From Regulatory Data to Quantitative Investment Signals in Equity Markets

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor rerum politicarum (Dr. rer. pol.)

By

Jan Lukas Schroeder

November, 2025

TU Dortmund University
Faculty of Business and Economics
Department of Finance

Acknowledgement

We would like to express our sincere gratitude to Peter Posch for his outstanding supervision and continuous support. He created an environment that encouraged exploration, failure, and learning, and constantly reminded us to ask the fundamental question: "*Why is that important?*" His openness, humility, and intellectual curiosity made every discussion both engaging and productive, fostering the exchange of diverse perspectives rather than rigid dogma.

We are deeply thankful to Maria Melzner, our closest discussion partner, for her countless brainstorming sessions, valuable feedback, and meticulous proofreading. Her insights and patience have been invaluable throughout this process.

We also extend our gratitude to Alexander Krause for his dedication to the field of finance and programming, which significantly influenced parts of the technical execution of this work, and to Elio Arturo Farina, whose expertise in LaTeX and constant support made the transition from draft to publication-ready manuscripts both efficient and enjoyable.

We thank all participants at the various conferences where the papers were presented for their invaluable feedback and insightful comments, as well as all others who provided valuable input outside of these events. We are especially grateful to the following scholars for their constructive input, discussions and support: Hasan Fehmi Baklacı (Yaşar University, Türkiye), Pradip Banerjee (Indian Institute of Management Indore), Seda Bilyay-Erdogan (Kadir Has University, Türkiye), Szabolcs Blazsek (Mercer University, U.S.A.), Marius Bommert (TU Dortmund University, Germany), Celine Daute (TU Dortmund University, Germany), Wei Gao (Widener University, U.S.A.), Conrado Diego García Gómez (University of Valladolid, Spain), Stefano Gubellini (San Diego State University, U.S.A.), Ben Jacobsen (TIAS Business School, Tilburg University, Netherlands), Jonathan Lewellen (Tuck School of Business at Dartmouth, U.S.A.), Christiane Pott (TU Dortmund University, Germany), and Nikolai Roussanov (Wharton School, University of Pennsylvania, U.S.A.).

Finally, we thank the anonymous journal reviewers for their critical and constructive feedback, which significantly improved both the analytical rigor of the results and the clarity of their presentation.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Introduction

The objective of this dissertation is to analyze how information in regulatory disclosures impact equity market prices and to determine how this information can be leveraged to construct quantitative investment strategies. Beyond its academic contribution, the practical goal of this research is to lay the foundations for new data-driven quantitative hedge funds.

We analyze more than 20 million regulatory disclosures spanning three decades, including over 1.5 million insider transactions, using datasets totaling more than 100 terabytes of structured and unstructured data. This dissertation investigates patterns, trends, and anomalies within legally mandated disclosures filed by regulated entities with the U.S. Securities and Exchange Commission (SEC). Specifically, it examines how features extracted from these filings affect stock price behavior. To operationalize these insights, we develop backtesting and portfolio simulation frameworks, enabling the translation of regulatory information into executable, data-driven investment strategies.

Throughout this dissertation, unless stated otherwise, references to "markets" pertain specifically to U.S. equity markets, meaning publicly traded stocks listed on the NASDAQ and NYSE exchange. References to "companies" therefore refer to publicly listed firms on these exchanges, whether headquartered in the U.S. or abroad. Other markets, such as currencies, bonds, or commodities, are outside the scope of this analysis. When referring to "entities," we mean any legal structure, such as a corporation, limited partnership, or individual, that is subject to U.S. securities regulations and therefore falls under the oversight and regulatory authority of the SEC. Similarly, references to "government" or "regulators" refer to U.S. government agencies, most prominently the SEC. The terms "filings" and "disclosures" are used interchangeably.

The dissertation comprises five empirical studies, each focusing on a different dimension of SEC filings:

1. A large-scale exploratory analysis of the EDGAR database.
2. Seasonal disclosure patterns and their link to the Halloween effect.
3. Hedge fund strategy replication using Form 13F filings.
4. Market reactions to non-reliance disclosures in Form 8-K filings.
5. Insider trading signals from Form 4 filings and their use in portfolio construction.

Our line of research has historically been limited by data accessibility and computational constraints. The SEC filings corpus represents over 1,000 terabytes of data of which the majority is available as text documents which require significant engineering capabilities and computational resources to collect, clean, and analyze at scale. To our knowledge, various datasets used in our studies are novel and previously unexplored in academic and institutional research, making this dissertation the first to provide systematic large-scale analysis of certain disclosure types.

The findings are relevant to a wide audience, including professional investment managers, quantitative hedge funds, equity research firms, and regulators such as the SEC. By bridging the gap between regulatory disclosure data and market behavior, this work offers both theoretical insights and direct practical applications.

1.1.1 Structure of Dissertation

Five empirical studies are performed, each leveraging regulatory disclosures filed with the SEC as a primary information source to explore and analyze structural patterns, anomalies, and signals for equity portfolio strategies.

1. Large-Scale Analysis of the EDGAR Database

The first study presents the most extensive exploratory analysis to date of the SEC's EDGAR database, covering over 15 million disclosures submitted by over 800,000 entities between 1994 and 2024. By examining filing activity across form types, filer categories, sectors, and time, the study uncovers long-term structural trends, patterns and anomalies. These include a persistent year-over-year decline in public company disclosures offset by rising private capital activity, recurring seasonal patterns in insider trading and general filing volume, and filing surges preceding major market events such as the dot-com bubble, the global financial crisis, and the 2022 downturn. Notably, just 2% of form types account for 80% of annual submissions. These findings provide a foundational reference point for understanding regulatory disclosure behavior in U.S. capital markets.

2. Seasonal Disclosure Patterns and the Halloween Effect

Building on insights from the first study, the second study presents seasonal fluctuations in disclosure behavior and their potential relationship with the "Halloween effect" – the observed market anomaly where stock market returns tend to outperform from November through April compared to May to October. Between 2004 and 2023, winter months exhibit significantly higher filing activity across multiple disclosure types: a 17% increase in total filing volume, 22% more insider trades, 13% more private capital raising activity, 12% more activist investor activity, and nearly fivefold more shareholder meetings coupled with 87% of audited financials being reported. These seasonal patterns are consistent with similar trends in European markets, suggesting a global phenomenon. The analysis positions disclosure seasonality as a previously underexplored explanatory component of the Halloween effect.

3. Hedge Fund Strategy Replication Using Form13F Filings

The third study evaluates the viability of replicating hedge fund portfolios disclosed via SEC Form 13F each quarter. Analyzing over 150,000 portfolios between 2013 and 2023, the study finds that cloned portfolios constructed from top-quartile hedge fund filings outperform the S&P 500 by 24.3% annually on a risk-adjusted basis. While differences exist in volatility, drawdowns, and tracking error, replicated portfolios closely mirror the returns of the original funds when rebalanced at the time of disclosure. The findings support the practical value of public hedge fund disclosures for systematic portfolio construction.

4. Market Reactions to Item 4.02 in Form 8-K Disclosures

The fourth study examines 8,006 disclosures reporting the non-reliance on previously issued financial statements under Item 4.02 of Form 8-K between 2004 and 2023. These filings, which signal material errors in prior financials, and precede financial restatements, are associated with significantly negative cumulative abnormal returns (-2.6% to -5.4%). Revenue recognition errors trigger more severe market reactions than net income errors, and disclosures that lack clarity about the nature or magnitude of the misstatement result in significantly harsher penalties. Interestingly, situations involving the SEC do not amplify negative reactions. Item 4.02 filings thus serve as a robust early-warning signal for investors ahead of financial restatements.

5. Insider Trading Signals in Form 4 Filings and Portfolio Construction Strategies

The final study proposes a novel equity portfolio construction strategy based on insider trading data disclosed in SEC Form 4 filings. Analyzing 1.8 million transactions from 2011 to 2023, the study uses Monte Carlo simulations to evaluate 1,600 feature dimensions for transaction selection. The strategy identifies high-conviction insider purchases that yield an average 5-day cumulative abnormal return (CAR) of 3.4%, compared to 1.2% for all insider buys. The best performing backtested portfolio outperforms the S&P 500 index across 11 of 14 performance metrics in- and out-of-sample, delivering a Sharpe ratio of 1.88, a Sortino ratio of 2.92, and an annualized excess return of 41.4%.

1.1.2 Limitations and Future Research Opportunities

The studies are subject to several limitations. First, in terms of price impact and portfolio strategy, only a subset of SEC form types was analyzed. For example, the Form 8-K study examined just one of the 33 event categories, leaving many others unexplored due to the absence of structured datasets and the predominance of unstructured text. Second, the insider trading strategy could be expanded to incorporate additional dimensions such as institutional trading from Form13F holdings, activist investor activity disclosed in Form 13D, or material events from Form 8-K that precede Form 4 insider transactions. Third, the hedge fund replication strategy focused on fund-level performance but did not analyze holdings from the perspective of the underlying companies. A complementary approach would reverse this perspective by starting with specific firms, such as recently IPOed companies, and testing whether the number of funds holding them correlates with subsequent stock performance.

Building on these limitations, multiple avenues for future research emerge. Expanding the scope of Form 8-K analysis to cover all 33 event categories, including director or auditor changes, entering into material agreements, and bankruptcy filings, would deepen the understanding of event-driven price impacts. Fine-tuning large language models offers promising potential for predicting IPO pricing, particularly by analyzing S-1 and 424B4 filings, while multi-modal analysis could extend research beyond text into audio, video, and image-based disclosures, such as live government briefings or corporate presentations. Future studies may also explore the expansion of information sources, including press releases and non-EDGAR disclosures, to capture a broader informational flow. The development of AI agents, combining multiple specialized models into interactive networks, represents a promising frontier for automating event studies, extracting structured data, and enhancing financial prediction.

Advancements in LLM fine-tuning, structured data extraction, and automated event-study frameworks will likely enable many of these directions.

1.1.3 Conclusion

The dissertation's central objective is to trace how mandated disclosures with the U.S. Securities and Exchange Commission (SEC) propagate through equity markets and to test whether these signals can be transformed into quantitative investment strategies. Using SEC filings as primary sources, five studies map the disclosure landscape at scale, uncover seasonal patterns that help understand the Halloween effect, evaluate hedge fund replication via Form 13F, quantify market reactions to Item 4.02 non-reliance Form 8-K filings, and design insider-driven portfolios grounded in Form 4 data. The contribution is both empirical and practical: novel datasets and large-scale text analytics reveal robust relationships between disclosure timing, content, and price impact, while the resulting rules and backtests are directly applicable to quantitative hedge fund design.

Our work is not exhaustive. Coverage is limited to three out of over 500 SEC form types, and future research might extend the Form 8-K taxonomy to derive new structured datasets, and exploit multi-modal language models via model fine-tuning to further improve signal quality.

Taken together, the results establish a reproducible framework for converting regulatory information into actionable investment insights, with relevance for professional asset managers, researchers, and regulators.

1.2 SEC Filings, EDGAR and Primary Information Sources

As this dissertation centers on regulatory disclosures published by entities regulated under the U.S. Securities and Exchange Commission (SEC), including public companies, insiders, investment funds, and financial institutions, it is essential to outline the SEC's role, its regulatory framework, and the reliability of its information systems.

The SEC serves as the cornerstone of market transparency and investor protection in the United States. Established in the aftermath of the 1929 stock market crash and the Great Depression, its mandate was to restore investor confidence, prevent fraud, and ensure fair and orderly markets. The Securities Act of 1933 introduced the first federal requirement for companies to register securities and disclose material information before offering them to the public[1]. This was followed by the Securities Exchange Act of 1934, which created the SEC and granted it broad authority to enforce securities laws, regulate secondary markets, and oversee disclosures by public companies, stock exchanges, broker-dealers, and corporate insiders [2].

[1] https://www.govinfo.gov/content/pkg/COMPS-1884/pdf/COMPS-1884.pdf
[2] https://www.govinfo.gov/content/pkg/COMPS-1885/pdf/COMPS-1885.pdf

Since 1933, companies and other regulated entities have been legally required to file disclosures with the SEC. Initially paper-based, filings have been submitted electronically through the EDGAR (Electronic Data Gathering, Analysis, and Retrieval) system since 1994. Since then, over 20 million disclosures have been published, accompanied by more than 100 million attachments, such as bond agreements, corporate bylaws, investor presentations, and other supporting documents. The vast majority of the 1,000 terabytes of data is unstructured text, with only a small fraction available in structured formats. These filings come from more than 800,000 regulated entities, spanning beyond publicly traded companies to include investment advisers, mutual funds, hedge funds, corporate insiders, broker-dealers, activist investors, issuers of asset-backed securities, and many more.

EDGAR serves as both a submission platform and a public access tool, making filings globally available at no cost, typically within minutes of submission. The SEC thus holds a legal monopoly over legally mandated disclosures: while companies may also release press statements, they cannot substitute for filing with the SEC.

Because disclosure requirements vary, the SEC uses more than 500 form types, each designed for specific purposes defining what information has to be published by whom, when. For example:

- Form 10-K[3]: annual reports including audited financials, risk factors, legal proceedings, management analysis, and governance details, predominantly filed by publicly traded companies on U.S. exchanges.
- Form 4[4]: insider trading disclosures, reporting share purchases or sales by executives, directors, or major shareholders, within two business days of the trade occurring.
- Form 8-K[5]: immediate disclosure of 33 defined material events, such as mergers, executive changes, major contracts, or bankruptcy, within four business days, predominantly filed by publicly traded companies on U.S. exchanges.
- Form 13F[6]: quarterly reports disclosing an investment company's portfolio holdings, predominantly filed by hedge funds.

The content and timing of disclosures depend on several factors, including an entity's purpose (e.g., operating company versus investment company), type of securities (e.g. publicly traded versus privately held), size (e.g., small-cap versus large-cap), domicile

[3] `https://www.sec.gov/files/form10.pdf`
[4] `https://www.sec.gov/files/form4data.pdf`
[5] `https://www.sec.gov/files/form8-k.pdf`
[6] `https://www.sec.gov/files/form13f.pdf`

and more. For example, U.S. domestic firms such as Microsoft file their annual report on Form 10-K within 60 to 90 days after fiscal year-end, depending on their public float, annual revenue and other factors. In contrast, companies traded on U.S. exchanges, but domiciled outside the U.S., so called foreign private issuers such as Sony, file their annual report on Form 20-F [7]. Form 20-F serves a similar purpose but differs in structure and content, and is published within four months of fiscal year-end. In this particular example, these distinctions create two layers of asymmetry: a timing asymmetry, as larger and domestic companies typically disclose earlier than smaller or foreign firms, and a content asymmetry, as the disclosure requirements for Form 10-K and Form 20-F differ in scope and detail.

1.2.1 Primary versus Secondary Information Sources

To understand the role of SEC filings, it is essential to first clarify where information originates when it becomes publicly available for the first time. We define primary information sources as channels through which information enters the public domain directly and in its original form. These sources contain unaltered data that has not been previously disclosed elsewhere. In contrast, secondary information sources aggregate, interpret, or repackage data derived from primary sources for broader consumption.

SEC filings submitted via the EDGAR system represent a primary source, as companies are legally required to disclose material information through this channel. Although a company may issue a press release via a distribution service such as BusinessWire, GlobeNewswire, or PR Newswire before filing with the SEC, EDGAR remains the official primary disclosure channel because filing with the SEC is a legal obligation. A company cannot choose to disclose material information solely through a press release without also submitting it to the SEC.

Other examples of primary sources include official releases by U.S. government agencies, such as the Federal Reserve (e.g., interest rate decisions [8]), the Department of Justice (e.g., enforcement actions [9]), or the U.S. Census Bureau (e.g., economic data releases [10]). Internationally, comparable systems include the UK's Financial Conduct Authority (FCA) with its Regulatory News Service (RNS [11]) and Europe's Euronext

[7] `https://www.sec.gov/files/form20-f.pdf`

[8] `https://www.federalreserve.gov/monetarypolicy/fomcminutes20250730.htm`

[9] `https://www.justice.gov/archives/opa/pr/twitter-agrees-doj-and-ftc-pay-150-million-civil-penalty-and-implement-comprehensive`

[10] `https://www.census.gov/construction/nrc/current/index.html`

[11] `https://data.fca.org.uk/\#/nsm/nationalstoragemechanism`

exchange network [12].

Secondary information sources reprocess and interpret primary data. Examples include financial data platforms such as S&P's Capital IQ, Refinitiv Eikon, or databases like CRSP accessed through WRDS. These systems extract, standardize, and structure information originally published through primary channels, such as SEC filings, to make it more accessible for research, analysis, and decision-making.

Most financial news articles (e.g., via Reuters, Bloomberg, The Wall Street Journal) also qualify as secondary sources, as they summarize and interpret primary data, such as quarterly earnings releases or Federal Reserve meeting minutes. Only rarely do such outlets publish original, previously undisclosed information, such as merger rumors or investigative findings. Sell-side equity research reports similarly fall under secondary sources, as they synthesize information from multiple primary disclosures and add forecasts or interpretations. However, when such reports include original insights gained from interviews, site visits, or private communications, they may contain limited elements of primary information.

Limitations of secondary sources include uncertainty regarding the timing and accuracy of the underlying information. The delay between a primary disclosure and its secondary interpretation can vary widely. For instance, the SEC may publicly announce an enforcement action in the morning, while media coverage appears hours later. Secondary sources, such as news articles, also introduce potential biases and interpretive errors, as they reflect the author's subjective framing. Key primary details may be omitted, altered, or misrepresented. Nonetheless, secondary sources serve a useful function in consolidating multiple primary materials, much like meta-analyses, and can help identify new or overlooked primary sources, as seen in research-oriented investigative reports (e.g., Hindenburg Research).

All studies in this dissertation focus exclusively on primary information sources, particularly those made available through mandatory SEC disclosures, ensuring analysis of original data rather than interpretations or summaries.

[12] `https://live.euronext.com/en/products/equities/company-news-archive`

Chapter 2: Anomalies, Trends and Patterns in Disclosure Activities: Understanding EDGAR

Author: Jan L. Schroeder (jan@sec-api.io)*

Abstract
We present the largest exploratory analysis to date of the SEC's EDGAR database, covering all 15 million regulatory disclosures filed by more than 800,000 entities—including public and private companies, funds, insiders, and other registrants—between 1994 and 2024. By examining filing volume and filer activity across entity types, form types, sectors, and time, we uncover structural trends and recurring patterns. Periods preceding major market events, such as the dot-com bubble, the 2007–08 global financial crisis, and the 2022 downturn, are characterized by shifts in disclosure behavior, including elevated activity in the asset-backed securities sector, the rise of SPAC filers, and increased registration of complex financial products. We document a steady decline in publicly traded company filers, offset by growing private capital formation. Insider trading activity consistently spikes in February, one month before the March release of annual reports. Filing activity also follows a seasonal cycle, peaking in February and reaching its lowest point in September. Finally, disclosure volume is highly concentrated: just 2% of form types account for 80% of all annual submissions. Our findings provide a foundation for understanding anomalies, trends and patterns in U.S. disclosure practices and offer insights for further investigation by researchers, regulators, and market participants.

Publication Details: Working Paper. Submitted to the International Journal of Disclosure and Governance.

Keywords: Market Anomalies, SEC Filings, Financial Data Analysis, Form 4, Form 8-K, Form 424B2, Form 10-K, Form D, Form NPORT-P, Form 13F-HR, SEC Correspondences

Acknowledgements: We would like to thank Antoinette Schoar, Professor of Finance at MIT Sloan School of Management, and Dushyantkumar Vyas, Associate Professor of Accounting at University of Toronto Mississauga, for their invaluable comments and insights. We extend our sincere gratitude to Fernando Becerra of Harvard University and Oleg Volkov of Queen Mary University of London for their invaluable feedback and contributions to the data visualization process.

The following is based on Schroeder (2025).

*Jan L. Schroeder is a Research Fellow at the Department of Finance, TU Dortmund, Germany, and the founder and CEO of SEC-API.io. The authors have no competing interests to declare. All errors are our own.

2.1 Introduction

Historically, research on the SEC's EDGAR database has concentrated on specific form types or subsets of such, predominantly 10-Ks and 10-Qs. A significant portion of this research emphasizes content analysis. For instance, Cohen et al. (2020) explored the textual variations in 10-K report sections, such as risk factor disclosures, and their subsequent impact on stock prices. Loughran and McDonald (2014) pioneered sentiment analysis in financial reports by formulating word lists that classify content into categories like positive and negative. The realm of financial report readability is extensive, with studies like Kim et al. (2019) delving into the relationship between readability and stock price crash risk, Bonsall et al. (2017) examining the clarity of language, and Li et al. (2013) proposing an extension of the Herfindahl Index to gauge competition in specific industries.

He and Plumlee (2020) utilized 8-K data spanning 2005 to 2016 to construct disclosure metrics, demonstrating that, despite their correlation, each metric uniquely represents firm disclosures. Notably, they identified the word-count-based 8-K measure as a particularly robust proxy for voluntary disclosure. Meanwhile, Cohen et al. (2015) identified a pattern where corporate insiders achieve abnormal returns by trading during the "8-K trading gap" — the interval between significant corporate events and their subsequent disclosure via Form 8-K.

In the context of S-1 filings and IPOs, Loughran and McDonald (2013) discerned that filings with higher levels of uncertainty in their Form S-1 language were associated with elevated initial returns, price adjustments, and subsequent volatility, emphasizing the influence of textual clarity on investor assessments.

Additionally, Köchling et al. (2021) analyzed SEC comment letters and observed that during periods of intense SEC workload related to IPO reviews, the comments tended to be more generalized initially, with swifter follow-up responses. Such periods also influenced offer price adjustments and underpricing.

Larocque et al. (2020) examined DEF 14A filings, determining that market participants leverage compensation details in proxy statements to predict future firm outcomes, with a notable positive correlation between sell-side analysts' forecast revisions and unexplained CEO compensation. Amel-Zadeh et al. (2016) scrutinized Form 4 footnotes, concluding that discretionary insider sales, as identified through footnotes, can foreshadow future negative stock returns, implying strategic use of these disclosures by insiders. Cheng et al. (2012) conducted a study on 13-D filings, establishing that firms targeted by hedge fund activists, particularly those well-versed in tax matters, tend to enhance tax efficiency without adopting high-risk tactics. Additionally, while some researchers have delved into analyzing EDGAR log files, IP addresses, and filing access patterns, García and Norli (2012) took a broader approach, evaluating the distribution of all EDGAR form types from 1993 to 2011, albeit without considering various dimensions such as sectors, industries, and regulatory shifts.

The analysis of EDGAR filing exhibits remains an under-explored area, possibly due to the challenges associated with data acquisition. However, Dyreng et al. (2020) utilized Exhibit 21 from 10-K filings and IRS data on multinational firms, discovering that while most firms comply with Exhibit 21's subsidiary disclosure requirements, certain nondisclosures occur, particularly in contexts involving tax havens, media attention, or other signs of suboptimal disclosure quality.

Previous research has not exhaustively analyzed the entire EDGAR corpus. This is likely due to the limitations of conventional research databases like COMPUSTAT and Thomson Reuters, which lack the detailed granularity our collection offers. A prevalent method in both research and industry involves navigating the master.idx files from EDGAR's index. Yet, these files offer limited metadata for each filing—filer CIK, form type, filing URL, and the accession number—precluding an in-depth analysis based on other filing properties, such as EDGAR acceptance dates and times, sectors, industries and SEC offices. Additionally, the computing power, particularly memory, needed to process the entire filings corpus has been a deterrent. However, recent advancements in in-memory processing and robust programming languages have made the analysis of such extensive data feasible.

Recognizing and interpreting filing trends is especially critical for regulators such as the SEC and self-regulating bodies, such as security exchanges like NASDAQ, for new policy formulation and updating policies. The analysis aids in pinpointing market participants who might need intensified oversight, especially when their activities deviate from norms identified in this research. For investment professionals, understanding these patterns is instrumental in risk management. By understanding filing volume baselines across industries and sectors, they can identify anomalies and emerging trends. Furthermore, these insights are invaluable for insurance industries when calibrating their risk premium models. For corporate entities, these insights assist in benchmarking filing practices against industry standards, affording a clearer perspective on their positioning relative to competitors. Such data can also guide strategic moves, especially when monitoring M&A (Form S-4) and IPO (Form S-1) activities, enabling better timing for acquisitions and capital raises.

2.2 Data & Methods

Our research methodology is segmented into three phases:

1. Data Collection: We sourced and curated our data of all SEC EDGAR filings via sec-api.io.
2. Data Pre-processing: The datasets underwent cleaning, merging, and augmentation to ensure they were primed for subsequent analysis.
3. Analysis: Through exploratory data analysis, statistical testing and time series analysis, we identified patterns and trends across multiple dimensions.

Throughout this research, Python 3.10 served as our primary programming language.

2.2.1 Data Collection & Pre-Processing

To obtain our data, we employed Python-based techniques for parsing and downloading all metadata of EDGAR filings from 1994 to 2024. We accessed this data via the APIs offered by sec-api.io, a commercial financial data service catering to global investment professionals. Specifically, we utilized the Query API[1] and Mapping API[2].

From our initial data collection, we constructed a primary dataset. This was subsequently divided into three distinct sub-datasets. Additionally, two supplementary datasets were acquired and merged with the three sub-datasets. This integration facilitated the mapping of a filer's CIK to other key dimensions, including sector, industry, SIC codes, and SEC offices.

Dataset 1 – Primary

The primary dataset encompasses metadata from over 15 million distinct EDGAR filings, identified by their individual accession numbers, spanning from January 1, 1994, to December 31, 2024. Additionally, it incorporates data from over 800,000 unique EDGAR filers. The main properties of the dataset used in our research are:

- The unique accession number of the filing.
- The EDGAR form type, such as "10-K" or "D/A".

[1] `https://sec-api.io/docs/query-api`
[2] `https://sec-api.io/docs/mapping-api`

- The "Accepted" timestamp denoting the date and time the filing was accepted by EDGAR.
- A list of all entities referenced in the filing, as categorized by the EDGAR system (e.g., filer, subject). Each of these entities comprises attributes such as CIK, name, SIC code, among others.

The filing metadata dataset builds the main resource for our analysis. This dataset allows us to create and analyze filing metrics across different dimensions, such as the date/time a filing was filed, clustering filings by form types, volume analysis by CIKs over time, and more.

Important Considerations The Accepted timestamp of a filing, as reported on the EDGAR system, assists in determining the associated year and month of disclosure. It's crucial to emphasize that the time a filing was accepted by the EDGAR system doesn't denote when a filing became publicly accessible. A prime illustration of this discrepancy is seen in EFFECT filings. While their Accepted timestamp is set at 12:15am, the SEC releases filings only between 6am and 10pm. Hence, while EFFECT filings display an Accepted timestamp of 12:15 am, they are actually made public at 6am. It's crucial to differentiate the "Accepted" timestamp from the "Filing Date" attribute visible on the EDGAR filing index page, which displays only the date. The "Accepted" and "Filing Date" attributes may not always coincide.

In the EDGAR system, filings are distinguished by their accession numbers, whereas filers are identified through their Central Index Key (CIK). It's essential to note that a single filing can reference multiple entities with different CIKs. Consequently, one filing might be linked to several EDGAR URLs. This often leads researchers to inadvertently count a single filing multiple times based on the number of associated URLs found in EDGAR index files. To avoid such discrepancies, it's imperative to uniquely identify and count filings by their accession numbers, not by their URLs. Further details on this can be found in the subsequent section titled "Understanding Multiple Entities."

The EDGAR system doesn't inherently normalize CIKs. According to official SEC documentation, CIKs are composed of 10 digits, with smaller CIKs padded with leading zeros for standardization. However, when accessing resources on EDGAR, these leading zeros are often omitted. For example, a CIK presented as "0000872243" is accessed as "872243" in URLs, such as: `https://www.sec.gov/Archives/edgar/data/869531/000149315223035704/0001493152-23-035704-index.htm`.

While accession numbers distinctly identify filings and even include a CIK, they cannot be directly mapped back to their respective CIKs.

Understanding Multiple Entities The relationship between an EDGAR filing and referenced EDGAR entities represents a one-to-many relationship. A single filing might pertain to just one entity or several. For instance, the example 8-K filing from Tesla (Figure 2.1) pertains exclusively to one entity: Tesla Inc, identified by CIK "0001318605".


Tesla, Inc. (Filer) CIK: 0001318605 (see all company filings)
IRS No.: **912197729** | State of Incorp.: **DE** | Fiscal Year End: **1231**
Type: **8-K** | Act: **34** | File No.: **001-34756** | Film No.: **231298092**
SIC: **3711** Motor Vehicles & Passenger Car Bodies

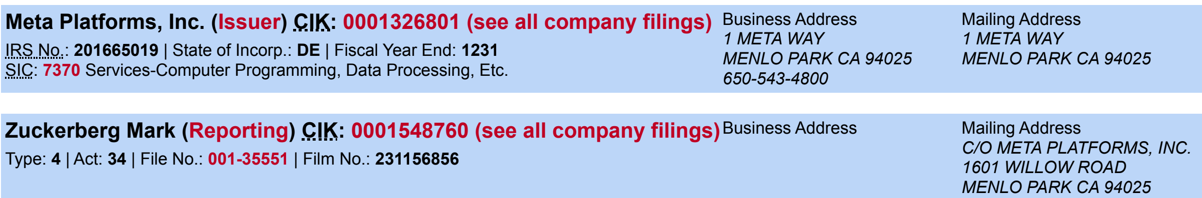
Business Address
3500 DEER CREEK RD
PALO ALTO CA 94304
650-681-5000

Mailing Address
3500 DEER CREEK RD
PALO ALTO CA 94304

Figure 2.1: Screenshot of a Form 8-K[3] Filing from EDGAR Featuring One Entity.

This image captures a part of the filings details page of one of Tesla's 8-K filing on EDGAR, exemplifying a one-to-one mapping where a single filing accession number is associated exclusively with the filer's CIK, indicating Tesla as the sole entity involved in this specific disclosure.

Conversely, Form 4 invariably references at least two entities: the issuer (the company issuing the securities) and the reporter (the individual or organization acquiring or disposing of the securities). As a result, a unique Form 4 filing, marked by its distinct accession number, produces two separate entries in the EDGAR index—one for each associated entity. Figure 2.2 represents an example showing a Form 4 filing that references both Meta (previously known as Facebook) and Mark Zuckerberg. This filing can be accessed through two separate EDGAR index URLs, both leading to the identical filing.


Meta Platforms, Inc. (Issuer) CIK: 0001326801 (see all company filings)
IRS No.: **201665019** | State of Incorp.: **DE** | Fiscal Year End: **1231**
SIC: **7370** Services-Computer Programming, Data Processing, Etc.

Business Address
1 META WAY
MENLO PARK CA 94025
650-543-4800

Mailing Address
1 META WAY
MENLO PARK CA 94025

Zuckerberg Mark (Reporting) CIK: 0001548760 (see all company filings)
Type: **4** | Act: **34** | File No.: **001-35551** | Film No.: **231156856**

Business Address

Mailing Address
C/O META PLATFORMS, INC.
1601 WILLOW ROAD
MENLO PARK CA 94025

Figure 2.2: EDGAR Screenshot of a Dual-Entity Form 4[4] Filing.

This image illustrates a part of the filings details page of a Form 4 filing that reports an insider transaction involving two entities: Meta Platforms, Inc., as the issuer, and Mark Zuckerberg, as the reporting individual. This example showcases a one-to-many mapping (specifically, one-to-two in this case) between the unique accession number of the filing and the two involved entities, identified by their respective CIKs. Consequently, this filing appears twice in the EDGAR index, under two different URLs – one for Meta Platforms and another for Zuckerberg. Despite the different URLs, the content of the filing remains identical in both instances.

[3] `https://www.sec.gov/Archives/edgar/data/1318605/000095017023050938/0000950170-23-050938-index.htm`

Additionally, certain filings like Form S-4 can be linked to an even greater number of entities. The example S-4 filing in Figure 2.3 connects to nine unique entities, each with its own access URL in the EDGAR index.

LKQ CORP (Filer) CIK: 0001065696 (see all company filings) IRS No.: **364215970** | State of Incorp.: **DE** | Fiscal Year End: **1231** Type: **S-4** | Act: **33** | File No.: **333-274311** | Film No.: **231232274** SIC: **5010** Wholesale-Motor Vehicles & Motor Vehicle Parts & Supplies	Business Address *500 WEST MADISON STREET* *SUITE 2800* *CHICAGO IL 60661* *312-621-1950*	Mailing Address *500 WEST MADISON STREET* *SUITE 2800* *CHICAGO IL 60661*
KEYSTONE AUTOMOTIVE OPERATIONS OF CANADA INC (Filer) CIK: 0001276076 (see all company filings) IRS No.: **232996445** | State of Incorp.: **DE** | Fiscal Year End: **1231** Type: **S-4** | Act: **33** | File No.: **333-274311-30** | Film No.: **231232296**	Business Address *44 TUNKHANNOCK AVENUE* *EXETER PA 18643* *5706032335*	Mailing Address
DRIVERFX COM INC (Filer) CIK: 0001276078 (see all company filings) IRS No.: **522204596** | State of Incorp.: **DE** | Fiscal Year End: **1231** Type: **S-4** | Act: **33** | File No.: **333-274311-05** | Film No.: **231232270**	Business Address *44 TUNKHANNOCK AVENUE* *EXETER PA 18643* *5706032335*	Mailing Address
A&A AUTO PARTS STORES INC (Filer) CIK: 0001276084 (see all company filings) IRS No.: **233001870** | State of Incorp.: **PA** | Fiscal Year End: **1231** Type: **S-4** | Act: **33** | File No.: **333-274311-08** | Film No.: **231232273**	Business Address *44 TUNKHANNOCK AVENUE* *EXETER PA 18643* *5706032335*	Mailing Address
Potomac German Auto, Inc. (Filer) CIK: 0001596255 (see all company filings) IRS No.: **521637030** | State of Incorp.: **MD** | Fiscal Year End: **1231** Type: **S-4** | Act: **33** | File No.: **333-274311-13** | Film No.: **231232279**	Business Address *500 WEST MADISON STREET* *SUITE 2800* *CHICAGO IL 60661* *(312) 621-1950*	Mailing Address *500 WEST MADISON STREET* *SUITE 2800* *CHICAGO IL 60661*

Figure 2.3: EDGAR Screenshot of a Condensed Form S-4[5] Filing Involving Nine Entities.

This figure depicts a part of the filings details page of a Form S-4 filing encompassing nine entities, each categorized as a Filer. The screenshot, for brevity, only shows five of these entities. This is a prime example of a one-to-many mapping (specifically one-to-nine in this case), where a single unique accession number is associated with nine different entities, each identified by their respective CIKs. As a result, this filing is represented nine times in the EDGAR index, under nine distinct URLs – one corresponding to each entity involved. However, the content of the filing remains consistent across all nine instances.

Dataset 2

Dataset 2 provides mappings between CIKs, sectors, and industries. We've segmented the data into 12 unique sectors, one of which is the 'NaN' category. This category encompasses entities that haven't been designated a specific sector, such as trusts or

[4] `https://www.sec.gov/Archives/edgar/data/1548760/000095010323011806/0000950103-23-011806-index.htm`

[5] `https://www.sec.gov/Archives/edgar/data/1984523/000106569623000087/0001065696-23-000087-index.htm`

insiders often documented in Form 4. This dataset retains historical mappings from 1994 onward, encompassing the associations of over 41,000 CIKs across 11 sectors and 153 industries.

We sourced this data from the CIK/CUSIP/Ticker Mapping API, specifically targeting mappings with a CIK, allowing us to ignore mappings for securities without an EDGAR CIK, like ETFs. The Mapping API offers access to a database detailing information about public companies since 1994, including entity name, ticker, CIK, CUSIPs, exchange, delisting status, security category, sector, industry, SIC code, SIC sector, SIC industry, currency, and location.

By integrating this dataset with Dataset 1, we can connect an EDGAR entity, using its CIK, to the corresponding sector and industry.

Dataset 3 – Supplementary

Dataset 3 comprises a complete list of SIC codes and their associated SEC oversight offices. This data is encapsulated in a single CSV file, structured with three columns: sic_code, office, and industry_title. These columns delineate the relationships between SIC codes, their corresponding industry titles, and the SEC Offices responsible for them.

The dataset includes 444 distinct SIC codes overseen by 12 unique SEC offices, such as the Office of Finance, Office of Manufacturing or the Office of Trade & Services. The mapping between the SIC codes and their SEC Offices was extracted from the SEC's official website[6]. By integrating this dataset with Dataset 1, we can associate EDGAR entities with the relevant SEC offices, using the SIC code as the connecting attribute.

2.2.2 Analysis & Research Workflow

We utilize exploratory and descriptive analyses, complemented by statistical testing, to discern trends and confirm their statistical significance. For time series, like annual filing volumes of specific form types, we apply the Augmented Dickey-Fuller (ADF) test to assess non-stationarity and determine the presence of statistically significant trends. Additionally, we employ the Analysis of Variance (ANOVA) to ascertain if there's a notable difference in the volume of EDGAR filings across months or other groups. To further pinpoint specific months with distinct differences, we implement the Tukey Honest Significant Difference (HSD) test for post-hoc pairwise comparisons.

