bars and hyphens within the text, relative to other structures. As HTML became the norm to ensure proper display on web browsers, in the early 2000's, there was a transition period that extended for almost a decade. During this period, most filings were not consistently generated in proper HTML. Instead, they were first authored as formatted text documents using word processors or similar tools. Subsequently, these documents were handed off to third-party companies or service providers that used algorithms to transform the text into HTML-tagged content. Despite the successful visual display of these files on screens, the underlying text became fragmented into an collection of tags lacking semantic significance. The were primarily geared towards replicating the formatting of the original documents. For instance, dashes or vertical slashes were sometimes embedded within tags to create lines, instead of using proper styling elements. Compounding the issue, eliminating the HTML tags doesn't restore a coherent "plain text" structure, making it challenging to reverse-engineer the transformation applied by the service providers to the raw text. As a result, even though these documents may appear visually cohesive on screens, extracting meaningful information automatically through traditional text manipulation methods can be somewhat demanding.

In addition to dealing with messy and non-semantic HTML, another aspect demanding attention is the absence of standardization. Section 101 of Regulation 13D outlines the content and disclosure requirements for filing a *Schedule 13D*. However, no mechanism is in place to enforce standardization, such as the use of electronic templates or machine-readable tags. Consequently, *filers* have the flexibility to introduce various modifications, including changes to wording and formatting. As each *filer* creates their document following general guidelines rather than a rigid template, occasional adjustments, typos, variations in date presentation, and diverse methods of representing multiple security classes and/or *filers* within a group are introduced. Even for filings generated after 2010, which tend to adopt a more semantic structure with hierarchical HTML tags and proper formatting, parsing remains a non-trivial task.

As result, the presence of HTML tags within *13D documents* doesn't imply a straightforward parsing process; on the contrary, there are numerous challenges that must be navigated for effective scraping. The first group of challenges is due to the individualized modifications introduced by *filers* as they tailor the content. Complicating matters further, older filings contains non-semantic HTML tags aimed solely at achieving the desired visual presentation. Hence, successful text-based automated extraction from *13D filings* requires first to recognize if the file is presented as plain text, non-semantic HTML, or semantic HTML and then address the special cases stemming from non-standardization. Those familiar with the process of scraping *13D filings* understand the necessity for numerous adjustments to improve accuracy and reliability of the extracted information. The challenges associated with parsing are discussed next, in Section A.3.6.

Before we proceed, it's worth noting that while ADU techniques, not implemented here, can assist with some of the hurdles just described, they do not provide a complete solution. Some challenges persist due to the lack of standardization in the filings, including issues such as varying item order, multiple date formats, and unpredictable data positioning, all of which continue to pose extraction difficulties. Note, though, that while the discussion provided in the rest of this appendix is based on text-based parsing, it is also useful in guiding those interested in implementing such solutions.

A.3.6 General challenges to scrape *13D filings*

Up to this point there were two blocks of requests to Edgar API and the subsequent work on the data extracted was trivial and not prone to errors. The first set of requests retrieves the *crawler.idx* files. The second block retrieves *index.htm* files. Information contained on these later files clearly identifies *filers* and *targets*. Although the last step involving scraping *index.htm* files required some coding to navigate the HTML structure, it was relatively simple. However, the same cannot be said for the subsequent step, where we parse individual *13D* *filing documents*. This phase presents two types of challenges: *general challenges*, arising from the diverse ways *filers* compose their documents without a standardized electronic template, and *specific challenges*, unique to different types of information (such as *event date*, block accumulation, and security identification). These challenges are briefly discussed in what follows.

A.3.7 CUSIP, investment objective, event date and ownership stakes

If one is only interested in *filing date*, *filer* (activist investor) and *target* company, the exercise can be done with minimal effort, using the procedure described until here. However, there are some caveats in using that that sample "as is". As we have shown in Section A.3.4, there are some records that seems to be duplicated (same filing submitted twice), or amendments incorrectly submitted as initial filings. In addition, many records refer to filings submitted by insiders, or they might refer to securities that are not common stocks. So even if one does not need specific information that is available exclusively in the *13D main filing document* (e.g. *event date*, *investment objective*), the sample obtained without scraping the *13D documents* is not of much use. Naively using all records obtained from Edgar "as is" might lead to spurious results. Many records do not represent a "initial filing", or do not represent a typical activist investor (one that deliberately finds a opportunity to extract value from exerting influence over a company business). We extend this discussion in this section and later supplement it in sections A.4.2, A.4.3 A.4.4, for data obtained from filings other than *13Ds*.

In this section, we also provide a brief discussion of the challenges related to acquiring CUSIPs (used for security identification) and extracting the *investment objective* (derived from item 4 of the filings) directly from the *13D filing*, along with gathering *ownership stakes* and *event date*. We demonstrate how we incorporated this information to further clean our dataset. In addition to the challenges discussed earlier, each of these items requires specific considerations, which we address below. This discussion sheds light on potential issues applicable to all datasets that use *13D filings*, including our own, and highlights areas that could benefit from additional collaboration.

- **CUSIP**: CUSIPs serve as the primary identifier for securities and act as the common attribute for linking Edgar with CRSP datasets and, subsequently, the Compustat dataset. The CUSIP matching process is vital, as it not only provides the features needed for regression analysis, be it related to fundamentals or the market, but also enables to identify whether the security is a US-incorporated common stock, the only type of security with should retain on our dataset.

Most filings include CUSIPs for owned securities, but some lack this information or have invalid or missing codes. Out of the 36,764 filings in our base, we removed 1,959 filings (5% of the cleaned sample) for the following cases: the CUSIP field is empty or flagged with "N/A" or an equivalent expression, the code presented is not a valid CUSIP number (even considering CUSIPs without verification digits), or whenever we could not find any reference to CUSIPs at all.[25] This led to 34,805 observations. At this point, we can resume the discussion started at the end of Section A.3.3, where we had *target*/*filer*/filing date duplicates but could not determine if *target* referred to the same security or to a different one. Now that we have CUSIPs (hence can differentiate among securities for a single company), we can further clean duplicates. We identified 620 duplicates with matching CUSIP/filer/filing date or CUSIP/filer/event date. After dropping them, our sample size reduced to 34,185 observations.

- **Investment objective**: In *13D filings*, there is a field called *Item 4: Investment Objective*, in which blockholders must disclose their intentions regarding exerting influence over the investee company. We use the text in this field to identify filings related to amendments and reorganizations that should not be included in our database.

  – **amendments** As described in A.3.4, there are multiple filings with the same *target* and *filer*. Before we had information on CUSIPs, we couldn't infer whether these were due to the acquisition of blocks of different securities issued by the same company. Now that we have CUSIPs, we can examine these cases. As an additional check, we search for references for the term *amendment* on the initial portion of the text

[25] Our CUSIP search initially targets a shorter text segment for efficiency, with a fallback mechanism to search the entire document using the term *CUSIP* combined with regular expression matching if not found in the assigned field.

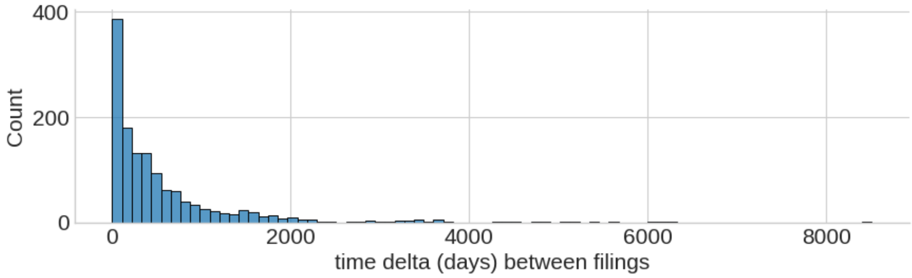


Figure A.4: Time delta between filings with coinciding CUSIP/filer

This figure shows a histogram of time delta in between filings where CUSIP and filer coincides, for which we could not detect the term 'amendment' in the text content of *investment objective*. The base sample from which those cases were extracted had been pre-cleaned for missing CUSIPs and coinciding "CUSIP/filer/event date" and "CUSIP/filer/filing date" and amounted to 32.803 observations.

of *investment purpose*.[26] Out of 3,218 duplicates with the same CUSIP and *filer*, we dropped 1,382 that were, in fact, amendments incorrectly filed as initial filings, resulting in a sample with 32,803 observations.