[6] https://www.sec.gov/corpfin/division-of-corporation-finance-standard-industrial-classification-sic-code-list

For our analysis, we started with variable selection, such as the total count of EDGAR filings over years or the number of entities filing in 2024 categorized by sector. After defining filter functions tailored to these variables, we organized our dataset accordingly, quantified our target variables, and then designed pivot tables. These tables were refined by excluding certain outliers and then visually represented through graphs and charts. Statistical tests further helped discern trends. Our analysis pivoted around variables like specific form types (e.g., 10-K, NT-10x), form type categories (like registration statements), sectors, industries, SEC offices, SIC codes, and various time intervals, as well as combining these variables for a multifaceted view.

2.3 Results & Discussion

In this section, we present the findings derived from our analysis of the EDGAR database over a span of three decades. We detail the trends in filing volumes and the distribution across sectors, industries, time and other factors. The data also provides insights into the usage and relevance of specific form types.

Throughout our analysis, where relevant, we employed statistical tests such as ANOVA and Kruskal-Wallis to verify the robustness of our findings and follow-up pairwise comparisons utilized the Tukey HSD test. Our results consistently rejected the null hypothesis, displaying H-statistics above 100, F-statistics above 200, and p-values below 0.05. We also applied the Augmented Dickey-Fuller test to assess the non-stationarity of our time series data. All identified trends were found to be non-stationary with p-values exceeding 0.05. To maintain clarity and ease of reading, we've refrained from repeatedly noting the statistical significance of our results.

2.3.1 EDGAR Filer Entities & Filing Volumes (1994–2024)

In our initial analysis, we focus on the annual count of filing entities referenced in EDGAR filings and the associated filing volume spanning from 1994 to 2024 (Figure 2.4, Table 2.1). An EDGAR entity is characterized by its reference in a filing and is identifiable via its CIK.

The year 2021 stands out as a pivotal period in EDGAR's history (Table 2.1). Remarkably, it marks the first year since the Sarbanes-Oxley Act enactment in 2002 to witness a pronounced rise in new EDGAR entities. The count of EDGAR entities submitting at least one filing during 2021 peaked at 160,186, surpassing the previous record set in 2006 by around 140,000 entities. Compared to 2020, 2021 saw a 53% increase in new filer registrations, making it the second-highest year for new entities submitting at least one filing. Notably, despite the adversities presented by the COVID-19 pandemic in 2020, the numbers for both new and total EDGAR filers exhibited growth relative to 2019.

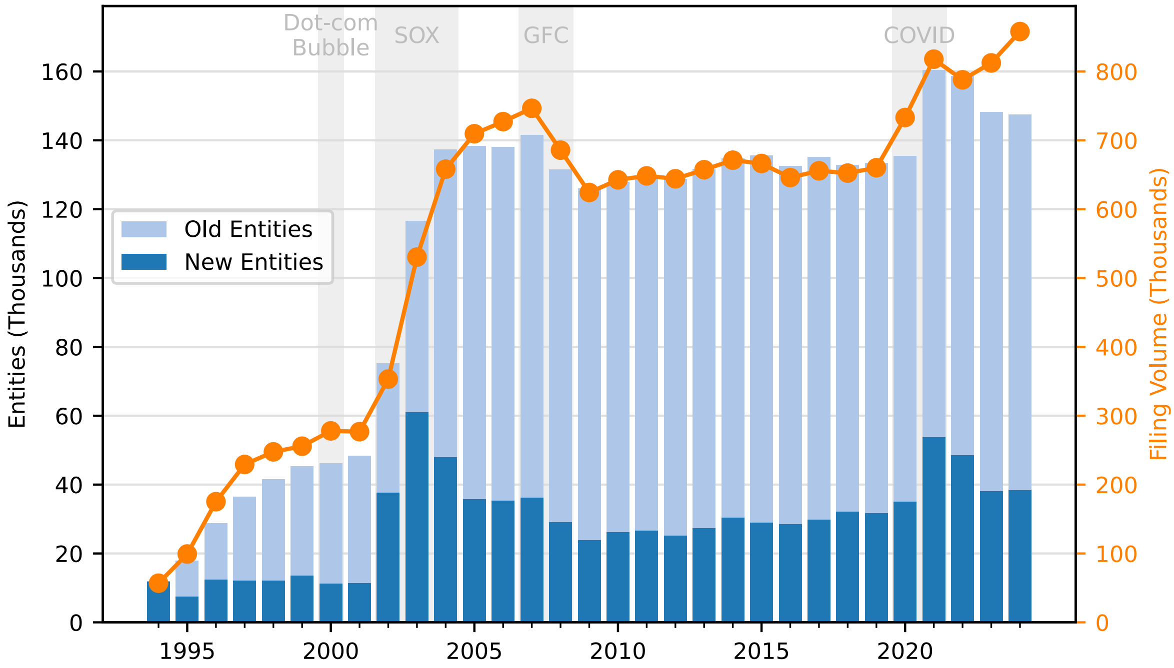


Figure 2.4: Historical EDGAR Filing Entities and Volumes (1994–2024).

This graph depicts the annual number of new and existing entities filing on EDGAR, measured against total annual filing volume. An "old entity" is defined as one that has filed in previous years and continued to file in the year counted, while a "new entity" refers to one filing for the first time in that year, with entities counted by their Central Index Key (CIK). Filing volume is determined by counting all filings by their unique accession number, not by the number of CIKs referenced. Key economic events, including the Dot-com Bubble, SOX enactment, GFC, and COVID-19 pandemic, are marked to demonstrate their correlation with changes in filing behavior and number of active entities on EDGAR. The orange line tracks the annual filing volume, revealing significant spikes associated with these events, and the bars distinguish the count of new versus existing entities filing each year.

Year	New Entities	Old Entities	Total Entities	Filing Volume
1994	11,933	0	11,933	56,831
1995	7,555	10,429	17,984	99,251
1996	12,431	16,336	28,767	175,218
1997	12,211	24,305	36,516	229,286
1998	12,240	29,365	41,605	247,565
1999	13,611	31,726	45,337	255,769
2000	11,347	34,878	46,225	278,020
2001	11,441	36,865	48,306	276,873
2002	37,717	37,459	75,176	353,275
2003	61,118	55,497	116,615	530,458
2004	47,948	89,317	137,265	658,117
2005	35,782	102,502	138,284	709,552
2006	35,336	102,718	138,054	727,144
2007	36,279	105,341	141,620	746,487
2008	29,203	102,391	131,594	685,742
2009	23,894	102,047	125,941	624,217
2010	26,173	101,678	127,851	642,662
2011	26,609	102,147	128,756	648,374
2012	25,263	103,467	128,730	644,156
2013	27,463	104,278	131,741	657,312
2014	30,472	104,237	134,709	671,145
2015	29,024	106,596	135,620	666,349
2016	28,641	103,916	132,557	645,884
2017	29,812	105,338	135,150	655,760
2018	32,172	100,654	132,826	652,411
2019	31,783	101,711	133,494	660,188
2020	35,056	100,338	135,394	733,112
2021	53,885	106,557	160,442	817,848
2022	48,583	110,022	158,605	787,868
2023	38,173	110,064	148,237	812,391
2024	38,415	109,069	147,484	857,926

Table 2.1: Annual Counts of Active SEC Filing Entities and Filing Volume (1994–2024).

The table reports the number of entities filing with the SEC each year, categorized as new entities (first-time filers) and old entities (filers also active in prior years). The total entities column represents the sum of both groups. The final column shows the total number of SEC filings submitted annually by all active entities, including by public companies, insiders, funds, private companies, and investment advisors, and others.

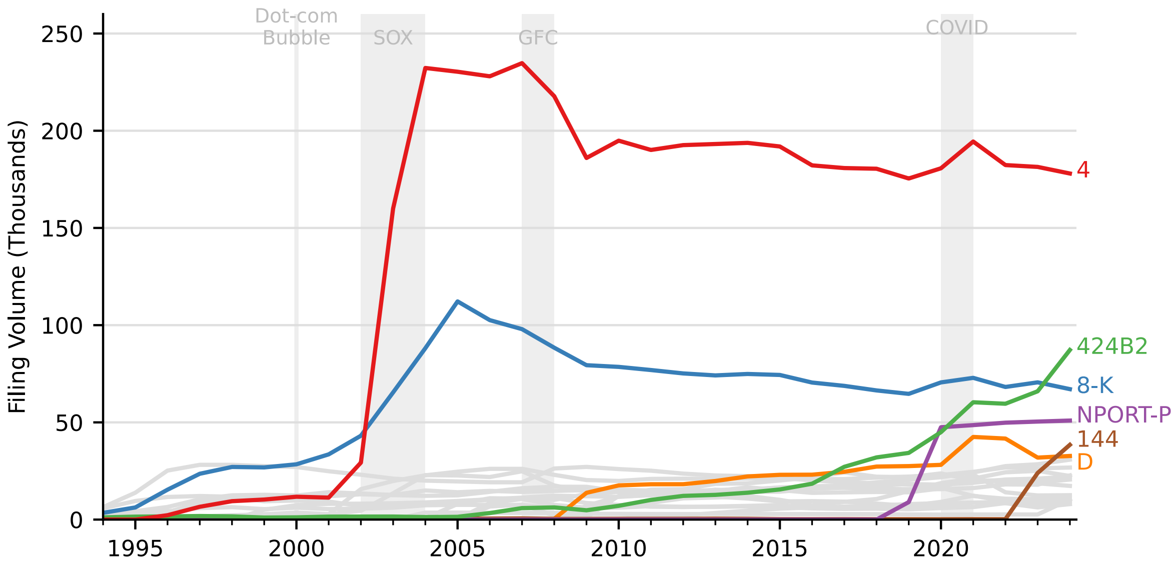


Figure 2.5: Trends in Annual Filing Volume for the Top 20 Most Commonly Filed Form Types (1994–2024).

This chart illustrates the annual filing volumes for the top 20 most commonly filed SEC forms, with significant economic periods and events marked to illustrate their impact on filing behaviors. While not all top 20 form types are colored, those without significant anomalies or trends are represented by gray lines. The Sarbanes-Oxley Act (SOX) has had a discernible impact on Form 4 volumes, related to insider trading disclosures, and Form 8-K, pertaining to material event disclosures, elevating them to the most frequently filed categories. Form 4 and Form 8-K are unscheduled forms, triggered by specific events such as insider stock purchases or significant corporate occurrences like leadership changes. Notably, filing volumes for these forms have been declining annually since their peak prior to the GFC. Meanwhile, Form 424B2 filings, which are prospectuses for complex financial products, have been on a steady rise since 2005, hitting a record high in 2024 and becoming the second largest form type category in EDGAR. These filings are also unscheduled, generated in response to the creation of new financial products. Form N-PORT, filed quarterly by mutual funds and replacing Form N-Q in 2019, emerges as the fourth largest category. Form D filings, representing private offerings as an alternative to IPOs, have shown a year-over-year upward trend since Regulation D's introduction in 2009, with a significant spike in 2022. This suggests an increasing preference for private capital raising over public offerings. The rest of the top 20 form types include Form 144, 3, 6-K, 10-Q, 10-K, 424B3, FWP, 497, 497K, SC 13G, 13F-HR, 10-D, and SEC correspondences.

	Form Type	2024	2023	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011
1	4	178	181	182	194	181	175	180	181	182	192	194	193	193	190
2	424B2	87	66	60	60	45	34	32	27	18	16	14	13	12	10
3	8-K	67	71	68	73	71	65	66	69	70	74	75	74	75	77
4	NPORT-P	51	50	50	49	48	9	0	0	0	0	0	0	0	0
5	144	38	24	0	0	0	0	0	0	0	0	0	0	0	0
6	D	33	32	42	42	28	28	27	25	23	23	22	20	18	18
7	13F-HR	31	28	28	24	22	20	19	18	17	17	15	14	13	13
8	6-K	27	26	27	25	23	22	22	21	21	21	21	21	21	21
9	SC 13G/A	23	18	18	16	16	16	16	16	17	16	16	15	15	14
10	FWP	22	25	24	21	19	15	11	9	7	10	12	11	11	9
11	D/A	22	21	21	19	18	18	17	17	17	16	16	14	13	12
12	497K	21	19	19	20	24	22	22	24	22	20	19	18	16	15
13	10-Q	17	19	20	20	18	18	19	19	20	22	22	23	24	25
14	3	13	12	14	23	16	14	14	14	14	15	17	15	14	15
15	497	11	10	11	12	16	15	16	16	16	14	14	15	14	13
16	N-PX	10	3	3	3	3	3	3	3	3	3	3	3	3	3
17	10-D	9	9	9	9	8	8	7	7	6	6	5	3	2	1
18	CORRESP	9	9	9	12	9	7	8	9	10	9	11	12	11	13
19	424B3	8	8	8	6	6	5	6	6	6	6	7	6	7	7
20	SC 13G	8	6	8	9	6	6	6	6	6	8	7	7	6	7

Table 2.2: Annual Filing Volume (in Thousands, Rounded) for the Top 20 Most Frequently Filed SEC Forms, Ranked by 2024 Volume (2011–2024).

	Form Type	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996
1	4	195	186	218	235	228	230	232	160	29	11	12	10	10	7	2
2	424B2	7	5	6	6	3	1	1	2	2	2	1	1	2	2	2
3	8-K	78	79	88	98	103	112	88	65	43	34	28	27	27	24	15
4	NPORT-P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	144	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6	D	18	14	0	0	0	0	0	0	0	0	0	0	0	0	0
7	13F-HR	12	12	12	11	10	10	9	8	8	8	7	5	0	0	0
8	6-K	20	20	23	26	26	25	23	20	16	3	2	1	0	0	0
9	SC 13G/A	15	17	17	16	14	12	12	13	13	12	12	13	12	10	6
10	FWP	7	6	8	11	8	0	0	0	0	0	0	0	0	0	0
11	D/A	12	7	0	0	0	0	0	0	0	0	0	0	0	0	0
12	497K	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0
13	10-Q	26	27	26	19	19	20	20	21	23	25	27	28	28	28	25
14	3	15	14	18	25	22	23	23	14	4	3	4	3	2	2	1
15	497	15	15	15	14	15	14	15	13	13	14	12	12	12	12	12
16	N-PX	3	3	3	3	3	4	4	0	0	0	0	0	0	0	0
17	10-D	1	1	2	8	7	0	0	0	0	0	0	0	0	0	0
18	CORRESP	14	13	10	11	11	8	1	0	0	0	0	0	0	0	0
19	424B3	6	6	6	8	8	8	7	6	6	5	6	5	6	6	4
20	SC 13G	8	8	11	11	10	9	8	7	7	9	10	9	11	9	4

Table 2.3: Continuation of Annual Filing Volume (in Thousands, Rounded) for the Top 20 Most Frequently Filed SEC Forms, Ranked by 2024 Volume (1996–2010).

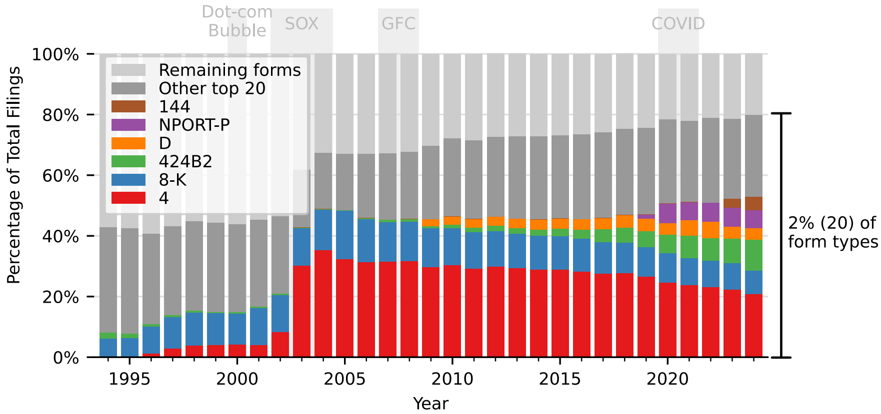


Figure 2.6: Annual Distribution of EDGAR Filings by Form Type (1994–2024).

This stacked bar chart represents the percentage distribution of total filings for each year, segmented by specific EDGAR form types. It highlights the information asymmetry with just 2% (20 different form types) accounting for the majority of filings, even 80% in 2022. The form types are color-coded, with Form 4 (insider trading) and Form 8-K (material events) dominating, especially after the enactment of SOX. The recent rise in Form D (private offerings) and 424B2 filings (complex financial products) is also evident. Key market periods like the Dot-com Bubble, SOX, GFC, and COVID-19 are marked to show their impact on filing behaviors. The remaining form types are grouped as 'Remaining forms' and 'Other top 20', depicted in shades of gray.

	2024	2023	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011
4	20.7%	22.3%	23.1%	23.8%	24.6%	26.6%	27.7%	27.6%	28.2%	28.8%	28.9%	29.4%	29.9%	29.3%
8-K	7.8%	8.7%	8.7%	8.9%	9.6%	9.8%	10.2%	10.5%	10.9%	11.2%	11.2%	11.3%	11.7%	11.9%
424B2	10.2%	8.1%	7.6%	7.4%	6.1%	5.2%	4.9%	4.1%	2.9%	2.3%	2.1%	1.9%	1.9%	1.6%
D	3.8%	3.9%	5.3%	5.2%	3.8%	4.2%	4.2%	3.7%	3.6%	3.5%	3.3%	3.0%	2.8%	2.8%
NPORT-P	5.9%	6.2%	6.3%	5.9%	6.5%	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
144	4.5%	3.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%
Other top 20	26.9%	26.4%	27.9%	26.8%	27.7%	28.6%	28.3%	28.3%	28.0%	27.4%	27.4%	27.1%	26.3%	25.9%
Remaining forms	20.1%	21.3%	21.1%	22.0%	21.5%	24.3%	24.7%	25.8%	26.4%	26.8%	27.2%	27.2%	27.3%	28.5%

Table 2.4: Annual Distribution of EDGAR Filings by Form Type (2011–2024).

	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996
4	30.3%	29.8%	31.8%	31.5%	31.4%	32.5%	35.3%	30.2%	8.3%	4.1%	4.2%	4.0%	3.8%	2.9%	1.3%
8-K	12.2%	12.7%	12.9%	13.1%	14.1%	15.8%	13.4%	12.3%	12.2%	12.1%	10.2%	10.5%	10.9%	10.3%	8.8%
424B2	1.1%	0.8%	0.9%	0.8%	0.5%	0.2%	0.2%	0.3%	0.4%	0.5%	0.4%	0.4%	0.7%	0.8%	0.9%
D	2.7%	2.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
NPORT-P	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
144	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Other top 20	25.8%	24.2%	22.1%	21.8%	21.2%	18.6%	18.4%	19.0%	25.6%	28.7%	29.0%	29.5%	29.4%	29.3%	29.8%
Remaining forms	27.8%	30.3%	32.3%	32.8%	32.8%	32.9%	32.6%	38.1%	53.4%	54.6%	56.1%	55.5%	55.1%	56.7%	59.2%

Table 2.5: Continuation of Annual Distribution of EDGAR Filings by Form Type (1996–2010).

2024 set a new benchmark with the most annual EDGAR filing volume recorded, exceeding 858,000 filings (Table 2.1). This surpassed the previous record established in 2021 and 2006. 2023 maintained this momentum, ranking third in terms of annual filing volume. The surge in filings between 2020 and 2024 can be largely attributed to the integration of N-PORT filings from 2020, replacing N-Q filings, a notable rise in prospectuses, predominantly the 424B2 filings for complex financial products and the introduction of sales notifications under Form 144.

While 2022 observed a marginal decline in new filer registrations, as well as in total filers and annual filing volume compared to 2021, it remains significant. Specifically, 2022 is positioned second in terms of new filer and total filer counts and fourth in annual filing volume statistics.

In 2022, of the 158,606 filers, a substantial 63% (or 100,474 entities) submitted only 1 to 2 filings. Leading the list was JPMorgan Chase (CIK: 19617) with an impressive count of over 18,000 filings. Notably, the top five filers for the year all originated from the banking sector: JPMorgan Chase (CIK: 19617), JPMorgan Chase Financial (CIK: 1665650), Morgan Stanley (CIK: 895421), Morgan Stanley Finance (CIK: 1666268), and CitiGroup (CIK: 831001).

While the top 20 EDGAR form types for 2024 represented merely 2% of the total 893 available form types, they accounted for a staggering 80% of all filings, with insider trading and material event disclosures and different types of prospectuses representing over 60% of the annual volume in 2024 (Figure 2.6, Table 2.4, Table 2.5). In contrast, the remaining 873 form types, which constituted 98%, contributed to only 20% of the year's total filings. This stark disparity highlights a significant information asymmetry: a small fraction of form types dominates the filing landscape.

The top form types can be categorized into the following groups:

- Insider trading disclosures: Form 4 and 3
- Material event disclosures: Form 8-K and 6-K
- Quarterly and annual reports: 10-Q, 10-K
- Prospectuses: 424B2, 424B3, FWP, 497, 497-K
- Private Offerings: Form D
- Institutional investment activity reports: Form SC 13G, 13F-HR, NPORT-P
- Asset-backed securities performance reports: Form 10-D
- Correspondences between the SEC and regulated entities: CORRESP, LETTER

Form 3[7] and Form 4[8] are related to the ownership and trading of securities by corporate insiders. Form 3 is filed by insiders (like officers such as CEOs and CFOs, directors such as board members, and owners of more than 10% of a class of securities) to initially report their holdings in the company's securities after an IPO or in situations where a company is already public and an individual becomes an insider at a later date. It must be filed within ten days of becoming an insider. Form 4 is filed by the same group of insiders to report changes in their ownership, such as purchases, sales, or gifts of company stock. It must be filed within two business days following the day on which a reportable transaction occurred.

Form 8-K[9] and Form 6-K[10] report significant events. Form 8-K is filed by U.S. publicly traded companies, among others, to report major events such as corporate acquisitions, changes in leadership, earnings announcements, and other material developments. It must be filed within four business days of the event. Form 6-K is filed by foreign private issuers whose securities are traded in the U.S. It reports material information, including changes in the company's financial condition and corporate events, that they have already disclosed or are required to disclose in their home countries. The filing frequency varies based on the company's home country reporting requirements and when material events occur.

Form 10-K and 10-Q represent scheduled financial reports filed by publicly traded companies and business development companies incorporated in the U.S. Form 10-K is an annual report providing a comprehensive overview of a company's financial performance, including audited financial statements, management's discussion and analysis (MD&A), information about the company's operations, market risks, internal controls, and legal proceedings. It must be filed within 60 to 90 days (depending on the company's size) after the end of the fiscal year. Form 10-Q is a quarterly report offering a less detailed, unaudited update on a company's financial situation, including interim financial statements and MD&A. This report is filed within 40 to 45 days after the end of each of the first three fiscal quarters. Publicly traded companies incorporated outside the United States but listed on U.S. exchanges are known as foreign private issuers (FPIs). FPIs file Form 20-F for their annual reports, which is analogous to Form 10-K filed by U.S.-incorporated companies. Specifically for Canadian entities listed on U.S. exchanges,

[7] https://www.sec.gov/files/form3.pdf

[8] https://www.sec.gov/files/form4data,0.pdf

[9] https://www.sec.gov/files/form8-k.pdf

[10] https://www.sec.gov/files/form6-k.pdf

Form 40-F[11] serves as the annual report, fulfilling a similar reporting function.

The Forms 424B1 to 424B5 under the 424B series and Form FWP are associated with securities offerings. Forms 424B1 to 424B5 are filed by companies conducting public securities offerings. Form 424B1 is used for the final prospectus in a new securities offering. Form 424B2 is for a preliminary prospectus for complex or structured financial products, such as notes linked to the performance of an unequally-weighted basket of different indices[12]. Form 424B3 is filed for an amended prospectus. Form 424B4 is the final prospectus for an initial public offering (IPO) and Form 424B5 is used for a prospectus supplement, often in shelf offerings. These forms disclose details about the offering, including financial statements, risk factors, and information about the issuer. Form FWP (Free Writing Prospectus) is filed by issuers engaged in public securities offerings, and provides additional or updated information beyond what is included in the prospectus. It can contain marketing materials, investor presentations, or new data about the offering. The timing of its filing depends on when the additional information becomes available during the offering process.

Form 497 and 497K are used by mutual funds and other registered investment companies. Form 497 includes a fund's prospectus and provides information about the fund's investment objectives, strategies, risks, performance, and fees. It is filed when new funds are launched, or there are significant or annual updates to the prospectus. Form 497K is specifically used for filing a summary prospectus with an overview of key fund information. It is filed in similar circumstances as Form 497.

Form D and D/A are related to private-market securities offerings exempt from registration under Regulation D. Form D is filed by private companies raising capital through offerings exempt from registration, such as those in venture capital rounds or offerings conducted by private equity firms. It includes information about the offering, such as the type of security, amount raised, and details about the company and its executives. Form D must be filed within 15 days after the first sale of securities in the offering. Form D/A is an amendment to a previously filed Form D, used to update or correct information. It may be filed as needed when there are significant changes to the offering or corrections to the original Form D information.

Form 10-D is used by issuers of asset-backed securities managing the asset pool (such as financial institutions or special-purpose vehicles), to report periodic distributions and the performance of the underlying asset pool. This includes details about payments of

[11] `https://www.sec.gov/files/form40-f.pdf`

[12] `https://www.sec.gov/Archives/edgar/data/70858/000148105723010339/form424b2.htm`

principal and interest, delinquency rates, prepayments, and any realized losses on the assets. It is typically submitted on a monthly basis, within a specified number of days after the payment or distribution date as outlined in the transaction agreements.

Form SC 13G, 13F-HR, and NPORT-P disclose information about ownership, holdings, and investment strategies of institutional entities (like mutual funds, ETFs, and large investment managers) or individuals/entities that have significant investments in public companies. SC 13D and SC 13D/A report significant ownership in a particular company, 13F-HR discloses broader portfolio holdings of institutional investors, and NPORT-P details the portfolio holdings of registered investment companies. The filing of these forms is triggered by specific events (like acquiring a certain percentage of a company's shares for SC 13D) or regular intervals (like quarterly for 13F-HR and monthly for NPORT-P while publicly disclosed quarterly). SC 13D/A is filed as needed when there are material changes to previously reported information.

Form CORRESP and UPLOAD represent correspondences during the SEC' review and comment process for communication between the SEC and entities under its regulation, such as publicly traded companies. The UPLOAD form type is used for SEC-originated letters sent to filers that may identify issues, concerns, or requests for additional information regarding specific forms, such as 10-K filings, while the CORRESP form type is used for filer response letters to SEC staff that are not part of amended filings. While there is sometimes confusion between UPLOAD and LETTER types, they represent the same information. UPLOAD represents the form type, whereas LETTER refers to the file type of the form type uploaded to EDGAR.

Recognizing these dominant form types holds relevance for credentialing bodies like FINRA and the CFA Institute, as well as for educational institutions. Their prominence underscores the need to update course materials for the Series X exams and the Chartered Financial Analyst programs, ensuring individuals are well-informed about the roles and implications these forms have in financial markets.

With the total number of annually active entities on EDGAR plateauing from 1999 to 2001, and filing volume also stagnating in the same timeframe, the introduction of the Sarbanes-Oxley Act of 2002 (SOX) can be seen as the number one growth catalyst for the EDGAR system for subsequent years (Figure 2.5).

From 2002 to 2004, SOX introduced a series of reforms that was enacted in response to a number of high-profile corporate and accounting scandals, including those involving Enron, WorldCom, and Tyco. These scandals, which cost investors billions of dollars when the share prices of the affected companies collapsed, eroded public trust in the nation's securities markets. SOX's primary objective is to restore investor confidence by implementing stricter regulations for publicly traded companies, their executives, and their auditors. The law introduced major changes to the regulation of financial practice and corporate governance.

During the introduction of SOX, a significant growth in the number of EDGAR filers is visible, which escalated from about 50,000 in 2001 to nearly 140,000 by 2004, marking a 180% increase in just three years. The peak in the influx of new entities on EDGAR was in 2003, reaching the highest level in the three decades from 1994 to 2024. Notably, 2002 saw a remarkable 230% growth in new filers compared to 2001. Concurrently, in the period from 2001 to 2004, the annual volume of disclosed filings on EDGAR surged from just under 300,000 in 2001 to over 650,000 by the end of 2004, reflecting a 116% increase in filing volume.

SOX had a particular strong impact on the number of Form 4 filings, increasing from just over 11,000 filings in 2001 to its peak of 232,282 filings in 2004, representing a 2000% increase. Since SOX, Form 4 filings represent the largest portion of filings per year, consistently representing between 25–30% of all filings disclosed to the public per year. Contrasting to the increasing number of EDGAR entities from 2009 to 2024, the number of Form 4 filings declines in the same timeframe (Figure 2.5, Table 2.2, Table 2.3).

The impact of SOX on Form 8-K filing volume is evident as the annual volume of 8-K filings, which stood at 33,543 in 2001, escalated to its all-time high of 112,000 in 2005, marking a 240% increase. Similar to the trends observed in Form 4 filings, the number of annual 8-K filings has been on a consistent decline since 2005.

The global financial crisis had a visible impact on EDGAR's metrics. For the first time since EDGAR's inception in 1994, the total number of entities declined during 2008 and 2009 (Table 2.1). A consistent year-on-year decrease in new filers was observed from 2003 to 2009. Post-2009, from 2010 onwards, there has been a gradual year-on-year increase in both new and overall entities on EDGAR.

In 2024, 424B2 prospectuses for complex financial products emerged as the second largest filing category, trailing only behind Form 4, and marking its highest annual filing count ever recorded for this category. This trend suggests that financial institutions are creating new complex financial products at an unprecedented rate. Such a significant increase in the registration of complex financial products serves as a compelling motivation for future research. It raises questions about whether this rapid proliferation of complex financial products could be a red flag, potentially harboring inherent risks to the financial markets.

In 2022, Form D notably emerged as the fifth largest form type category by annual volume, following closely behind N-PORT filings. This is particularly interesting given that Form D is utilized for raising capital privately, circumventing the need for an IPO. This method offers a faster, more streamlined approach with less administrative burden for securing funds from investors such as venture capital funds or private equity firms. The surge in Form D filings since its introduction in 2009, under Regulation D, indicates a significant shift in capital-raising strategies. This trend is underscored by the declining number of S-1 filings annually from 2011 to 2019, with 2019 marking the lowest volume of S-1 filings recorded. Notably, the decrease in S-1 filings, starting in 2011, coincides with the initial adoption of Form D filings in 2010. This correlation suggests that companies are increasingly favoring private funding rounds over public offerings, reflecting a broader shift in how companies are choosing to finance their growth and operations. This change in capital-raising patterns merits further investigation to understand its implications for market dynamics, investor behavior, and the future of public versus private capital markets.

2.3.2 Seasonal Patterns

In our examination of seasonal filing patterns, we analyze the monthly filing volumes of the top 20 most frequently filed EDGAR form types, as previously introduced, spanning the years 2004 to 2024 (Figure 2.7, Table 2.6, Table 2.7). For the sake of clarity and focus in our analysis, we have excluded the years 1994 to 2003. This decision is due to the potential for outliers in these earlier years and their reduced relevance to our study, particularly in light of the substantial impact that SOX had on the disclosure landscape, as highlighted earlier.

Figure 2.7 presents a visualization of the monthly distribution of filings from 2004 to 2024. It consists of 20 subplots, each representing the distinct monthly filing pattern of a specific form type. The topmost subplot is dedicated to illustrating the overall monthly filing volume across the entire EDGAR system. In each subplot, two standard deviation bands are displayed, representing the lower and upper boundaries that encompass 95% of all values per month (with the lower boundary being above 5% and the upper boundary below 95%). A blue line in each subplot indicates the average number of filings filed per month, facilitating the identification and analysis of trends.

Our analysis of the entire EDGAR system's filings reveals a distinct seasonal pattern: February consistently emerges as the month with the highest number of filings disclosed

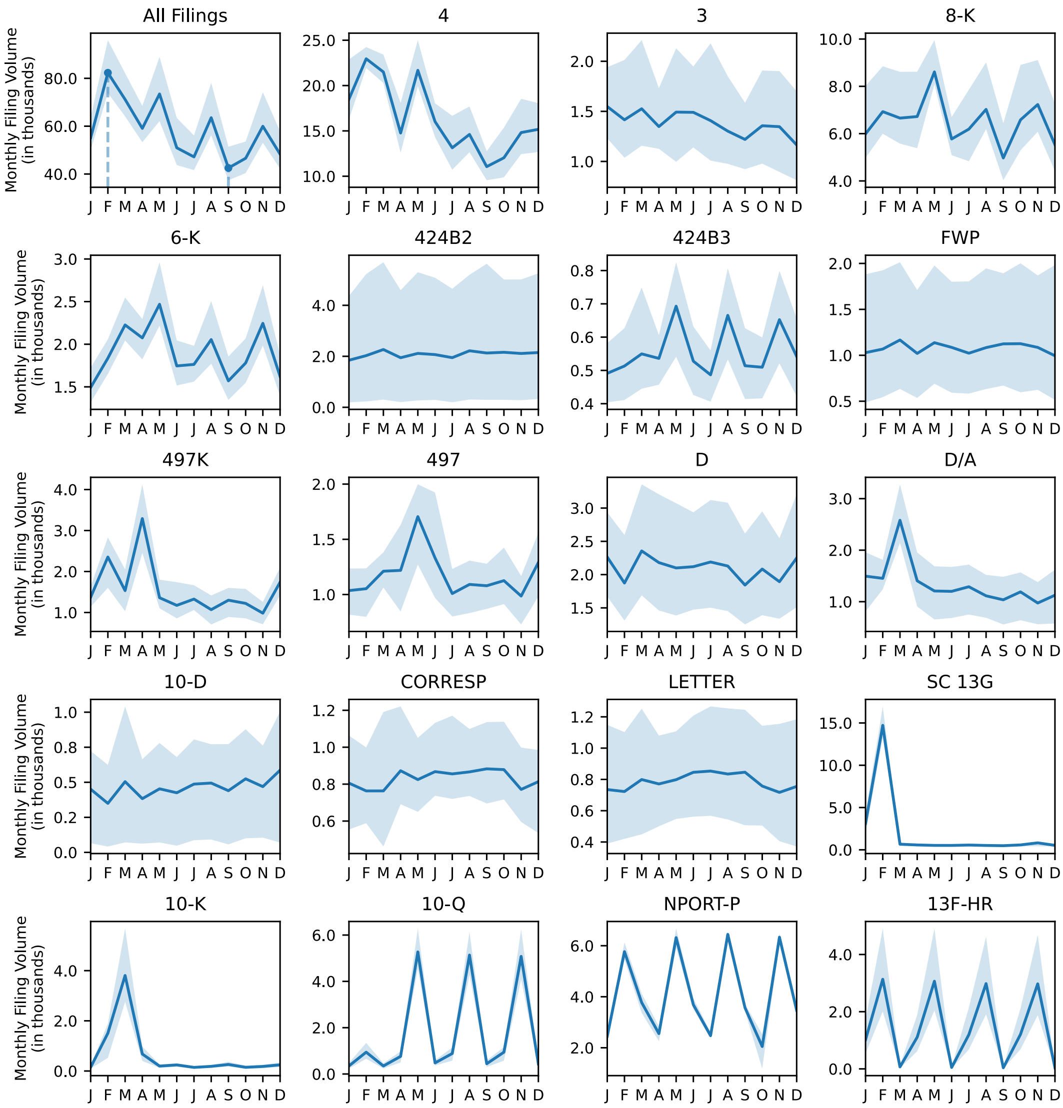


Figure 2.7: Seasonal Trends in Monthly Filing Volumes for the Top 20 Most Frequently Filed Form Types (2004–2024).