What about the remaining filings with coinciding CUSIP/*filer*? Figure A.4 displays a histogram of days in between filings for such remaining cases (1,338 observations). For very short intervals, it is most likely that these are re-submissions of filings, not configured as amendments in the sense of the regulation (they are not *13D/A*), but possibly submitted to correct missing information. Mid-sized intervals (3-6 months) are most likely amendments that were not detected in our previous approach, using text search. For longer periods, one year and beyond, there should still be some amendments, but there might also be cases where investors have acquired a block, divested at some point, and much later invested again in the same company.

After manually verifying a random sample of those remaining filings with longer time intervals, we found that they mainly belong to two cases. Firstly, some cases were indeed amendments that we could not detect with the techniques we have employed until now. Secondly, and most importantly, some groups such as GAMCO, that

[26] When a filing is an amendment but was incorrectly filed as an initial filing, it can typically be identified by examining *item 4*, which often mentions that it's an amendment right at the beginning of the text. However, we need to be careful because boilerplate text that refers to various generic investor scenarios also have mentions to "amendments" in situations other than amendments of the *13D filing*. To avoid capturing such references, we limited our search to the initial portion of the *investment purpose* text.

manages the Gabelli Mutual Funds,[27] might file a "second" initial filing, usually within a long time span from the previous ones because the group of funds holding the securities has substantially changed from the ones that appeared in the "first" *SC 13D*. In other words, the allocation has changed within vehicles of the same group.

Hence, we decided to drop outright all remaining duplicates that referred to coinciding *CUSIP/filer*, resulting in a dataset with 31,465 observations, each one representing a distinct pair *security/filer*. However, the uniqueness of pairs does not hold for *target/filer*, as one *filer* might submit one filing for distinct securities issued by the same company.[28] In fact, at this point, our sample has 412 of such cases without conditioning to the same *filing date*. If we match *filing date/CUSIP/filer*, then there are 49 extra filings beyond the unique combinations. We will postpone addressing these records until we obtain information about their security types when we merge CRSP/Compustat data. This is because if we were to randomly select and retain only one record for each coinciding pair of *target/filer*, without knowledge of the security type, there is a risk of unintentionally removing records related to the acquisition of common stock in favor of those related to non-common stocks.

– **bankruptcy and reorganizations** We follow the literature on blockholders and exclude observations related to bankruptcy cases. Initially, our approach to identify these cases involved a combination of textual searches within the content of *"Item 4 - Investment Objectives"* of the *13D filings*, coupled with *8K* filings that contained *"Item 1.03 — Bankruptcy or Receivership"*. These were extracted over a 20-day window preceding the *13D's filing date*. However, we observed that the *8K* procedure, as detailed in Section A.4.2, significantly outperformed the textual search. It captured cases mentioned in the text and others that couldn't be identified through the text alone. Thus, in practice, this specific textual search can be omitted. However, note that we only have convenient access to *8K items* from August 2004[29] onwards. Hence,

[27]GAMCO is one of the largest groups that acts as an activist blockholder, founded by Mario Gabelli.

[28]This is not a rule, as it is also common for a single filing to refer to the acquisition of more than one type of security, especially regarding different share classes.

[29]By "convenient access," we mean information that is presented in a structured form and doesn't require parsing from unstructured data. Prior to that date, obtaining this data would necessitate a scraping procedure

the decision to use these flags restricts our dataset, limiting the data available to us starting from 2005.

- **Event date**: One of the primary pieces of information featured on the cover page of a *13D filing* is the *event date*—the date when the investor surpasses the regulatory threshold of 5%, also referred to in the literature as the *trigger date*. This date is typically indicated above the section titled *"Date of Event Which Requires Filing of This Statement"*.[30] In the majority of cases, this information is readily available and can be parsed effortlessly using a standard date parser. However, there will inevitably be common challenges typical of when parsing dates from free-text, such as variations in formats, ambiguity, incomplete entries (e.g., only the year or only year+month) and typos. Hence, it is crucial to take precautions to maximize the number of successfully parsed results while mitigating potential systematic errors that could arise due to the default behavior of the parser when encountering incomplete dates.[31] In terms of increasing parsing results, several corrective measures should be implemented, including ensuring that months with typos are not discarded, a task facilitated by pre-processing dates to recognize date patterns and correct typos before the parser is executed. Additionally, we've introduced fallback mechanisms to locate *event dates* on a broader section of the cover page if the data is not found in the typically designated field.

  To address potential biases introduced by the parser, we've configured it to exclusively accept complete dates. However, this area can be improved, increasing the overall number

on *8K* filings, which we chose not to pursue in this study.

[30] As mentioned earlier, the text proposed in the regulation is not always the same as the wording adopted in the *filing document*. Instead, the latter often contains typos and various changes in wording. For example, one might use *"Date Requiring Filing of This Statement"*, *Date of Event"*, *"Event Date"*, among others. This implies that when scraping data based on its relative positioning to those texts as a reference and using regular expression matching, it's important to employ a variety of patterns that can effectively match most variations introduced by *filers*. This applies not exclusively to *event dates* but rather to all parsing elements of a *13D filing*.

[31] Many parsers default to assigning the current year when only the day and month are provided in the input, and this is particularly relevant when handling filings from previous years. For instance, consider a filing from 2004 with only the day and month mentioned (let's say "24-April"), and if the parser is executed in the current year, 2023, the parsed date would be interpreted as "24-April-2023". Furthermore, if the day is missing in the input, parsers frequently assume the day as 01, and in cases where the month is absent, they presume January (the first month of the year). Thus, if only the year is given—such as "2020"—the parsed date will be January 1, 2020. These issues can be readily identified using basic data analytics techniques: accumulation of entries occurring on the first day of the year, the first day of each month, or within the current year without corresponding filings signals that the parser's assumptions are introducing date biases.

of records since those without successfully parsed *event dates* are ultimately dropped. For instance, a potential enhancement could involve the imputation of eventual missing year based on the *filing date*, before running the parser, although this suggestion remains unimplemented in our current approach.

In addition to the aforementioned issues, another challenge arises due to the representation of dates as free-text: multiple *event dates* within a single filing. Although these dates likely correspond to transaction dates, this approach deviates from the mandated guidelines. The regulatory framework specifies that *any* trading activity involving the security addressed in the filing within the 60 days leading up to the *event date* should be disclosed in *"Item 5 - Interest in Securities of the Issuer"*:

> *"Item 5. Interest in Securities of the Issuer. (c) Describe any transactions in the class of securities reported on that were effected during the past sixty days or since the most recent filing of Schedule 13D (§ 240.13d–101), whichever is less, by the persons named in response to paragraph (a).*
>
> *Instruction. The description of a transaction required by Item 5(c) shall include, but not necessarily be limited to: (1) The identity of the person covered by Item 5(c) who effected the transaction; (2) the date of transaction; (3) the amount of securities involved; (4) the price per share or unit; and (5) where and how the transaction was effected."*

When faced with multiple *event dates*, a choice has to be made regarding which date to use. Should it be the first, the last, or somewhere in between? We opted to select the first date from the list. Our choice is equivalent to assuming that later dates refer to trading during the pre-disclosure accumulation period, a topic that is often debated among scholars and practitioners. However, note that this selection is arbitrary, because *filers* could be, instead, referring to transactions that occurred before reaching the 5% threshold.[32] Furthermore, that are cases with more than two *event dates* (although rare)

[32]For example, a *filer* might accumulate 4% ownership one month before and then add 1% in the last month. In such a case, the last transaction would be the one characterizing the threshold-crossing event.

that could well represent transactions both before and after the *event date* (pre-disclosure). Although we haven't explored the issue of selecting a date among multiple *event dates* in depth, discussing this challenge and sharing our choice leads to two important takeaways. Firstly, it illustrates the numerous decisions made when handling *13D data*, decisions that often remain undisclosed, whether dealing with commercial datasets or constructing researcher datasets. By transparently disclosing our choices, we shed light on potential limitations of the dataset that might otherwise remain hidden. Secondly, making these issues evident opens the door for potential improvements, among the various areas we've mentioned in this appendix. One potential approach involves creating rules using data from *item 5* with conjunction with the *filing date* to objectively determine the *event date* among the many dates presented.