The series of line graphs presents the average monthly filing volume for the top 20 most common form types filed with the SEC over two decades, with the top left graph providing an overview of the entire EDGAR system's monthly volume, not just the aggregate of the top 20. Each of the other subplots represents the seasonal pattern of a specific form type, including insider trading forms (Form 4 and Form 3), material event disclosures (Form 8-K and 6-K), quarterly and annual reports (10-K and 10-Q), prospectuses (424B2, 424B3, FWP, 497), private offerings (Form D), institutional investment reports (13F-HR and SC 13G), asset-backed securities performance reports (Form 10-D), and SEC correspondences (CORRESP and LETTER). SC 13G includes SC 13G/A filings. The blue line in each plot represents the mean monthly filing volume, while the blue-tinted shaded areas represent the range within two standard deviations, capturing 95% of all data points. The timeframe from 1994 to 2003 is excluded due to significant shifts in filing patterns following the introduction of SOX. Forms are categorized as either scheduled, with fixed SEC reporting deadlines, including 10-K, 10-Q, NPORT-P, 13F-HR, and SC 13G, or unscheduled, triggered by specific events like a CEO change disclosed via Form 8-K. A temporal asymmetry is observed, with February seeing the highest volume of disclosures, and September the least. Notably, insider trading activity, as indicated by the peak in Form 4 filings during February, suggests insiders may strategically time their trades ahead of the March spike in Form 10-K filings, despite the regulations of Rule 10b5-1.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2004	51.9	73.8	66.5	56.1	61.0	47.4	44.5	58.2	41.0	45.1	61.4	51.1
2005	55.7	75.6	72.2	58.0	72.1	52.4	49.0	68.7	45.9	46.0	61.8	51.9
2006	60.5	78.6	71.9	56.1	78.1	54.7	47.4	69.3	43.7	50.7	65.5	50.6
2007	63.8	82.2	70.6	62.1	79.9	53.5	52.9	71.8	41.6	52.9	67.2	48.1
2008	59.7	84.4	66.6	58.5	72.7	47.9	48.9	60.5	40.6	46.7	53.8	45.6
2009	48.4	74.0	69.8	52.2	62.1	43.9	42.7	54.5	38.4	41.4	52.8	44.0
2010	49.9	70.4	71.9	55.7	65.0	47.4	42.6	56.7	38.7	40.8	56.4	47.1
2011	52.0	76.7	70.3	55.2	71.1	46.4	42.6	59.9	38.8	39.1	53.8	42.5
2012	50.1	80.6	64.1	55.9	70.0	43.4	42.6	58.8	37.4	42.3	54.1	45.1
2013	52.7	77.3	66.0	58.1	71.5	43.8	46.5	57.1	40.0	46.1	54.4	43.7
2014	54.6	79.9	66.3	57.9	69.8	47.9	46.9	57.5	41.2	46.1	53.9	49.0
2015	52.5	82.5	70.5	56.9	68.1	50.5	47.5	57.0	39.4	42.8	53.6	45.1
2016	50.1	83.0	65.9	53.3	67.7	46.6	41.6	58.4	41.2	40.4	54.4	43.2
2017	52.6	78.6	70.7	53.3	72.3	49.2	40.5	58.4	39.6	43.5	54.8	42.2
2018	54.1	80.4	65.3	55.5	71.0	48.5	42.8	58.2	36.5	46.3	53.0	40.9
2019	49.9	79.7	64.1	57.3	74.1	45.2	45.8	56.2	39.7	43.4	58.7	46.0
2020	53.6	88.5	74.1	61.0	72.7	56.9	49.5	64.7	46.0	49.4	61.6	55.2
2021	58.5	96.0	84.4	68.5	80.6	63.5	56.1	72.8	51.8	53.6	74.2	57.9
2022	61.7	95.3	82.2	67.5	83.8	58.0	48.5	74.8	49.4	49.0	67.0	50.8
2023	59.2	93.1	81.3	65.5	90.4	64.1	52.0	78.1	50.7	53.3	68.7	55.8
2024	66.6	98.0	80.0	75.8	89.0	59.1	59.4	82.9	51.4	59.2	78.0	58.6
Total	1,158.1	1,728.6	1,494.7	1,240.4	1,543.0	1,070.3	990.3	1,334.5	893.0	978.1	1,259.1	1,014.4
Mean	55.1	82.3	71.2	59.1	73.5	51.0	47.2	63.5	42.5	46.6	60.0	48.3
Median	53.6	80.4	70.5	57.3	72.1	48.5	46.9	58.8	41.0	46.1	56.4	47.1
SD	5.1	7.7	6.1	5.9	7.8	6.3	4.9	8.3	4.8	5.2	7.5	5.3
Min	48.4	70.4	64.1	52.2	61.0	43.4	40.5	54.5	36.5	39.1	52.8	40.9
Max	66.6	98.0	84.4	75.8	90.4	64.1	59.4	82.9	51.8	59.2	78.0	58.6

Table 2.6: Total Number of Regulatory Filings on EDGAR (in Thousands) by Year and Month, with Monthly Descriptive Statistics (2004–2024).

each year from 2004 to 2024, with a median of 80.4 thousand filings published that month. Conversely, the disclosure activity across all entities, as indicated by monthly filing volume, is consistently at its lowest in September for the analyzed time frame, followed by subdued activity in October and July. Further peaks in filing volume are discernible in May, August, and November.

In analyzing the top 20 most frequently used form types on EDGAR, we can discern the underlying reasons for the high filing activity in February and the low in September. Notably, Form 4, which accounts for 29% of the annual filing volume, exhibits the highest activity in February and the lowest in September. This is particularly interesting because Form 4 represents an unscheduled disclosure type, triggered by insider trading activity rather than a recurring regulatory deadline. This observed trend, where corporate insiders exhibit heightened activity in February – one month preceding the release of annual reports on Form 10-K in March – implies that insiders, who are privy to annual results, might be timing their buying and selling activities accordingly. This pattern occurs

Form Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
4	18.5	23.0	21.5	14.8	21.7	16.0	13.1	14.6	11.1	12.0	14.8	15.2
	(17.7)	(23.1)	(21.0)	(14.0)	(20.9)	(16.1)	(12.8)	(14.1)	(10.8)	(11.5)	(14.3)	(14.4)
3	1.5	1.4	1.5	1.3	1.5	1.5	1.4	1.3	1.2	1.4	1.3	1.2
	(1.4)	(1.3)	(1.4)	(1.2)	(1.4)	(1.4)	(1.2)	(1.2)	(1.1)	(1.3)	(1.1)	(1.1)
8-K	6.0	6.9	6.7	6.7	8.6	5.8	6.2	7.0	5.0	6.6	7.2	5.5
	(5.5)	(6.6)	(6.5)	(6.5)	(8.5)	(5.6)	(5.9)	(6.9)	(4.5)	(6.0)	(6.9)	(5.0)
6-K	1.5	1.8	2.2	2.1	2.5	1.7	1.8	2.1	1.6	1.8	2.2	1.6
	(1.4)	(1.8)	(2.1)	(2.0)	(2.4)	(1.7)	(1.7)	(2.0)	(1.5)	(1.7)	(2.1)	(1.5)
424B2	1.8	2.0	2.3	1.9	2.1	2.1	1.9	2.2	2.1	2.2	2.1	2.1
	(1.0)	(1.2)	(1.2)	(1.1)	(1.1)	(1.2)	(1.1)	(1.2)	(1.3)	(1.1)	(1.2)	(1.2)
424B3	0.5	0.5	0.5	0.5	0.7	0.5	0.5	0.7	0.5	0.5	0.7	0.5
	(0.5)	(0.5)	(0.5)	(0.5)	(0.7)	(0.5)	(0.5)	(0.7)	(0.5)	(0.5)	(0.7)	(0.5)
FWP	1.0	1.1	1.2	1.0	1.1	1.1	1.0	1.1	1.1	1.1	1.1	1.0
	(0.9)	(0.9)	(0.9)	(1.0)	(1.0)	(0.9)	(0.9)	(0.8)	(0.9)	(1.0)	(0.9)	(0.7)
497K	1.4	2.4	1.5	3.3	1.4	1.2	1.3	1.1	1.3	1.2	1.0	1.7
	(1.4)	(2.5)	(1.5)	(3.4)	(1.4)	(1.2)	(1.4)	(1.1)	(1.4)	(1.3)	(1.0)	(1.8)
497	1.0	1.1	1.2	1.2	1.7	1.3	1.0	1.1	1.1	1.1	1.0	1.3
	(1.0)	(1.1)	(1.2)	(1.2)	(1.8)	(1.2)	(1.0)	(1.1)	(1.1)	(1.1)	(1.0)	(1.3)
D	2.3	1.9	2.4	2.2	2.1	2.1	2.2	2.1	1.8	2.1	1.9	2.2
	(2.2)	(1.8)	(2.2)	(1.9)	(1.9)	(2.0)	(2.0)	(2.0)	(1.8)	(2.0)	(1.9)	(2.2)
D/A	1.5	1.5	2.6	1.4	1.2	1.2	1.3	1.1	1.0	1.2	1.0	1.1
	(1.7)	(1.5)	(2.4)	(1.4)	(1.2)	(1.3)	(1.3)	(1.2)	(1.2)	(1.3)	(1.0)	(1.2)
10-D	0.5	0.3	0.5	0.4	0.5	0.4	0.5	0.5	0.4	0.5	0.5	0.6
	(0.5)	(0.4)	(0.5)	(0.4)	(0.5)	(0.5)	(0.5)	(0.7)	(0.5)	(0.6)	(0.5)	(0.7)
CORRESP	0.8	0.8	0.8	0.9	0.8	0.9	0.9	0.9	0.9	0.9	0.8	0.8
	(0.8)	(0.8)	(0.7)	(0.9)	(0.9)	(0.8)	(0.8)	(0.9)	(0.9)	(0.9)	(0.8)	(0.8)
LETTER	0.7	0.7	0.8	0.8	0.8	0.8	0.9	0.8	0.8	0.8	0.7	0.8
	(0.6)	(0.6)	(0.7)	(0.7)	(0.7)	(0.8)	(0.9)	(0.8)	(0.9)	(0.7)	(0.6)	(0.7)
SC 13G + /A	3.1	14.7	0.7	0.6	0.5	0.5	0.6	0.5	0.5	0.6	0.8	0.5
	(3.3)	(14.4)	(0.6)	(0.6)	(0.5)	(0.5)	(0.5)	(0.5)	(0.4)	(0.5)	(0.5)	(0.5)
10-K	0.2	1.5	3.8	0.7	0.2	0.2	0.1	0.2	0.3	0.1	0.2	0.2
	(0.1)	(1.7)	(3.5)	(0.6)	(0.2)	(0.3)	(0.1)	(0.2)	(0.3)	(0.1)	(0.2)	(0.2)
10-Q	0.3	0.9	0.3	0.8	5.3	0.5	0.9	5.1	0.4	0.9	5.1	0.4
	(0.3)	(0.9)	(0.3)	(0.7)	(5.2)	(0.5)	(0.9)	(5.1)	(0.4)	(1.0)	(5.0)	(0.4)
NPORT-P	2.5	5.8	3.8	2.6	6.3	3.7	2.5	6.4	3.6	2.0	6.3	3.5
	(2.5)	(5.7)	(3.8)	(2.6)	(6.4)	(3.6)	(2.4)	(6.4)	(3.6)	(2.4)	(6.4)	(3.5)
13F-HR	1.0	3.1	0.1	1.1	3.1	0.0	1.2	3.0	0.0	1.2	3.0	0.0
	(0.8)	(2.9)	(0.1)	(1.0)	(2.9)	(0.0)	(1.0)	(2.8)	(0.0)	(1.0)	(2.8)	(0.0)

Table 2.7: Mean and Median (in Brackets) Monthly Filing Volume (in Thousands) for the Top 20 Most Frequently Filed SEC Form Types (January–December, 2004–2024).

despite the existence of regulations like Rule 10b5-1 and established insider trading plans, suggesting that insiders may strategically align their transactions with their knowledge of impending annual financial disclosures.

The spike in filings during February can also be attributed to certain scheduled filings. For instance, NPORT-P and 13F-HR forms, which are filed on a quarterly basis, peak every February, May, August, and November. Additionally, SC 13G and SC 13G/A forms, filed annually, also contribute to the increased filing volume in February. These findings highlight a blend of both scheduled and unscheduled filing activities influencing the distinct seasonal pattern observed in the EDGAR filing data.

Form 10-Q filings disclosing quarterly financial results peak in May, August and November as expected due to their scheduled nature by the related regulation. These months are particularly significant as they correspond with the release of quarterly financial information, potentially leading to heightened market volatility – an area ripe for future research.

Form 8-K and 6-K, used for material event disclosures, both exhibit a similar monthly activity pattern from 2004 to 2024, with May experiencing the highest number of filings and September the least, contributing to the overall reduced disclosure activity in that month. Given that both Form 8-K and 6-K are unscheduled and triggered only by the occurrence of specific events, it becomes evident that a deeper investigation is warranted to uncover the potential underlying reasons for their filing patterns. These forms are pivotal in disclosing material events and developments in a company's operations, and understanding the nuances behind their filing frequency and timing could reveal significant insights into corporate behavior and market dynamics.

In contrast, filings of 424B2 prospectuses, which cover complex financial products issued by financial institutions, display remarkable stability throughout the year. The volume of these filings remains consistent across all months, showing no significant peaks or troughs. This indicates a continuous and steady issuance of complex financial products.

On the other hand, 424B3 prospectuses exhibit a pronounced pattern, with increased filing volumes in May, August, and November. This pattern mirrors the 10-Q filing schedule, suggesting that companies frequently update their prospectuses in line with the publication of new quarterly results. This finding illustrates how corporate disclosures are often closely tied to and influenced by their financial reporting cycles.

The stability observed in SEC correspondences, as evidenced by the consistent monthly filings of CORRESP and LETTER forms, stands in contrast to the variability noted in other filings. This constancy is unexpected, particularly considering that the SEC reviews various filings and contends with their monthly fluctuations. Given this context, one might anticipate that SEC correspondences would reflect some degree of correlation with major form types like 10-K, 10-Q, or 8-K filings. The absence of such a pattern in SEC correspondences suggests a need for further exploration to understand how the SEC manages its review processes amidst the monthly volatility of other filings. The lack of correlation between the monthly pattern of SEC correspondences and major form types may be attributed to potential resource constraints within the SEC staff. If the SEC staff levels do not scale to accommodate the peaks in filing volume observed in our analysis, it would mean that the same workforce is handling varying volumes of filings throughout the year. Consequently, with a constant staff size, the SEC produces a consistent volume of correspondences, regardless of the significant fluctuations in filing volumes. This hypothesis suggests that the SEC's capacity to engage in correspondence is more a function of its internal resources than the external ebb and flow of filing volumes. This insight points towards the need for a deeper investigation into the SEC's resource allocation and operational strategies in managing its review and correspondence processes.

A pattern of temporal information dissemination asymmetry is observed, wherein 30.2% of EDGAR filings typically appear in the first quarter of the year, 26.2% in the second quarter, 21.6% in the third, with the residual 21.9% distributed over the fourth quarter.

Since information is a key driver of stock prices, future studies might explore the correlation between monthly EDGAR filing volumes, specific filing types, and their influence on broader indices such as the S&P 500, as well as more sector-specific indices. A pertinent question would be whether indices display amplified volatility—potentially triggered by heightened information disclosures—in the year's first half compared to its latter half. With September, July, and October earmarked as months of reduced activity, it would be insightful to determine if these months also correspond to diminished volatility in major indices.

2.3.3 Sectoral Analysis

In analyzing the breakdown of annually active EDGAR entities and filing volumes by sector (Figure 2.8, Figure 2.9), we can discern trends and patterns across different sectors. However, it's crucial to understand the context behind the data before delving into the results. Firstly, it's important to note that while in Figure 2.9 the annual filing volume across all sectors barely exceeded 500,000 in 2021, and the total number of filings indicated in Figure 2.4 exceeded 800,000 for that same year, these values are not directly comparable. The filing volume per sector, though a part of the total filing volume on EDGAR, doesn't equate to a 500,000 to 800,000 ratio. This is because the sector-wise filing volume is calculated by associating CIKs with specific sectors, rather than using accession numbers, which uniquely identify each filing. As mentioned in the Data & Methods section, a single filing can be linked to multiple CIKs.

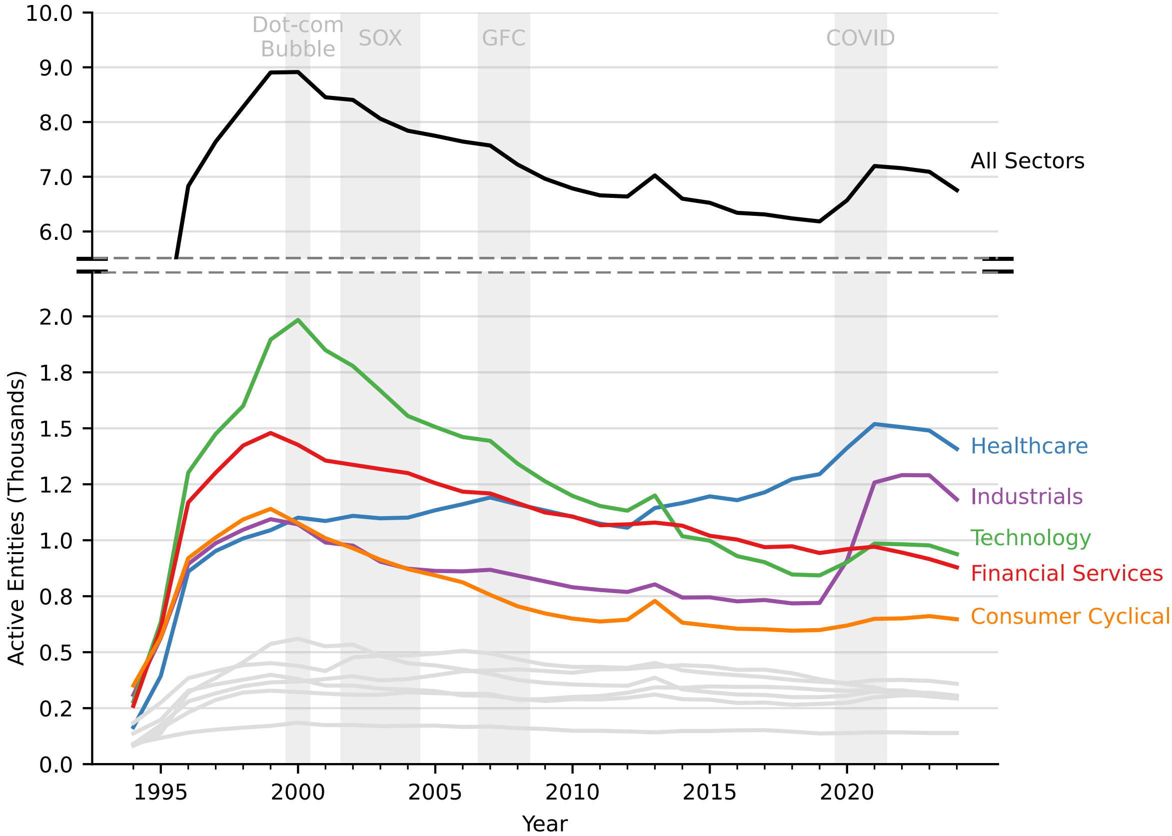


Figure 2.8: Distribution of Active Public Company Filing Entities by Sector (1994–2024).

This line chart illustrates the trends in the number of entities within 11 key sectors over a period marked by significant economic events, with each entity represented by its unique CIK and counted upon filing at least once in the respective year. The black line represents the aggregate count across all sectors, which shows a peak around the Dot-com era followed by a general decline. Individual sectors such as Healthcare, Industrials, Technology, Financial Services, and Consumer Cyclical are traced, highlighting the unique trajectories of each. Notably, the Healthcare sector demonstrates a steady rise in entity numbers, contrasting with the overall trend of declining entities across most sectors. The gray lines represent other sectors not highlighted, which collectively contribute to the total number of entities.

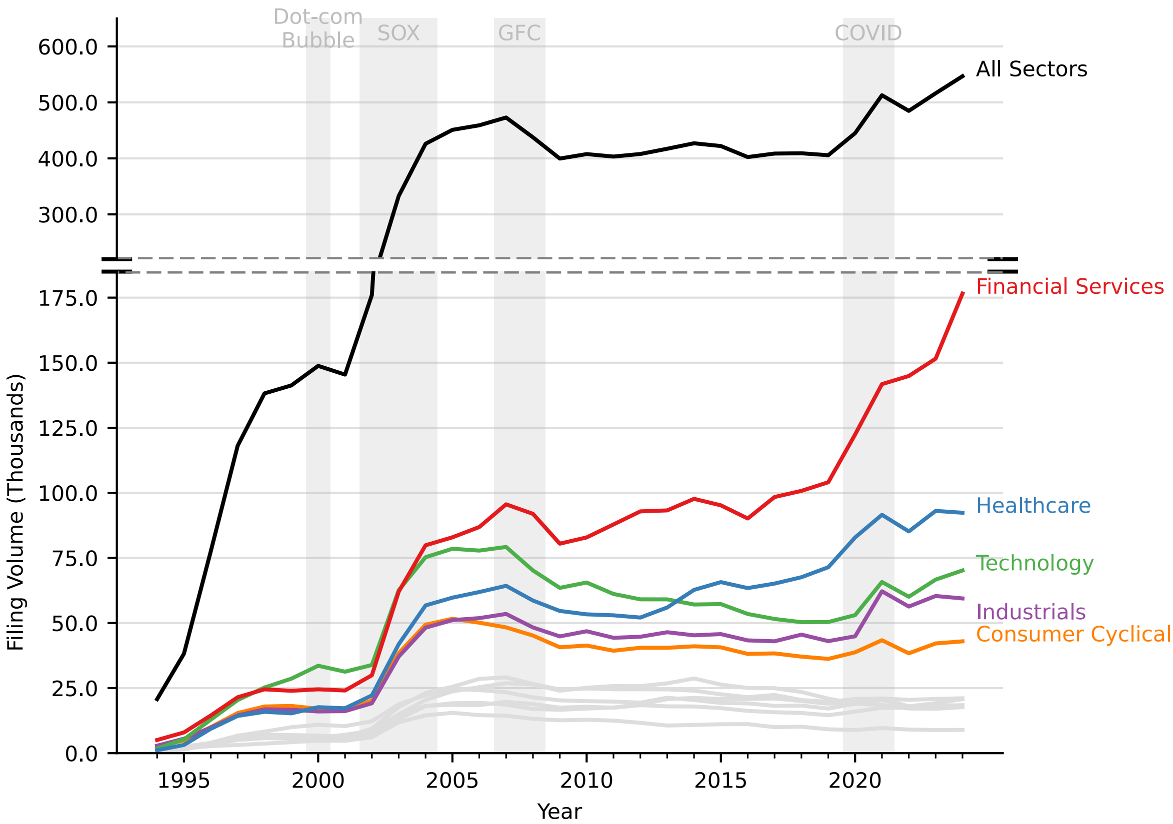


Figure 2.9: Annual Filing Volume of Public Companies by Sector (1994–2024).

This line graph showcases the filing volume trends across various sectors over nearly three decades. Filings are tallied based on the referenced entities' Central Index Keys (CIKs) within the documents, rather than their unique accession numbers to allow mapping of filings to their respective sectors by using the CIK of the entity as link between filing and sector. The black line represents the combined annual filing volume for all sectors, with significant growth observed during key economic events, marked as the Dot-com Bubble, SOX, GFC, and COVID-19 pandemic. The financial services sector, represented in red, displays a marked increase in filing volume over time, becoming the most active sector. The healthcare, technology, industrials, and consumer cyclical sectors are also plotted, each following unique trajectories through the years. The technology sector, while initially leading pre-dot-com bubble, experienced a post-bubble decline and by 2024 is seen as the third-largest contributor. The figure highlights the dynamic nature of filing volumes in response to economic and regulatory changes.

	2024	2023	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011
Healthcare	1,408	1,490	1,505	1,519	1,412	1,295	1,273	1,214	1,179	1,196	1,166	1,144	1,056	1,074
Industrials	1,182	1,290	1,291	1,258	910	720	718	733	727	745	744	803	769	778
Technology	938	977	982	985	902	843	847	902	929	998	1,018	1,200	1,132	1,153
Financial Services	879	916	945	971	960	943	973	969	1,003	1,020	1,065	1,079	1,071	1,067
Consumer Cyclical	647	661	651	649	619	599	596	602	605	618	632	729	645	637
Basic Materials	359	372	376	375	362	368	376	389	397	406	419	452	430	434
Communication Services	306	319	314	326	306	299	299	309	311	321	334	386	350	352
Consumer Defensive	302	309	307	299	274	269	265	275	273	288	290	311	296	289
Real Estate	302	315	329	330	328	332	341	345	345	347	341	343	319	304
Energy	293	303	314	342	358	379	406	422	421	437	442	435	425	423
Utilities	139	139	142	142	139	137	145	152	151	148	148	142	146	149
Total	6,755	7,091	7,156	7,196	6,570	6,184	6,239	6,312	6,341	6,524	6,599	7,024	6,639	6,660

Table 2.8: Number of Active Public Company Filers by Sector and Year, Ranked by 2024 Count (2011–2024).

	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996
Healthcare	1,105	1,133	1,160	1,191	1,161	1,134	1,101	1,098	1,109	1,086	1,101	1,045	1,008	952	860
Industrials	790	816	842	868	861	863	873	904	976	990	1,071	1,094	1,047	987	894
Technology	1,198	1,263	1,342	1,444	1,461	1,506	1,555	1,668	1,778	1,849	1,984	1,896	1,600	1,475	1,302
Financial Services	1,106	1,124	1,166	1,209	1,217	1,255	1,300	1,318	1,337	1,356	1,426	1,479	1,423	1,302	1,169
Consumer Cyclical	650	673	705	756	812	843	871	913	964	1,009	1,076	1,140	1,093	1,012	920
Basic Materials	434	445	468	495	506	494	485	484	477	416	439	451	442	415	384
Communication Services	356	363	376	403	423	441	451	484	534	526	560	537	455	385	320
Consumer Defensive	290	282	292	306	307	326	333	337	351	351	380	399	377	356	328
Real Estate	300	291	287	313	314	317	321	310	310	315	322	328	320	287	231
Energy	408	416	424	418	415	397	380	374	393	380	369	365	348	316	280
Utilities	149	157	161	168	166	172	171	170	175	174	185	171	163	154	141
Total	6,786	6,963	7,223	7,571	7,643	7,748	7,841	8,060	8,404	8,452	8,913	8,905	8,276	7,641	6,829

Table 2.9: Continuation of Number of Active Public Company Filers by Sector and Year, Ranked by 2024 Count (1996–2010).

	2024	2023	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011
Financial Services	176,568	151,548	144,896	141,774	122,423	104,072	100,765	98,399	90,149	95,209	97,701	93,258	92,887	87,863
Healthcare	92,362	93,070	85,181	91,571	82,874	71,420	67,591	65,151	63,415	65,710	62,723	55,957	52,094	52,958
Technology	70,193	66,676	60,079	65,761	53,026	50,379	50,320	51,551	53,477	57,273	57,114	59,081	59,097	61,160
Industrials	59,442	60,377	56,324	62,185	44,897	43,040	45,540	42,975	43,320	45,734	45,276	46,455	44,721	44,318
Consumer Cyclical	42,959	42,154	38,358	43,387	38,740	36,212	37,086	38,323	38,121	40,652	41,054	40,475	40,485	39,351
Basic Materials	21,093	20,831	20,483	21,025	20,801	19,716	20,399	22,366	21,512	22,602	24,018	24,489	24,547	24,538
Communication Services	20,720	19,213	17,975	20,437	19,004	17,130	18,276	18,166	19,124	19,288	20,396	21,293	19,402	19,805
Real Estate	18,440	18,018	17,928	21,040	19,460	18,986	20,030	20,979	20,808	20,804	21,220	20,754	18,778	17,520
Consumer Defensive	18,305	18,447	17,393	17,456	15,943	14,577	15,435	15,725	16,238	17,299	17,936	17,986	18,417	17,467
Energy	17,688	17,041	17,154	18,525	18,990	20,949	23,501	24,961	25,034	26,356	28,696	26,814	25,740	25,781
Utilities	8,937	8,897	9,066	9,630	8,889	9,165	10,160	10,000	11,158	11,113	10,843	10,610	11,651	12,505
Total	546,707	516,272	484,837	512,791	445,047	405,646	409,103	408,596	402,356	422,040	426,977	417,172	407,819	403,266

Table 2.10: Number of Regulatory Disclosures by Sector and Year, Ranked by 2024 Count (2011–2024).

	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996
Financial Services	82,880	80,480	91,936	95,593	86,866	82,919	79,864	62,040	29,928	24,068	24,517	23,955	24,513	21,471	14,473
Healthcare	53,329	54,634	58,628	64,292	61,903	59,737	56,755	41,854	22,177	17,249	17,659	15,297	15,882	14,297	9,364
Technology	65,556	63,505	70,184	79,219	77,824	78,522	75,288	62,575	33,856	31,295	33,578	28,613	25,214	20,389	12,880
Industrials	46,836	44,856	48,292	53,487	51,875	51,118	48,207	37,081	19,166	16,140	16,021	16,674	16,823	14,503	9,808
Consumer Cyclical	41,334	40,681	45,176	48,319	50,127	51,635	49,338	38,403	20,241	16,259	17,051	18,142	17,926	15,394	10,001
Basic Materials	24,834	24,443	26,584	29,101	28,539	25,512	23,106	17,689	9,409	6,212	6,719	6,958	7,024	5,825	3,930
Communication Services	19,976	20,284	21,437	23,392	24,227	24,550	22,285	18,770	12,245	10,394	10,905	9,974	8,276	6,718	4,006
Real Estate	17,132	16,653	16,965	18,639	19,319	19,177	18,067	13,184	7,143	6,266	5,654	6,100	6,986	6,107	3,332
Consumer Defensive	17,811	17,498	18,825	19,696	18,438	18,543	18,224	14,473	7,151	5,723	5,955	5,673	6,148	5,155	3,631
Energy	25,130	23,993	26,391	26,861	25,351	23,730	20,458	14,936	8,586	7,064	5,926	5,675	5,753	5,062	3,461
Utilities	12,817	12,632	13,197	14,389	14,611	15,485	14,502	11,873	6,117	4,764	4,803	4,214	3,668	3,139	2,696
Total	407,635	399,659	437,615	472,988	459,080	450,928	426,094	332,878	176,019	145,434	148,788	141,275	138,213	118,060	77,582

Table 2.11: Continuation of Number of Regulatory Disclosures by Sector and Year, Ranked by 2024 Count (1996–2010).

When we analyze entities by sector, it is evident that the number of entities peaked during the height of the dot-com bubble in 2000, with 8,913 entities (Table 2.9). Since then, there has been a consistent year-over-year decline in the number of publicly traded companies, reaching an all-time low in 2019 (Table 2.8). However, following the COVID pandemic and the associated economic stimulus, there was a revival, with a 15% increase in entities from 6,184 in 2019 to 7,196 in 2021. This resurgence also saw a significant 26% jump in filing volume, from 406,000 in 2019 to 513,000 in 2021 (Table 2.10). Despite this increase, the number of entities declined again in 2022.

Before the dot-com bubble, from 1994 to 2000, the technology sector included the most active entities, followed by the financial services sector. However, post-bubble, the tech sector experienced the largest decline in entities. In contrast, the healthcare sector has been on an upward trajectory since 2012, reaching a high in 2021 with 1,519 entities. This increase aligns with the introduction of the Affordable Care Act, often referred to as Obamacare, which expanded the healthcare market, particularly benefiting pharmaceutical companies.

We find that the utilities, real estate, energy, basic materials, and communications services sectors exhibit minimal fluctuations in their annual filing volumes and entities. This stability might suggest a lower availability of new information compared to the other sectors and potentially a reduced likelihood of unexpected informational surprises, potentially leading to reduced stock price volatility within these sectors – a promising avenue for future research.

Our findings reveal that the utilities, real estate, energy, basic materials, and communications services sectors demonstrate minimal fluctuations in their annual filing volumes and number of entities. Notably, these sectors are characterized by the lowest levels of filing activity and entity count. This relative stability in filing activity, coupled with a lesser volume of information dissemination compared to other sectors, indicates a lower frequency of new information release and potentially reflecting a lower frequency of significant corporate events or changes that necessitate filings. Consequently, this consistently lower level of disclosure activity may lead to a reduced likelihood of unexpected informational surprises within these sectors. Such a trend could, in turn, result in lower stock price volatility for companies within these groups, particularly influencing the relevant sector ETFs. This area presents a promising opportunity for future research to explore the correlation between the consistency of filing activity, the availability of new information, and the resultant impact on stock price volatility.

Despite the decline in entities across sectors, 2021 marked a record high in filing volume, surpassing the previous record set before the financial crisis. The financial services sector, despite a decrease in the number of entities since 2000, has consistently been the largest in terms of annual filings, particularly post-SOX. In 2024, it accounted for 32% of all sector-linked filings. The healthcare sector followed with 17% of the total volume in 2024. Combined, these two sectors constituted 49% of the entire filing volume (Table 2.10).

The sustained decline in the number of publicly traded companies, especially in the tech sector, could be attributed to sectoral consolidation and a shift towards raising capital through private rounds, as indicated by the rise in Form D filings. Large tech companies, with substantial cash reserves, have been acquiring smaller firms, often after their IPOs, due to reduced administrative challenges and compliance with SOX. This consolidation trend is exemplified by acquisitions such as Google's purchase of FitBit, Salesforce's acquisition of Tableau, and Microsoft's takeover of Activision Blizzard.

When analyzing the number of entities that can be assigned to a sector versus those that cannot, we find that only 5% of all entities in 2024 represent entities linked to a sector, implying they are publicly traded companies (Figure 2.10). This indicates that a vast majority, approximately 95%, of all active entities on EDGAR represent non-publicly traded companies.

In terms of filing volume, this 5% of sector-linked entities account for around 40% of the annual filings on EDGAR (Figure 2.11). Conversely, the remaining 60% of the filing volume is generated by the other 95% of entities. This distribution underscores a notable concentration of filing activity within a relatively small segment of entities – publicly traded companies – which represent a significant portion of the overall filing volume. In contrast, the majority of entities on EDGAR, which are non-publicly traded, account for only a marginally larger share of the total filing volume. This observation highlights the substantial impact that a small number of publicly traded entities have on the filing landscape, compared to their non-publicly traded counterparts.

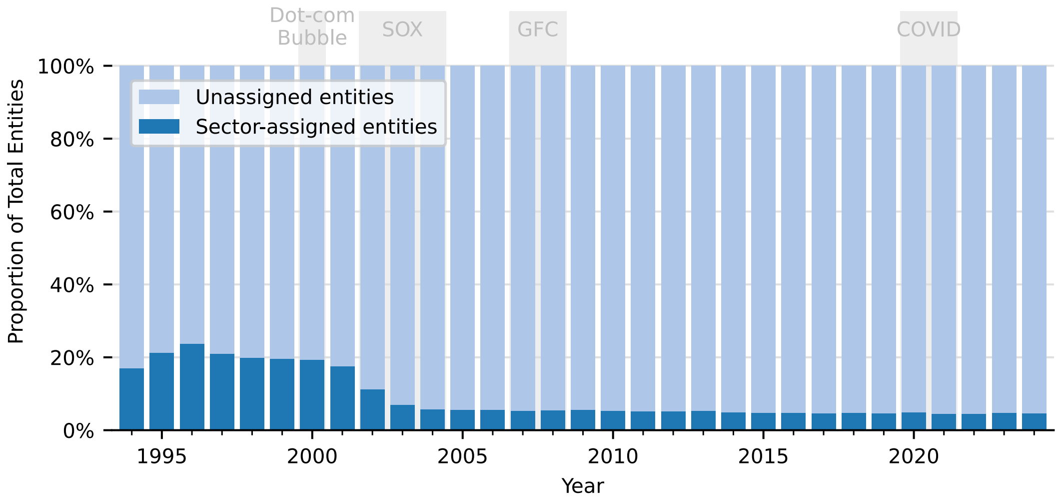


Figure 2.10: Proportion of Total Entities with and without Sector Assignment (1994–2024).

This bar chart compares the annual percentage share of entities classified by sectors to those unclassified within the EDGAR database, set against the timeline of notable economic events. The chart highlights a persistent majority of entities without sector assignment throughout the years, underscoring a significant portion of the database that remains categorized outside traditional sector classifications.

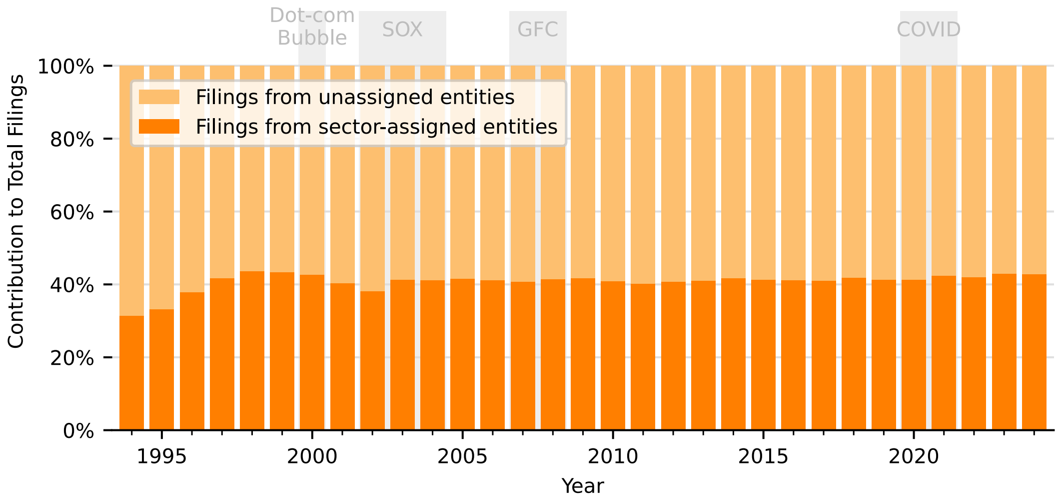


Figure 2.11: Filing Volume of Sector-Assigned vs. Unassigned Entities (Yearly % Share, 1994–2024).

This bar graph compares the yearly percentage contribution of filings from entities with and without sector assignments within the EDGAR system. The persistent prominence of filings from unassigned entities throughout the timeline reflects the extensive activity within EDGAR from a diverse array of entities beyond traditional sector divisions.

2.3.4 SIC Codes & SEC Offices

The SEC uses Standard Industrial Classification (SIC) codes to help organize and delegate the oversight and review of filings based on industry specialization within the Division of Corporation Finance. The SEC's Division of Corporation Finance[13] is responsible for ensuring that investors are provided with material information to make informed investment decisions, primarily through the review of documents that publicly traded companies are required to file with the SEC. Within the division, there are multiple offices responsible for the review of filings. These offices are often organized based on industries. To determine which office reviews which filings, the SEC frequently uses the company's SIC code to categorize it into the appropriate industry. The SIC codes system classifies industries with a three- or four-digit code. Every publicly traded company has an associated SIC code that represents its primary business activity.