While the regulatory text is clear about mandating a maximum 10-day interval between the triggering event and its disclosure, there were uncertainties among practitioners regarding the nature of these 10 days. This included questions about whether they referred to calendar days, working days, or even trading days. Furthermore, there was ambiguity about whether the counting should start from the trading date or the settlement date, which occurs three days later. In November 2009, the SEC provided clarification on these issues by specifying that the 10-day requirement refers to calendar days, with the counting starting from the day immediately following the transaction.

> *"Question: Rule 13d-1(a) states that a Schedule 13D must be filed within 10 days after the acquisition of more than five percent of a class of equity securities registered under Section 12 of the Exchange Act. Is the Schedule 13D due 10 days after the trade date or the settlement date of a securities transaction that creates the reporting obligation?*
>
> *Answer: The Schedule 13D beneficial ownership report must be filed within 10 days of the trade date of the securities transaction. Although under contract law the date on which the ownership of the shares is transferred may be the settlement date, an investor may, at a minimum, exercise investment power*

> *over the securities that were acquired through the trade as of the trade date. For purposes of calculating the 10-day time period, the first calendar day after the trade date counts as day number one.*"

n the series of *13D filings*, there are instances where disclosure significantly exceeds the mandated time interval from the 'event date.' Despite the SEC's clear guidance in November 2009 on how to count these dates, cases of delayed filings, often exceeding the 10-day limit, continue to occur. Bebchuk et al. (2013) rightfully suggests that such extended intervals should undergo regulatory scrutiny. However, note that cases of delayed filings often exceeding the 10-day limit are primarily related to amendments and other *non-core events* mistakenly filed as initial submissions. Therefore, this issue is less about enforcement and more about the need for robust filing validation processes to prevent incorrect submissions.

Another less common issue pertains to negative time deltas, which can emerge due to errors but are frequently linked to derivatives where the *"event date"* is effectively the "expiration date". As expiration dates are in the future, these records will have negative time deltas. Notably, although we haven't specifically investigated the reasons for large or negative time deltas, we observed a substantial reduction in such cases after cleaning the dataset to retain only common stocks and eliminating amendments incorrectly filed as initial submissions.

Finally, we chose to exclude filings with negative time deltas or time deltas exceeding 20 working days. This choice might appear somewhat lenient, as it involves retaining records with time deltas significantly beyond the regulatory limit. However, we deemed this approach prudent for three primary reasons. First, the maximum number of days is not strictly enforced. Second, despite the clarification in 2009, the original text does not explicitly mention calendar days, so it's plausible that some filers continue to misinterpret how to count them. Third, some filings are submitted shortly after prior submissions for corrections or to present additional information.[33] By retaining time deltas up to 20

[33] A potential refinement (not implemented) is to retain the information of the correction (last filings) but replace the *filing date* with the one for the first filing

working days, our aim was to eliminate obvious outliers while still accommodating cases where individuals may have inadvertently submitted filings later than usual.

Before we proceed, there is a practical consideration to address when matching any dataset with CRSP data, especially concerning *event dates*. As *filers* can use any day in free-text to refer to *event dates*, and these may not correspond to trading days, it's important to ensure that the matching process takes this into account to avoid skipping filings unnecessarily. To address this issue, we can use the nearest trading day after the given *event date* for matching purposes.

- **Ownership stakes** Extracting the *ownership stake*, a mandatory item in *13D filings*, may initially seem straightforward: retrieve the percentage ownership from the designated fields and, in cases involving multiple *filers*, sum the amounts for each *filer* to obtain the overall ownership percentage. Subsequently, we approximate the dollar amount by multiplying this percentage by the stock price on the *event date*.[34]

  The previously described approach falls short when dealing with multiple filers. To clarify, let's explore scenarios involving multiple filers. The term *"multiple filers"* might suggest the notion of independent investors who have collectively pooled their resources with a shared goal of acquiring a block of shares to exert influence over a company. We'll refer to this characterization as *collaborative investor groups"*.[35] However among events with *multiple filers*, the ones that involve *collaborative investor groups* are more the exception than the rule: most cases involve distinct legal entities that share a common affiliation, typically falling into two primary categories: corporate conglomerates or investment

[34]Notice that this approach is an approximation that, on average, overstates the amount actually paid by the investor because the *targeted* stock price often increases over the span of days when an activist investor purchases the stock, with a significant change on the *event date*. However, this isn't necessarily problematic, as long as one is aware of this fact and uses the information appropriately. For instance, these numbers aren't suitable for calculating returns for the activist investor, as they are likely to be understated. However, for purposes such as using the dollar stake in a regression as a dependent variable, the overstatement is less of an issue when evaluating the partial effects.

[35]This definition could be associated with *"wolf pack activism"*, a strategy where a group of investors coordinates to accumulate shares in a *target company*. Although wolf packs might openly declare their collective intentions, the term is most often associated with behind-the-scenes action without public disclosure of their collaboration . That's why we prefer to use the term *"collaborative investor groups"* here, to distinguish cases where these intentions are openly declared. See Coffee and Palia (2015) and Brav, Dasgupta, and Mathews (2022) for more information.)

management companies and their associated funds. These entities may maintain separate legal identities, yet they often operate in coordination or under the influence of a shared overarching strategy or ownership. We will refer to these as *"affiliated investor groups"*.

Affiliated investor groups typically submit a single filing. Double counting arises because regulation requires both direct and indirect beneficiaries to disclosure ownership.

> *"(a) Any person who, after acquiring directly or indirectly the beneficial ownership of any equity security of a class which is specified in paragraph (i) of this section, is directly or indirectly the beneficial owner of more than five percent of the class shall, within 10 days after the acquisition, file with the Commission, a statement containing the information required by Schedule 13D (§ 240.13d–101)."*

In several instances, indirect beneficiaries, including controlling shareholders and feeder funds,[36] are listed alongside direct beneficiaries, often without clear distinction. If ownership stakes are simply summed up, they will be overstated, as seen in many earlier datasets.

Another issue contributing to double counting, particularly for *collaborative investor groups*, arises from the flexibility provided by the regulation that allows the *filer* to choose between submitting a single joint filing or individual filings by each group member. Furthermore, the requirement for including information about other group members is rather loosely defined, stating that it should only reflect information known or reasonably known by the filing person:

> *"(k)(1) Whenever two or more persons are required to file a statement containing the information required by Schedule 13D or Schedule 13G with respect to the same securities, only one statement need be filed(...).*
>
> *(k)(2) A group's filing obligation may be satisfied either by a single joint filing or by each of the group's members making an individual filing. If the group's members elect to make their own filings, each such filing should identify all members of the group but the information provided concerning the other persons*

[36] Feeder funds pool money from investors and invest it in a master fund that, in turn, makes various asset or security investments.

> *making the filing need only reflect information which the filing person knows or has reason to know".*

Handling these variations in filing methods will determine the appropriate approaches for consolidating the parsed data on ownership stakes. For example, when multiple filers submit individual filings containing the same list of all other filers with their respective ownership percentages, we should discard duplicated filings. Conversely, when filers file individually without specifying the ownership percentages of other group members, we should compute the aggregate. However, there are intermediary situations for which the solution is not that simple. For instance, when some group filers file individually and mention the ownership of certain group members, but not for all of them. Moreover, group investors may trade independently on varying dates, resulting in varying *event dates* (and likely *filing dates*) if they choose to file separately. Distinguishing between coordinated group efforts and actions influenced by common knowledge or seller-driven market events can be challenging. We address this issue by flagging such filings into two groups based on whether the trades occurred less or more than 6 months apart.

While we have taken measures to address double counting, we recognize that our ownership stake data may still be overstated due to the automated data collection process. In this context, we refer to the analysis conducted by Dlugosz et al. (2006)[37] to inform our approach to using this statistic. Their analysis revealed that when ownership stakes affected by double counting are used as independent variables, the regression coefficients become biased. However, when ownership stake serves as the dependent variable, the coefficients remain consistent. Therefore, we should only employ this statistic in the latter setup.