For example, a company primarily involved in the software industry has a different SIC code than a company in the pharmaceutical industry. When these companies file documents with the SEC, the filings of the software company are reviewed by an office specializing in technology companies, while the filings of the pharmaceutical company are reviewed by an office specializing in healthcare companies.

It is worth noting that not every filer has an associated SIC code, and therefore some entities are not linked to a specific SEC office. Additionally, a single filing can reference multiple entities, each having distinct SIC codes. In these instances, the filing contributes to the annual volume count for each entity, influencing multiple SIC codes simultaneously.

For this part of our analysis we counted the number of active entities per industry classification (SIC code) that filed at least one SEC filing in a year from 1994 to 2024. We then identify the top 20 industries with the most active entities for each year, combine these lists into one, and select the top 20 industries from this aggregated list. This process results in a group classified as the "Top 20 Industries with the Most EDGAR Entities." A similar approach is used for filing volume, where we count the annual number of filings published by entities within their respective industry groups. This data is visualized in Figure 2.12.

A significant anomaly noted in Figure 2.12 is the number of active entities in the asset-backed securities (ABS) industry with SIC code 6189. Beginning in 1997, this industry consistently had the highest number of active entities annually, even surpassing industries related to the software sector before the dot-com bubble. From 2000 to 2006,

[13] `https://www.sec.gov/page/corpfin-section-landing`

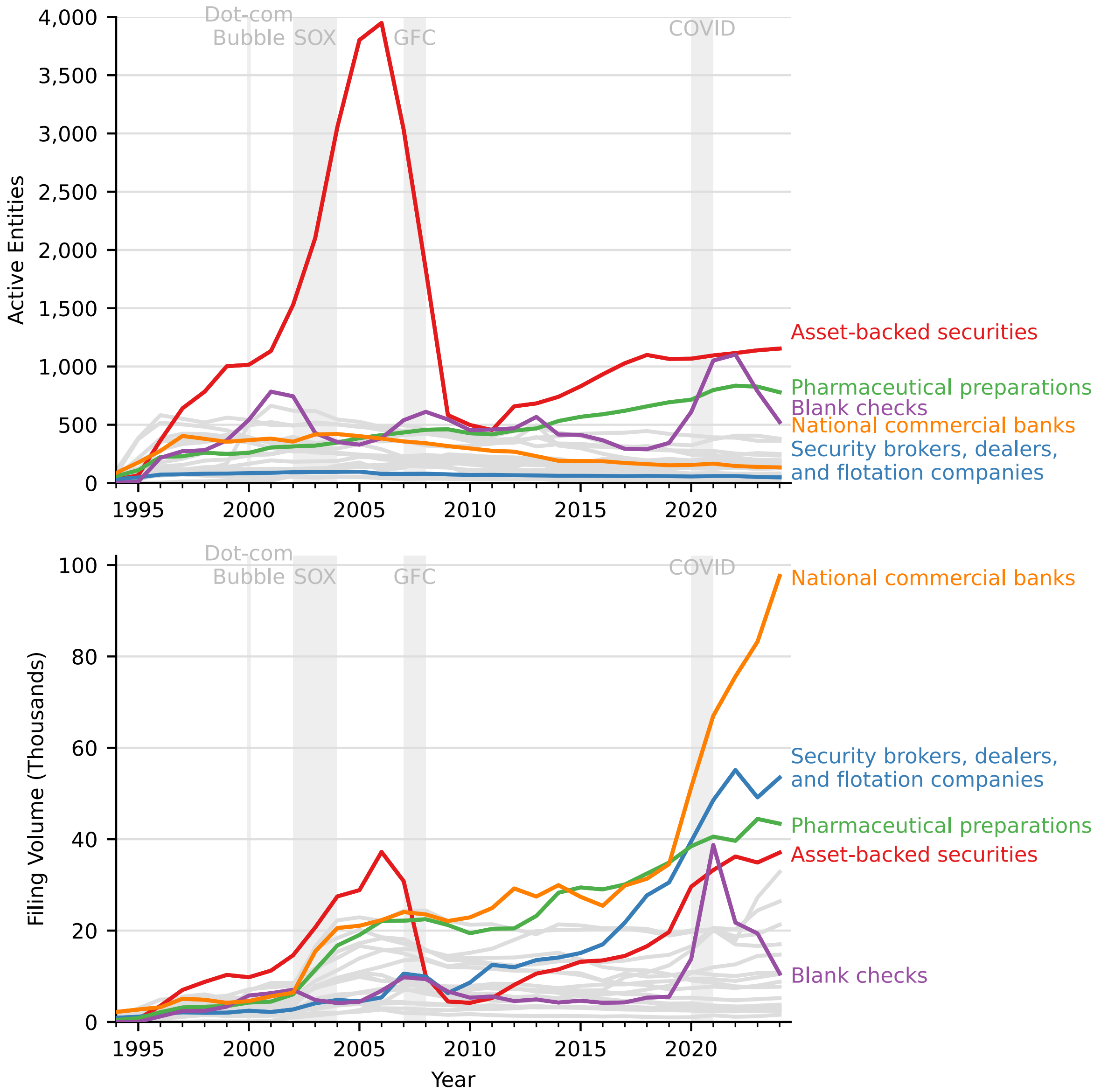


Figure 2.12: Active Entities and Filing Volume per Top SIC Codes (1994–2024).

The top graph traces the number of active entities classified under top Standard Industrial Classification (SIC) codes with the highest number of active entities over a period marked by the Dot-com Bubble, SOX, GFC, and COVID-19, while the lower graph details the corresponding filing volumes for these industries. The visual comparison highlights the activity levels and regulatory filing behaviors within key sectors of the economy across different market conditions. Notably, the asset-backed securities industry experienced unprecedented growth in entities and filing volume leading up to the GFC. Recent years have seen a significant uptick in filings from national commercial banks and security brokers, dealers, and flotation companies, with the pandemic period marking a particular surge. This period also witnessed a notable increase in activity for pharmaceutical preparations and asset-backed securities sectors. The rise of blank check companies, or SPACs, during the COVID era is clearly reflected.

the ABS industry experienced a 297% increase in entities, from 1,000 in 2000 to nearly 4,000 in 2006. In 2006, the ABS industry, with 3,971 entities, was the largest industry group, 670% larger than the second largest group, which had only 512 entities. In retrospect, the significant disproportion in size observed in the ABS industry in 2006 can be considered a major early warning sign of the GFC. With the ABS industry disproportionately larger than any other industry group, a factor that, in hindsight, pointed towards potential systemic risks accumulating in the financial system. Such an imbalance, especially in a sector directly linked to financial markets and credit instruments, could have been indicative of underlying vulnerabilities that eventually contributed to the GFC. This observation underscores the importance of closely monitoring and analyzing sectoral trends and disparities in financial data, as they can often provide early signals of broader market stresses.

Post-GFC, the ABS industry saw a dramatic reduction, with a low of 487 entities in 2011, marking an 88% decrease and the largest shrinkage recorded in EDGAR's history. Despite this decline, the ABS industry remained the largest group from 1997 to 2024, even reaching a new high in 2024 with 1,189 entities. When analyzing filing activity by industry classification, the ABS group also stands out, though not as prominently as in the entity count. In 2006, as the industry with the highest filing volume, it surpassed the second most active industry by 60%. This margin, while smaller than in entity count, was still substantial enough to signal early warning signs.

It's important to note that the peak in the number of ABS entities per year does not appear in the sector-based entities or filing volume charts, rendering this anomaly invisible across the "entities and volume per sector" dimension. This underscores the need for risk monitoring tools to expand their scope and consider multiple dimensions of analysis. Conversely, the tech bubble was not detectable using the "entities by SIC code" dimension or even in the "filing volume by SIC code." The anomaly becomes apparent only when examining entities by sector, highlighting the importance of a multi-dimensional approach in identifying market trends and risks.

The "blank checks" phenomenon is prominently reflected in the annual entity activity and filing volume per industry, even ranking as the second highest in terms of the number of active entities in 2022. This trend showcases the significant rise in Special Purpose Acquisition Companies (SPACs), often referred to as blank check companies, which have become increasingly popular as vehicles for public listings during the economic stimulus period triggered by the COVID pandemic.

Confirming our earlier sectoral analysis, the pharmaceutical preparations industry exhibits a consistent year-over-year increase in the number of entities. This trend is more clearly highlighted in the analysis by industry classification than in the entities per sector analysis. Remarkably, since the introduction of EDGAR in 1994, the pharmaceutical preparations industry is the only one across the entire universe to demonstrate constant year-over-year growth over 30 years. Post-implementation of Obamacare, from 2010 to 2012, there was a notable acceleration in filing volume within this industry, suggesting an increase in information flow. The impact of Obamacare on this industry can be attributed to several factors. The organizations in this group predominantly produce pharmaceutical drugs ready for direct consumption, such as tablets and liquids, targeting both professionals and the general public. The expansion of healthcare coverage under Obamacare increased the patient base, thereby enlarging the market for these pharmaceutical products. This expansion not only created more demand but also possibly led to increased investment and development in new pharmaceuticals, as companies sought to capitalize on the growing market, potentially with more extensive capital-raising efforts and increased marketing activities. The need to communicate more frequently and effectively with investors and the market, particularly in light of the evolving healthcare landscape, could explain the increased volume of filings. As a result, this period saw heightened activity in terms of both entity growth and more so filling volume, underscoring the significant influence of healthcare policy on industry dynamics and corporate behavior.

Moving onto different industries, in recent years, particularly from 2020 to 2024, the national commercial banks (SIC code: 6021) and security brokers, dealers, and flotation companies (SIC code: 6221) emerged as the two industries with the highest filing volume, both reaching all-time highs in 2024. The filing volume from national commercial banks exhibited a remarkable 115% increase from 2019 to 2022, a significant rise within just three years. This surge mirrors a similar, yet more pronounced anomaly that was observed in the ABS category prior to the GFC. Such a substantial increase in filing volume within a relatively short period signals a potential area of concern and definitely warrants further in-depth research to understand the underlying factors and implications. Furthermore, national commercial banks accounted for 20% of all filings in 2022, while brokers, dealers, and flotation companies contributed 15% to the annual disclosure volume. This high volume of filings can be partly attributed to the prominence of complex financial products detailed in 424B2 prospectuses that represented the third-largest EDGAR form type category by volume in 2022. The brokers/dealers/flotation companies industry includes investment bankers, mutual fund agents, and various dealers and brokers involved in diverse financial instruments. Flotation companies, while typically associated with assisting in IPOs, also play a role in floating complex financial products detailed in 424B2 prospectuses while helping structure these products to meet investment objectives and regulatory standards.

Confirming our results from the sectoral analysis, only 7.6% of entities were assigned a SIC code in 2024, a stark contrast to the 50/50 ratio in 2000 (Figure 2.13). To provide context, persons and trusts reporting security acquisitions or disposals on Form 3, 4 and 5 are not associated with a SIC code. However, they are still mandated to file with the SEC and are subsequently included as an entity in the EDGAR filer data. Given that Form 3/4/5 filings account for approximately 30% of all filings filed annually, this might shed light on the high number of entities not associated with a SIC code. Yet, it only accounts for a fraction of the 92.4% of entities without a SIC code. This significant decrease in the proportion of entities with an assigned SIC code raises questions about the oversight responsibilities within the SEC, as entities are linked to SEC offices via SIC codes. Despite representing a small fraction of total entities, these sector-classified entities were responsible for 54% of the total filing volume on EDGAR in 2024 (Figure 2.14). This trend highlights a potential gap in oversight and underscores the need for a more comprehensive understanding of the entities operating within the unassigned categories.

While SIC codes offer insights into industry anomalies and trends, the SEC offices provide an organizational perspective on how the SEC manages and oversees these entities.

The anomaly in the number of entities overseen by the Office of Structured Finance, which mirrors the trend observed in the ABS industry prior to the GFC, is notable (Figure 2.15). Since the Office of Structured Finance primarily oversees ABS entities, this parallel trend is expected. However, an intriguing contrast emerges when examining the filing volume associated with entities overseen by this office. Unlike the entity count, the filing volume does not show any significant anomalies and is relatively unremarkable, even ranking as the fourth smallest category pre-GFC. As highlighted earlier, this discrepancy underscores the importance of monitoring multiple dimensions, such as entity count and filing volume across sectors, industries, and offices, among others, to effectively detect early warning signs of potential market disruptions or misalignments.

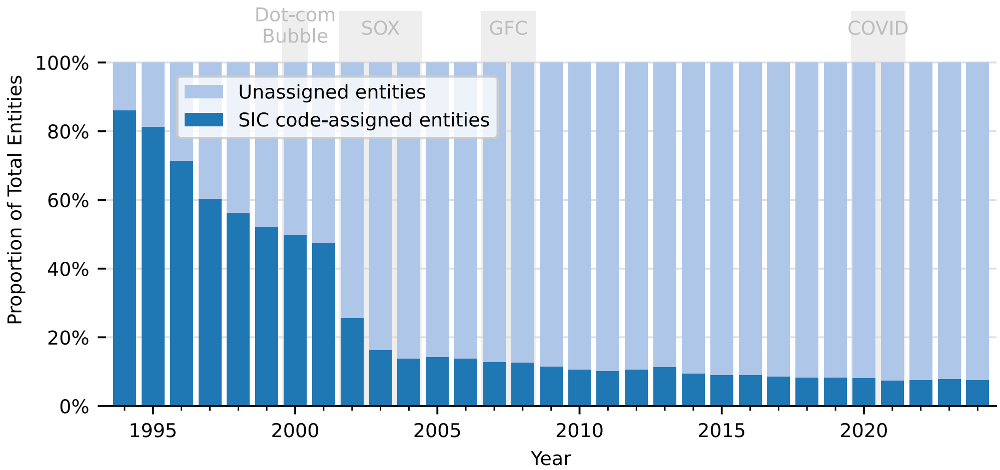


Figure 2.13: Yearly Proportion of SIC Code-Assigned vs. Unassigned Entities (1994–2024).

This bar chart illustrates the annual distribution of entities based on their Standard Industrial Classification (SIC) code assignment within the EDGAR system. It compares the percentage share of entities that have been assigned SIC codes to those without such designation over significant economic periods, including the Dot-com Bubble, SOX legislation, the Global Financial Crisis (GFC), and the COVID-19 pandemic. The chart reveals a notable trend of increasing proportions of unassigned entities, indicating a growing diversity of entities engaging with the EDGAR system that do not fall into the conventional sector-based classification system.

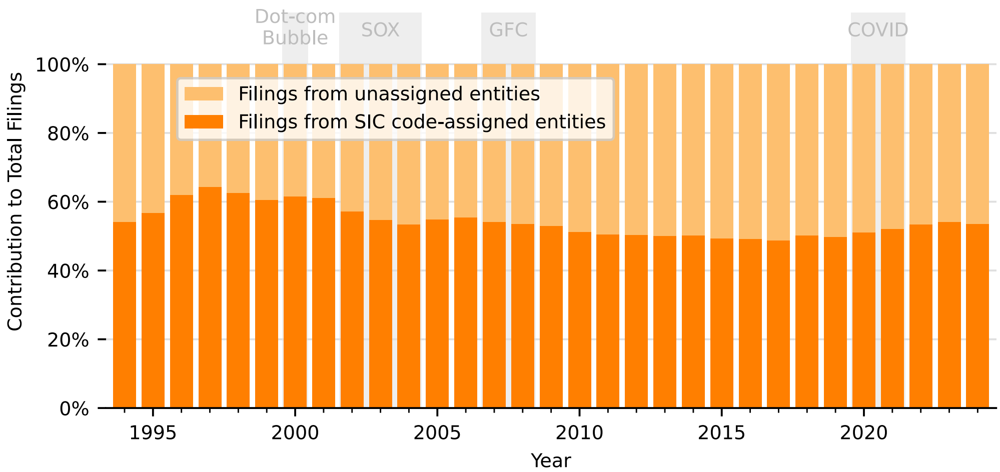


Figure 2.14: Filing Volume Share of Assigned vs. Unassigned Entities (Yearly % Share, 1994–2024).

This bar chart depicts the annual proportion of total filings made by entities that are assigned SIC codes compared to those without such codes. Over the timeline, including the Dot-com Bubble, SOX, GFC, and COVID-19 pandemic, a consistent 50/50 ratio between submissions filed by entities with assigned SIC codes and those unassigned is visualized.

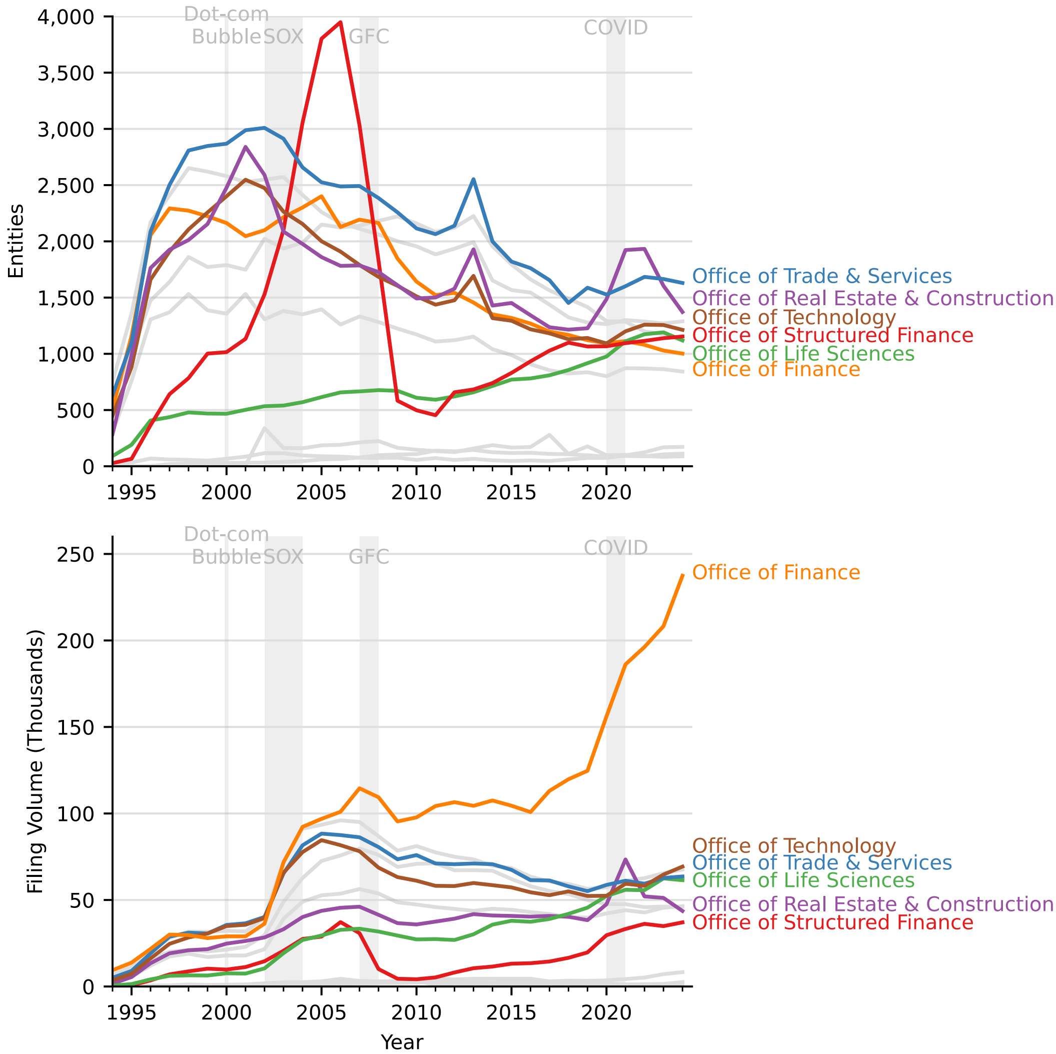


Figure 2.15: Active Filing Entities and Filing Volume by SEC Office per Year (1994–2024).

The top graph quantifies the number of active entities annually monitored by respective SEC offices, linked through the Standard Industrial Classification (SIC) codes that define the jurisdiction of each office. The bottom graph matches filing volumes to these offices using the same SIC code assignments. Across significant economic phases like the Dot-com Bubble, SOX, GFC, and COVID-19, the graphs highlight the dynamics within sectors that fall under the purview of different SEC offices. Notably, the Office of Structured Finance, which supervises asset-backed securities entities, exhibited a pronounced surge in active entities preceding the GFC. However, the filing volume associated with this office during the same period only represented a relatively small fraction, not signaling any discernible anomalies or signs of risk. This emphasizes the critical need to examine multiple dimensions to effectively monitor and assess risks in financial markets. Conversely, the Office of Finance displayed a steady decline in the count of active entities year over year, yet paradoxically, it experienced a substantial escalation in filing volume, illustrating the divergent trends between entity activity and regulatory submissions.

Further supporting our earlier findings, the outstanding filing activity in 2024 by national commercial banks and brokers/dealers/flotation companies is also reflected in the broader perspective provided by filing volume by entities associated with SEC offices. The Office of Finance, which oversees entities in these and other related industries, monitored over 240,000 filings in 2024. This represents 241% more annual filings overseen compared to the second-largest office, the Office of Technology, which was responsible for just under 70,000 filings that year. The magnitude and rapidity of this increase in filing volume are unprecedented for any SEC office over the last 30 years. Such a substantial and abrupt rise in filing volume should be viewed as a significant anomaly and necessitates further investigation by regulators. The goal is to assess potential risks to financial markets and ensure timely oversight and regulatory intervention.

2.4 Conclusion

Our analysis of over 15 million filings from the SEC's EDGAR database, covering a span from 1994 to 2024 and more than 800,000 filers, has provided unique insights into market dynamics and potential early warning signs preceding significant market events. We have identified distinct anomalies preceding events like the dot-com bubble, the Global Financial Crisis, and the 2022 market downturn. These findings underscore the critical importance of employing diverse metrics and dimensions for effective market risk monitoring. Our study reveals that while certain anomalies are detectable across one dimension, they may remain hidden in another. For instance, the surge in entities grouped by the ABS industry SIC code prior to the GFC was significant, yet a similar increase was not observable in the entities grouped by respective SEC office.

Particularly noteworthy are the anomalies detected in the national commercial banks and brokers/dealers/flotation companies sectors, as well as the unprecedented rise in 424B2 filings for the registration of complex financial products in 2024. These trends present compelling opportunities for future research, especially in analyzing 424B2 prospectuses to decipher the structure and implications of these complex financial products.

The observed seasonal filing patterns, peaking in February and dipping in September, suggest a temporal asymmetry in information dissemination that could be linked to market volatility. Future studies should delve into the correlation between these filing trends and market fluctuations.

Furthermore, our analysis highlights a profound shift in capital-raising strategies, with a marked decline in public offerings and a surge in private capital raising through

Form D filings. This trend, coupled with the steady rise in entities within the health-care sector following the introduction of Obamacare, points to significant sector-specific developments.

Lastly, our findings that a small fraction of entities on EDGAR represents publicly traded companies, yet accounts for a substantial portion of filing volume, along with the observation that a majority of entities lack an assigned SIC code, raise critical questions about oversight and market dynamics. The peak in insider trading activity one month before annual report disclosures also opens avenues for further research, particularly regarding the nature of these transactions and their strategic timing.

In conclusion, our study not only highlights the evolving landscape of financial market disclosures but also emphasizes the need for multi-dimensional monitoring approaches to better understand and anticipate market trends and risks.

Bibliography

Amel-Zadeh, A., Faasse, J., and Wutzler, J. (2016). Are All Insider Sales Created Equal? New Evidence from Form 4 Footnote Disclosures. *Social Science Research Network*.

Aragon, G. O., Hertzel, M. G., and Shi, Z. (2013). Why Do Hedge Funds Avoid Disclosure? Evidence from Confidential 13F Filings. *Journal of Financial and Quantitative Analysis*, 48(5):1499–1518.

Bonsall, S. B., Leone, A. J., Miller, B. P., and Rennekamp, K. M. (2017). A Plain English Measure of Financial Reporting Readability. *Journal of Accounting and Economics*, 63(2–3):329–357.

Brown, S., Tian, X., and Tucker, J. W. (2018). The Spillover Effect of SEC Comment Letters on Qualitative Corporate Disclosure: Evidence from the Risk Factor Disclosure. *Contemporary Accounting Research*, 35(2):622–656.

Cao, J., Calderon, T. G., Chandra, A., and Wang, L. (2010). Analyzing Late SEC Filings for Differential Impacts of IS and Accounting Issues. *International Journal of Accounting Information Systems*, 11(3):189–207.

Cassell, C. A., Cunningham, L. M., and Lisic, L. L. (2019). The Readability of Company Responses to SEC Comment Letters and SEC 10-K Filing Review Outcomes. *Review of Accounting Studies*, 24(4):1252–1276.

Cassell, C. A., Cunningham, L. M., and Myers, L. A. (2011). The Determinants and Costs of Non-Compliance with SEC Reporting Requirements: Evidence from SEC 10-K Comment Letters. *Social Science Research Network*.

Cheng, C. S. A., Huang, H., Li, Y., and Stanfield, J. (2012). The Effect of Hedge Fund Activism on Corporate Tax Avoidance. *The Accounting Review*, 87(5):1493–1526.

Cohen, A., Jackson, R. J., and Mitts, J. (2015). The 8-K Trading Gap. *Social Science Research Network*.

Cohen, L., Malloy, C., and Nguyen, Q. (2020). Lazy Prices. *The Journal of Finance*, 75(3):1371–1415.

Dechow, P. M., Sloan, R. G., and Sweeney, A. P. (1996). Causes and Consequences of Earnings Manipulation: An Analysis of Firms Subject to Enforcement Actions by the SEC. *Contemporary Accounting Research*, 13(1):1–36.

Dyreng, S. D., Hoopes, J. L., Langetieg, P., and Wilde, J. H. (2020). Strategic Subsidiary Disclosure. *Journal of Accounting Research*, 58(3):643–692.

Ege, M., Glenn, J. L., and Robinson, J. R. (2019). Unexpected SEC Resource Constraints and Comment Letter Quality. *Contemporary Accounting Research*, 37(1):33–67.

Engelberg, J., Reed, A. V., and Ringgenberg, M. C. (2012). How are Shorts Informed? Short Sellers, News, and Information Processing. *Social Science Research Network*.

Flugum, R., Lee, C., and Souther, M. E. (2022). Shining a Light in a Dark Corner: Does EDGAR Search Activity Reveal the Strategically Leaked Plans of Activist Investors? *Journal of Financial and Quantitative Analysis*, 58(7):2820–2851.

García, D. and Norli, O. (2012). Crawling EDGAR. *The Spanish Review of Financial Economics*, 10(1):1–10.

Gunny, K. and Hermis, J. (2019). How Busyness Influences SEC Compliance Activities: Evidence from the Filing Review Process and Comment Letters. *Contemporary Accounting Research*, 37(1):7–32.

He, J. and Plumlee, M. (2020). Measuring Disclosure Using 8-K Filings. *Review of Accounting Studies*, 25(3):903–962.

Hering, J. (2017). The Annual Report Algorithm: Retrieval of Financial Statements and Extraction of Textual Information. In *Fourth International Conference on Computer Science and Information Technology (CoSIT 2017).*

Huddart, S., Ke, B., and Shi, C. (2007). Jeopardy, Non-public Information, and Insider Trading around SEC 10-K and 10-Q Filings. *Journal of Accounting and Economics*, 43(1):3–36.

Jiang, F., Lee, J., Martin, X., and Zhou, G. (2019). Manager Sentiment and Stock Returns. *Journal of Financial Economics*, 132(1):126–149.

Kearney, C. and Liu, S. (2014). Textual Sentiment in Finance: A Survey of Methods and Models. *International Review of Financial Analysis*, 33:171–185.

Kim, C., Wang, K., and Zhang, L. (2019). Readability of 10-K Reports and Stock Price Crash Risk. *Contemporary Accounting Research*, 36(2):1184–1216.

Köchling, G., Schmidtke, P., and Posch, P. N. (2021). SEC Workload, IPO Filing Reviews, and IPO Pricing. *Social Science Research Network*.

Larocque, S., Martin, M., and Walther, B. R. (2020). Are Earnings Forecasts Informed by Proxy Statement Compensation Disclosures? *Contemporary Accounting Research*, 37(2):741–772.

Li, F. (2008). Annual Report Readability, Current Earnings, and Earnings Persistence. *Journal of Accounting and Economics*, 45(2–3):221–247.

Li, F., Lundholm, R. J., and Minnis, M. (2013). A Measure of Competition Based on 10-K Filings. *Journal of Accounting Research*, 51(2):399–436.

Loughran, T. and McDonald, B. (2013). IPO First-Day Returns, Offer Price Revisions, Volatility, and Form S-1 Language. *Journal of Financial Economics*, 109(2):307–326.

Loughran, T. and McDonald, B. (2014). Measuring Readability in Financial Disclosures. *The Journal of Finance*, 69(4):1643–1671.

Loughran, T. and McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54(4):1187–1230.

Ryans, J. P. (2020). Textual Classification of SEC Comment Letters. *Review of Accounting Studies*, 26(1):37–80.

Schroeder, J. (2025). Anomalies, Trends and Patterns in Disclosure Activities: Understanding EDGAR. Working Paper.

Chapter 3: A New Puzzle Piece for the "Sell in May, and Go Away" Anomaly: Regulatory Disclosures

Author: Jan L. Schroeder (jan@sec-api.io)*

Abstract
We propose a new puzzle piece for the Halloween effect ("Sell in May and Go Away") by identifying a seasonal pattern in SEC regulatory disclosures that aligns with the effect's summer and winter periods. From 2004 to 2023, SEC filing volumes are 17% higher in winter (November–April) than in summer (May–October). Winter also sees a 22% rise in insider trading, 13% more private securities offerings, 12% more activist investor activity, a 96% increase in shareholder meetings, and 473% more annual reports. February consistently shows the highest number of disclosures, while September shows the lowest. Similar patterns across European markets suggest global consistency. As regulatory filings contain material price-relevant information, this seasonal disclosure pattern offers a new contributing factor to the Halloween effect puzzle.

Publication Details: Published in the International Journal of Financial Studies, 2025, 13(4), 208. https://doi.org/10.3390/ijfs13040208

Keywords: Halloween Effect, Sell in May and Go Away, SEC Filings, Market Anomaly

Acknowledgements: This paper was accepted for the 37th Australasian Finance and Banking Conference (AFBC), Australia, 10–13 December 2024. We are sincerely grateful for the constructive feedback and comments from Ben Jacobsen (Professor of Finance, TIAS Business School, Tilburg University), author of the original Halloween effect paper, Jonathan W. Lewellen (Professor of Finance, Tuck School of Business at Dartmouth), and Peter N. Posch (Professor of Finance, TU Dortmund).

The following is based on Schroeder (2025a).

*Jan L. Schroeder is a Research Fellow at the Department of Finance, TU Dortmund, Germany, and the founder and CEO of SEC-API.io. The authors have no competing interests to declare. All errors are our own. Corresponding author: Jan L. Schroeder.

3.1 Introduction

The Halloween effect, also known as "Sell in May, and Go Away", describes an anomaly in which stock markets tend to outperform from November to April (winter) compared to May to October (summer). While numerous studies have explored potential explanations for this phenomenon, a conclusive cause remains outstanding. In this paper, we identify a previously unknown pattern that aligns closely with the Halloween effect: a recurring seasonality in regulatory information disclosures filed with the U.S. Securities and Exchange Commission (SEC). This pattern repeats consistently year after year over our two-decade analysis period (2004–2023) and exhibits a similar structure in European markets.

The phrase "Sell in May and go away" first appeared in U.S. news coverage in a *Financial Times* article from 1964, and a trading strategy based on this effect was shown to outperform a buy-and-hold strategy with lower risk (Bouman and Jacobsen, 2002). The phenomenon is observable in equity markets of 36 countries, from at least 1970 to the present (Bouman and Jacobsen, 2002; Zhang and Jacobsen, 2021), and the existence of the effect was further validated by Jacobsen and Visaltanachoti (2009), who found that the Halloween effect is significant in more than two-thirds of U.S. sectors and industries when examining returns for 17 sectors from July 1926 to December 2006. Haggard and Witte (2010) confirm that the Halloween effect in U.S. returns is significant in the period from 1954 to 2008, but not 1926 to 1953. Andrade et al. (2013), using both Bouman and Jacobsen's (2002) original sample period (May 1970 to October 1998) and an extended out-of-sample period (November 1998 to April 2012), confirmed that the sell-in-May effect persists out-of-sample, and is observable across various strategies, including size, value, equity volatility risk, and credit risk premiums.

The investigated reasons behind the effect include data mining inconsistencies, shifts in risk profiles, financial news sentiment and patterns, interest rates, trading volume, sector effects, vacations (Bouman and Jacobsen, 2002), weather (Jacobsen and Marquering, 2008), the January effect (Lucey and Zhao, 2008), political climate (Powell et al., 2009), and more, while a full explanation for the anomaly is still outstanding.

We introduce a new piece to the Halloween effect puzzle: a robust, recurring seasonal pattern in the volume of regulatory disclosures filed with the SEC by market participants. Analyzing publication dates of all regulatory filings submitted by public companies, corporate insiders, funds, brokers, and other regulated entities from 2004 to 2023, we identify a consistent 17% increase in information volume during the winter months compared to the summer. This pattern recurs every year across our two-decade sample. Specifically, the winter period exhibits a 22% rise in insider trading activity, 13% more private securities offerings, 12% more activist investor activity, a 96% increase in shareholder meetings, and a 473% surge in annual reports and audited financial statements. September consistently shows the lowest disclosure volume–mirroring its historical distinction as the only month with a negative average market return over the past century.

Most of the information disclosed in SEC filings is legally mandated under U.S. securities regulations—companies, funds, and insiders are required by law to submit these disclosures rather than doing so voluntarily. The content of these filings includes financial results, executive changes, insider trading activities, fund holdings, and other material corporate events. Because of their legal and informational significance, SEC filings represent a primary source of information for investors, analysts, and the media. Prior research has consistently demonstrated that various types of SEC filings have a significant impact on stock prices (Cohen et al., 2020; Kim et al., 2019; Loughran and McDonald, 2014; You and Zhang, 2008; Hirshleifer et al., 2009; Brown and Tucker, 2011; Li, 2010; Merkley, 2013; Mayew et al., 2014; Muslu et al., 2015; Cohen et al., 2012; Cline et al., 2017; Brav et al., 2008; Dimitrov and Jain, 2011).

Extending our analysis to European markets, we find similar seasonal disclosure patterns among London Stock Exchange and EuroNext companies. A related, but less pronounced, temporal asymmetry appears in the press releases of U.S. listed firms.

To our knowledge, this is the first study to analyze all regulatory disclosures published through the SEC's EDGAR system and identify the recurring, year-over-year temporal disclosure patterns presented here. Our study requires aggregating and analyzing over 10 million regulatory filings, 3.5 million press releases, over 3 million IBES analyst announcements, and 160,000 institutional investment manager portfolios. Analyzing such large volumes of data and the technical and resource-intensive nature of this research has likely deterred prior investigations. Furthermore, analyzing the entire EDGAR filing stack including more than 800 different form types requires extensive knowledge of the various filing types, their purpose, and governing regulations and relationships between them.

We document a previously unreported, robust seasonal pattern in regulatory disclosures that aligns with the winter and summer cycles of the Halloween effect. We do not attempt to establish a causal relationship between disclosure volume and the Halloween effect, nor do we claim a direct cause-and-effect link. Our findings provide a new empirical foundation and motivation for future research to explore how this recurring information cycle may contribute to seasonal market anomalies.

While we do not test the mechanism directly, it is reasonable to connect Bouman and Jacobsen's (2002) finding of winter outperformance with the seasonal surge in regulatory disclosures identified here. Given the well-documented material impact of SEC filings on stock prices, our results offer a compelling new empirical perspective that adds to the broader understanding of the Halloween effect.

3.2 Data and Methods

Since 1933, all public companies and a wide range of financial entities operating in U.S. financial markets—such as mutual funds, hedge funds, investment advisers, insiders, broker-dealers, and more—have been legally required to disclose information to the Securities and Exchange Commission (SEC). Since 1994, these filings have been submitted electronically through the EDGAR system, which is managed by the SEC and holds over 10 million disclosures and 100 million attachments from more than 800,000 filing entities. EDGAR provides public access to this information universe, covering everything from financial statements and insider trades to material events and corporate actions. While companies may also issue press releases, only SEC filings, across more than 800 specialized form types, fulfill regulatory requirements and serve as the authoritative source of regulatory disclosures.

The terms "filing" and "regulatory disclosure" are used interchangeably throughout this paper. References to "markets" refer exclusively to equity markets; other markets—such as currencies, bonds, or commodities—are beyond the scope of this study.

We used the commercially available SEC filings database provided by SEC-API.io, and downloaded, with its Query API [1], all historical, survivorship bias-free filing metadata of the last 30 years from 1994 to 2023, including the EDGAR form type, and the date and time each filing was submitted to the EDGAR system. The dataset includes every SEC filing ever digitally published, encompassing disclosures from all entities required to file with the SEC. This includes private and public companies, both domestic (e.g., Microsoft) and foreign (e.g., Alibaba), listed on U.S. exchanges, corporate executives and insiders, various types of funds (such as mutual funds and hedge funds), institutional investors, business development companies, asset-backed securities issuers, and more. The dataset captures filings from both currently operating and defunct entities, providing a

[1] `https://sec-api.io/docs/query-api`

comprehensive historical record of SEC disclosures. The filings were filtered by unique accession numbers rather than unique URLs to prevent double-counting.

Regulatory disclosures submitted to the London Stock Exchange's (LSE) RNS information system, pertaining to UK entities regulated by the Financial Conduct Authority (FCA), were obtained via the National Storage Mechanism for the period 2013 to 2023. Disclosures for other European entities listed on the EuroNext stock exchanges were sourced directly from EuroNext, covering the period 2017 to 2023. Data from 2021 was excluded due to a mid-year regulatory change that caused a tenfold increase in disclosure volume, rendering the year an outlier. Unlike LSE RNS disclosures, EuroNext data did not include disclosure type information in its metadata, limiting the granularity of the analysis.

To obtain all survivorship bias-free company press releases of NASDAQ and NYSE listed companies, we used LexisNexis, covering 15 years of data from 2009 to 2023. We excluded all alerts from law firms, research reports, book announcements on Amazon, and any other press releases not published by the companies themselves. This ensured that only press releases issued by the listed companies were included in our analysis.

To further test our findings, we utilized Refinitiv Eikon to obtain IBES analyst ratings for any company ever listed on the NASDAQ or NYSE stock exchange from 2009 to 2023 and the SEC-API.io 13F Institutional Ownership API to analyze quarterly fluctuations in the gross market values of institutional investment managers, such as hedge funds, between 2013 and 2023. Additionally, we used the SEC-API.io Extractor API to access all item sections in 8-K filings, allowing us to extract and determine the timing of specific material events over the past decades.

We used the data aggregation and transformation strategies as described in Schroeder 2023. Python tutorials to assist in replicating our results are available online[2].

To validate the robustness and statistical significance of the seasonal pattern in disclosure volumes and other information types, we applied two complementary statistical approaches. First, a cross-sectional analysis was performed by grouping monthly filing volumes from 2004 to 2023 into two seasonal categories: summer (May–October) and winter (November–April). After testing for normality using the Shapiro–Wilk test and for homogeneity of variance using Levene's test, we applied a two-sided Mann–Whitney U test due to the non-normal distribution of the data. If results showed a statistically significant difference between the two periods ($p < 0.01\%$), we concluded that disclosure

[2] https://sec-api.io/resources/analyzing-sec-edgar-filing-trends-and-patterns-from-1994-to-2022, https://sec-api.io/resources/edgar-filer-analysis

volumes are systematically higher in winter months. Second, we performed a year-by-year comparison by summing the number of filings disclosed during the summer and winter periods for each year from 2004 to 2023. Using these paired seasonal totals, we applied a Wilcoxon signed-rank test to assess whether winter volumes consistently exceeded summer volumes across years. If the test produced a T-statistic of 0 and a p-value smaller than 0.1%, we concluded that there is clear evidence that the observed seasonal difference is robust and unlikely to be due to chance. We applied these approaches for the periods May–October and November–April, and for various time-lagged variations. These tests were performed on all SEC filings, including the two most common EDGAR form types—insider trading and material event disclosures, which represent approximately 45% of the annual filing volume—and other SEC filings with pronounced temporal patterns and press releases.

As we primarily compare filing volumes between different periods within a given year and across various years, where the periods might have different time scales (e.g., monthly vs. 6-month), the sample groups can be seen as dependent rather than independent. This comparison resembles a matched-pairs analysis, where we compare the number of filings disclosed under a specific form type, such as Form 4 (insider trading activity), in the summer months (May to October) to the number of filings disclosed in the winter months (November to April) of the same form type. Since we are measuring the same form type across different times, it is reasonable to assume that the groups are not independent.

To ensure the validity of our findings, we selected the methods based on the underlying characteristics of the data. We tested for normality using the Shapiro–Wilk test and Q–Q plots, and for equal variance using Levene's test. For normally distributed data with equal variance, we applied paired t-tests. When data has equal variance but is not normally distributed, we use the Wilcoxon Signed-Rank Test. For data that is neither normally distributed nor has equal variance, we also use the Wilcoxon Signed-Rank Test. When data is normally distributed but exhibits heteroscedasticity, we apply a generalized least squares model to account for unequal variance, suitable for paired data.

Our SEC filings data series begins in 1994, meaning we cannot validate our explanation for the period from 1954 to 1994. Before the introduction of the EDGAR system and electronic filings in 1994, the SEC regulated the market following the 1929 stock market crash with the introduction of the Securities Exchange Act of 1934, with information being disclosed on paper. As demonstrated by Schroeder (2025b), the Sarbanes–Oxley Act (SOX) of 2002 significantly transformed the filing landscape, increasing the annual volume of Form 4 (insider trading disclosures) and Form 8-K (material event disclosures) filings by more than tenfold, with both filing types representing approximately 45% of all disclosures filed annually. Insider trading regulations were first introduced in 1934 through Sections 10(b) and 16 of the Securities Exchange Act, requiring insiders to report changes in their ownership of company securities by the 10th day of the month following the transaction. SOX amended this requirement, mandating that filings be made within two business days after the transaction. Given the substantial increase in information volume following the completion of the SOX rollout in 2004, our analysis focuses on filings from 2004 to 2023. Although the volume of Form 4 and Form 8-K filings was significantly lower pre-SOX, it can be assumed that the same information disclosed in these forms post-2004 was still reaching the market and investors through other means prior to the regulatory changes. This assumption supports the validity of our explanation. For example, 8-K filings provide a structured vehicle for material event disclosures such as mergers, bankruptcies, or shareholder matters. While SOX introduced various new triggering events for 8-K filings, it did not change the frequency of such events occurring. Companies entered into and terminated material agreements before SOX, but the act now required companies to disclose these events in SEC filings. The same applies to the departure of directors, creation of financial obligations, and other new trigger events introduced by SOX.

3.3 Results

Our research reveals a significant temporal pattern in disclosure volume by entities regulated by the SEC, robust across two decades from 2004 to 2023, with a 17% increase in average monthly SEC filing volume during the November–April winter period compared to May–October summer period (Table 3.4, Table 3.3). Monthly filing volume peaks in February and reaches its lowest point in September, aligning with the Halloween effect periods from winter (November–April) to summer (May–October). We observe that private securities offerings, insider trading, material event disclosures, activist investor announcements, and annual mutual fund performance and operating reports, among others, all have their highest volume months during the winter period, with all disclosures consistently seeing their lowest monthly filing volume in summer. Additionally, 84% of all annual reports and audited financial statements and 64% of all annual shareholder voting material are disclosed during the winter period.

Figure 3.1 shows the mean number of SEC filings published per month from 2004 to 2023, with the 2.5% and 97.5% percentiles highlighted and the high and low months marked accordingly. We find that most SEC filings are consistently published in February each year, while also observing that September is the month with the least information disclosed, followed by October (Table 3.1).

In Figure 3.2, when filtering for the top 2% most commonly disclosed form types, we find that only 20 out of more than 700 different types represent 86% of the total annual information volume (Table 3.4). Examining the monthly filing volume, February consistently has the highest volume, while June and July show a sharp decline (Table 3.2). September is the least active month, with August being an exception due to SEC-mandated quarterly reports by companies (Form 10-Q) and quarterly disclosures by mutual funds and hedge funds (Form NPORT-P, Form 13F-HR), which peak in August.

By calculating the average monthly volume of all SEC filings for the periods May to October and November to April from 2004 to 2023, we find an average of 52,797 filings per month for the May–October period, compared to 61,526 filings per month for the November–April period—representing a statistically significant increase of 16.5% (Table 3.4).

The average cumulative filing volume between November and April is 371,828, reflecting a 16.3% increase compared to the summer period, which averages 320,169 filings from 2004 to 2023 (Table 3.5). When focusing on the top 2% of the most commonly disclosed form types, we find that just 20 out of more than 700 distinct types account for 86% of the total annual information volume. Within this group, the mean monthly filing volume rises by 19% during the winter months compared to the summer months (Table 3.4).

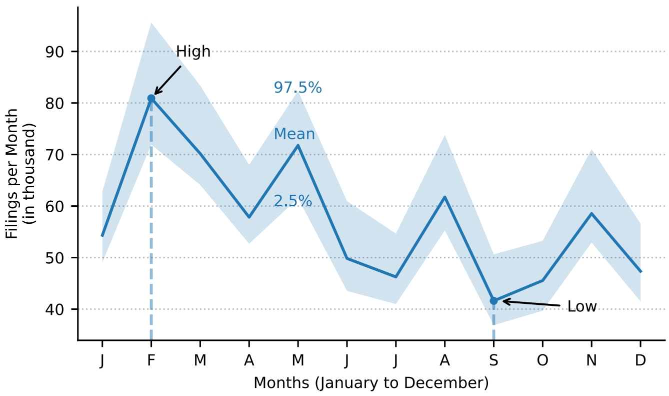


Figure 3.1: Average Monthly Disclosure Volume of All SEC EDGAR Filings (in Thousands, 2004–2023).

The solid blue line represents the mean filing volume for each month, while the shaded area indicates the range between the 2.5th and 97.5th percentiles. February consistently exhibits the highest average filing volume ("High"), while September records the lowest volume ("Low").

Monthly Filing Volume	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	54,564	**81,516**	70,721	58,213	72,678	50,532	46,518	62,524	**41,982**	45,935	59,048	47,766
Std. Dev.	4,493	6,986	5,892	4,517	7,062	6,098	4,123	7,102	4,246	4,371	6,436	4,823
Min	48,386	70,381	64,057	52,171	61,034	43,366	40,476	54,532	36,481	39,148	52,819	40,879
Median	53,136	**80,142**	70,398	57,073	71,809	48,222	46,738	58,587	**40,778**	46,070	55,606	46,544
Max	63,782	95,947	84,426	68,447	90,076	63,727	56,111	77,149	51,773	53,537	74,150	57,816

Table 3.1: Monthly Disclosure Statistics of All SEC EDGAR Filings (2004–2023)

*February consistently exhibits the highest mean and median monthly filing volume (*81,516 *filings), while September shows the lowest (*41,982 *filings).*

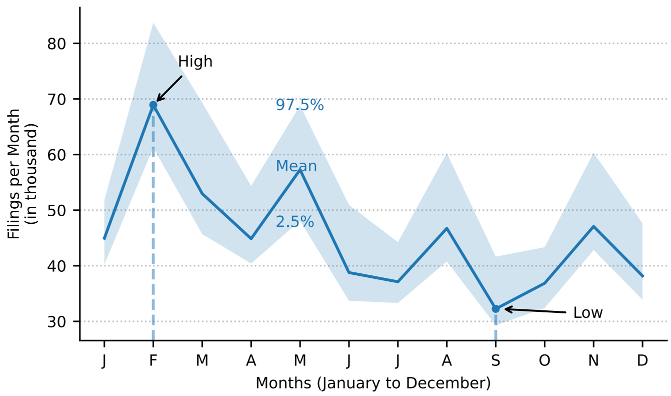


Figure 3.2: Average Monthly Filing Volume (January–December, 2004–2023) for the Top 20 Most Frequently Used SEC Form Types.

The solid blue line represents the mean filing volume for each month, while the shaded area indicates the range between the 2.5th and 97.5th percentiles. February consistently exhibits the highest average filing volume ("High"), while September records the lowest volume ("Low").

Monthly Filing Volume	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	45,258	**69,553**	53,662	45,114	58,008	39,378	37,361	47,373	**32,550**	37,175	47,498	38,481
Std. Dev.	3,606	6,796	7,336	4,071	6,107	5,788	3,429	6,214	3,955	3,549	5,702	4,226
Min	39,379	61,061	45,309	39,711	46,346	33,601	32,836	40,049	29,161	31,955	42,742	33,788
Median	44,355	**68,449**	50,505	43,978	57,314	37,582	37,337	44,516	**30,676**	36,130	45,202	37,820
Max	53,598	84,116	69,489	55,016	71,811	52,484	46,636	61,127	42,453	44,677	63,129	48,904

Table 3.2: Monthly Disclosure Statistics of the Top 20 Most Frequently Used SEC Form Types (2004–2023).

Average number of filings disclosed per month (January to December, 2004–2023) for the top 20 most commonly used form types, which represent just 2% of all form types but account for 86% of the total information volume disclosed on the SEC EDGAR system.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	May–Oct	Nov–Apr	Diff (%)
2004	51.9	73.8	66.5	56.1	61.0	47.4	44.5	58.2	41.0	45.1	61.4	51.1	297.3	360.8	+21.4
2005	55.7	75.6	72.2	58.0	72.1	52.4	49.0	68.7	45.9	46.0	61.8	51.9	334.2	375.3	+12.3
2006	60.5	78.6	71.9	56.1	78.1	54.7	47.4	69.3	43.7	50.7	65.5	50.6	343.9	383.2	+11.4
2007	63.8	82.2	70.6	62.1	79.9	53.5	52.9	71.8	41.6	52.9	67.2	48.1	352.5	394.0	+11.8
2008	59.7	84.4	66.6	58.5	72.7	47.9	48.9	60.5	40.6	46.7	53.8	45.6	317.1	368.6	+16.2
2009	48.4	74.0	69.8	52.2	62.1	43.9	42.7	54.5	38.4	41.4	52.8	44.0	283.0	341.2	+20.5
2010	49.9	70.4	71.9	55.7	65.0	47.4	42.6	56.7	38.7	40.8	56.4	47.1	291.3	351.4	+20.6
2011	52.0	76.7	70.3	55.2	71.1	46.4	42.6	59.9	38.8	39.1	53.8	42.5	298.0	350.4	+17.6
2012	50.1	80.6	64.1	55.9	70.0	43.4	42.6	58.8	37.4	42.3	54.1	45.1	294.4	349.8	+18.8
2013	52.7	77.3	66.0	58.1	71.5	43.8	46.5	57.1	40.0	46.1	54.4	43.7	305.1	352.2	+15.5
2014	54.6	79.9	66.3	57.9	69.8	47.9	46.9	57.5	41.2	46.1	53.9	49.0	309.6	361.6	+16.8
2015	52.5	82.5	70.5	56.9	68.1	50.5	47.5	57.0	39.4	42.8	53.6	45.1	305.2	361.1	+18.3
2016	50.1	83.0	65.9	53.3	67.7	46.6	41.6	58.4	41.2	40.4	54.4	43.2	295.9	350.0	+18.3
2017	52.6	78.6	70.7	53.3	72.3	49.2	40.5	58.4	39.6	43.5	54.8	42.2	303.4	352.3	+16.1
2018	54.1	80.4	65.3	55.5	71.0	48.5	42.8	58.2	36.5	46.3	53.0	40.9	303.3	349.1	+15.1
2019	49.9	79.7	64.1	57.3	74.1	45.2	45.8	56.2	39.7	43.4	58.7	46.0	304.4	355.8	+16.9
2020	53.6	88.5	74.1	61.0	72.7	56.9	49.5	64.7	46.0	49.4	61.6	55.1	339.1	393.9	+16.1
2021	58.5	95.9	84.4	68.4	80.6	63.5	56.1	72.7	51.8	53.5	74.2	57.8	378.2	439.3	+16.2
2022	61.7	95.2	82.1	67.4	83.8	57.9	48.4	74.6	49.3	48.9	66.8	50.6	362.8	423.9	+16.8
2023	59.1	92.9	81.0	65.3	90.1	63.7	51.5	77.1	49.0	53.1	68.6	55.6	384.6	422.5	+9.9

Table 3.3: Monthly SEC Filing Volumes (in Thousands) per Year from 2004 to 2023.

The table reports the number of filings published each month per year, along with total filing volumes for the summer period (May–October) and winter period (November–April) for each year. The final column shows the percentage difference between the two periods, indicating that every year exhibits higher filing activity during the winter months.

	Share of Total Filing Volume	Period	Mean Monthly Filing Volume	Years	Difference (%)	Highest Volume Month	Lowest Volume Month	Peak Monthly Volume in Winter	Lowest Monthly Volume in Summer
All SEC Filings	100%	May – Oct	52,797	20	+16.53%***	February	September	Yes	Yes
		Nov – Apr	61,526						
2% (20) of all EDGAR Form Types	85.6%	May – Oct	40,554	20	+19.11%***	February	September	Yes	Yes
		Nov – Apr	48,303						

Table 3.4: Comparison of Mean Monthly Filing Volumes Between Summer (May–October) and Winter (November–April) Periods from 2004 to 2023.

The table displays the mean monthly filing volumes, the percentage differences, and the highest and lowest volume months for all SEC filings and the top 2% of EDGAR form types, which represent 85.6% of the total filing volume. The table also indicates whether the peak monthly volume occurred in the winter and whether the lowest monthly volume occurred in the summer period of the Halloween effect. Significance levels are denoted as follows: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Type	Share of Total Filing Volume	Years	May–Oct Mean Period Filing Volume	May–Oct Standard Deviation	Nov–Apr Mean Period Filing Volume	Nov–Apr Standard Deviation	Difference (%)
All SEC Filings	100%	20	371,828	28,680	320,169	30,243	+16.33%***
Top 2% of all EDGAR Form Types	85.6%	20	299,567	29,050	251,845	27,127	+19.1%***
Unscheduled Filings Triggered by Specific Events Rather Than a Pre-Defined Regulatory Timeline							
Insider Trading (Form 4)	37.1%	20	108,109	9,394	89,058	10,587	+21.85%***
Activist Investor Ownership (Form SC 13D+/A)	1.1%	20	3,283	588	2,948	523	+11.49***
Material Events (Form 8-K)	14.9%	20	39,377	6,916	39,393	5,900	−0.35%
Material Events by Foreign Issuers (Form 6-K)	4.2%	20	11,410	1,098	11,276	1,182	+1.28%*
Private Securities Offerings (Form D+/A, adopted in 2009)	5.6%	15	20,491	7,834	18,214	7,376	+13.37%***
Mutual Fund Summary Prospectus and Pros. Materials (Form 497, 497K adopted in 2009)	5.2%	20	17,763	3,567	14,712	2,687	+20.58%***
Filings with Annual Seasonal Volume Spikes							
Annual Reports and Audited Financial Statements by Companies (Form 10-K)	1.5%	20	6,609	645	1,194	268	+473.47%***
Shareholder Voting Information, Proxy Statements (Form DEF 14A, DEFA14A)	1.9%	20	6,930	715	3,578	510	+96.17%***
Ownership Report by Passive Investors (Form SC 13G+/A)	4.2%	20	20,139	1,838	3,120	568	+558.38%***
Annual Report of Changes in Insiders' Ownership (Form 5)	0.8%	20	4,015	1,842	619	452	+673.50%***
Annual Report by Mutual Funds (Form N-CEN since 2019, N-SAR prior)	0.1%	5	2,101	812	816	250	+145.14%*
Filings Without a Temporal Pattern							
Complex Financial Product Prospectuses (Form 424B2)	3.6%	20	10,762	10,262	11,100	10,658	−2.01%
Information for Investors during Securities Offerings (Form FWP)	1.9%	18	6,139	2,990	6,315	2,925	−3.57%
SEC Correspondence with Filers (CORRESP, LETTER)	3.5%	19	9,524	2,682	10,350	2,538	−7.78%**

Table 3.5: Comparison of Mean Filing Volumes for Various SEC Filing Types Between Winter (November–April) and Summer (May–October) Periods from 2004 to 2023.

The table presents the share of total filing volume, mean filing volumes per period (November–April vs. May–October), standard deviations, and percentage differences between the means of the two periods, for all SEC filings, the top 2% of all EDGAR form types, unscheduled filings triggered by specific events, filings with annual seasonal volume spikes, and filings with consistent information flow without a temporal pattern. The data demonstrates significant increases in filing volumes during the winter period, with notable spikes in annual reports, shareholder voting information, and ownership reports by passive investors and insiders. Significance levels are denoted as follows: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

In order to understand the origins of the temporal pattern found in the volume of monthly SEC filings, we decompose the pattern into its constituent parts by grouping form types into the following categories:

1. Unscheduled Filings Triggered by Specific Events Rather Than a Pre-defined Regulatory Timeline (Figure 3.3): These are mostly unscheduled filings and include eight of the most frequently disclosed form types, namely insider trading forms (Form 3 and 4), material event disclosures (Forms 8-K and 6-K), mutual fund prospectuses (Form 497, 497K), ownership reports by activist investors (Form SC 13D, D/A), and private security offerings (Form D, D/A). The data reveals distinct seasonal patterns, with peaks typically observed between February and May, and noticeable declines during the summer months.
2. Once-Annually Occurring Volume Spikes (Figure 3.4): These are filings required to be filed once per year and include annual reports and audited financial statements on Form 10-K, shareholder proxy statements (DEF 14A, DEFA14A), mutual fund annual reports (N-CEN), reports disclosing of over 5% passive ownership (Form SC 13G, 13G/A), and annual statements of ownership by insiders (Form 5). The data reveals peaks in February and March, with a decline in filings observed throughout the remainder of the year.
3. Quarterly Recurring Volume Spikes (Figure 3.5): Quarterly recurring pattern driven by regulatory schedules. Form 10-K and Form 10-Q are classified as the same group as companies are required to disclose quarterly reports four times a year – one annual report on Form 10-K and three quarterly updates on Form 10-Q. Annual reports peak in March, and quarterly in May, August, and November. Similarly, mutual fund and hedge fund reports (Form NPORT-P and 13F-HR) also peak in February, May, August, and November every year.
4. Consistent Information Flow Without a Pattern (Figure 3.6): These filings exhibit a stable average monthly volume throughout the year, showing no discernible pattern. Examples include complex financial product prospectuses (Form 424B2) and supplemental information provided during a securities offering (Form FWP), as well as SEC comment letters (CORRESP, LETTER) used by the SEC to comment on filings published by companies and other entities, for example, to request more information.

Consistent with the overall pattern observed in SEC filings, we find a similar temporal variation in insider trading, activist investor activity, mutual fund summary prospectuses, and private securities offerings (Table 3.6). The results indicate a significant increase in the mean monthly filing volume of 21.3%, 11.7%, 14.8% and 13.4%, respectively. For these event-driven filings not disclosed on a pre-defined regulatory timeline but rather triggered by specific events, six of the eight filing types reach their highest volume month during the winter period (November–April). Material event disclosures and mutual fund prospectus materials peak just one month after the official Halloween winter period ends, in May. Not a single filing type reaches its peak volume during any summer month (Table 3.6). The majority of event-driven filings reach their lowest volume in September during the summer period of the Halloween effect. This includes filings disclosing insider trading activities, material events, activist investor announcements, and private securities offerings. When treating May as an extended winter month, we find that 10 out of 12 high and low peak volume months align with the summer and winter cycle of the Halloween period. The other two observations represent the second lowest month, and one is unrelated.

When analyzing the quarterly gross market values and quarterly buy and sell activities of all hedge funds from 2013 to 2023, we did not find any discernible temporal pattern that aligns with the winter and summer periods discussed here, nor did we uncover any meaningful additional insights. The same applies to analyst recommendations and their price target upgrades and downgrades.

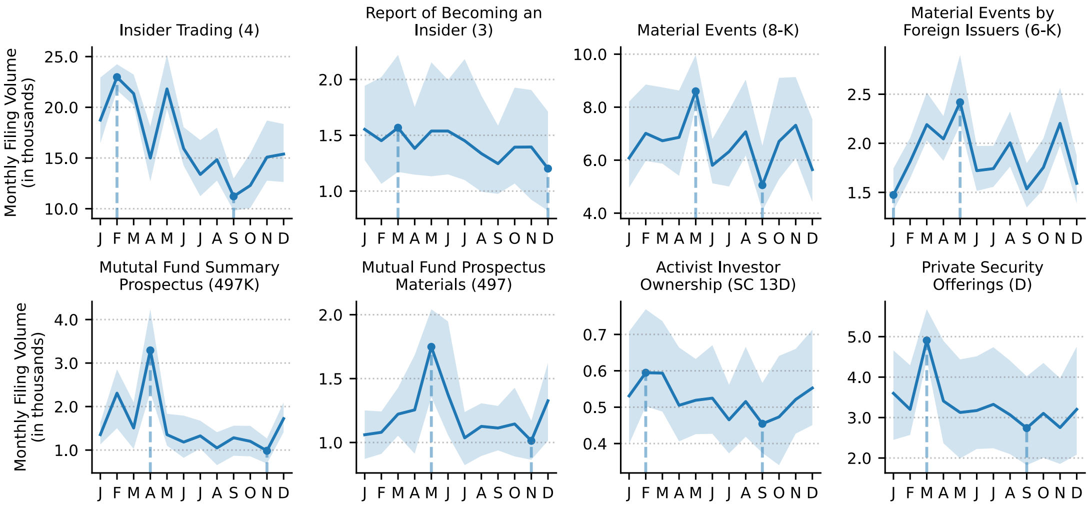


Figure 3.3: Patterns in Average Monthly Filing Volumes from 2004 to 2023 for Eight Event-Driven EDGAR Form Types, Not Caused by Pre-Defined Regulatory Timelines but by Triggering Events.

The forms include insider trading reports (Form 3 and 4), material event disclosures (Forms 8-K and 6-K), mutual fund and investment company filings (Form 497, 497K), ownership reports by activist investors (Form SC 13D and /A), and private security offerings (Form D and /A).

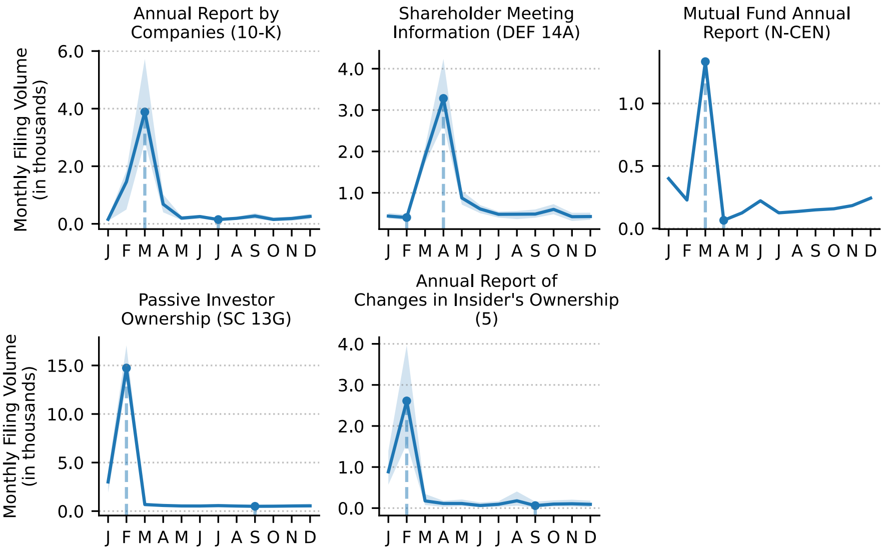


Figure 3.4: Once-Annually Occurring Volume Spikes of Five EDGAR Filing Types from 2004 to 2023.

The forms include annual reports by companies (Form 10-K), shareholder meeting information as proxy statements (Forms DEF 14A and DEFA14A), mutual fund annual reports (Form N-CEN), passive investor ownership disclosures (Forms SC 13G and SC 13G/A), and annual statements of ownership structure by insiders (Form 5).

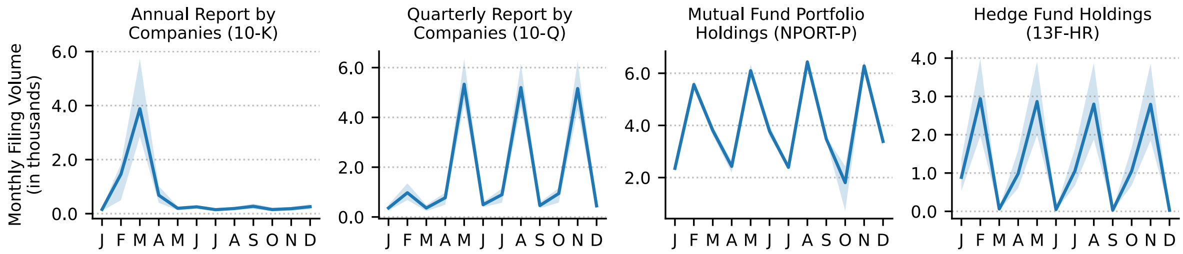


Figure 3.5: Quarterly Recurring Disclosure Pattern, Driven by Regulatory Schedules, as Measured by Monthly Average Filing Volume (January to December, 2004 to 2023).

Form 10-K and 10-Q filings are related as companies are required to disclose quarterly reports four times a year—one annual report on Form 10-K and three quarterly updates on Form 10-Q. Quarterly reports by companies peak in March, May, August, and November. Similarly, quarterly holding reports of mutual funds (investment companies) and hedge funds (institutional investment managers) on Forms NPORT-P and 13F-HR peak in February, May, August, and November each year.

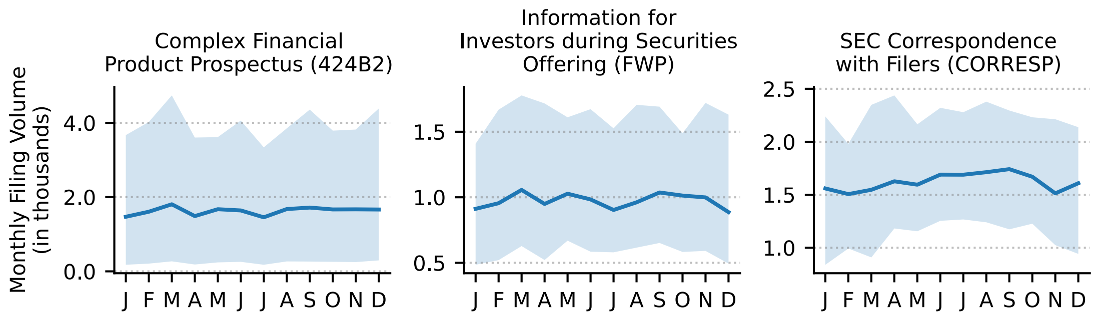


Figure 3.6: Consistent Flow of Information with No Discernible Pattern and Stable Average Monthly Filing Volumes Throughout the Years from 2004 to 2023.

These filings include complex financial product prospectuses (Form 424B2), information for investors during securities offerings (Form FWP), as well as correspondences between the SEC and its filers (CORRESP, LETTER).

	Share of Total Filing Volume	Period	Mean Monthly Filing Volume	Years	Difference (%)	Highest Volume Month	Lowest Volume Month	Peak Monthly Volume in Winter	Lowest Monthly Volume in Summer
Unscheduled Filings Triggered by Specific Events Rather Than a Predefined Regulatory Timeline									
Insider Trading (Form 4)	37.1%	May – Oct	14,911	20	+21.31%***	February	September	Yes	Yes
		Nov – Apr	18,089						
Activist Investor Ownership (Form SC 13D +/A)	1.1%	May – Oct	492	20	+11.73%***	February	September	Yes	Yes
		Nov – Apr	550						
Material Events (Form 8-K)	14.9%	May – Oct	6,593	20	+0.21%	May	September	No	Yes
		Nov – Apr	6,607						
Material Events By Foreign Issuers (Form 6-K)	4.2%	May – Oct	1,863	20	+1.25%	May	January, followed by September	No	No, September is 2nd lowest month
		Nov – Apr	1,886						
Mutual Fund Summary Prospectus and Pros. Materials (Form 497, 497K starting in 2009)	5.2%	May – Oct	2,166	20	+14.76%*	May	November	No	No
		Nov – Apr	2,486						
Private Security Offerings (Form D + /A) Introduced in 2009	5.6%	May – Oct	3,157	15	13.37%**	March	September	Yes	Yes
		Nov – Apr	3,579						

Table 3.6: Summary Statistics of Unscheduled Filings Triggered by Specific Events, Not by Regulatory Deadlines (2004–2023).

This table compares mean monthly volumes of selected unscheduled SEC filings between the winter (November–April) and summer (May–October) periods from 2004 to 2023. Filing types include insider trading (Form 4), activist investor activity (Form SC 13D/A), material events (Forms 8-K and 6-K), mutual fund summary prospectuses (Forms 497 and 497K, from 2009), and private offerings (Forms D and D/A, from 2009). It reports each form's share of total filings, period averages, percentage differences, and peak/low months, indicating whether peaks occur in winter and lows in summer. Significance levels: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

	Share of Total Filing Volume	Period	Mean Monthly Filing Volume	Years	Difference (%)	Highest Volume Month	Lowest Volume Month	Peak Monthly Volume in Winter	Lowest Monthly Volume in Summer
Filings with Annual Seasonal Volume Spikes									
Annual Reports by Companies (Form 10-K)	1.5%	May – Oct	202	20	+446.21%***	23047	16.41%	March	Yes
		Nov – Apr	1,104			117485	83.59%		
Shareholder Meeting Information (DEF 14A, DEFA14A)	1.9%	May – Oct	589	20	+94.21%***	67132	37.01%	April	Yes
		Nov – Apr	1,144			114249	62.99%		
Ownership Report by Passive Investors (SC 13G + /A)	4.2%	May – Oct	524	20	+537.92%***	59759	14.21%	February	Yes
		Nov – Apr	3,344			59759	85.79%		
Annual Report of Changes in Insiders' Ownership (Form 5)	0.8%	May – Oct	101	20	+556.37%***	11522	13.82%	February	Yes
		Nov – Apr	663			71865	86.18%		
Annual Report by Mutual Funds (Form N-CEN) (since 2018, N-SAR prior)	0.1%	May – Oct	153	5	+167.31%***	3665	31.12%	March	Yes
		Nov – Apr	408			8111	68.88%		

Table 3.7: Statistics of Filings with Once-Annually Occurring Volume Spikes (2004–2023).

This table compares mean monthly filing volumes between the winter (November–April) and summer (May–October) periods for SEC form types that exhibit once-annual volume spikes. Included are company annual reports (Form 10-K), shareholder meeting and proxy materials (Forms DEF 14A, DEFA14A), passive ownership disclosures (Forms SC 13G, SC 13G/A), insider ownership statements (Form 5), and mutual fund annual reports (Forms N-CEN and formerly N-SAR). The table reports each form's share of total filings, average monthly volumes, percentage differences, total volume distribution, peak months, and whether peaks fall within the winter period. Significance levels: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

In summary, our findings indicate an information surge during the winter periods (November to April) compared to the summer periods (May to October). Bouman and Jacobsen (2002) observed that winter periods consistently outperformed summer periods. This suggests that the information surge in winter coincides with observed outperformance, while the information drought in summer aligns with underperformance, indicating a correlation between filing volume and stock price performance. Another correlation between filing volume and performance is evident in the month of September. In the United States, September is the only month of the year that exhibits a negative average return over the past century (Fang et al., 2017; Hirsch, 2023). Coincidentally, September also has the lowest average monthly filing volume, a pattern observed consistently over the past 20 years.

Jeon et al. (2022) found that the influence of news on stock returns has increased over time, particularly after the adoption of the EDGAR system and the widespread use of the Internet. We therefore discuss in the next section how the surge in filing volume can result in the outperformance observed in the winter period compared to the summer period. We focus on filing types that exhibit a strong temporal pattern aligning with the Halloween effect periods. These filings also belong to the group of the top 2% of the most commonly disclosed information each year, and make up 86% of the total annual filing volume. Coincidentally, these filing types have shown the strongest results in studies examining the impact of SEC filings on stock prices. This convergence of factors makes them particularly relevant to our analysis of the Halloween effect.

3.3.1 Annual Reports (Form 10-K)

We begin with the most pronounced surge in volume: 10-K filings (annual reports with audited financial statements). The majority, 84%, of all annual reports in 10-K filings are disclosed during the winter period, from November to April (Figure 3.7, Table 3.8). March consistently marks the month with the highest disclosure volume, accounting for 49.56% of all annually published 10-Ks.

Cohen et al. (2020) found that changes to the language and construction of 10-Ks predict future returns, earnings, profitability, news announcements, and even firm-level bankruptcies. Portfolios that short companies which changed key sections of their 10-K reports year-over-year and buy those with no changes earn up to 188 basis points per month in risk-adjusted returns. Similarly, Loughran and McDonald (2014) discovered that 10-K file size correlates with subsequent stock return volatility. Specifically, a one-standard-deviation increase in file size, number of words, common words, and vocabulary leads to respective increases of 9%, 7%, 6%, and 6% of the absolute standardized unexpected earnings (SUE) standard deviation. They found that larger 10-K file sizes are associated with significantly higher post-filing date abnormal return volatility and higher absolute SUE.

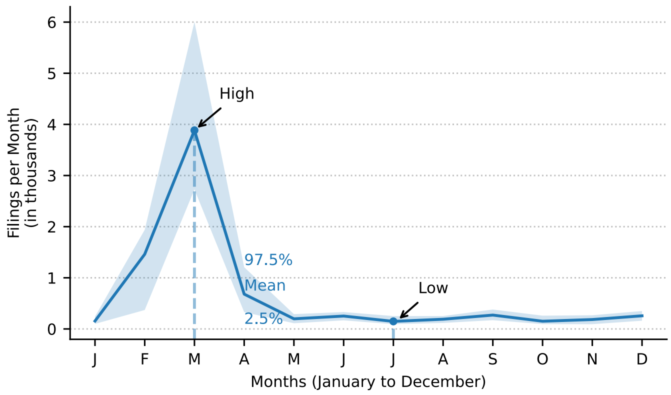


Figure 3.7: Average Monthly Filing Volume (January to December, 2004–2023) for All Annual Reports with Audited Financial Statements Filed on Form 10-K.

The blue line indicates the mean number of filings per month, with the light-blue shaded area indicating the interval between the 2.5% and 97.5% percentiles.