[37] 2006 conducted a comprehensive examination of ownership stakes in commercial databases, uncovering biases resulting from double counting. They compared these reported stakes with *13D filings* and proxy materials, revealing instances of overstated ownership. Notably, the likelihood of overstatement increased with higher reported stakes.

A.4 Information from other filings (not-13Ds)

Datasets related to activist blockholders predominantly focus on events where activist investors identify a *target* company and acquire a stake with the intention to exert influence. However, within the universe of *13D filings*, numerous instances deviate from this ideal scenario. Prior studies commonly acknowledge the exclusion of events involving companies undergoing bankruptcy reorganization, non-US-incorporated firms, and merger arbitrage. Importantly, while these exclusions are theoretically applied, the specific methodology for their implementation is often left undisclosed.

To address this, we discuss our strategy for identifying non-core filings by leveraging additional data from SEC Edgar filings. Our approach involves collecting data on other filings for the same filing entity, the same *targeted* company, or both the *filer* and *targeted* company when applicable within a specified time window leading up to the *13D filing* trigger date.

A.4.1 20-F, 40-F or 6-K

Traditionally, researchers have excluded non-US companies from their datasets by filtering out securities with CRSP share codes that are not 10 or 11. However, while this step has to be conducted at a certain point, it involves the time-consuming process of parsing *filing documents* to extract CUSIPs, followed by matching those CUSIPs with CRSP data. As previously discussed, a more efficient strategy would involve avoiding the parsing of filings that are likely to be discarded in later stages. Parsing and processing filings significantly consume more time compared to alternative methods, such as filtering based on non-US addresses, which can be extracted from standardized structured data obtained from index.htm files. This is why we chose to first exclude non-US companies based on their addresses and only then parse filings, merge them with CRSP data, to finally remove filings associated with share codes other than 10 or 11. However, the address-based approach is not foolproof. For instance, some addresses may correspond to US offices of offshore entities, while in other cases, the address field might be either missing or inaccurately parsed.

Another alternative method for efficiently filtering out non-US companies before parsing *13D documents* is to assess whether companies are linked to filings like 20-F, 40-F, or 6-K. The presence of such filings serves as an indicator of their non-US status. While we initially did not include this step in our empirical work before the parsing stage, we recommend incorporating this identification process at the beginning of data collection, especially when combining it with the address-based method, to enhance overall efficiency.

In our study, after excluding share codes other than 10 or 11, we were left with 5,952 distinct subject companies. Out of these 5,952, we subsequently removed 23 observations using the offshore filing criteria, resulting in a refined sample size of 9,793 events.

A.4.2 8Ks: bankruptcy and notice of delistings

In the literature on blockholder activism, authors often refer to excluding filings associated with companies that have undergone bankruptcies or reorganizations, as well as *filers* who acquired blocks primarily for risk arbitrage purposes. As quoted in Brav, Jiang, and Li (2022), and consistent with their approach since their earlier work in Brav et al. (2008):

> *"we also exclude 13D filings whose Item 4 indicates a purpose of M&A risk arbitrage, bankruptcy reorganization, or distress financing. All these transactions involve an intention to influence corporate control and hence trigger the filing of Schedule 13D, but do not represent an activist strategy in our context."*

Brav and his co-authors conducted manual checks of *Item 4* section to spot these specific filings, which is a resource intensive approach. While our work does incorporate some manual checks, they are sparingly deployed, primarily for random verification aimed at ensuring the consistency of results produced by our automated scripts. Consequently, we've devised an alternative strategy for identifying such cases, leveraging *8K* filings, which will be detailed in the forthcoming section.

Now, we proceed to outline the practical steps involved in extracting the *8K* filings' *items*. As explained in Section A.2.1, there are typically two ways to initiate the data scraping process for Edgar filings. While the choice between these methods is generally inconsequential, there are specific cases, such as when dealing with *8K* items, where one method clearly emerges as the preferred choice.

When the objective is to extract the *items* referred to in the *8K* filings, it's advantageous to begin the process using the *submission.zip* file. This file already contains a comprehensive list of the items within each *8K* filing. This practical approach streamlines the process, simplifying the identification of the items referenced in the filing by merely looping through the content of the *.zip* file, instead of downloading all the *8K* documents for scraping.

We loop over the JSON files contained in the *submission.zip* that refer to *13D targeted companies* and, from them, we extract all records that related to *8K* filings. We are particularly interested two specific items: *1.03*, which pertains to disclosures about bankruptcy, and *3.01*, which concerns notices of delistings.[38] We create dummies for each of these items whenever they are filed prior to the *13D filing* date. We consider three specific timeframes for this assessment: 3 months, 6 months, and 1 year prior to the *13D filing date*.

A.4.3 Merger-related proxy filings

We control for mergers that took place before the *13D filing* by flagging events where we could identify filings like *PREM14*, *DEFM14*, and their variations[39] within the same *pre-13D* filing windows (3 months, 6 months, and 1 year) as we did for bankruptcies. This is a situation where the choice of starting point (*crawler.idx* or *submissions.zip*) is irrelevant as we just need to spot if there was such a filing, and the work is done - no need to extract another specific information from it. In our case, since we were already going through JSON files of *targeted companies* to collect *8K* filings, we used the same iteration to extract the records related to

[38] Notices of delistings notices of delistings do not equate to actual delistings but rather refer to stock exchange notifications regarding non-compliance with exchange policies (e.g., failure to publish regulatory filings like *10Q* and *10K*).

[39] *PREM14* and *DEFM14* are proxy statements that provide detailed information about proposed mergers or acquisitions (indicated by the "M"), along with voting instructions for shareholders. "PREM" stands for "Preliminary Proxy Statement", while "DEF" represents "Definitive Proxy Statement". Additionally, the inclusion of an "A" at the end of the filing name (*PREM14A* and *DEFM14A*) signifies amendments to previously filed documents. Conversely, the "C" designation indicates consent solicitations that go beyond the typical proxy voting process (i.e. amendments to the merger agreement, changes in terms or consents for actions not covered in the initial voting proxy.)

merger-related proxies.

A.4.4 Form 4: mapping insiders to *targeted companies*

Another concern in activist blockholder studies is to identify filings in which the *filer* is a related party. We have already eliminated 25% of the events, as they shared the same addresses for *filers/targets* in the *index.htm* files, under the assumption that these were insider traders. However, while this approach helps mitigate such issues, we can adopt a more robust technique to label *remaining* suspicious cases. Since insiders are required to submit a *Form 4* filing whenever they engage in a transaction, we leverage this information to flag cases where *13D events* are preceded by insider transactions (Form 4 filings). This flagging occurs when the *filer/subject* pair is the same for both the *Form 4* and the *13D filing*.

An important distinction needs to be highlighted at this point. Our intention is not to flag *every* insider transaction conducted within the pre-13D filing window. *Form 4* filings are quite prevalent, as they are submitted within a two-day window following transactions where insiders (such as company officers, managers, or beneficial owners with over 10% ownership) buy or sell any quantity of the company's common stocks, warrants, or convertibles. If we were to match *13D* and *Form 4* filings based solely on the *target company*, we would inevitably end up matching instances where insiders are selling to *13D filers* who are not insiders, as well as numerous other unrelated cases that occur within the same time window. While it might be intriguing to analyze the correlation between ownership stakes and excess returns when *filers* buy from insiders, this is not our primary objective here. Our focus lies solely in identifying insiders among *13D filers*, so we can drop filings submitted by them, because these filings deviate from what we consider in our study as "blockholder activist events".

A.5 Conclusion

This Appendix detailed the data extraction process from *13D filings* to obtain a suitable dataset of blockholder activism events, complementing the more focused content of the paper main text. It addressed specific limitations and offered practical solutions for parsing *historical raw static datasets*, notably *13D main filing documents* (non-machine-readable and non-standardized) and for identifying and labeling individual *core events*. We provided practical guidance for data gathering, dataset management, and script refinement, that are simple, feasible, reproducible and that prioritize use of open-source data. This procedural approach enhances transparency, and ensures methodological rigor, ultimately fostering trust and encouraging collaboration within the research community.