Monthly Filing Volume	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	153	1,480	**3,869**	672	197	248	**145**	188	266	150	183	251
Std. Dev.	49	503	1,026	262	54	49	50	42	75	55	50	77
Min	94	351	2,706	275	102	167	90	106	154	94	91	133
Median	133	1,716	**3,524**	595	190	254	130	190	266	**122**	180	242
Max	251	1,981	6,072	1,223	295	348	274	255	380	273	269	353

Table 3.8: Monthly Disclosure Statistics of Annual Reports with Audited Financial Statements on Form 10-K Filings (January to December, 2004–2023).

Two decades before Loughran and McDonald (2014), Bernard and Thomas (1989) replicated prior findings of a significant post-earnings-announcement drift. They showed that a long position in the highest SUE decile and a short position in the lowest decile yielded an abnormal return of approximately 4.2% over 60 days, annualized to about 18%. Most of the drift occurred within the first 60 trading days after the earnings announcement, with little evidence of significant drift beyond 180 trading days.

Another important aspect to consider is the impact of announcements on companies with customer–supplier relationships with the disclosing company. Cohen and Frazzini (2008) examined the effect of significant news, such as earnings announcements, disclosed by customer firms on their supplier firms' stock prices. They found that a long–short strategy based on customer returns yields significant abnormal returns (1.55% per month or 18.6% annualized). While this is specifically applicable to customer–supplier relationships, their findings provide additional support for our thesis. The information surge not only directly impacts the disclosing entities but also affects stock price movements in related supplier firms.

Similar to Cohen et al. (2020), You and Zhang (2008) highlighted the gradual incorporation of earnings information, noting that investors tend to underreact to information in complex 10-K filings, resulting in significant stock price drifts. Hirshleifer et al. (2009) confirmed these findings, showing that the immediate price and volume reaction to a firm's earnings surprise is much weaker on high-news days, with post-announcement drift being much stronger when a greater number of same-day earnings announcements are made by other firms.

The incorporation of information from annual reports into prices can continue for days to weeks after the release of the report. The timeframe of 60 days (as observed by Bernard and Thomas (1989)) coincides with the two-month gap between the March peak and May, marking the beginning of the underperforming summer cycle.

The importance of 10-K filings and their impact on the market is further validated by Kim et al. (2019), who found that more readable 10-K reports are associated with lower stock price crash risk. Additionally, Brown and Tucker (2011) discovered that while price reactions to modifications in the Management Discussion & Analysis (MD&A) section of 10-K filings have weakened, investors still respond positively to such changes. Other studies, such as those by Li (2010), Merkley (2013), Mayew et al. (2014), and Muslu et al. (2015), also highlight the impact of 10-K content, particularly the MD&A section, on stock prices.

Given that 84% of all annually published 10-Ks are disclosed during the winter period, combined with the observed effects that 10-Ks have on stock prices, we find strong support for the SEC filing volume surge directly impacting the overperformance.

One might argue that 10-Q filings, which disclose financial performance and operating results three times a year, should be equally important as 10-K filings. However, it is crucial to understand that 10-Qs contain significantly less information compared to 10-Ks. 10-Q filings include only unaudited financial results, whereas 10-Ks provide audited financial results and auditor opinions, among other material information. Additionally, 10-K filings consist of 21 sections, while 10-Qs include only 11 sections. Furthermore, the market reacts more strongly to 10-K filings than to 10-Qs (Griffin, 2003).

3.3.2 Ownership Reports: Insider Trading (Form 4) and Activist Investors (Form SC 13D)

Insider trading reports on Form 4 are filed by officers, directors, and significant shareholders who own more than 10% of a company's shares whenever they buy or sell shares in the company. The form includes details such as the insider's name, their relationship to the company, the date of the transaction, the number of shares traded, and the price at which the shares were bought or sold, and must be filed within two business days of the transaction.

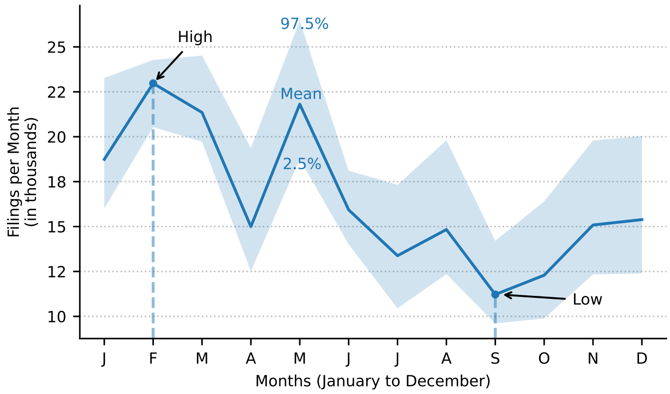


Figure 3.8: Average Monthly Filing Volume (January to December, 2004–2023) of the Largest Regulatory Information Class: Insider Trading Disclosures (Form 4).

Monthly Filing Volume	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	18,635	**22,948**	21,454	14,860	21,768	16,018	13,238	14,746	**11,133**	12,154	14,935	15,276
Std. Dev.	2,505	1,088	1,424	2,370	2,268	1,500	2,241	2,190	1,459	2,199	2,413	2,391
Min	15,789	20,437	19,618	12,293	18,268	13,979	10,229	12,112	9,482	9,551	12,063	12,292
Median	17,688	**23,153**	20,896	14,087	20,980	15,756	12,870	14,120	**10,900**	11,577	14,336	14,492
Max	23,377	24,285	24,965	20,313	26,972	18,151	17,440	20,338	14,487	16,840	20,197	20,450

Table 3.9: Monthly Disclosure Statistics of Insider Trading Disclosures on Form 4 (January to December, 2004–2023).

Figure 3.8 illustrates the temporal pattern in insider trading activity reported on Form 4. February and May consistently see the most insider trading, followed by a sharp decline from June to September, with September marking the month with the lowest disclosures.

The 22% increase in insider trading activity during the winter period compared to the summer period could additionally contribute to the outperformance, as explained by findings from Cohen et al. (2012), who found that opportunistic trades are significant predictors of future returns, while routine trades are not. A portfolio focusing on opportunistic trades yields higher abnormal returns compared to one focusing on routine trades. The increased insider trading activity, particularly opportunistic trades, likely contributes to the better performance observed in the winter months.

Cline et al. (2017) found that persistently profitable insider buys are associated with significant abnormal returns of 2.52%, with the market reaction to such trades being more pronounced in the post-electronic reporting period. The increased insider trading activity during the winter period, coupled with the abnormal positive returns observed by Cline et al. (2017), provides another plausible contributing factor to the outperformance during the winter period.

Further contributing to the better performance during the winter period is the market's reaction to activism announcements, with their frequency measured by the monthly volume of Forms SC 13D and D/A. Activist investors who intend to influence or control a company must file these forms within 10 days after acquiring more than 5% of the company's shares. Brav et al. (2008) identified positive abnormal returns of 7% to 8% around activism announcements. These filings reach their highest monthly volume in February and March and their lowest volume in September.

3.3.3 Material Event Disclosures (Form 8-K)

Material event disclosures on Form 8-K are filed by companies to inform about significant events that might materially impact a company's financial condition or operations. It covers a wide range of events, including entry into or termination of material agreements; completion of acquisition or disposition of assets; bankruptcy; shareholder voting matters; legal proceedings; changes in the company's executives, board, and certifying accountant; corporate changes such as mergers and reorganizations; or amendments to the articles of incorporation or bylaws. Form 8-K must be filed within four business days of the occurrence of the event.

As shown in Figure 3.9, a similar pattern is observable in material event disclosures, which peak in May and decline over the summer months, reaching their lowest volume in September (Table 3.10). It is important to note that some events in 8-K filings are disclosed on a repetitive schedule. For instance, Item 2.02 "Results of Operations" is often used by companies to disclose preliminary quarterly or annual results before filing their 10-K or 10-Q reports. Therefore, the August volume spike can be attributed to the quarterly results published in August, aligning with the 10-Q pattern shown further below.

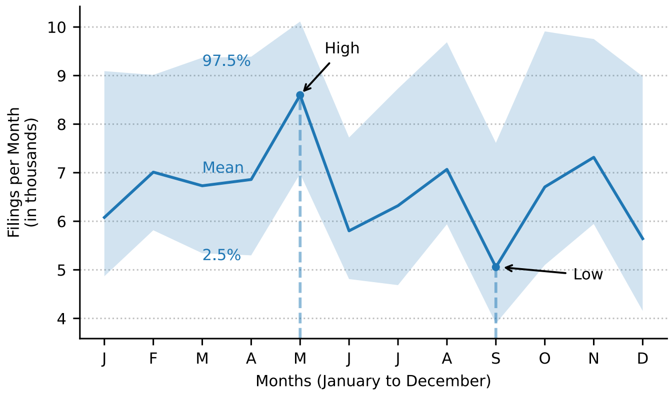


Figure 3.9: Average Monthly Filing Volume (January to December, 2004–2023) of the Second Largest Regulatory Information Class: Material Event Disclosures (Form 8-K).

Monthly Filing Volume	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	6,029	6,970	6,722	6,780	**8,617**	5,807	6,244	7,066	**5,018**	6,642	7,290	5,588
Std. Dev.	1,336	1,043	1,146	1,295	843	784	1,300	1,106	1,119	1,561	1,179	1,412
Min	4,825	5,778	5,151	5,230	6,410	4,703	4,565	5,789	3,811	4,991	5,935	4,084
Median	5,570	6,663	6,490	6,606	**8,497**	5,694	6,042	6,886	**4,552**	6,146	6,946	5,064
Max	9,183	9,118	9,443	9,974	10,155	8,149	9,323	10,108	8,140	9,914	10,180	9,344

Table 3.10: Monthly Disclosure Statistics of Material Event Disclosures on Form 8-K (January to December, 2004–2023).

For material event disclosures, no significant difference is observed between the November–April and May–October periods, with a negligible increase of 0.21%. However, when comparing the November–May period to the June–October period, we found a statistically significant increase of 11.31% (p-value < 0.01), highlighting a higher filing volume during the winter months.

The interaction between material event disclosures (Form 8-K), which see an increase in average monthly filing volume of 11.3% during the extended winter period, and insider trading was demonstrated by Cohen et al. (2015). Insiders seem to enjoy systematic abnormal returns of 42 basis points on average per trade during the "8-K Trading Gap," the days between a material event occurring and the company disclosing it in a Form 8-K, up to four days later. A trading strategy based on positive 8-K filings yields abnormal returns of 35.4 basis points, with insider trading during the 8-K Trading Gap predicting the directional impact of 8-K filings on stock prices with high accuracy. This led to the US

Congress passing the "8-K Trading Gap Act of 2021," requiring certain publicly traded companies to create policies designed to prevent executives from trading their securities after a significant event but before disclosing that event through a public filing.

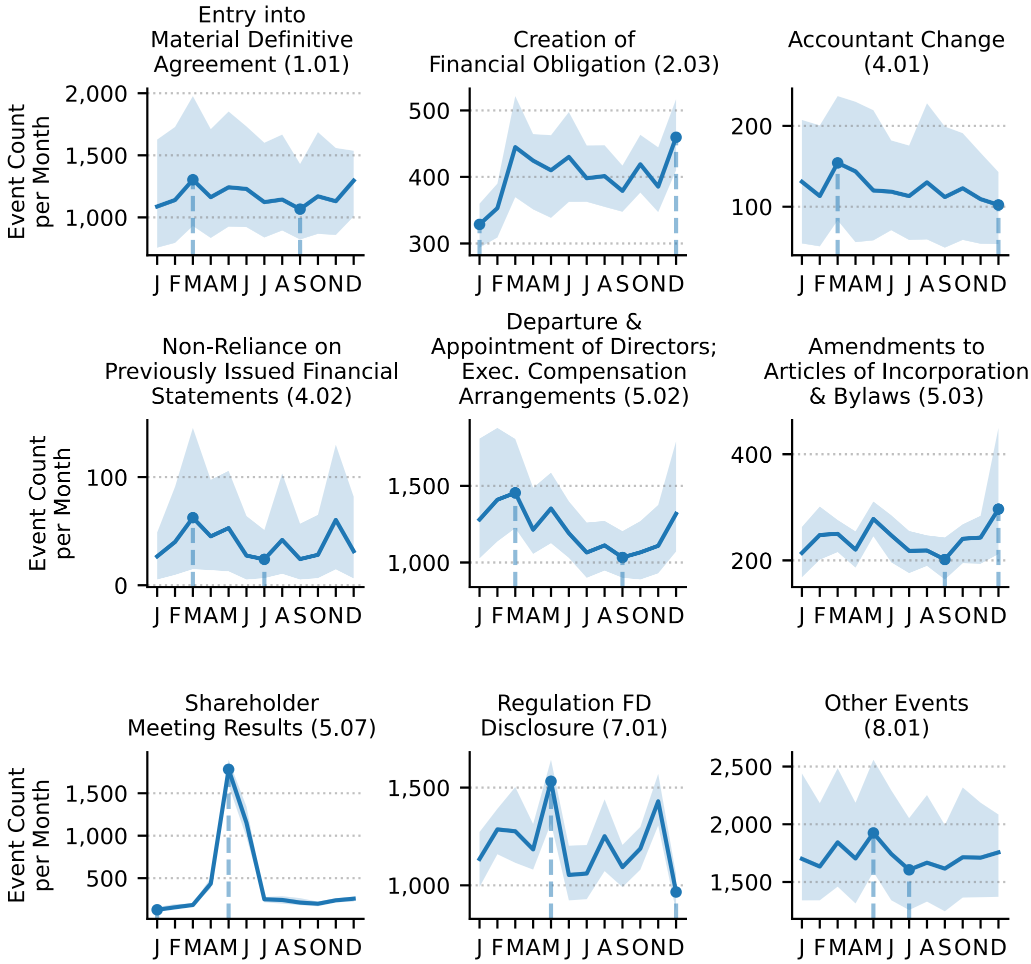


Figure 3.10: Average Monthly Occurrences of Selected Material Events from 2005 to 2023, as Extracted from Form 8-K Filings.

The events include Entry into a Material Definitive Agreement (Item 1.01), Creation of a Financial Obligation (Item 2.03), Accountant Change (Item 4.01), Non-Reliance on Previously Issued Financial Statements (Item 4.02), Departure and Appointment of Directors or Executive Compensation Arrangements (Item 5.02), Amendments to Articles of Incorporation & Bylaws (Item 5.03), Shareholder Meeting Results (Item 5.07), Regulation FD Disclosure (Item 7.01), and Other Events (Item 8.01). Each panel shows the mean event count per month (blue line), with the light blue shaded area indicating the interval between the 2.5% and 97.5% percentiles. High and low months are marked to highlight the seasonality and frequency of these corporate actions over the given period.

Additionally, Form 8-Ks are used to announce special dividends. Beladi et al. (2016) found that firms are more likely to announce special dividends at the end of the year, particularly in November and December, resulting in short-term abnormal returns that are higher during the Halloween period (November–April).

The majority of shareholder meetings occur between April and June, peaking in May each year, as evidenced by the average monthly filing volume of Item 5.07 disclosed in Form 8-K (Figure 3.10). Item 5.07, titled "Submission of Matters to a Vote of Security Holders," reports the outcomes of matters such as elections of directors, advisory votes on executive compensation, and other significant corporate actions, with a maximum delay of four days after the vote. Although the April to June period spans the transition from the winter to the summer cycle, we still assume shareholder meetings positively impact the winter cycle. This is because of the distribution of various information to shareholders weeks and months in advance to prepare for decision-making. Proxy statements (Form DEF 14A), which provide shareholders with important information to be voted on and solicit their votes on various proposals and matters, reach their highest disclosure volume in April. Brickley (1986) found significant positive abnormal returns around stockholder meeting dates, therefore suggesting that the surge of shareholder meetings during the winter-to-summer transition period (April to June) partially contributes to the Halloween effect.

In addition, Dimitrov and Jain (2011) identified that managers generally report more positive news before annual shareholder meetings to reduce shareholder discontent. They found that firms exhibit significantly positive abnormal returns during the 40 days leading up to the annual shareholder meeting, which falls within the end of winter cycle. Notably, approximately two-thirds of these positive pre-meeting returns are reversed in the post-meeting period, which falls within the beginning of the summer cycle.

3.4 Similar Patterns in European Markets

While we do not observe a similar seasonal pattern across the total number of regulatory disclosures submitted to the London Stock Exchange's (LSE) RNS system by UK-regulated entities, we do find notable patterns in specific disclosure types between 2013 and 2023. For instance, the number of annual reports disclosed by companies listed on the LSE during the winter period (November–April) is 48% higher compared to the summer period (May–October), with the majority of these reports consistently published in March each year (Table 3.11, Figure 3.11). This March peak mirrors the pattern observed for US-listed companies, which also show the highest volume of annual report disclosures in March. Similarly, the winter period sees a 61% increase in the issuance of proxy voting forms, aligning with the majority of annual general meetings occurring in April and May each year.

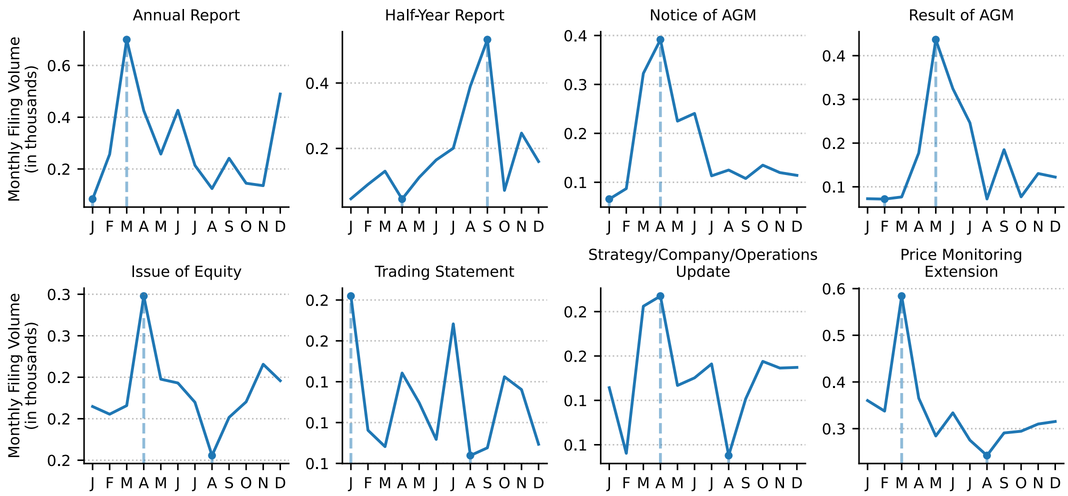


Figure 3.11: Average Number of Regulatory Disclosures per Month by Form Type Filed by Publicly Listed Companies with the LSE's RNS System (2013–2023).

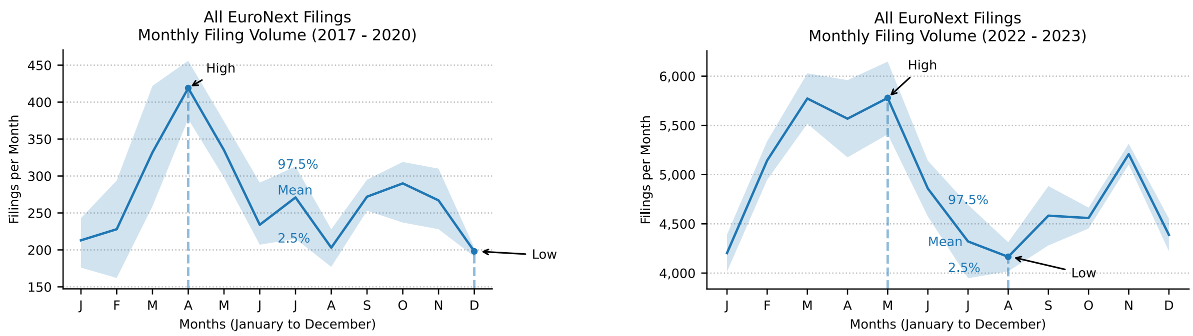


Figure 3.12: Monthly Average Number of Regulatory Disclosures from Companies Listed on the EuroNext Exchange (2017–2023).

The left panel shows the monthly average number of disclosures filed from 2017 to 2020, while the right panel focuses on the period from 2022 to 2023. Data from 2021 was excluded due to a new regulation implemented mid-year that caused a tenfold increase in disclosure volume, creating an outlier.

Type	Nov – April Mean Period Filing Volume	Nov – April Standard Deviation	May – Oct Mean Period Filing Volume	May – Oct Standard Deviation	Difference (%)
All RNS Filings	179,823	65,802	169,329	61,747	−5.84%*
By Filing Type					
Annual Report	2,091	486	1,409	416	+48.41%***
Half-Year Report	719	146	1,471	226	−51.16***
Proxy Form	278	49	173	31	+60.95%***
Notice of AGM	618	187	569	239	+8.7%*
Result of AGM	579	211	1,195	415	−51.54%***
Trading Statement	790	324	709	239	+11.42%
Issue of Equity	1,498	775	1,394	521	+7.43%
Strategy & Operations Updates	941	442	867	269	+8.54%
Price Monitoring Extensions	1,257	1,318	958	977	+31.22%

*Table 3.11: Statistics of Disclosure Volumes via the LSE's RNS System (2013–2023). Average disclosure volumes of regulatory filings published by U.K.-regulated entities through the London Stock Exchange's RNS system, comparing the winter (November–April) and summer (May–October) periods. Data covers 2013 to 2023. Seven out of nine information categories show significantly higher disclosure volumes in winter. Statistical significance is denoted as follows: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

Year	Nov – Apr	May – Oct	Difference (%)
2017	1,887	1,671	+12.93
2018	1,674	1,711	−2.16
2019	1,562	1,474	+5.97
2020	1,514	1,575	−3.87
2022	31,660	29,930	+5.78
2023	28,909	26,602	+8.67
Total	67,206	62,963	+6.74
Mean	19,202	17,989	+4.87
Std. Dev.	23,253	21,755	
Median	19,202	17,989	+5.78

Table 3.12: Total EuroNext Company Disclosures by Year and Period (2017–2023). Total number of regulatory disclosures by companies listed on EuroNext, split by winter (November–April) and summer (May–October) periods. Data spans 2017 to 2023, excluding 2021 due to a mid-year regulatory change that caused a tenfold increase in disclosure volume, making it an outlier.

Additionally, we find an 11% increase in trading statements during the winter months. Trading statements provide periodic updates from publicly listed companies on key financial performance indicators, including revenue and profitability metrics, market conditions, operational highlights (e.g., contract wins), future guidance and outlook, and significant events such as litigation, acquisitions, or divestments.

The volume of issuance of equity disclosures increases by 7.4% during the winter compared to the summer, reflecting a seasonal preference for raising capital during this period. Price Monitoring Extensions (PMEs) represent regulatory announcements that are triggered automatically by the LSE to highlight significant volatility or unusual price movements in listed securities during the auction periods at the start or end of trading sessions, and also exhibit a seasonal pattern. During the winter months, 31% more PMEs are issued, suggesting increased market volatility and unusual price movements during this time.

While September consistently marks the month with the lowest number of regulatory disclosures in the US, August represents the lowest filing volume for companies listed in the UK and on EuroNext exchanges (Figure 3.12, Table 3.12). In the UK, August sees the lowest number of filings across all disclosure types, including equity issuance announcements, trading statements, operational updates, and PMEs, marking August as the least active month for regulatory disclosures.

3.5 Press Releases vs. Financial News

Bouman and Jacobsen (2002) did not find a discernible pattern in general financial news volume. Although journalists and news outlets like Reuters, Bloomberg and Wall Street Journal use SEC and other regulatory filings as an information source, their job is to provide a constant flow of news. As Shiller (2016) states, "many news stories in fact seem to have been written under a deadline to produce something – anything – to go along with the numbers from the market."

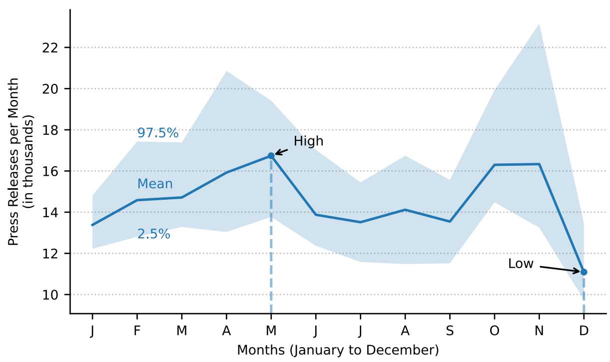


Figure 3.13: Average Number of Press Releases Published per Month by NASDAQ- and NYSE-Listed Companies (2009–2023).

The graph shows the mean number of press releases per month, with the 2.5% and 97.5% percentiles shaded.

Monthly PR Volume	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	13,158	14,526	14,701	15,923	**16,554**	13,827	13,469	13,964	13,394	16,046	16,101	**11,093**
Std. Dev.	1,210	1,494	1,432	2,255	1,965	1,397	1,097	1,773	1,232	1,957	3,019	1,096
Min	10,045	12,716	13,255	12,580	13,531	12,335	11,308	11,355	11,066	12,499	12,850	9,722
Median	13,419	14,184	14,369	15,486	**17,216**	13,355	13,382	13,790	13,675	15,909	15,695	**10,849**
Max	14,978	17,749	17,517	22,267	19,594	17,493	15,702	17,065	16,089	20,800	25,735	13,646

Table 3.13: Descriptive Statistics of Monthly Press Releases Published by NASDAQ- and NYSE-Listed Companies (2009–2023).

Period	Years	Mean Monthly PR Volume	Difference
May – Oct	15	14,542	−2.01%
Nov – Apr	15	14,250	
Jun – Sep	15	13,663	+8.05%**
Oct – May	15	14,763	
Jun – Sep	15	13,663	+11.88%***
Oct – May (ex. Dec)	15	15,287	

Table 3.14: Comparison of Mean Monthly Press Release Volumes Across Periods from 2009 to 2023 for All NASDAQ- and NYSE-Listed Companies.

The table displays the mean monthly PR volumes, percentage differences between periods, and t-test results to indicate statistical significance.

We find a similar, albeit less pronounced, temporal pattern in the number of press releases (PRs) companies publish each month. From 2009 to 2023, PRs of all NASDAQ and NYSE listed companies exhibit the highest average volume in May, followed by a sharp drop over the summer months from June to September (Figure 3.13, Table 3.13). Unlike SEC filings, the fewest PRs are published in December. This can be attributed to the Christmas and New Year holidays, which result in fewer working days and a potential decrease in non-material PRs being disclosed, leaving mainly material PRs to be published.

During periods of low SEC filing volume, journalists still have a job and need to report news, which may involve recycling old stories or covering less significant topics. This constant supply of content means that the volume of general financial news includes a lot of noise, which may obscure the signal of the underlying SEC filing patterns. The varied motivations of journalists to publish news can further obscure the patterns. In some cases, journalists have accepted payments for stock recommendations, and certain newspapers have even taken a share of profits from traders in exchange for favorable coverage (Quinn and Turner, 2020). When comparing the average monthly volume for the summer period (May to October) to the winter period (November to April), we do not find a statistically significant difference. However, when comparing the summer months (June to September) to an extended winter period (October to May), we find a statistically significant increase in average monthly PR volume during the winter of 8.05% (Table 3.14).

Unlike journalists, companies have strong incentives to report positive news in their PRs and are not obligated to report when there is nothing significant to disclose. Journalists, on the other hand, must continuously produce content to maintain their employment. This explains the high noise level in general financial news compared to corporate press releases.

3.5.1 Extended Reasoning Beyond Regulatory Filings

We will now indulge in a thought experiment trying to logically derive through inversion why it is reasonable to assume that an information surge leads to positive momentum.

For the information surge in winter to cause outperformance during this period, information generally needs to have a stronger positive impact on stock prices than a negative one. In other words, an increase in information disclosure tends to generate more positive sentiment, leading to stronger stock price momentum during the November–April period. Information can create positive stock price momentum, negative momentum, or have a neutral impact on prices (i.e., no impact at all). To validate the assumption that more information leads to improved performance and strengthen our thesis, let's apply Charlie Munger's famous approach of "invert, always invert," a principle originating from the German mathematician Carl Gustav Jacob Jacobi (Munger et al., 2023), and consider the opposite scenario: the majority of information causes negative stock price momentum.

From 1926 to 2012, the S&P 500 index exhibited an average positive annual return of 9.7%. The compound annual real return after inflation across all U.S. stocks from 1802 to 2012 was 6.6%, and 3.6% for bonds, including major market crashes such as the Wall Street Crash of 1929, Black Monday in 1987, the Dot-Com bubble burst in 2000, and the Global Financial Crisis in 2008 (Siegel, 2014).

Over the last 100 years, the volume of information has undoubtedly increased year-over-year. If information would have had a dominant negative impact, we would expect to see long-term negative stock price momentum, with prices declining as information accumulated. However, this is not what we observe. Instead, we see positive annual returns over the long run. This suggests that while prices may drop temporarily due to the negative impact of information, there must be other factors driving the long-term positive momentum.

Potential factors include the following:

- Weather: Studies by Saunders (1993) and Hirshleifer and Shumway (2003) found that stock prices tend to be higher on sunny days and lower on cloudy days. Keef and Roush (2016) identified that temperatures and wind speeds also impact stock prices, while trading volume drops significantly during blizzards (Loughran and Schultz, 2004). However, these effects are short-lived, impacting only the day and the subsequent day of the corresponding weather, and cannot explain longer-term trends.
- Daylight Saving Time: Kamstra et al. (2000) showed that daylight-saving weekends are typically followed by large negative returns on financial market indices, which is statistically significant for the U.S., Canadian, and U.K. indices across various metrics. However, the impact is only short-lived.
- Sports Events: Edmans et al. (2007) analyzed 1,100 international soccer matches and 1,500 additional games from cricket, rugby, ice hockey, and basketball from January 1973 to December 2004. They found that national stock markets experience significant negative returns following losses in major soccer matches. Losses in other sports also impact returns, though the effect is generally smaller, and there is no significant impact on positive returns following wins.
- Day of the Week: Cross (1973) found that stock prices rose more often on Fridays than on any other day of the week and least often on Mondays. Lakonishok and Smidt (1988) confirmed this weekend effect, noting significantly negative mean stock returns on Mondays and positive returns on the last trading day of the week. Dellavigna and Pollet (2009) also identified that the immediate stock price response to earnings announcements is 15% lower on Fridays compared to other weekdays. However, these effects are short-lived and impact only single-day returns.
- Holidays: Bergsma and Jiang (2015) found that stocks earn 2.5% higher abnormal returns in the month of a cultural New Year relative to other months. Fang et al. (2017) discovered a strong link between school holidays and market returns, with stock market returns in the month after major school holidays being 0.6% to 1% lower than in other months. Ariel (1990) found that stocks tend to advance with disproportionate frequency and high mean returns on the trading day prior to holidays, averaging nine to fourteen times the mean return for the remaining days of the year. Lakonishok and Smidt (1988) observed the Holiday Effect, with significantly higher returns before holidays compared to regular days. Kim and Park

(1994) confirmed these findings, noting abnormally high returns on the trading day before holidays in the stock markets of the UK and Japan, independent of the US holiday effect.

In conclusion, while temporary factors such as weather, daylight saving time, sports events, day of the week, and holidays can influence short-term stock price movements, they do not account for the long-term positive momentum observed in the market. One could argue that emotions, psychology, and biases might play a role (Pring, 1995).

However, this would imply that those factors cause positive stock momentum in the long run, which is unlikely to be the primary driver. Therefore, the assumption that the majority of information causes negative momentum cannot be true, supporting our hypothesis that information has a stronger positive impact on stock prices.

The third and final scenario is that information generally has no impact on stock prices. If this were true, markets would not be driven by information but by other factors, primarily emotions, behavioral biases and psychology, as the factors outlined above are only short-lived. This would imply that market prices over the last 100 years are driven solely by investment psychology rather than fundamental, material information. If this were the case, any data-based valuation, statistical investment model, or number-driven approach would be rendered useless. Since these models are not universally considered ineffective, it follows that information cannot be entirely neutral in its impact on stock prices.

Thus, it is reasonable to conclude that information does impact stock prices, supporting the idea that information in general has a stronger positive influence on price momentum than a negative one. Thereby strengthening our findings that the information surge in winter is a plausible contributing factor for the outperformance during the November-to-April period.

3.6 Limitations and Future Research Opportunities

Our study does not attempt to establish a causal explanation for the Halloween effect, nor does it claim a direct cause-and-effect relationship between disclosure volume and seasonal stock market performance. Rather, we identify a new recurring seasonal pattern in regulatory disclosures that aligns with the Halloween effect periods over the analyzed period from 2004 to 2023. While prior research has demonstrated that various SEC form types significantly impact stock prices, we do not empirically test the direct relationship between disclosure volume and stock returns. The reason such analysis was not conducted is straightforward: it would require an entirely new research project, effectively consolidating and replicating over 20 prior studies on the most common form types and their impact on stock prices into a single comprehensive study. Our findings provide a compelling reason to perform such a new project.

Our analysis is descriptive in nature, highlighting strong temporal correlations but not establishing causality. The mechanism through which SEC filings may influence the Halloween effect remains unexplored and provides a promising avenue for future research. To validate a causal link, more rigorous empirical testing, potentially at the firm or event level and incorporating return data, is needed.

Additionally, while the Halloween effect is often more pronounced in European markets, our analysis of European regulatory disclosures is limited. We obtained London Stock Exchange data from 2013 to 2023 and EuroNext data from 2017 to 2023. However, the EuroNext dataset excludes 2021 due to a mid-year change in disclosure requirements, which caused an anomalous increase in filings. Despite these constraints, we observe similar seasonal disclosure patterns in both markets, with EuroNext data showing peak volumes in April and May.

Overall, while our findings offer a compelling new angle on the Halloween effect, the empirical scope remains limited, and further research is required to assess causality and extend analysis across broader international markets.

3.7 Conclusion

We provide a new puzzle piece to help decipher the Halloween effect anomaly, known as "Sell in May and Go Away," which suggests that stocks perform better between November and April than they do between May and October. By analyzing all regulatory SEC disclosures from 2004 to 2023, including filings from publicly listed companies, insiders, institutional investors, funds, and other market participants, we identify an annually recurring pattern in disclosure volume. This pattern aligns with the Halloween effect periods, with an increase in disclosure volume during the winter (November to April) and a decline during the summer (May to October). September stands out as the month with the lowest disclosure volume in the U.S., aligning with prior findings that September is the only month in the past century to exhibit a negative average return.

Across all information categories, we find a 16% increase in disclosure volumes during the winter period compared to the summer. Insider trading, activist investor activity, and private securities offerings exhibit similar patterns, with filing volumes rising by 22%, 11%, and 13%, respectively, during the winter period. The publication of annual reports with audited financial statements increases by 473% during the winter compared to the summer.

We also observe similar seasonal disclosure patterns in European markets. For example, companies listed on the London Stock Exchange publish 48% more annual reports during the winter than in the summer, with March consistently being the peak month. Price Monitoring Extensions, triggered by market volatility and unusual price movements, increase by 31% during the winter months, while trading statements, which include updates on company performance, rise by 11% during the same period. In contrast, August consistently marks the lowest filing volume for companies listed on the London Stock Exchange and EuroNext exchanges.

We do not attempt to establish a causal explanation for the Halloween effect, nor do we claim a direct cause-and-effect relationship between disclosure volume and seasonal stock market performance. With prior research demonstrating the significant impact of SEC filings on stock prices, and our finding of the alignment of regulatory information surges and droughts with the periods of stock overperformance and underperformance during the winter and summer, respectively, we offer a compelling new piece to the Halloween effect puzzle. Future research could empirically test the direct relationship between disclosure volume and stock returns across the most common SEC form types to evaluate the mechanism through which SEC filings may influence the Halloween effect.

Bibliography

Andrade, S. C., Chhaochharia, V., and Fuerst, M. E. (2013). “Sell in May and go away” just won’t go away. *Financial Analysts Journal*, 69(4):94–105.

Ariel, R. A. (1990). High Stock Returns before Holidays: Existence and Evidence on Possible Causes. *The Journal of Finance*, 45(5):1611–1626.

Beladi, H., Chao, C.-C., and Hu, M. (2016). The Christmas effect—Special dividend announcements. *International Review of Financial Analysis*, 43:15–30.

Bergsma, K. and Jiang, D. (2015). Cultural New Year Holidays and Stock Returns around the World. *Financial Management*, 45(1):3–35.

Bernard, V. L. and Thomas, J. K. (1989). Post-Earnings-Announcement drift: Delayed price response or risk premium? *Journal of Accounting Research*, 27:1–36.

Bouman, S. and Jacobsen, B. (2002). The Halloween indicator, “Sell in May and go away”: Another puzzle. *American Economic Review*, 92(5):1618–1635.

Brav, A., Jiang, W., Partnoy, F., and Thomas, R. (2008). Hedge fund activism, corporate governance, and firm performance. *The Journal of Finance*, 63(4):1729–1775.

Brickley, J. A. (1986). Interpreting common stock returns around proxy statement disclosures and annual shareholder meetings. *Journal of Financial and Quantitative Analysis*, 21(3):343–355.

Brown, S. V. and Tucker, J. W. (2011). Large-sample evidence on firms’ year-over-year MD&A modifications. *Journal of Accounting Research*, 49(2):309–346.

Cline, B. N., Gokkaya, S., and Liu, X. (2017). The persistence of opportunistic insider trading. *Financial Management*, 46(4):919–964.