While we offer a comprehensive breakdown of the data extraction process, providing practical insights into API access and data management, catering to both Corporate Finance researchers and practitioners, we acknowledge that this work is far from exhaustive. Our current approach is limited due to its scale and should be further refined through the collective efforts of fellow researchers and practitioners interested in this field. Potential refinements can vary from simple rule additions, cross-checks, or even complete paradigm shift, where rule-based techniques used here, are either replaced or complemented by more sophisticated methods, as for example those employed in ADU. This work should be viewed as the first step in a collaborative effort and a useful resource for those interested in activist investor events datasets.

Appendix B

Additional Tables and Figures

Table B.1: Regression: abnormal return over flags
(full table)

Dependent variable: *abnormal return (CAPM), ± 20 days, t_0=event date*

	2006 to 2022					*2010 to 2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	0.1116***	0.1258***	0.1396***	0.1292**	0.1390**	0.1102***	0.1769***	0.1876***	0.2366***	0.2450***
	(0.0089)	(0.0463)	(0.0519)	(0.0607)	(0.0637)	(0.0103)	(0.0597)	(0.0615)	(0.0713)	(0.0717)
book-to-market		-0.0042	-0.0053	-0.0058	-0.0070		-0.0290	-0.0315	-0.0343	-0.0376
		(0.0186)	(0.0186)	(0.0185)	(0.0186)		(0.0247)	(0.0257)	(0.0246)	(0.0258)
cash-to-assets[1]		1.2712*	1.2403	1.3051*	1.2641*		1.3239*	1.3105*	1.3479*	1.3338*
		(0.7421)	(0.7561)	(0.7482)	(0.7612)		(0.7570)	(0.7610)	(0.7578)	(0.7614)
ROA		-0.2643*	-0.2560	-0.2483*	-0.2470		-0.4441*	-0.4366*	-0.4202*	-0.4171*
		(0.1529)	(0.1559)	(0.1496)	(0.1530)		(0.2379)	(0.2387)	(0.2320)	(0.2335)
ln market capitalization[1]		-4.2514	-5.8244	-3.5391	-4.8482		-9.5124	-10.7161	-13.1968*	-14.2994**
		(6.2526)	(6.0494)	(7.0869)	(6.7947)		(7.0470)	(6.6754)	(7.5264)	(7.1340)
tobin's Q		-0.0076	-0.0076	-0.0081	-0.0078		-0.0023	-0.0024	-0.0038	-0.0039
		(0.0103)	(0.0103)	(0.0101)	(0.0101)		(0.0126)	(0.0125)	(0.0123)	(0.0122)
profit margin[1]		0.0140	0.0154	0.0115	0.0125		0.0240	0.0287	0.0227	0.0274
		(0.0152)	(0.0158)	(0.0147)	(0.0154)		(0.0165)	(0.0177)	(0.0165)	(0.0178)
cash flow		0.0984	0.0924	0.0998	0.0953		0.3162**	0.3176**	0.2956**	0.2975**
		(0.1136)	(0.1168)	(0.1111)	(0.1137)		(0.1426)	(0.1437)	(0.1339)	(0.1346)
market leverage[1]		0.0075	-0.0284	-0.0054	-0.0014		0.5670	0.3839	0.6689	0.5299
		(0.6331)	(0.6510)	(0.6158)	(0.6323)		(0.6988)	(0.6985)	(0.7167)	(0.7186)
book leverage[1]		0.1108	0.0952	0.1024	0.0880		-0.1780	-0.1523	-0.1536	-0.1359
		(0.2353)	(0.2391)	(0.2305)	(0.2336)		(0.3065)	(0.3205)	(0.2940)	(0.3091)
dividend yield[1]		-0.7512	-1.0398	-0.6826	-0.8781		-3.9328	-4.5702	-4.4399	-4.9549
		(2.7002)	(2.7208)	(2.5092)	(2.5366)		(3.3114)	(3.3680)	(3.3652)	(3.4028)
payout ratio[1]		1.0228	0.6286	0.7813	0.3498		-0.3407	-0.5647	-0.5972	-0.7634
		(1.6547)	(1.5300)	(1.6617)	(1.5360)		(1.7132)	(1.7555)	(1.7027)	(1.7416)
sales growth[1]		0.0243	0.0173	0.0133	0.0080		-0.6043***	-0.5504***	-0.6551***	-0.6002***
		(0.0197)	(0.0190)	(0.0184)	(0.0185)		(0.1584)	(0.1429)	(0.1632)	(0.1475)
amihud liquidity measure		0.0174	0.0107	0.0133	0.0070		-0.0121	-0.0131	-0.0143	-0.0143
		(0.0203)	(0.0187)	(0.0195)	(0.0180)		(0.0173)	(0.0176)	(0.0180)	(0.0183)
group filings flag				0.0494	0.0492				0.0370	0.0406
				(0.0423)	(0.0411)				(0.0539)	(0.0538)
multiple *(1^{st} occurrence)*				-0.0374*	-0.0338				-0.0753***	-0.0656**
				(0.0214)	(0.0214)				(0.0274)	(0.0273)
multiple *(2^{nd} within 6MO)*				-0.0182	-0.0134				-0.0547	-0.0459
				(0.0273)	(0.0269)				(0.0361)	(0.0357)
multiple *(2^{nd} after 6MO)*				0.0233	0.0245				-0.0152	-0.0152
				(0.0260)	(0.0262)				(0.0270)	(0.0268)
F4 flag				-0.0960***	-0.1053***				-0.0462	-0.0489
				(0.0369)	(0.0368)				(0.0492)	(0.0484)
merge flag				-0.1113***	-0.1025***				-0.1327***	-0.1237***
				(0.0261)	(0.0264)				(0.0340)	(0.0340)
notice of delisting flag				0.0611	0.0526				-0.0421	-0.0508
				(0.0586)	(0.0565)				(0.0469)	(0.0489)
bankruptcy flag				-0.7071***	-0.7178***				-0.5967***	-0.6032***
				(0.1658)	(0.1654)				(0.2023)	(0.1953)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.0144	0.0228	0.0263	0.0341	0.0000	0.0212	0.0324	0.0376	0.0474
R-squared adj.	0.0000	0.0098	0.0127	0.0190	0.0213	0.0000	0.0134	0.0191	0.0250	0.0296
number of observations	3176	2822	2822	2822	2822	1823	1633	1633	1633	1633

This table shows the coefficients and standard errors (in parenthesis) for the flags (dummies) when the dependent variable is abnormal returns with reference to CAPM. The estimation window covers 100 trading days (t-120, t-20) that precedes the observation window. The observation window spans over 40 trading days centered around the trigger date (t-20, t+20). Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).
[1] These regressors were multiplied by 10^{-3} for better visualization of the coefficients and standard errors.

Table 2.3: Regression: logarithm of ownership stake (dollars) over flags *(full table)*

Dependent variable: *dollar ownership stake (natural logarithm)*

	2006 to 2022					*2010 to 2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	8.3043***	8.5032***	8.2838***	8.8751***	8.6100***	8.3444***	8.6720***	8.4122***	9.0790***	8.8020***
	(0.0345)	(0.1443)	(0.2037)	(0.1468)	(0.2047)	(0.0451)	(0.1455)	(0.2085)	(0.1641)	(0.2195)
book-to-market		-0.1897***	-0.1880***	-0.1925***	-0.1926***		-0.2643***	-0.2588***	-0.2889***	-0.2877***
		(0.0631)	(0.0611)	(0.0621)	(0.0606)		(0.0862)	(0.0839)	(0.0878)	(0.0858)
cash-to-assets[1]		-10.1055***	-9.1075***	-10.2642***	-9.3267***		-8.4600***	-7.3620***	-9.2063***	-8.0851***
		(1.9003)	(1.8968)	(1.8481)	(1.8461)		(2.5154)	(2.4438)	(2.4270)	(2.3580)
ROA		0.4998	0.6964*	0.5599	0.7177*		0.7551	0.8279*	0.7481	0.8183
		(0.4009)	(0.3992)	(0.3995)	(0.3998)		(0.4848)	(0.4869)	(0.4989)	(0.5035)
tobin's Q		0.2538***	0.2402***	0.2318***	0.2200***		0.2003***	0.1899***	0.1752***	0.1646***
		(0.0345)	(0.0343)	(0.0341)	(0.0342)		(0.0432)	(0.0424)	(0.0437)	(0.0435)
profit margin[1]		-0.0292	-0.0268	-0.0180	-0.0161		-0.0001	0.0045	0.0299	0.0345
		(0.0434)	(0.0435)	(0.0394)	(0.0398)		(0.0550)	(0.0557)	(0.0455)	(0.0465)
cash flow		0.7771**	0.6862**	0.6573*	0.5939*		0.6535	0.6748	0.5016	0.5296
		(0.3511)	(0.3423)	(0.3541)	(0.3487)		(0.4374)	(0.4341)	(0.4533)	(0.4542)
market leverage[1]		-2.1202	-2.2374	-1.7057	-1.7096		-2.5498	-2.8382	-1.6472	-1.9025
		(2.0568)	(2.0573)	(2.0172)	(2.0169)		(2.5814)	(2.5624)	(2.4387)	(2.4107)
book leverage[1]		1.3733	1.1326	1.2743	1.0328		1.1526	1.1991	1.0912	1.1234
		(0.9536)	(0.9510)	(0.9401)	(0.9554)		(1.3621)	(1.3570)	(1.2231)	(1.2244)
dividend yield		-0.0112*	-0.0129*	-0.0138**	-0.0149**		0.0285	0.0280	0.0194	0.0191
		(0.0068)	(0.0068)	(0.0070)	(0.0071)		(0.0252)	(0.0249)	(0.0255)	(0.0253)
payout		0.0306***	0.0309***	0.0299***	0.0299***		0.0285**	0.0282**	0.0279**	0.0276**
		(0.0067)	(0.0069)	(0.0067)	(0.0069)		(0.0121)	(0.0121)	(0.0125)	(0.0125)
sales growth[1]		-0.2680***	-0.3006***	-0.2105***	-0.2408***		3.8575***	3.8631***	3.3960***	3.3675***
		(0.0481)	(0.0496)	(0.0490)	(0.0498)		(0.4586)	(0.5205)	(0.5007)	(0.5498)
amihud liquidity measure		-1.2540***	-1.2357***	-1.2253***	-1.2089***		-1.4605***	-1.4480***	-1.3897***	-1.3762***
		(0.1137)	(0.1129)	(0.1121)	(0.1116)		(0.1623)	(0.1585)	(0.1640)	(0.1602)
group filings flag				0.9850***	0.9585***				0.9559***	0.9309***
				(0.1652)	(0.1620)				(0.2179)	(0.2125)
multiple *(1st occurrence)*				-0.4589***	-0.4189***				-0.4147***	-0.3946***
				(0.0882)	(0.0898)				(0.1137)	(0.1162)
multiple *(2nd within 6MO)*				-0.6118***	-0.5794***				-0.8525***	-0.8409***
				(0.1126)	(0.1138)				(0.1369)	(0.1402)
multiple *(2nd after 6MO)*				-0.4700***	-0.4720***				-0.4683***	-0.4819***
				(0.0686)	(0.0681)				(0.0879)	(0.0885)
F4 flag				-0.0777	-0.1269				-0.0668	-0.0897
				(0.2001)	(0.1980)				(0.2773)	(0.2752)
merge flag				-0.1153	-0.0623				-0.0410	-0.0183
				(0.1254)	(0.1243)				(0.1761)	(0.1707)
notice of delisting flag				-0.3452***	-0.3428***				-0.6504***	-0.6374***
				(0.1248)	(0.1227)				(0.1723)	(0.1717)
bankruptcy flag				-0.5757	-0.6053				-0.4987	-0.5849
				(0.7259)	(0.7083)				(0.7711)	(0.7811)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.3225	0.3348	0.3480	0.3586	0.0000	0.3274	0.3390	0.3623	0.3729
R-squared adj.	0.0000	0.3197	0.3281	0.3433	0.3503	0.0000	0.3224	0.3304	0.3544	0.3616
number of observations	3166	2822	2822	2822	2822	1823	1633	1633	1633	1633

This table shows the coefficients and standard errors (in parenthesis) for the flags (dummies) when the dependent variable is the natural logarithm of ownership stake in dollars. Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).
[1] These regressors were multiplied by 10^{-3} for better visualization of the coefficients and standard errors.

Table 2.4: Regression: logarithm of ownership stake (dollars) over flags
controled for size - (full table)

Dependent variable: *dependent variable: dollar ownership stake* *(natural logarithm)* *controlled for size*

	2006 to 2022					*2010 to 2019*				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*
Intercept	8.3043***	2.6610***	2.6494***	2.6144***	2.5977***	8.3444***	2.6540***	2.6607***	2.6304***	2.6340***
	(0.0345)	(0.0866)	(0.1081)	(0.0940)	(0.1062)	(0.0451)	(0.1253)	(0.1388)	(0.1414)	(0.1490)
book-to-market		0.0031	0.0002	0.0134	0.0080		-0.0318	-0.0317	-0.0259	-0.0279
		(0.0279)	(0.0287)	(0.0280)	(0.0289)		(0.0456)	(0.0462)	(0.0460)	(0.0469)
cash-to-assets[1]		-1.2651	-1.3077	-1.5046*	-1.6402*		0.3991	0.5810	-0.3181	-0.1459
		(0.9335)	(0.9423)	(0.8549)	(0.8621)		(1.2484)	(1.2401)	(1.2077)	(1.2042)
ROA		-0.1685	-0.1966	-0.1817	-0.2287		-0.0030	0.0027	-0.0511	-0.0474
		(0.1809)	(0.1833)	(0.1669)	(0.1688)		(0.2600)	(0.2636)	(0.2558)	(0.2598)
ln market capitalization		1.0074***	1.0069***	1.0153***	1.0158***		1.0026***	0.9999***	1.0076***	1.0055***
		(0.0129)	(0.0134)	(0.0128)	(0.0133)		(0.0175)	(0.0182)	(0.0177)	(0.0184)
tobin's Q		-0.0377**	-0.0397**	-0.0379**	-0.0401**		-0.0534**	-0.0557**	-0.0524**	-0.0553**
		(0.0158)	(0.0159)	(0.0158)	(0.0159)		(0.0230)	(0.0234)	(0.0241)	(0.0246)
profit margin[1]		0.0256	0.0243	0.0280	0.0257		0.0212	0.0214	0.0308	0.0311
		(0.0238)	(0.0244)	(0.0194)	(0.0196)		(0.0299)	(0.0315)	(0.0213)	(0.0221)
cash flow		-0.2701*	-0.2487*	-0.1894	-0.1546		-0.3474	-0.3347	-0.2644	-0.2521
		(0.1485)	(0.1484)	(0.1419)	(0.1415)		(0.2136)	(0.2133)	(0.2137)	(0.2135)
market leverage[1]		-0.6262	-0.5848	-0.6256	-0.6121		0.3568	0.3760	0.0854	0.1471
		(1.0264)	(1.0465)	(1.0021)	(1.0107)		(1.3644)	(1.3792)	(1.3546)	(1.3561)
book leverage[1]		0.8164**	0.8194*	0.4797	0.5026		0.7067	0.7140	0.5047	0.5044
		(0.4045)	(0.4185)	(0.3654)	(0.3779)		(0.5470)	(0.5604)	(0.5019)	(0.5182)
dividend yield		0.0000	0.0002	-0.0005	-0.0002		0.0021	0.0025	-0.0010	-0.0008
		(0.0046)	(0.0047)	(0.0047)	(0.0047)		(0.0134)	(0.0137)	(0.0134)	(0.0137)
payout ratio		0.0056	0.0047	0.0049	0.0038		0.0060	0.0067	0.0057	0.0063
		(0.0043)	(0.0042)	(0.0042)	(0.0042)		(0.0074)	(0.0073)	(0.0075)	(0.0074)
sales growth[1]		0.2179***	0.2098***	0.1923***	0.1880***		-0.8321***	-0.8901***	-0.6592***	-0.7619***
		(0.0228)	(0.0237)	(0.0218)	(0.0221)		(0.2342)	(0.2542)	(0.2077)	(0.2380)
amihud liquidity measure		-0.0245	-0.0359	-0.0399	-0.0514*		-0.0613	-0.0604	-0.0772**	-0.0753**
		(0.0276)	(0.0286)	(0.0264)	(0.0273)		(0.0386)	(0.0394)	(0.0372)	(0.0382)
group filings flag				1.1501***	1.1537***				1.0641***	1.0674***
				(0.0826)	(0.0834)				(0.1154)	(0.1155)
multiple *(1^{st} occurrence)*				0.0053	-0.0100				0.0038	0.0011
				(0.0516)	(0.0531)				(0.0654)	(0.0665)
multiple *(2^{nd} within 6MO)*				-0.1323**	-0.1458**				-0.2253***	-0.2324***
				(0.0615)	(0.0626)				(0.0855)	(0.0875)
multiple *(2^{nd} after 6MO)*				-0.0348	-0.0412				0.0046	-0.0025
				(0.0402)	(0.0405)				(0.0519)	(0.0530)
F4 flag				0.0910	0.0903				0.1152	0.1151
				(0.1058)	(0.1060)				(0.1724)	(0.1717)
merge flag				-0.1970**	-0.2028***				-0.1424	-0.1392
				(0.0772)	(0.0765)				(0.0977)	(0.0960)
notice of delisting flag				0.2300***	0.2267***				0.2027*	0.1982*
				(0.0710)	(0.0708)				(0.1100)	(0.1087)
bankruptcy flag				0.0131	-0.0213				-0.2237	-0.2795*
				(0.3363)	(0.3325)				(0.1381)	(0.1527)
year fx effects	N	N	Y	N	Y	N	N	Y	N	Y
clustered se	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R-squared	0.0000	0.7795	0.7810	0.7945	0.7959	0.0000	0.7823	0.7836	0.7959	0.7972
R-squared adj.	0.0000	0.7784	0.7787	0.7929	0.7932	0.0000	0.7805	0.7806	0.7933	0.7934
number of observations	3166	2822	2822	2822	2822	1823	1633	1633	1633	1633