Cohen, A., Jackson, R. J., and Mitts, J. (2015). The 8-K trading gap. Technical report, Columbia Law and Economics Working Paper No. 524.

Cohen, L. and Frazzini, A. (2008). Economic links and predictable returns. *The Journal of Finance*, 63(4):1977–2011.

Cohen, L., Malloy, C., and Nguyen, Q. (2020). Lazy Prices. *The Journal of Finance*, 75(3):1371–1415.

Cohen, L., Malloy, C., and Pomorski, L. (2012). Decoding inside information. *The Journal of Finance*, 67(3):1009–1043.

Cross, F. (1973). The behavior of stock prices on Fridays and Mondays. *Financial Analysts Journal*, 29(6):67–69.

Dellavigna, S. and Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *The Journal of Finance*, 64(2):709–749.

Dimitrov, V. and Jain, P. C. (2011). It's showtime: Do managers report better news before annual shareholder meetings? *Journal of Accounting Research*, 49(5):1193–1221.

Edmans, A., García, D., and Norli, O. (2007). Sports sentiment and stock returns. *The Journal of Finance*, 62(4):1967–1998.

Fang, L., Lin, C., and Shao, Y. (2017). School holidays and stock market seasonality. *Financial Management*, 47(1):131–157.

Griffin, P. A. (2003). Got information? Investor response to form 10-K and form 10-Q EDGAR filings. *Review of Accounting Studies*, 8(4):433–460.

Haggard, K. S. and Witte, H. D. (2010). The Halloween effect: Trick or treat? *International Review of Financial Analysis*, 19(5):379–387.

Hirsch, J. A. (2023). *Stock Trader's Almanac 2024*. John Wiley & Sons.

Hirshleifer, D., Lim, S. H., and Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, 64(5):2289–2325.

Hirshleifer, D. and Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3):1009–1032.

Jacobsen, B. and Marquering, W. (2008). Is it the weather? *Journal of Banking & Finance*, 32(4):526–540.

Jacobsen, B. and Visaltanachoti, N. (2009). The Halloween effect in U.S. sectors. *The Financial Review*, 44(3):437–459.

Jeon, Y., McCurdy, T. H., and Zhao, X. (2022). News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics*, 145(2):1–17.

Kamstra, M. J., Kramer, L. A., and Levi, M. D. (2000). Losing sleep at the market: The daylight saving anomaly. *The American Economic Review*, 90(4):1005–1011.

Keef, S. P. and Roush, M. L. (2016). The Weather and Stock Returns in New Zealand. *Quarterly Journal of Business and Economics*, 41:61.

Kim, C. and Park, J. (1994). Holiday effects and stock returns: Further evidence. *Journal of Financial and Quantitative Analysis*, 29(1):145–157.

Kim, C., Wang, K., and Zhang, L. (2019). Readability of 10-K Reports and Stock Price Crash Risk. *Contemporary Accounting Research*, 36(2):1184–1216.

Lakonishok, J. and Smidt, S. (1988). Are seasonal anomalies real? A Ninety-Year perspective. *The Review of Financial Studies*, 1(4):403–425.

Li, F. (2010). The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. *Journal of Accounting Research*, 48(5):1049–1102.

Loughran, T. and McDonald, B. (2014). Measuring Readability in Financial Disclosures. *The Journal of Finance*, 69(4):1643–1671.

Loughran, T. and Schultz, P. (2004). Weather, stock returns, and the impact of localized trading behavior. *Journal of Financial and Quantitative Analysis*, 39(2):343–364.

Lucey, B. M. and Zhao, S. (2008). Halloween or January? Yet another puzzle. *International Review of Financial Analysis*, 17(5):1055–1069.

Mayew, W. J., Sethuraman, M., and Venkatachalam, M. (2014). MD&A disclosure and the firm's ability to continue as a going concern. *The Accounting Review*, 90(4):1621–1651.

Merkley, K. J. (2013). Narrative disclosure and earnings performance: Evidence from R&D disclosures. *The Accounting Review*, 89(2):725–757.

Munger, C., Kaufman, P., Collison, J., and Buffett, W. (2023). *Poor Charlie's Almanack: The Essential Wit and Wisdom of Charles T. Munger*. Stripe Matter Incorporated.

Muslu, V., Radhakrishnan, S., Subramanyam, K. R., and Lim, D. (2015). Forward-looking MD&A disclosures and the information environment. *Management Science*, 61(5):931–948.

Powell, J. G., Shi, J., Smith, T., and Whaley, R. E. (2009). Political regimes, business cycles, seasonalities, and returns. *Journal of Banking & Finance*, 33(6):1112–1128.

Pring, M. J. (1995). *Investment psychology explained: Classic strategies to beat the markets.* John Wiley & Sons.

Quinn, W. and Turner, J. D. (2020). *Boom and Bust: A Global History of Financial Bubbles.* Cambridge University Press.

Saunders, E. M. (1993). Stock prices and the Wall Street Weather. *The American Economic Review*, 83(5):1337–1345.

Schroeder, J. (2025a). A New Puzzle Piece for the "Sell in May, and Go Away" Anomaly: Regulatory Disclosures. *International Journal of Financial Studies.*

Schroeder, J. (2025b). Anomalies, Trends and Patterns in Disclosure Activities: Understanding EDGAR. Working Paper.

Shiller, R. J. (2016). *Irrational Exuberance: Revised and Expanded Third Edition.* Princeton University Press.

Siegel, J. J. (2014). *Stocks for the Long Run 5/E: The Definitive guide to financial market returns & Long-Term investment Strategies.* McGraw Hill Professional.

You, H. and Zhang, X.-J. (2008). Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies*, 14(4):559–586.

Zhang, C. Y. and Jacobsen, B. (2021). The Halloween indicator, "Sell in May and Go Away": Everywhere and all the time. *Journal of International Money and Finance*, 110:102268.

Chapter 4: Outperforming the Market: Portfolio Strategy Cloning from SEC 13F Filings

Author: Jan L. Schroeder (jan@sec-api.io)*

Abstract
Can mirroring of hedge fund strategies lead to market-outperforming returns? Our findings demonstrate that cloned portfolios in the top quartile, derived from SEC Form 13F filings, replicate the funds' performances and exceed the S&P 500 index by 24.3% on an annualized risk-adjusted basis. Analyzing over 150,000 portfolios between 2013 and 2023, we compare original versus replicated strategies across twelve performance metrics. Except for annualized volatility, maximum drawdown and tracking error, we demonstrate that cloned portfolios rebalanced on the disclosure date of filings, successfully mirror the performance of the original funds, including both market-underperforming and -outperforming funds.

Publication Details: Working Paper. Submitted to the Journal of Investing.

Keywords: Alpha Cloning, Portfolio Replication, SEC 13F Filings, Institutional Investment Managers, Quantitative Analysis of Fund Performances and Portfolio Cloning

Acknowledgements: We would like to thank Jonathan Lewellen, Professor of Finance at Tuck School of Business at Dartmouth and Research Associate at the National Bureau of Economic Research, and Nikolai Roussanov, Professor of Finance at Wharton, University of Pennsylvania, for their invaluable comments and insights.

The following is based on Schroeder (2025b).

*Jan L. Schroeder is the founder and CEO of SEC-API.io and Research Fellow at the Department of Finance, TU Dortmund. The authors have no competing interests to declare. All errors are our own.

4.1 Introduction

Inspired by Frank et al. (2004) and Verbeek and Wang (2013), we analyze the unexplored viability of replicating the investment strategies of top-performing institutional investment managers, commonly known as hedge funds, to determine whether such replication can emulate the performance of the original funds and ultimately achieve returns that surpass the performance of the SP500 benchmark index. By creating cloned portfolios from the holdings data disclosed in SEC 13F filings, we endeavor to mimic the performance characteristics of the source funds. Our analysis encompasses over 150,000 portfolios, while investigating more than 13,000 institutional investment managers from 1998 to 2023.

4.1.1 Registered Investment Companies vs. Institutional Investment Managers

Clarifying the distinctions between registered investment companies and institutional investment managers—and by extension, mutual funds and hedge funds—is essential before discussing our paper. Mutual funds are classified as registered investment companies (RICs) under Section 3(c)(1) and Section 3(c)(7) of the Investment Company Act of 1940. There are three types of RICs: open-end management investment companies, closed-end management investment companies, and unit investment trusts (UITs). Open-end funds, commonly known as mutual funds, continuously offer new shares and redeem existing shares at their net asset value. They do not trade on exchanges like publicly listed companies or ETFs; instead, investors can purchase and redeem shares directly from the fund. An example of an open-end fund is the Vanguard Federal Money Market Fund (ticker: VMFXX). Hedge funds, while also being investment companies, are exempt from the above regulations and are classified as institutional investment managers (IIMs). They are primarily regulated under the Securities Exchange Act of 1934, particularly Section 13(f). An IIM is defined as an entity that exercises investment discretion over an account holding equity securities with an aggregate fair market value of at least $100 million on the last trading day of any month of a calendar year. Unlike mutual funds, hedge funds are not identified by ticker symbols. Although all hedge funds with assets under management (AUM) exceeding $100 million are IIMs, not all IIMs are hedge funds. Form ADV distinguishes seven types of IIMs: hedge funds, liquidity funds, private equity funds, real estate funds, securitized asset funds, venture capital funds, and private funds.

For readability, we will use the term "mutual fund" to refer specifically to "open-end management investment companies," a type of registered investment company. When we refer to "hedge funds" or simply "funds," we mean all "institutional investment managers" that are regulated under Section 13(f) and thus disclose their holdings through Form 13F filings on the SEC EDGAR system. Our analysis focuses solely on institutional investment managers.

Both mutual funds and hedge funds must disclose their holdings publicly on a quarterly basis through the SEC's EDGAR system. Mutual funds use Form N-PORT from 2019 (previously N-Q), while hedge funds use Form 13F (13F-E until 1999, and 13F-HR starting in 1998). It is important to note that hedge funds do not disclose regulated information such as their gross market values and portfolio holdings except through the SEC's EDGAR system and FINRA. Consequently, any database that claims to provide monthly performance data on hedge funds is inherently unreliable, as there is no verifiable source of truth for such data. The only dependable sources are the quarterly published 13F filings and the annual Form ADV filings. Therefore, researchers should exercise caution and avoid relying on databases that provide hedge fund performance data with a resolution higher than quarterly. Our analysis exclusively focuses on Form 13F data, providing insights into hedge funds, rather than Form N-PORT data associated with mutual funds.

4.1.2 Prior Research

We analyze the viability of replicating the buy and sell activities of over 150,000 hedge fund portfolios to compare the performance of original and cloned portfolios across twelve performance and risk factors. Our study diverges from existing research by focusing exclusively on institutional investment managers, excluding mutual funds, and replicating entire portfolios on a quarterly basis rather than using subsets or factor-based replicas. Previous research on performance replication falls into three categories: hedge fund performance cloning, mutual fund performance cloning, and straightforward fund performance evaluation without cloning.

Table 4.1 provides an overview of previous research on hedge fund performance cloning, detailing each study's sample size (number of funds analyzed), historical data period, cloning strategy employed, and authors.

Interestingly, all prior research in the hedge fund performance cloning category focuses on replicating specific aspects of portfolios, either through deriving particular factors (Hasanhodzic and Lo, 2007; Hayes, 2012; Weber and Peres, 2013; Hayes and Ba, 2014) or

Sample Size	Data History (Years)		Cloning Strategy	Authors
1,610	1986 to 2005	(19)	Factor models using six factors to replicate individual fund exposures to produce linear clones	Hasanhodzic and Lo (2007)
1,662	1983 to 2018	(35)	Portfolio creation based on "best ideas" positions of hedge funds	Antón et al. (2010)
6	1990 to 2009	(19)	Estimate factor exposures of hedge fund indices to construct factor-based clones	Hayes (2012)
13	1994 to 2012	(18)	Clone returns using futures contract returns as explanatory variable	Bollen and Fisher (2013)
942	1999 to 2007	(8)	Return evaluation of positions disclosed after confidential treatment expires	Agarwal et al. (2013)
7,000	2006 to 2012	(6)	Clone returns using investible risk factors derived from futures contract prices	Weber and Peres (2013)
250	1999 to 2006	(7)	Individual position performance based on confidential treatment filings	Aragon et al. (2013)
21	Unknown		Returns captured using a beta regime-switching model	Hayes and Ba (2014)
1,331	2013 to 2020	(7)	Machine learning based to construct portfolios from a set of funds	DiPietro (2019)
1,331	2013 to 2020	(7)	Machine learning based factor models used to replicate fund returns	Fleiss et al. (2021)
1,500	2004 to 2017	(13)	Individual positions of funds selected based on crowding metrics	Brown et al. (2021)
643	2003 to 2017	(14)	Copycatted trades identified from digital footprints left on SEC EDGAR website	Cao et al. (2021b)
13,035	2013 to 2023	(11)	Time-lagged replication of entire hedge fund portfolios	Our paper (2024)

Table 4.1: Overview of Prior Work in Hedge Fund Performance Cloning.

This table summarizes various studies on hedge fund performance replication, detailing the sample size, data history, cloning strategy employed, and the authors. The studies range from using factor models and subset portfolio creation based on "best ideas" positions to machine learning-based constructions and beta regime-switching models. Our study extends this body of work by implementing time-lagged replication of entire hedge fund portfolios.

by copying only selected parts of portfolios (Antón et al., 2010; Bollen and Fisher, 2013; Agarwal et al., 2013; Aragon et al., 2013; DiPietro, 2019; Fleiss et al., 2021; Brown et al., 2021; Cao et al., 2021b). Our paper is the first to replicate entire hedge fund portfolios comprehensively. While this approach is novel for hedge funds, similar strategies have been applied to mutual funds with success (Frank et al., 2004; Verbeek and Wang, 2013), demonstrating that disclosed holdings by active mutual funds can be effectively used by copycats to achieve comparable or superior net returns. Our methodology is primarily inspired by Frank et al. (2004) and Verbeek and Wang (2013), who examined the implications of mandatory portfolio disclosure by mutual funds, particularly the feasibility of "copycat" funds replicating disclosed portfolios to attain similar returns. These studies created hypothetical "copycat" funds that adjusted their holdings with each semi-annual and quarterly disclosure, respectively, revealing efficiencies in mimicking disclosed strategies. It's crucial to distinguish between factor cloning and portfolio cloning. Prior research on hedge fund strategy cloning is primarily focused on factor-based cloning strategies. For example, Hasanhodzic and Lo (2007) developed "linear clones" to evaluate how well six distinct factors could predict the funds' expected returns. Other examples of factor-based approaches include Hayes (2012), Weber and Peres (2013), Hayes and Ba (2014), DiPietro (2019), and Fleiss et al. (2021). Portfolio cloning can be further categorized into entire portfolio cloning and selected position cloning. While Frank et al. (2004) and Verbeek and Wang (2013) are examples of cloning entire mutual fund portfolios, no such work has been undertaken for hedge funds. All prior research on hedge fund portfolio cloning focuses on selected position cloning (Agarwal et al., 2013; Antón et al., 2010; Aragon et al., 2013; Brown et al., 2021; Cao et al., 2021b). The landscape of prior research focusing on the evaluation of fund performances and comparing performances across different types of funds, rather than replicating such performances through copycat portfolios, is much larger than the replication studies. Examples include comparing hedge fund and mutual fund performance (Ackermann et al., 1999; Griffin and Xu, 2009; Agarwal et al., 2009; Lewellen, 2011) and analyzing specific performance characteristics (Malkiel, 1995; Fung and Hsieh, 2000, 2001; Agarwal and Naik, 2003; Martin and Puthenpurackal, 2008; Sandvik et al., 2011; Sias et al., 2016; Amir-Ghassemi et al., 2022; Cao et al., 2021a; Brown et al., 2021; Lazo-Paz et al., 2023).

Our paper ventures into an area not extensively explored previously due to several reasons. Anderson and Brockman (2018) shed light on inconsistencies within Form 13F institutional holdings, highlighting discrepancies between reported holdings and the SEC's official list, including errors in market valuations. Therefore, one significant barrier has been the inaccuracies within the 13F filings themselves, which necessitate outlier detection algorithms and considerable time investment to clean and standardize the data for analysis. Furthermore, accurately mapping CUSIPs to ticker symbols, especially in a manner that avoids survivorship bias, poses a problem given that CUSIPs and tickers change over time and may not always have a direct equivalent. Securing historical pricing data that is free from survivorship bias presents another hurdle, as does obtaining accurate pricing for companies that have experienced changes in their ticker symbols or CUSIPs. These factors contribute to the complexity of assembling a comprehensive dataset that accurately reflects historical market movements and institutional behaviors. Lastly, the computational power and memory required to simulate the performance of over 150,000 portfolios spanning more than a decade is substantial. This computational demand has likely deterred prior attempts to undertake such an analysis.

4.2 Data, Hypothesis & Methods

4.2.1 Data

Our data collection and consolidation process leverages multiple Application Programming Interfaces (APIs) for the acquisition of unbiased datasets. We use data that is free from survivorship bias, incorporating both historical hedge fund information and security pricing data that account for all entities throughout the time period studied, regardless of their current status.

We sourced structured fund holdings data from SEC-API.io by utilizing their 13F Institutional Ownership API[1]. This platform enabled us to systematically retrieve and organize holdings information from 13F-HR filings submitted to the SEC's EDGAR system. Additionally, to facilitate the tracking of securities, we employed SEC-API.io's Mapping API[2], which translates CUSIP identifiers into EDGAR Central Index Keys (CIK) and ticker symbols, including instances of ticker symbol changes. The extraction of gross market values reported in 13F filings was carried out through SEC-API.io's Filing Download API[3], which allows for the efficient bulk retrieval of 13F cover pages. Historical equity pricing data was sourced from AlgoSeek[4], specifically their equity prices database and Security Master File. These resources provided not only historical adjusted and unadjusted price data for all holdings of each fund, but also facilitated the mapping of CUSIP identifiers to AlgoSeek's internal security IDs over time, ensuring accurate price tracking even when a company has undergone ticker changes.

13F filings include two core files that are of interest to our analysis. The first file represents the cover page detailing the number of holdings and the aggregated value of reported assets, among others. The second file is the holdings information table attached to each 13F filing. It lists each invested company or instrument, specifying the security type (e.g., stock, note), quantity held, and position value. Figure 4.1 illustrates an example of a cover page of Ritholtz Wealth Management's (CIK 1633901) 13F filing, while Figure 4.2 shows an example of a 13F information table.

[1] https://sec-api.io/docs/query-api/13f-institutional-ownership-api
[2] https://sec-api.io/docs/mapping-api
[3] https://sec-api.io/docs/sec-filings-render-api
[4] https://www.algoseek.com/products.html\#us_equity_market_data

Report Summary:

Number of Other Included Managers:	1
Form 13F Information Table Entry Total:	753
Form 13F Information Table Value Total:	2,577,869,068
	(round to nearest dollar)

Figure 4.1: Ritholtz Wealth Management's 13F Cover Page Sample.

This image showcases a sample cover page from a 13F filing by Ritholtz Wealth Management, highlighting the disclosed number of holdings (753) and the aggregate portfolio value denominated in USD ($2.58 billion). The cover page of the filing can be accessed through the provided link in the footnote[5]*.*

NAME OF ISSUER	TITLE OF CLASS	CUSIP	FIGI	VALUE (to the nearest dollar)	SHRS OR PRN AMT	SH/ PRN	PUT/ CALL	INVESTMENT DISCRETION	OTHER MANAGER	VOTING AUTHORITY SOLE	SHARED	NONE
3M CO	COM	88579Y101	BBG001S5T7X2	2,144,228	19,614	SH		SOLE		97	0	19,518
ABBOTT LABS	COM	002824100	BBG001S5N9M6	4,077,602	37,046	SH		SOLE		76	0	36,969
ABBVIE INC	COM	00287Y109	BBG0025Y4RZ3	5,636,477	36,371	SH		SOLE		372	0	35,999
ACACIA RESH CORP	ACACIA TCH COM	003881307	BBG001SJLJT9	92,445	23,583	SH		SOLE		0	0	23,583
ACCENTURE PLC IRELAND	SHS CLASS A	G1151C101	BBG001SCXK90	1,693,320	4,826	SH		SOLE		138	0	4,687
ACME UTD CORP	COM	004816104	BBG001S5NCC0	223,066	5,205	SH		SOLE		0	0	5,205
ADOBE INC	COM	00724F101	BBG001S5NCQ5	11,026,958	18,483	SH		SOLE		86	0	18,397
ADTALEM GLOBAL ED INC	COM	00737L103	BBG001S8M5S3	972,086	16,490	SH		SOLE		175	0	16,315

Figure 4.2: Detail of Ritholtz Wealth Management's Holdings From a 13F Filing

This image is a detailed excerpt from Ritholtz Wealth Management's 13F information table file, showcasing the firm's holdings. It includes columns for the issuer name, class of securities, identifiers such as CUSIP and FIGI, the total value of each holding rounded to the nearest dollar, and the number of shares held. The type of security is also indicated. Access to the complete source filing can be obtained through the provided link[6]*.*

4.2.2 Hypothesis Formulation & Testing

Null Hypothesis (H0): Cloned portfolios, sourced from 13F filings of institutional investment managers with a track record of market outperformance on a risk-adjusted basis from 2013 to 2023 and rebalanced on the filing disclosure date rather than quarter end date, do not match or exceed the performance of their source funds or the market benchmark.

Alternative Hypothesis (H1): Cloned portfolios that are rebalanced on the filing disclosure date, reflecting the strategies of institutional investment managers with historical market outperformance on a risk-adjusted basis, duplicate or surpass the

[5] `https://www.sec.gov/Archives/edgar/data/1698218/000108514624000293/xslForm13F_X02/primary_doc.xml`

[6] `https://www.sec.gov/Archives/edgar/data/1698218/000108514624000293/xslForm13F_X02/infotable.xml`

market performance achieved by the source funds.

To test these hypotheses, we conduct an analysis of the descriptive statistics for the differences in performance metrics between the cloned and original funds. We compare the distribution of each factor's values for the cloned funds against those of the original funds. Using the Alpha factor as an illustrative case, we compute the difference (delta) between the cloned and original fund alphas to form a pairwise delta distribution. We then calculate and evaluate the descriptive statistics of this distribution, including the mean and the 25th, 50th, and 75th percentiles. For factors where higher values indicate superior performance, such as the Sharpe ratio and CAGR, we reject the null hypothesis for that factor if all calculated statistical metrics (mean, 25/50/75th percentiles) of the pairwise delta distribution are zero or positive. Conversely, for factors where lower values are preferable, such as maximum drawdown and annualized volatility, the null hypothesis for that factor is rejected if all metrics are zero or negative. The overall null hypothesis is rejected if the delta distributions for at least 8 out of the 12 factors show that the cloned portfolios perform at least as well as, or better than, the original portfolios.

To examine whether a delta distribution aligns with a normal distribution, we employ a Q-Q (quantile-quantile) plot. This graphical tool plots the actual quantiles from our dataset against the expected quantiles of a standard normal distribution. Ideally, if the data is normally distributed, the plot points should align closely with the diagonal 45-degree reference line. Any significant deviations from this line may indicate non-normality in the data distribution. For a more formal test of normality, we apply the Shapiro-Wilk test, which calculates a test statistic and an associated p-value. A p-value greater than 0.05 typically suggests that the dataset does not significantly diverge from normality, while a p-value below this threshold indicates non-normal distribution characteristics in the data. We conduct Levene's test and an F-test to verify the assumption of equal variances across the delta distributions, which is a critical prerequisite for selecting the appropriate statistical follow-on test. Both tests serve to determine if multiple groups have similar levels of variance, with a high p-value supporting the hypothesis of equal variances. The Wilcoxon signed-rank test, a non-parametric method, is utilized to ascertain whether the delta distributions of each factor significantly deviate from zero. Given the real-world nature of financial data, it is reasonable to presume that the distributions of factors for both original and cloned funds – and consequently their delta distributions – are not normally distributed. Additionally, the assumption of equal variances and the criteria for independence are not satisfied. The connection between the cloned and original portfolios complicates the assumption of independent observations, a key requirement for several statistical tests, including the Mann-Whitney U test. Cloned portfolios, by design, substantially duplicate the composition of their source portfolios. As a result, their performance is intrinsically tied to the performance of the shared holdings. This replication process generates a dependency, as changes in the value of shared positions simultaneously affect both the cloned and original portfolios.

Inspired by the methodologies of Brinson et al. (1986), Ackermann et al. (1999), Hayes (2012), and Agarwal et al. (2013), we have selected the S&P 500 index (ticker: SPY) as the benchmark for measuring the performance of the funds. Our analysis encompasses 12 factors, all annualized based on the approach of Paleologo (2021). In contrast to Verbeek and Wang (2013), our study addresses several key areas they missed: We consider volatility and other risk metrics, not just gross and net returns, our analysis goes beyond mean results to include the distribution of outcomes and we evaluate risk-adjusted alpha returns to provide a more comprehensive assessment of performance.

Factor and Description	**Sources**
Alpha: Fund's ability to outperform the SPY on a risk-adjusted basis. Higher = superior. $$\alpha = R_p - (R_f + \beta(R_b - R_f))$$ Where: • α is the annualized alpha of the portfolio based on quarterly returns. • R_p represents the annualized return of the portfolio, calculated as $R_p = \frac{1}{n}\sum_{i=1}^{n} r_{p_i} \times 4.$ • R_b represents the annualized return of the benchmark, calculated as $R_p = \frac{1}{n}\sum_{i=1}^{n} r_{b_i} \times 4.$ • r_{p_i} and r_{b_i} represent the quarterly returns of the portfolio and SPY, respectively, of the i-th quarter. • R_f is the risk-free rate of 2%. • β is the beta of the portfolio.	Jensen (1968), Ackermann et al. (1999), Fama and French (2004)
Beta: Fund's market volatility in relation to SPY. A value closer to 0 suggests lower market-related volatility and is considered more favorable. $$\beta = \frac{\text{Cov}(R_p, R_b)}{\text{Var}(R_b)}$$ Where: • β is the beta of the portfolio. • R_p represents the quarterly returns of the portfolio. • R_b represents the quarterly returns of SPY. • $\text{Cov}(R_p, R_b)$ is the covariance between the portfolio returns and the SPY returns. • $\text{Var}(R_b)$ is the variance of the SPY returns.	Jensen (1968), Blume (1975), Wallerstein et al. (2009), Sandvik et al. (2011)

<table>
<tr><th>Factor and Description</th><th>Sources</th></tr>
<tr><td>Arithmetic Mean Return (AMR): Fund's average annual return. Higher = superior.

$$\text{AMR} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) \times 4$$

Where:
x_i represents the quarterly returns of the portfolio.
n is the number of quarters the fund reported 13F filings that can be cloned.</td><td rowspan="2">Mitchell and Pulvino (2001), O'Shaughnessy (2005), Wallerstein et al. (2009)</td></tr>
<tr><td>Geometric Mean Return (GMR): Fund's compound average annual return. Higher = superior.

$$\text{GMR} = \left(1 + \frac{1}{n} \sum_{i=1}^{n} x_i \right)^4 - 1$$

Where x_i and n are the same as for AMR.</td></tr>
<tr><td>Cumulative Average Growth Rate (CAGR): Fund's mean annual growth rate. Higher = superior.

$$\text{CAGR} = \left(\prod_{i=1}^{n} (1 + r_i) \right)^{\frac{4}{n}} - 1$$

Where:
r_i represents the portfolio's quarterly returns.
n is the number of quarterly returns.
$\prod$ denotes the product over all quarterly returns.</td><td>Ackermann et al. (1999)</td></tr>
<tr><td>Excess Return: Difference between the fund's annualized return and that of the SPY benchmark. Higher = superior.

$$\text{Annualized Excess Returns} = R_p - R_b$$

Where R_p and R_b are the same as for alpha.</td><td rowspan="3">Antón et al. (2010), Hayes (2012), Bollen and Fisher (2013)</td></tr>
<tr><td>Tracking Error: Standard deviation of the fund's annualized excess returns relative to SPY. Lower tracking error reflects more consistent performance in relation to the benchmark.

$$\text{Annualized Tracking Error} = \sigma_{(R_p - R_b)} \times \sqrt{4}$$

Where:
R_p represents the individual quarterly returns of the portfolio.
R_b represents the individual quarterly returns of the benchmark.
$\sigma_{(R_p - R_b)}$ is the standard deviation of the difference between the portfolio returns and the benchmark returns.
The time factor for annualization is $\sqrt{4}$ to account for quarterly returns.</td></tr>
<tr><td>Information Ratio: Annualized excess return divided by the tracking error to evaluate the consistency of outperformance relative to the benchmark. A higher information ratio is indicative of a fund manager's efficient active management.

$$\text{Information Ratio} = \frac{\text{Excess Return}}{\text{Tracking Error}}$$</td></tr>
</table>

Factor and Description	Sources
Sharpe Ratio: A metric for risk-adjusted return, computed as the fund's annualized return over its annualized volatility with a risk-free rate set at 2%. Higher = superior. $$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p}$$ • R_p is the annualized return of the portfolio, calculated as $R_p = \frac{1}{n}\sum_{i=1}^{n} r_{p_i} \times 4$. • R_f is the risk-free rate of 2%. • σ_p is the annualized standard deviation of the portfolio returns, calculated as $\sigma_p = \sigma \times \sqrt{4}$. • r_{p_i} represents the individual portfolio returns. • σ represents the standard deviation of the portfolio returns.	Sharpe (1994), Ackermann et al. (1999), Lo (2002), Hasanhodzic and Lo (2007), Wallerstein et al. (2009)
Sortino Ratio: Focuses on downside volatility with a minimum acceptable return set at 2%. Higher Sortino Ratios indicate better downside risk management. $$\text{Sortino Ratio} = \frac{R_p - \text{MAR}}{\sigma_{\text{downside,annualized}}}$$ Where: • R_p is the annualized return of the portfolio, calculated as $R_p = \frac{1}{n}\sum_{i=1}^{n} r_{p_i} \times 4$. • r_{p_i} represents the quarterly portfolio returns. • MAR is the annualized minimum acceptable return of 2%. • $\sigma_{\text{downside,annualized}}$ is the annualized downside deviation, calculated as $\sigma_{\text{downside,annualized}} = \sigma_{\text{downside}} \times \sqrt{4}$. • σ_{downside} is the standard deviation of the negative excess returns.	Sortino and Price (1994), Pedersen and Satchell (2002), Chaudhry and Johnson (2008)
Volatility: Standard deviation of the fund's quarterly returns, providing an understanding of return variability. Lower annualized volatility represents more stable performance. $$\sigma_{\text{annualized}} = \sigma \times \sqrt{4}$$ Where: • σ is the standard deviation of the returns. • $T_{\text{return period}} = 4$ for quarterly returns.	Hasanhodzic and Lo (2007), Markowitz and Blay (2013), Brown et al. (2021)
Maximum Drawdown per Quarter (Worst Single Quarter): The largest peak-to-trough loss in the fund's value over each quarter. A lower maximum drawdown is indicative of lower risk and better performance during downturns. $$\text{Max Drawdown} = \min(R_{\text{quarterly}})$$ Where: • $R_{\text{quarterly}}$ represents the series of quarterly returns. • $\min(R_{\text{quarterly}})$ denotes the minimum value in the series of quarterly returns.	Ackermann et al. (1999), Markowitz and Blay (2013), Brown et al. (2021)

Methodology

Our analysis starts with the computation of performance metrics for each fund, utilizing the gross market value (GMV) of their holdings as indicated on the 13F cover pages.

To ensure the integrity of our dataset, we excluded funds with less than six consecutive 13F filings, thereby confirming that each included fund was operational and complied with SEC disclosure regulations for a minimum of one year. Additionally, we corrected identifiable inaccuracies within the 13F disclosures. Such corrections were necessary for instances where funds reported GMvs. in incorrect magnitudes – e.g., denominating dollar values in millions instead of thousands – or inaccurately swapped the figures for the number of holdings and GMVs. Funds that omitted a quarterly report or failed to disclose their GMV in any 13F filing were also excluded from our analysis.

To calculate Alpha and Beta, we utilized the percentage change in reported GMvs. across each reporting period to represent a fund's quarterly returns. Specifically, quarterly returns were computed for the periods of January 1st to March 31st (Q1), April 1st to June 30th (Q2), July 1st to September 30th (Q3), and October 1st to December 31st (Q4) annually. It is crucial to note that fund returns were calculated based on these fixed quarterly periods, rather than the intervals between 13F disclosures, to prevent inaccuracies in performance measurement that could skew Alpha and Beta calculations. The S&P 500 index, represented by the ticker SPY, served as the benchmark for these calculations, with SPY's quarterly returns computed over the identical periods. To accommodate quarters beginning or ending on weekends or public holidays, we adjusted the starting or ending price of SPY by carrying forward the last available closing price. For instance, if Q2 begins on a Saturday, we use the previous Friday's adjusted closing price as the starting point for Q2's return calculation:

$$\text{Q2 Return (SPY)} = \left(\frac{\text{Adj Close (June 30th)} - \text{Adj Close (April 1st)}}{\text{Adj Close (April 1st)}} \right)$$

Here, the adjusted closing price for April 1st equals that of March 31st if the markets were closed on April 1st.

To replicate a fund's portfolio, we gathered all relevant holdings data (CIKs, CUSIPs, reported value, and number of shares for each position) from the 13F information tables, referencing the public disclosure date of the filing rather than the period-end date. Holdings deemed unreplicable – such as options, warrants, or notes, which either lack necessary disclosure for replication (e.g., strike prices and expiration for options) or are insufficiently liquid for trading simulation – were excluded from the cloning process. The replication of a portfolio was then straightforwardly achieved by duplicating all its clonable holdings.

We conducted backtesting simulations that involved executing buy and sell orders for each holding based on their reported shares at the disclosure date, using unadjusted closing prices. For instance, if Fund A reported holding 100 shares of Stock S in Q1 and then reported an increase to 150 shares in Q2 on the disclosure date T, indicating a purchase of 50 shares, we simulated buying an additional 50 shares of S at its unadjusted closing price on date T. Conversely, if Fund A's holdings of Stock S decreased from 100 shares in Q1 to 50 shares in Q2, we simulated selling 50 shares at Stock S's unadjusted closing price on date T. Following Verbeek and Wang (2013) approach, we compute the quarterly return of a cloned portfolio as the return of a simulated buy-and-hold portfolio that invests in the most recently disclosed equity positions and is rebalanced at the next 13F disclosure date. The quarterly return QR for a cloned portfolio targeting fund j is defined as:

$$\mathrm{QR}_t^j = \sum_{i=1}^{N} \tilde{w}_{i,t-1}^j R_{i,t}$$

Where $R_{i,t}$ denotes the return on the equity position i, and the value weights are given by

$$\tilde{w}_{i,t-1}^j = \frac{N_{i,t-1}^j P_{i,t-1}}{\sum_{i=1}^{N} N_{i,t-\tau}^j P_{i,t-1}}$$

Where $N_{i,t-\tau}^j$ denotes the number of shares of stock i held by fund j disclosed in the fund's most recent 13F filing at time $t-\tau$, and $P_{i,t-1}$ is the stock price at the date when the portfolio was previously rebalanced, i.e. approximately three months prior to t. Unlike the original portfolios, which calculate quarterly returns based on the calendar quarter's first and last day, cloned portfolios' returns were calculated based on the disclosure dates to avoid data leakage or foresight bias, as cloned positions could not have been known at the quarter's end.

To evaluate our hypothesis, we calculated the distributions of all factors for both original and cloned portfolios, categorizing the number of funds within specific ranges (e.g., 10 funds with an alpha of 0.2%, 18 funds with an alpha of −0.03%). Given that the distributions did not adhere to normality and equal variance as well as independence was not observed, we opted for the Wilcoxon Signed-Rank Test to assess the statistical significance of differences in all factor distributions between the original and cloned portfolios. This approach ensures a rigorous examination of the performance differential without the constraints of parametric assumptions.

Limitations of Our Methodology

One significant constraint of using 13F data in our analysis is its exclusion of cash flow activities, such as cash inflows from additional capital raised by funds or cash redemptions. In scenarios where a fund acquires extra securities through new capital, our method replicates these purchases without accounting for the cloned portfolio's cash levels. Adapting our cloning strategy to scenarios with less or limited capital would necessitate a recalibration of portfolio weights, possibly incorporating fractional shares for smaller positions to maintain proportional accuracy. Contrary to Kacperczyk et al. (2006) and Verbeek and Wang (2013), our analysis does not incorporate trading costs, such as execution costs for buying or selling shares on exchanges like NASDAQ or NYSE. This exclusion is based on the significant decline in execution costs over the past decades and the availability of market-making executions, where entities can earn fees rather than incur costs by buying at the bid and selling at the ask. Additionally, we do not account for liquidity constraints. Our methodology involves buying or selling the delta of positions in a single day at the closing price, without considering the spread or liquidity. This approach may not accurately reflect the true nature of executing large-scale positions, as funds must adhere to internal risk parameters that limit the proportion of average daily dollar volume they can trade. Consequently, if a position size exceeds these thresholds, a fund would typically scale in or out of the position over several days to weeks, depending on its size. 13F filings are disclosed after the quarter ends, typically a few days to several weeks later. Consequently, our analysis compares the quarterly returns of source portfolios with the lagged quarterly returns of cloned portfolios, leading to a slight mismatch in the return calculation periods. This discrepancy introduces a potential distortion, as the return periods for the original and cloned portfolios do not align perfectly. However, we anticipate that any resulting error will remain relatively stable across the board due to the regulatory consistency in the timing of 13F filings. Given our extensive sample size of over 150,000 portfolios, we expect that such discrepancies will not significantly affect the overall results of our study.