This table shows the coefficients and standard errors (in parenthesis) for the flags (dummies) when the dependent variable is the natural logarithm of ownership stake in dollars. Columns 1 to 5 refers to the full period for which we have extracted flags from 8K filings (2006 to 2022). Columns 6 to 10 refers to the period in between crisis (2008 financial crisis and the pandemics). Firm specific controls are pre-determined, as they refer to the last period available before the evaluation window. Standard errors are clustered at SIC level. Standard errors are presented in parenthesis. *, **, and *** indicate statistical significance of the coefficients at the 10%, 5%, and 1% significance levels, respectively. The base year for the time-fixed effects is 2010 (dropped dummy).
[1] These regressors were multiplied by 10^{-3} for better visualization of the coefficients and standard errors.

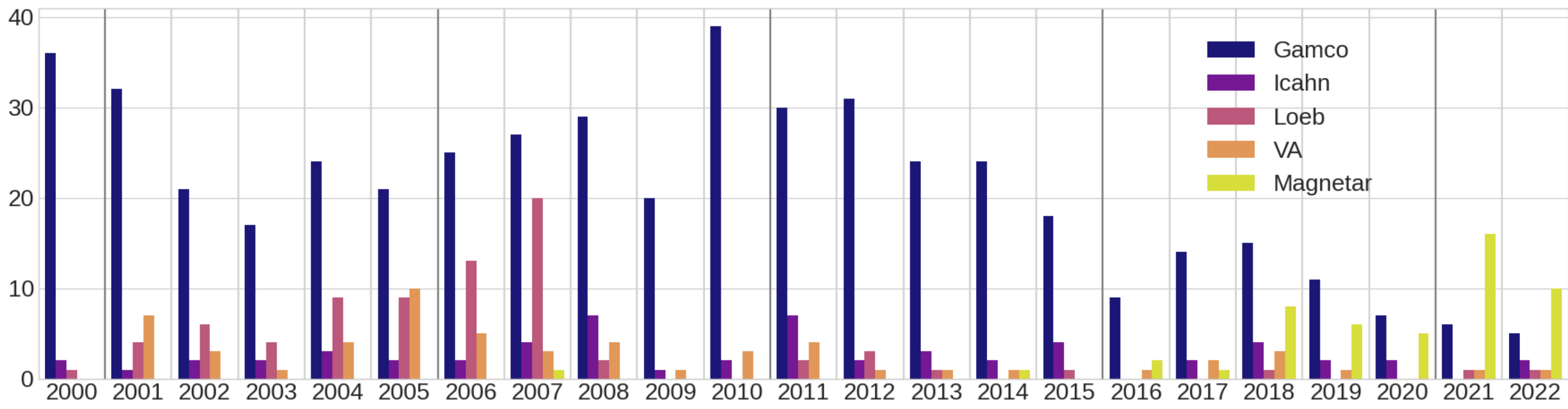


Figure 2.1: Events for the top 5 13D filers by year

This figure shows amount of unique 13D events (unique *permno-date*) for the top 5 filers, aggregated by year. The ranking to determine the *top 5* was calculated using aggregate numbers since 1994 (start of teh sample). Data is shown from 2000 to facilitate viewing. Counts consolidate CIKs for different vehicles used by the same *filer entity*. The sample has been cleaned for duplicated records, companies outside the contiguous continental US and securities that are not common stocks. We dropped observations in the following cases: CUSIP, ownership stake or event date missing (or not parsed by our algorithm), no match CUSIP/PERMNO, delta days between filing date and event date that is either negative or superior to 20 days.

Table 2.5: TOP 100 13D filers with filing amounts

filer name	amount	filer name	amount
GAMCO INVESTORS, INC / GABELLI FUNDS	565	ZWEIG DIMENNA PARTNERS LP	12
LOEB PARTNERS CORP / THIRD POINT LLC	89	SPECIAL SITUATIONS FUND III LP	12
ICAHN CARL C	62	STILWELL JOSEPH	12
VA PARTNERS LLC	57	PRAESIDIUM INVESTMENT MANAGEMENT COMPANY, LLC	11
MAGNETAR FINANCIAL LLC	54	CHAP CAP PARTNERS LP	11
BEAR STEARNS & CO INC /NY/	48	MILL ROAD CAPITAL, LP	11
STEEL PARTNERS II LP	47	RGM CAPITAL, LLC	11
STARBOARD VALUE LP	44	NEUBERGER BERMAN LLC /ADV	11
ELLIOTT ASSOCIATES, LP	44	ANCORA ADVISORS, LLC	11
FARALLON CAPITAL MANAGEMENT LLC	43	SANDELL ASSET MANAGEMENT CORP	11
MILLENCO LLC	42	CLINTON GROUP INC	11
DISCOVERY GROUP I, LLC	38	TPG GROUP HOLDINGS (SBS) ADVISORS, INC	11
ORBIMED ADVISORS LLC	38	LANE ALTMAN & OWENS	11
WYNNEFIELD PARTNERS SMALL CAP VALUE LP	35	PEQUOT CAPITAL MANAGEMENT INC	11
JANA PARTNERS LLC	35	RA CAPITAL MANAGEMENT, LP	11
SOROS FUND MANAGEMENT LLC	31	HARBERT DISCOVERY FUND, LP	10
BLUM CAPITAL PARTNERS LP	30	EDENBROOK CAPITAL, LLC	10
BIOTECHNOLOGY VALUE FUND LP	25	PERSHING SQUARE CAPITAL MNGT, LP / ACKMAN WILLI...	10
RCG HOLDINGS LLC	25	NEW ENTERPRISE ASSOCIATES 10 LP	10
GOLDMAN SACHS GROUP INC	24	FLYNN JAMES E	10
HUMMINGBIRD MANAGEMENT LLC	23	COOPERMAN LEON G	10
NOONDAY ASSET MANAGEMENT, LP	22	EMINENCE CAPITAL, LP	10
CANNELL CAPITAL LLC	21	CARLSON CAPITAL LP	10
VGH PARTNERS LLC	20	BAKER BROS ADVISORS LP	10
FRANKLIN RESOURCES INC	19	PRIVET FUND LP	10
VIEX CAPITAL ADVISORS, LLC	18	RELATIONAL INVESTORS LLC	10
KOPP INVESTMENT ADVISORS LLC	17	RED MOUNTAIN CAPITAL PARTNERS LLC	10
SOUTHEASTERN ASSET MANAGEMENT INC/TN/	17	POLEN CAPITAL CREDIT, LLC	9
FEINBERG STEPHEN	15	PHILOTIMO FUND, LP	9
PESSIN NORMAN H	15	ALPINE ASSOCIATES A LTD PARTNERSHIP /NJ	9
MARXE AUSTIN W & GREENHOUSE DAVID M	15	BARINGTON COMPANIES EQUITY PARTNERS LP	9
ATLANTIC INVESTMENT MANAGEMENT, INC	15	HICKS THOMAS O	9
RAGING CAPITAL MANAGEMENT, LLC	15	MARCATO CAPITAL MANAGEMENT LP	9
MMI INVESTMENTS, LP	14	LUXOR CAPITAL GROUP, LP	9
SHAMROCK ACTIVIST VALUE FUND LP	14	ROUMELL ASSET MANAGEMENT, LLC	9
MILLER LLOYD I III	14	STADIUM CAPITAL MANAGEMENT LLC	9
LEGION PARTNERS ASSET MANAGEMENT, LLC	14	SCOTT RICHARD L	9
GENERAL ATLANTIC, LP	14	SC FUNDAMENTAL INC	9
COLISEUM CAPITAL MANAGEMENT, LLC	14	HBK INVESTMENTS LP	9
NIERENBERG DAVID	13	CIBELLI MARIO	9
GARDNER LEWIS ASSET MANAGEMENT LP	13	ALTAI CAPITAL MANAGEMENT, LP	9
NORTHERN RIGHT CAPITAL MANAGEMENT, LP	13	KKR GROUP PARTNERSHIP LP	8
PERRY CAPITAL	13	SPRINGOWL ASSOCIATES LLC	8
PIRATE CAPITAL LLC	13	FRANKLIN MUTUAL ADVISERS LLC	8
SAC CAPITAL ADVISORS LLC	12	HEARTLAND ADVISORS INC	8
ENGAGED CAPITAL LLC	12	KING LUTHER CAPITAL MANAGEMENT CORP	8
P2 CAPITAL PARTNERS, LLC	12	ARES MANAGEMENT LLC	8
CITIGROUP INC	12	TANG CAPITAL PARTNERS LP	8
BLUE HARBOUR GROUP, LP	12	GLENVIEW CAPITAL MANAGEMENT, LLC	8
ROBOTTI ROBERT	12	DAWSON CAPITAL MANAGEMENT INC /CT	8

This table shows the names of the top 100 13D filers, in decreasing order of filing events. Data covers all EDGAR 13D filings, since 1994 until May 30, 2023, after being cleaned for the following: remove non-US incorporated targets, non-US companies, securities that are not common stocks, target companies from financials and utility sectors. We excluded filings for which we could not parse CUSIPs or event dates, and the ones for which the interval between event date and filing date was negative or superior to 20 consecutive days. Filings coinciding *event date/permno* or *filing date/PERMNO* were collapsed to represent single events. We collapsed into a single number the a mounts concerning the same filer that have used different vehicles (with different CIKs).

Table 2.6: TOP 100 13D filers in alphabetical order

filer name	filer name
ALPINE ASSOCIATES A LTD PARTNERSHIP /NJ	MARCATO CAPITAL MANAGEMENT LP
ALTAI CAPITAL MANAGEMENT, LP	MARXE AUSTIN W & GREENHOUSE DAVID M
ANCORA ADVISORS, LLC	MILL ROAD CAPITAL, LP
ARES MANAGEMENT LLC	MILLENCO LLC
ATLANTIC INVESTMENT MANAGEMENT, INC	MILLER LLOYD I III
BAKER BROS ADVISORS LP	MMI INVESTMENTS, LP
BARINGTON COMPANIES EQUITY PARTNERS LP	NEUBERGER BERMAN LLC /ADV
BEAR STEARNS & CO INC /NY/	NEW ENTERPRISE ASSOCIATES 10 LP
BIOTECHNOLOGY VALUE FUND LP	NIERENBERG DAVID
BLUE HARBOUR GROUP, LP	NOONDAY ASSET MANAGEMENT, LP
BLUM CAPITAL PARTNERS LP	NORTHERN RIGHT CAPITAL MANAGEMENT, LP
CANNELL CAPITAL LLC	ORBIMED ADVISORS LLC
CARLSON CAPITAL LP	P2 CAPITAL PARTNERS, LLC
CHAP CAP PARTNERS LP	PEQUOT CAPITAL MANAGEMENT INC
CIBELLI MARIO	PERRY CAPITAL
CITIGROUP INC	PERSHING SQUARE CAPITAL MNGT,LP / ACKMAN WILLIAM
CLINTON GROUP INC	PESSIN NORMAN H
COLISEUM CAPITAL MANAGEMENT, LLC	PHILOTIMO FUND, LP
COOPERMAN LEON G	PIRATE CAPITAL LLC
DAWSON CAPITAL MANAGEMENT INC /CT	POLEN CAPITAL CREDIT, LLC
DISCOVERY GROUP I, LLC	PRAESIDIUM INVESTMENT MANAGEMENT COMPANY, LLC
EDENBROOK CAPITAL, LLC	PRIVET FUND LP
ELLIOTT ASSOCIATES, LP	RA CAPITAL MANAGEMENT, LP
EMINENCE CAPITAL, LP	RAGING CAPITAL MANAGEMENT, LLC
ENGAGED CAPITAL LLC	RCG HOLDINGS LLC
FARALLON CAPITAL MANAGEMENT LLC	RED MOUNTAIN CAPITAL PARTNERS LLC
FEINBERG STEPHEN	RELATIONAL INVESTORS LLC
FLYNN JAMES E	RGM CAPITAL, LLC
FRANKLIN MUTUAL ADVISERS LLC	ROBOTTI ROBERT
FRANKLIN RESOURCES INC	ROUMELL ASSET MANAGEMENT, LLC
GAMCO INVESTORS, INC / GABELLI FUNDS	SAC CAPITAL ADVISORS LLC
GARDNER LEWIS ASSET MANAGEMENT LP	SANDELL ASSET MANAGEMENT CORP
GENERAL ATLANTIC, LP	SC FUNDAMENTAL INC
GLENVIEW CAPITAL MANAGEMENT, LLC	SCOTT RICHARD L
GOLDMAN SACHS GROUP INC	SHAMROCK ACTIVIST VALUE FUND LP
HARBERT DISCOVERY FUND, LP	SOROS FUND MANAGEMENT LLC
HBK INVESTMENTS LP	SOUTHEASTERN ASSET MANAGEMENT INC/TN/
HEARTLAND ADVISORS INC	SPECIAL SITUATIONS FUND III LP
HICKS THOMAS O	SPRINGOWL ASSOCIATES LLC
HUMMINGBIRD MANAGEMENT LLC	STADIUM CAPITAL MANAGEMENT LLC
ICAHN CARL C	STARBOARD VALUE LP
JANA PARTNERS LLC	STEEL PARTNERS II LP
KING LUTHER CAPITAL MANAGEMENT CORP	STILWELL JOSEPH
KKR GROUP PARTNERSHIP LP	TANG CAPITAL PARTNERS LP
KOPP INVESTMENT ADVISORS LLC	TPG GROUP HOLDINGS (SBS) ADVISORS, INC
LANE ALTMAN & OWENS	VA PARTNERS LLC
LEGION PARTNERS ASSET MANAGEMENT, LLC	VGH PARTNERS LLC
LOEB PARTNERS CORP / THIRD POINT LLC	VIEX CAPITAL ADVISORS, LLC
LUXOR CAPITAL GROUP, LP	WYNNEFIELD PARTNERS SMALL CAP VALUE LP
MAGNETAR FINANCIAL LLC	ZWEIG DIMENNA PARTNERS LP

This table shows the names of the top 100 13D filers, in alphabetical order. Data covers all EDGAR 13D filings, since 1994 until May 30, 2023, after being cleaned for following adjustments: remove non-US incorporated targets, non-US filers, securities that are not common stocks, target companies from financials and utility sectors. We excluded filings for which we could not parse CUSIPs or event dates, and the ones for which the interval between event date and filing date was negative or superior to 20 consecutive days. Filings coinciding *event date/PERMNO* or *filing date/permno* were collapsed to represent single events. We collapsed into a single number the a mounts concerning the same filer that have used different vehicles (with different CIKs).