We observed that some funds invest in highly correlated or identical securities, such as allocating 3% of the fund's GMV to the SP500 ETF (SPY) and another 2% to a similar index under a different ticker (IVV). Similarly, investments in NASDAQ 100 (QQQ) alongside its largest constituents like APPL, AMZN, and META result in overlapping exposures. While such strategies might be rational for large institutional managers adhering to specific risk parameters, they are less efficient for cloning with significantly less capital. In these instances, our cloning algorithms could be refined to filter out redundancies, reallocating capital from omitted positions to enhance the portfolio's diversification and efficiency.

Additionally, the selective inclusion of source positions in the cloned portfolio introduces a potential selection bias. This bias may reflect the underlying criterion (tradeability at the time of 13F disclosure) that could inadvertently misalign the cloned portfolio's performance with the strategic choices made for the source portfolio. Our criteria for excluding positions were solely based on the tradability of a security, assessed through its linkage to a ticker on a U.S. exchange, which may influence the comparability and performance outcomes of our cloned portfolios.

In practice, we successfully replicated a high-performing portfolio with over 700 holdings, scaling down its GMV from \$7 billion to \$700K for simulation in a paper trading account through Interactive Brokers' APIs. However, manually replicating and rebalancing portfolios, especially those with over 100 holdings – which accounts for 50% of all funds – every quarter is not feasible without automated software assistance, as we've developed for this research.

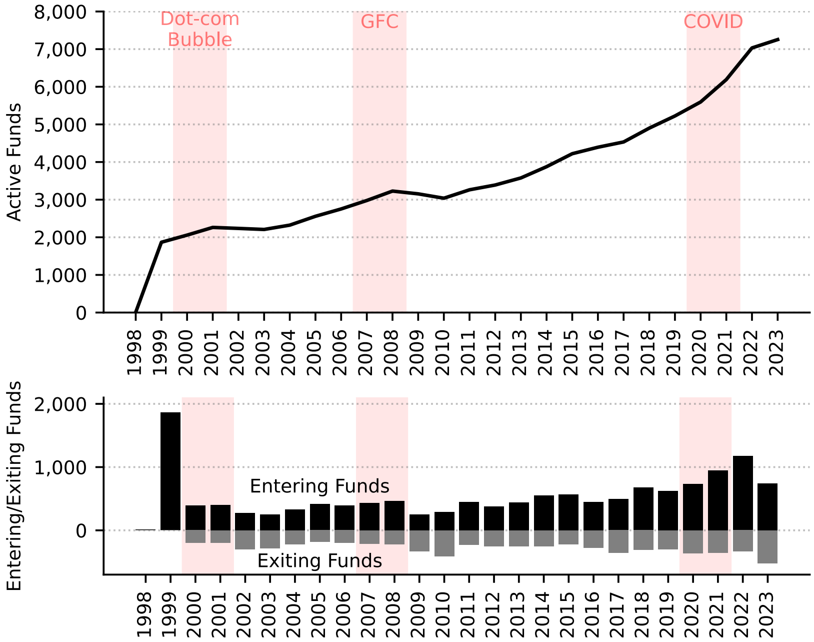
4.3 Results & Discussion

4.3.1 Fund Lifespans and Survival Rates

We present the results of our analysis on the total number of active funds from 1998 to 2023 in Figure 4.3. This analysis allows for discerning trends in fund creation, offering insights into whether the asset management industry is experiencing growth or decline. Over the past 26 years, from 1998 to 2023, our analysis reveals a dynamic landscape in the fund management industry, involving 13,035 hedge funds. Notably, 2023 emerged as a landmark year, showcasing the highest number of active funds, with a remarkable surge in new entrants from under 500 in 2016 to over 1,000 by 2022. This period witnessed the greatest influx of newly registered funds, highlighting a significant shift in the industry's composition. The year 2022 stood out not only for its record-setting new fund registrations but also for signaling a pronounced market adjustment. The repercussions of this strong market correction became particularly evident in 2023, a year that saw over 500 funds withdrawing from the investment arena, surpassing the exit rates observed in the aftermath of the Global Financial Crisis (GFC) in 2009 and 2010. This trend underscores a notable correlation between market downturns or recessions and an increase in the number of funds ceasing operations, with the peak of exits often coinciding with the transition from a bear to a bull market. This suggests that many funds tend to exit the market just as conditions begin to improve. For clarity, "entering funds" are identified as those reporting a Form 13F-HR for the first time, whereas "exiting funds" are recognized as those with previous 13F-HR filings but none in the given year. Given the SEC's mandate requiring quarterly disclosures (four times per year), the absence of any filings from a fund within a year is a strong indicator of its operational cessation.

Figure 4.5 illustrates the survival rates and operational longevity of funds from 1998 to 2023. We quantify the duration of activity for each fund by tracking the number of years it reported Form 13F filings to the SEC via EDGAR. This step involves plotting the data on a histogram where the x-axis represents the number of active years, and the y-axis shows the count of funds for each duration of activity. Utilizing the entire period from 1998 to 2023, we create 26 distinct buckets to categorize the funds based on the number of years they were operational. This approach allows us to visualize the distribution of fund lifespans across the observed timeframe. Descriptive statistics capture the central tendency and dispersion of fund lifespans. These statistics include the mean (average number of active years), standard deviation (variation in lifespan), and the first, second (median), and third quantiles.

Our survival rate analysis, encompassing 13,035 funds operating between 1998 and 2023, reveals an average fund lifespan of 7 years and a median lifespan of 5 years. This contrasts with Ackermann et al. (1999), who analyzed 547 hedge funds from 1987 to 1995 and found a mean fund age of 5.22 years and a median age of 4.01 years. Comparing their results with our more recent data, we conclude that hedge funds have successfully extended their lifespan by an average of 2 years, with a median increase of 1 year. In contrast to the hedge fund industry, the mutual fund industry demonstrates longer lifespans. Kacperczyk et al. (2006) found that mutual funds had a mean age of 13 years and a median age of 8 years during the period from 1984 to 2002. Similarly, Verbeek and Wang (2013) analyzed 3,046 mutual funds from 1985 to 2008 and reported a mean age of 12.63 years and a median age of 7.75 years. These comparisons highlight that mutual funds typically enjoy longer operational periods compared to hedge funds. The standard deviation of 6 years highlights a significant variance in fund lifespans, indicating that while some funds operate for a brief period, others demonstrate much longer endurance. A closer look at the distribution shows that the minimum active duration is just 1 year, reflecting the challenges some funds face in sustaining operations. The 25th percentile, at 2 years, implies that a quarter of the funds ceased filings within two years of operation.

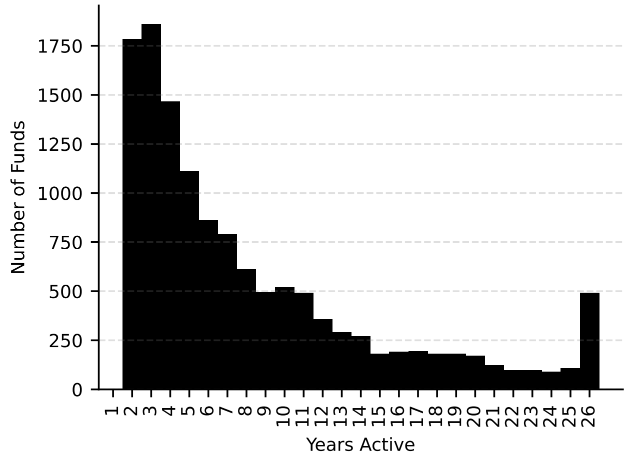


Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Active Funds	8	1,868	2,058	2,262	2,236	2,208	2,323	2,558	2,753	2,978	3,228	3,154	3,038
New Funds	8	1,860	388	398	271	252	331	418	395	433	466	253	291
Exiting Funds	0	0	198	194	297	280	216	183	200	208	216	327	407

Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Active Funds	3,262	3,389	3,577	3,876	4,221	4,393	4,533	4,901	5,224	5,596	6,191	7,032	7,254
New Funds	448	379	436	552	564	445	497	677	618	734	949	1,175	739
Exiting Funds	224	252	248	253	219	273	357	309	295	362	354	334	517

Figure 4.3: Trends in Active, Entering, and Exiting Investment Managers (1998–2023).

The top plot tracks the annual count of active investment managers filing at least one Form 13F with the SEC, identified by unique CIK numbers, from 1998 to 2023. Highlighted periods indicate significant market events: the Dot-com Bubble, the Global Financial Crisis (GFC), and the COVID-19 pandemic. The corresponding bar chart below shows the dynamics of market entry and exit among funds. "Entering Funds" denotes new managers filing a 13F for the first time, while "Exiting Funds" represents those that did not file a 13F in a given year after having filed in the preceding year. The two accompanying tables quantify the visual data, summarizing the number of active, new, and exiting funds per year, providing a detailed numerical basis for the graphical trends observed above.



Metric	Value
Number of Funds	13,035
Mean (Active Years)	7
Std. Dev. (Active Years)	6
Min. Years Active	1
25th Percentile	2
Median	5
75th Percentile	10
Max. Years Active	26

Figure 4.4: Fund Life Span Distribution and Descriptive Statistics (1998–2023).

The histogram displays the distribution of fund longevity, detailing the number of years funds remained active out of the 13,035 *that have filed Form 13F-HR with the SEC from 1998 to 2023. Each fund's activity duration is identified by its unique CIK, with a fund counted as active in any year it filed a 13F-HR form. The accompanying table provides descriptive statistics on fund survival, revealing that the median operational lifespan of a fund is only 5 years. Notably, about 3.8% of the funds exhibit exceptional longevity, operating for over 26 years, demonstrating significant endurance in the investment landscape.*

At the higher end, the 75th percentile reveals that three-quarters of the funds were active for 10 years or less, and 25% of all funds were active for 11 years and more, underscoring that a significant portion of the sector does not surpass a decade of activity. Remarkably, the maximum lifespan extends to 26 years, showcasing that just under 500 funds have managed to navigate the asset management industry successfully over the entire period analyzed.

4.3.2 Evolution of Asset Values

In this part, we delve into the progression of asset values from 2013 to 2023 by analyzing the holdings values reported in 13F filings and report our findings in Figure 4.4. This analysis aims to shed light on the performance of institutional portfolios over time, revealing how they've adapted to shifting market conditions.

Figure 4.4 presents the time series of the aggregated gross market values of all portfolios in USD reported quarterly by investment managers, delineating the standard reporting periods that conclude at the end of March, June, September, and December. The trend from the beginning of 2013 to the onset of 2020 is characterized by a relative stagnation in holding values, suggesting a period of flat growth across the combined portfolios of the investment industry. During this timeframe, the growth in total holdings value does not exhibit significant upward momentum. A pivotal shift in the trajectory of holding values is marked by the COVID-19 pandemic. From 2020 to the end of 2021, there is an escalation in holding values, reflecting perhaps a reactive surge in investment during the pandemic period. In the fourth quarter of 2021, the aggregated assets reached an unprecedented peak, surpassing \$60 trillion. This surge is subsequently countered by a downturn that coincides with the market correction in 2022, indicating a retraction of holding values in response to changing economic conditions.

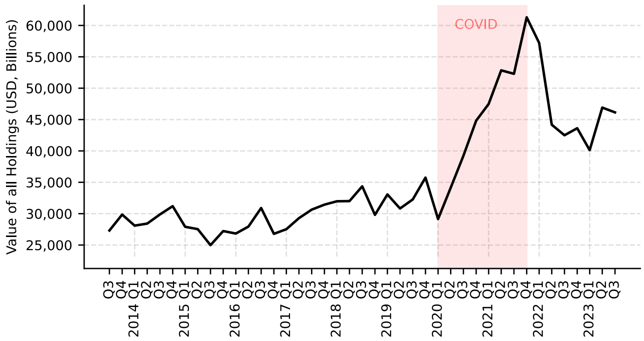


Figure 4.5: Aggregate Gross Market Value of 13F Portfolios (Q3 2013 – Q3 2023).

This time series graph traces the aggregate gross market value (GMV) of all portfolios reported in 13F filings from the third quarter of 2013 to the third quarter of 2023, measured in USD billions. Quarterly GMvs. remained relatively stable from Q3 2013 until the end of Q1 2020, fluctuating between \$25 trillion and \$35 trillion. However, during the COVID-19 period, marked by unprecedented quantitative easing measures from the Federal Reserve, the GMvs. soared, reaching a peak of over \$60 trillion. This spike reflects the infusion of liquidity into the markets, leading to an all-time high in the aggregated market value of the portfolios. The timeline demonstrates a retraction in 2022, indicative of the market correction and reduction in combined holdings across reported funds.

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
N	3,243	3,600	3,879	3,971	4,195	4,563	4,800	5,179	5,798	6,424	6,661
Mean	9,206	8,666	6,857	6,745	7,493	6,533	7,448	8,654	10,573	6,790	6,928
Median	497	466	413	413	461	361	414	459	475	323	326
Std. dev.	84,340	80,370	46,367	48,468	64,454	58,665	70,888	84,746	108,680	77,890	81,822
25th	205	195	170	177	193	157	186	202	219	149	152
75th	1,883	1,709	1,463	1,505	1,510	1,205	1,302	1,417	1,538	1,004	1,017
Min	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1
Max	3,005,627	3,393,645	1,487,144	1,792,633	2,286,847	2,225,644	2,932,328	3,447,554	4,404,582	3,672,651	4,070,961

Table 4.3: Descriptive Statistics of the Gross Market Values (GMV) in Millions of Dollars per Fund from 2013 to 2023.

The table shows the number of funds ("N") each year, along with the mean, median, and selected percentiles of GMV. The statistics indicate a heavily right-tailed distribution with significant disparities between the mean and median GMVs, influenced by a small number of funds managing exceedingly large assets.

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
N	3,243	3,600	3,879	3,971	4,195	4,563	4,800	5,179	5,798	6,424	6,661
Mean	1,090	741	697	692	649	592	708	782	709	795	417
Median	103	101	94	98	101	93	99	105	107	94	92
Std. dev.	15,751	9,226	10,189	11,687	8,486	6,969	11,590	11,190	11,210	17,594	3,110
25th	41	39	34	36	36	33	37	41	41	35	36
75th	259	250	238	244	250	240	245	246	264	224	220
Min	1	1	1	1	1	1	1	1	1	1	1
Max	537,814	323,481	522,844	665,096	442,915	288,190	555,527	430,862	696,810	952,754	157,613

Table 4.4: Descriptive Statistics for the Number of Holdings per Fund from 2013 to 2023.

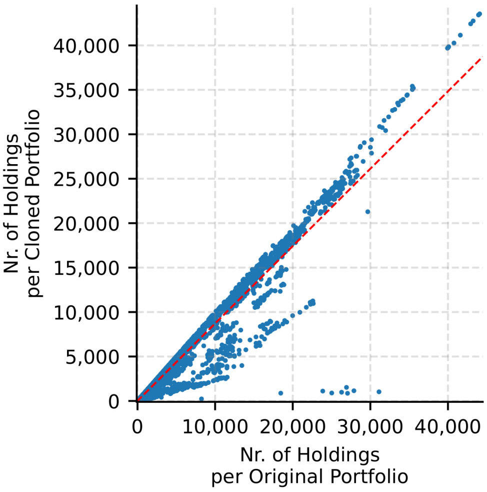
The table shows the number of funds ("N") each year, alongside the mean, median, and selected percentiles for the number of holdings. The statistics reveal variations in portfolio sizes, with some funds holding a vast number of securities while others maintain a more concentrated portfolio, also indicating a heavily right-tailed distribution with significant disparities between the mean and median number of holdings per fund.

Table 4.3 and Table 4.4 reveal that the mean value is insufficient to accurately describe the GMV's and number of holdings per fund. For instance, in 2023, the mean fund GMV was $6.93 billion, whereas the median GMV was significantly lower at $0.33 billion. The same heavy right-tail distribution is observed in the number of holdings, with the mean number of holdings per fund at 417 positions in 2023, while the median number of positions is just 92. This distortion arises because a small fraction of funds manage disproportionately large assets, with some exceeding $40 billion. Between 2013 and 2023, 75% of all funds managed GMvs. between $1 billion and $1.8 billion, while only 2.5% exceeded the $32 to $48 billion threshold. We also observed a downward trend in the median GMV, which declined from $496 million in 2013 to $326 million in 2023. This trend is mirrored in the top 2.5% of funds, where the threshold for inclusion in this elite group dropped from $45 billion in 2013 to $34 billion in 2023. These trends can be attributed to the doubling of the number of active funds from 2013 to 2023. Newer, younger funds typically start with less capital, and the number of new funds reached an all-time high in 2022 with 1,175 entries. Consequently, while the total GMV remained relatively stable over the decade, the increased number of funds meant that more entities shared the overall asset pool, except for the period affected by COVID-19 from 2020 to 2022.

4.3.3 Performance of Cloned vs. Original Portfolios

To confirm the effectiveness of our replication process for portfolios from 2013 to 2023 and ensure a sufficiently large sample size for statistically significant conclusions, we analyzed the correlation between the number of holdings disclosed in 13F filings and those in the replicated portfolios. This comparison was visualized in a scatterplot in Figure 4.6, with the original portfolio holdings on the x-axis and the replicated portfolio holdings on the

y-axis. A linear regression analysis yielded an R^2 value of 0.9535, indicating that 95.35% of the variance in the number of holdings between original and replicated portfolios can be explained by the linear relationship, with a statistically significant p-value demonstrating the regression's reliability.



	Original	Cloned
Number of Portfolios	161,408	161,408
Mean (Holdings per Port.)	398	358
Std. Dev. (Holdings per Port.)	1,508	1,344
Min	3	3
25%	42	38
50%	102	95
75%	246	230
Max	44,083	43,546

Figure 4.6: Replication Accuracy of Portfolio Holdings from 13F Filings (2013–2023).

The left side of the figure presents a scatterplot comparing the number of holdings for each original portfolio, as reported in 13F filings by the source fund, with the number of holdings successfully replicated, across 161,408 *portfolios between 2013 and 2023. The tight clustering around the linear regression line (in red), with an R^2 value of 0.9535, illustrates a high level of replication accuracy—indicating that 95.35% of the variance in the number of holdings is accounted for by the replication process. The slope of the regression line is 0.87, which is close to 1, suggesting that the cloned portfolios closely match their original counterparts in terms of the number of holdings. Outliers in the plot are clearly discernible and are due to holdings such as options and warrants, which were not replicated as part of our methodology. The right-hand table provides supporting descriptive statistics, highlighting that the median number of holdings in original portfolios is 102 compared to 95 in the cloned portfolios, further confirming the efficacy of the replication approach.*

The regression's standard error was notably low at 0.0005, and the slope of the regression line was 0.8705, closely approaching 1, alongside an intercept of 11.92. These figures collectively suggest a high degree of accuracy in the replication process across 161,408 portfolios examined.

Descriptive statistics provide further evidence of the accuracy of our replication process. Specifically, at the 25th percentile, the original portfolio contained 42 holdings, while their replicated counterpart had 38, indicating a slight deviation. Moreover, at the 75th percentile, there was a noticeable, yet modest, difference with the original portfolio having 246 holdings compared to 230 in the replicated portfolio. This disparity, as anticipated, is due to the exclusion of certain assets such as options, notes, and warrants from the replication process.

The Spearman's rank correlation coefficient calculation resulted in a value of 0.9926, accompanied by a statistically significant p-value of less than 0.01. This outcome underscores the efficacy of our cloning approach, revealing a near-perfect rank-order correlation between the original and replicated portfolios, which indicates a successful replication strategy.

4.3.4 Example of a Single Replication Process & Analysis

To illustrate and explain our methodology, we examine the case of Ritholtz Wealth Management (Ritholtz) with CIK 1698218, identified as one of the top-performing funds in our analysis. Ritholtz initiated its reporting with a 13F filing for Q2 2017 (April 1st to June 30th, 2017), submitted on August 4th, 2017. This date also marks our first replication of the fund's holdings. Over 25 quarters, spanning from 2017 to 2023, the fund submitted 25 Form 13F filings, each revealing an updated portfolio.

For each new 13F disclosure, we analyzed the fund's securities holdings (e.g., 200 shares of Apple) and calculated adjustments needed between the current and preceding filings to rebalance the replicated portfolio accordingly. Throughout this period, our replication focused solely on reproducible positions, excluding instruments like options.

The GMV of both the source and the replicated portfolios was determined for each reporting period. The GMV of the source portfolio was directly extracted from its 13F filing, while the GMV of the cloned portfolio was calculated by multiplying the number of shares by the unadjusted closing prices of the respective securities.

The bottom right time series plot in Figure 4.7 visually compares the normalized GMV of the original fund's portfolios, the cloned portfolios, and the SPY benchmark, each standardized to start at $100,000 in Q2 2017. This comparison effectively showcases the source fund's consistent outperformance of the SPY benchmark throughout the analysis period and the cloned portfolio's accurate reflection of the source fund's performance trajectory.

The cloneability of the holdings is depicted in the bottom left dual bar chart, comparing the number of holdings in the original and cloned portfolios for each 13F filing. Apart from the COVID-19 period, the cloned portfolios matched the original in terms of the number of holdings, demonstrating our strategy's efficacy.

Additionally, the asset allocation breakdown by security type is presented in the top right plot, highlighting the portfolio's investment distribution across ETFs, US and Canadian stocks, CEFs, and ADRs. Notably, in the initial two years, the fund achieved benchmark outperformance primarily through a 95% allocation in ETFs, with the remainder in US stocks. This figure demonstrates our cloning strategy's ability to accurately replicate diverse asset classes, thereby closely mirroring the performance of the source fund.

	Original	Cloned
General		
CIK	1698218	1698218
Asset Value (Most Recent)	$2,227,118,451	$2,215,443,852
Asset Value (Start)	$238,244,000	$211,002,396
# Most recent Holdings	702	700
Periods	25	25
First PeriodOfReport	2017-06-30	2017-08-04
% Holding Value invested in ETFs	55.31%	55.31%
Factors		
Beta	0.5805	0.8845
Alpha	30.71%	30.69%
Annualized Arithmetic Mean Return	38.92%	41.37%
Annualized Geometric Mean Return	44.98%	48.25%
Annualized Median Return	53.53%	49.58%
Annualized CAGR	42.99%	45.68%
Annualized Sharpe Ratio (RFR=2%)	2.0484	1.9070
Annualized Sortino Ratio (MAR=2%)	5.3719	6.0044
Annualized Volatility	18.02%	20.65%
Max DD (Qtr)	−8.98%	−7.76%
Annualized Excess Return	26.22%	29.55%
Tracking Error	8.46%	8.33%
Information Ratio	3.0993	3.5474
1 Yr Return	18.26%	19.21%
3 Yr Return	150.13%	155.17%
5 Yr Return	463.99%	584.55%
Total Return	834.81%	949.96%
Skewness (Qtr Returns)	−0.2730	0.2110
Kurtosis (Qtr Returns)	−0.0420	−0.2410

Table 4.5: Comparative Performance Metrics of Original and Cloned Portfolio.

The table details metrics of the original fund, Ritholtz Wealth Management, and its cloned counterpart, illustrating the effectiveness of the cloning strategy. General information, such as recent and initial asset values, number of holdings, reporting periods, and percentage of holdings in ETFs, provides a foundation for comparison. Performance and risk factors, such as Sharpe (Risk Free Rate: 2%) and Sortino (Minimum Acceptable Return: 2%) ratios, are detailed. Notably, the cloned fund exhibits a marginally higher Alpha, annualized geometric mean return, and total return, demonstrating its capability to closely emulate and even slightly enhance the performance of the original portfolio. Risk metrics like annualized volatility and maximum quarterly drawdown (Max DD) are also compared, showing a modest increase in risk for the cloned fund. Additionally, the 1, 3, and 5-year returns highlight the cloned portfolio's outperformance over time.

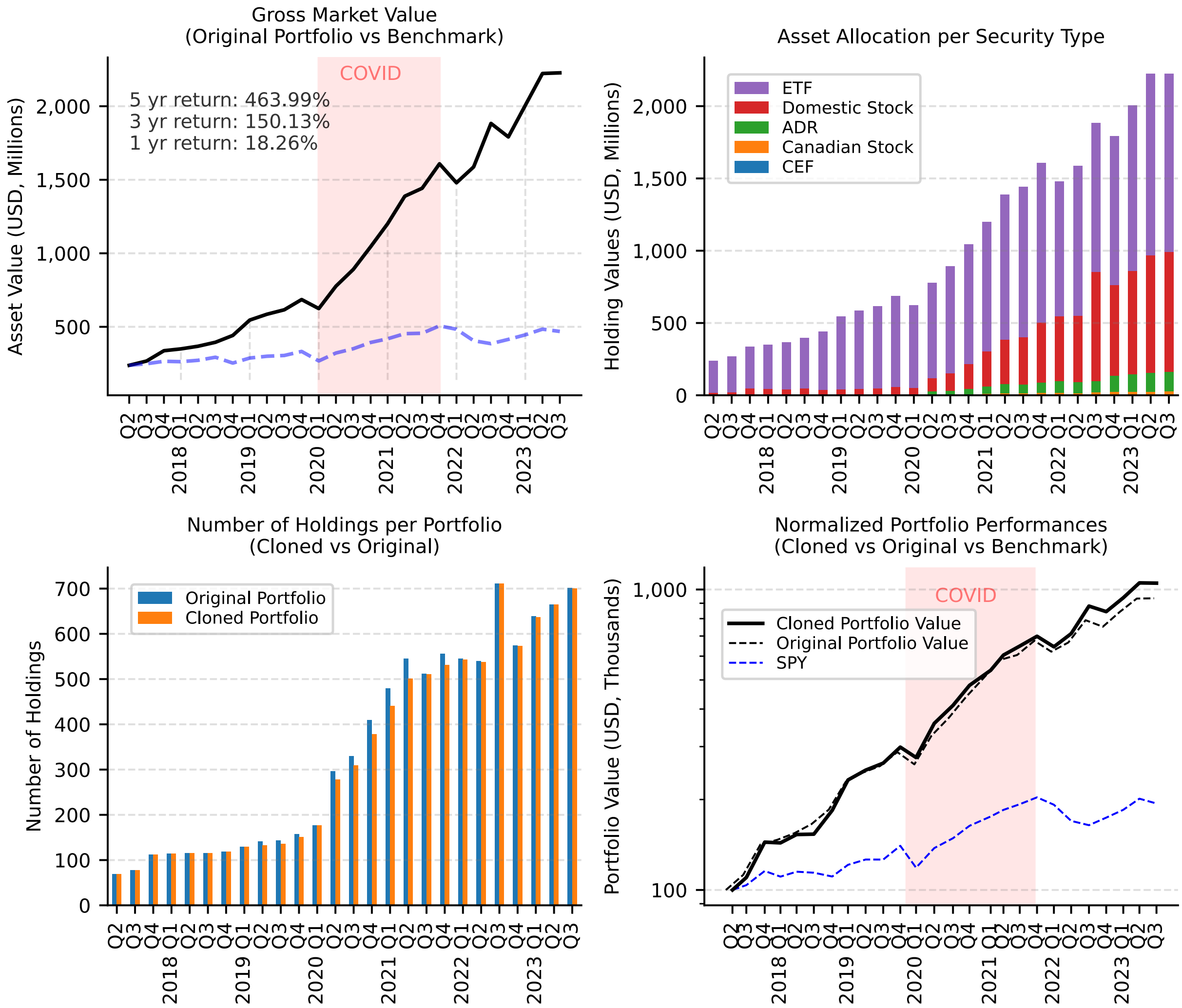


Figure 4.7: Representations of Original and Cloned Portfolio Metrics and Asset Allocation across Time.

The figure features four visualizations that compare the performance and asset composition of source portfolios of Ritholtz Wealth Management (Ritholtz) and cloned counterparts, track their holdings over time, and benchmark them against the SPY index. Top Left: Showcases the time series of gross market value (GMV) of the original portfolios disclosed in 13F filings by Ritholtz versus the SP500 index (ticker: SPY), based on equal initial investments, that being the starting value of the source fund. This subplot, covering Q2 2017 to Q3 2023, reveals returns for 1-year (18.26%), 3-year (150.13%), and 5-year periods (463.99%), with values expressed in millions of USD. Top Right: Displays a stacked bar chart detailing quarterly asset allocation by security type within each source portfolio. Categories include ETFs, US and Canadian stocks, CEFs, and ADRs, with the total asset value per quarter matching the GMV from the top left plot. Bottom Left: A comparative bar chart illustrates the number of instruments (holdings) per portfolio as reported in each 13F filing, for both original and cloned portfolios. The chart traces the evolution from an identical count in Q2 2017 through a divergence starting in early 2019 due to unreplicable options, a widening gap during COVID, and a recent convergence in Q4 2022, where the cloned portfolio achieves 99.7% replication of the source's holdings. Bottom Right: This subplot presents a time series of normalized GMVs, each starting at $100,000, for cloned and original portfolios alongside SPY, serving as a benchmark. It visually underscores the effectiveness of the cloning strategy in mirroring the original portfolios' GMV and, by extension, their performance, with both consistently outperforming SPY.

4.3.5 Factor Comparison for all Source and Cloned Portfolios

Upon constructing both the source and cloned portfolios, we calculated their respective performance factors as delineated in the Data & Methods section. Figure 4.5 serves as an exemplar for Ritholtz Wealth Management, exhibiting the performance and risk metrics of the original fund with those of its cloned counterpart. To derive these metrics for Ritholtz and its replicate, we conducted a backtest that involved simulating the trading activities of the cloned portfolios. This entailed executing buy and sell orders to rebalance the clone in tandem with the disclosure dates of Ritholtz's subsequent 13F filings.

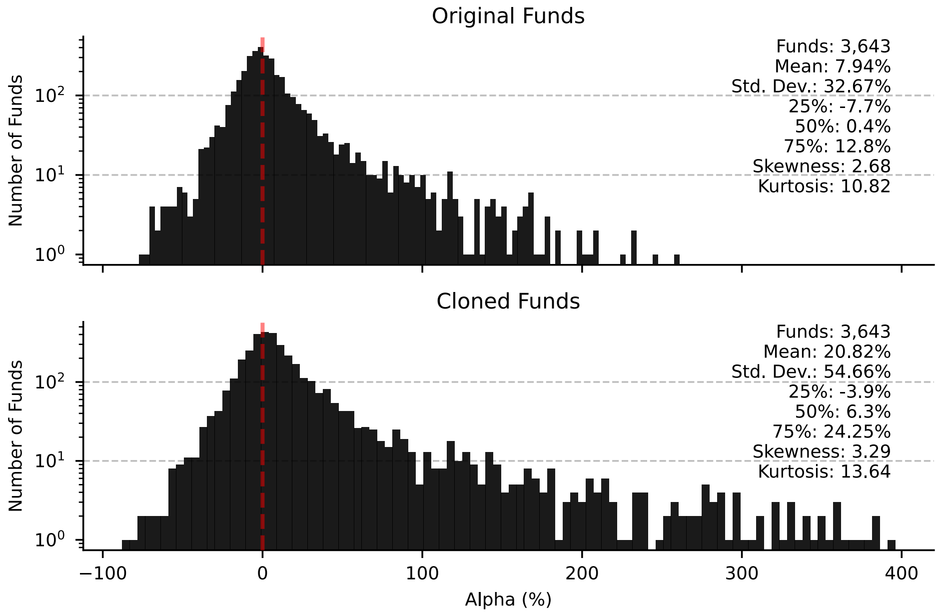


Figure 4.8: Alpha Distribution Comparison between Original and Cloned Funds.

The top graph illustrates the alpha distribution for 3,643 original funds from 2013 to 2023, with a mean of 7.94% and notable right skewness, as reflected by a skewness value of 2.68. The distribution exhibits heavy tails with a kurtosis of 10.82, indicating a propensity for extreme outliers. Quartile values highlight a wide dispersion of alpha, with the median fund slightly outperforming the SP500 benchmark by 0.4%. The bottom graph depicts the alpha distribution for the cloned funds, mirroring the same number of funds. The mean alpha significantly increases to 20.82%, and the distribution becomes more positively skewed (skewness of 3.29) and exhibits even heavier tails (kurtosis of 13.64) than the original funds. The cloned funds demonstrate a stronger median performance with an alpha of 6.3%, and a more substantial upper quartile performance, with 75% of funds attaining an alpha above −3.9%. Both distributions show that the cloned funds not only replicate the performance of the original funds but also amplify their positive traits, achieving higher mean returns and exhibiting a tendency towards higher outperformance, as indicated by their increased alpha values.

This methodology was consistently applied across the dataset to generate the suite of the 12 performance and risk factors for all original and cloned entities. Focusing on the Alpha factor as an illustrative case, Figure 4.8 graphically contrasts the alpha distribution of all original funds with that of the cloned funds.

To address the null hypothesis—that portfolios replicated from 13F filings between 2013 and 2023, and rebalanced according to the disclosure dates, do not achieve the performance of the source funds or outperform the market—we employed the Wilcoxon Signed-Rank Test on the delta distributions. This test evaluates each factor's delta distribution to ascertain statistically significant discrepancies from 0. The results, including the descriptive statistics detailing the differences for each factor across all funds, are systematically presented in Figure 4.7, providing a robust statistical foundation to reject the null hypothesis. Figure 4.6 allows for a direct comparison between the factors of source and cloned funds with Figure 4.9 visualizing the correlation between each factor in a scatterplot.

		Mean		Std. Dev.		Min		25%		50%		75%		Max	
Factor	Samples	Original	Cloned	Original	Cloned	Original	Cloned	Original	Cloned	Original	Cloned	Original	Cloned	Original	Cloned
Reporting Periods	3,643	29.0	29.0	10.21	10.21	6.0	6.0	19.0	19.0	31.0	31.0	39.0	39.0	40.0	40.0
Recent Holdings	3,643	247.16	247.16	448.24	448.24	6.0	6.0	43.0	43.0	101.0	101.0	225.0	225.0	4,428.0	4,428.0
Recent Asset Value ($M)	3,643	7,700.17	1,866.43	77,463.51	5,227.44	2.05	1.55	223.84	183.62	511.92	424.38	1,667.16	1,268.68	4,070,960.99	147,491.05
Alpha (%)	3,643	7.94	20.82	32.67	54.66	−77.4	−88.0	−7.7	−3.9	0.4	6.3	12.8	24.25	261.1	396.0
Beta	3,643	1.14	1.42	1.18	2.2	−6.88	−18.49	0.8	0.81	1.03	1.13	1.33	1.68	17.94	33.58
Ann. Arithmetic Mean Return (%)	3,643	23.1	38.02	33.82	60.64	−74.9	−72.3	6.1	10.3	14.5	20.4	28.0	40.2	262.6	710.8
Ann. Geometric Mean Return (%)	3,643	30.67	67.44	56.66	211.92	−56.3	−54.9	6.2	10.7	15.3	22.1	31.05	46.7	652.8	5848.0
Ann. Median Return (%)	3,643	14.73	13.96	18.4	20.55	−53.8	−67.8	4.9	3.3	13.5	12.9	23.0	23.2	214.3	237.1
CAGR (%)	3,643	9.99	13.08	19.01	23.12	−64.6	−62.8	0.75	1.2	8.8	11.6	17.8	21.9	121.2	220.2
Ann. Sharpe (RFR=2%)	3,643	0.4	0.48	0.49	0.46	−2.93	−2.29	0.17	0.26	0.4	0.46	0.66	0.7	4.13	3.5
Ann. Sortino (MAR=2%)	3,635	1.4	2.2	6.95	11.87	−2.87	−3.07	0.33	0.55	0.84	1.16	1.63	2.24	364.39	628.88
Ann. Volatility (%)	3,643	51.93	79.56	67.09	125.31	6.2	7.5	19.7	22.5	28.8	38.9	50.0	75.95	446.3	1324.0
Max Drawdown - Qtr Returns (%)	3,643	−35.53	−39.82	21.01	24.33	−91.44	−98.25	−46.96	−56.82	−28.81	−34.52	−20.1	−19.26	−0.04	0.67
Ann. Excess Return (%)	3,643	9.69	25.48	33.49	60.46	−88.2	−84.9	−6.7	−1.8	1.2	8.0	14.4	27.6	237.2	697.6
Tracking Error	3,643	23.33	38.07	34.23	63.08	1.4	0.8	6.3	8.9	11.9	17.8	22.9	36.85	224.9	661.7
Information Ratio	3,643	−0.08	0.38	1.4	1.11	−6.99	−6.46	−0.86	−0.18	0.1	0.52	0.77	0.97	10.19	6.44

Table 4.6: Descriptive Statistics of Performance and Risk Metrics for Original and Cloned Funds.

This table provides a side-by-side statistical summary of various performance and risk factors for 3,643 *original and cloned funds for the period 2013 to 2023. It contrasts key metrics such as Alpha, Beta, annualized mean returns, Compound Annual Growth Rate (CAGR), Sharpe Ratio with a Risk Free Rate (RFR) of 2%, Sortino Ratio with a Minimum Acceptable Return (MAR) of 2%, volatility, maximum drawdown of quarterly returns, excess return, tracking error, and information ratio. The statistics underscore the cloned funds' enhanced mean Alpha and returns across different measures, a slightly elevated Beta, and increased volatility compared to original funds. Notable too is the cloned funds' higher maximum drawdown and excess return, with a tracking error and information ratio that suggest differentiated performance relative to the original funds.*

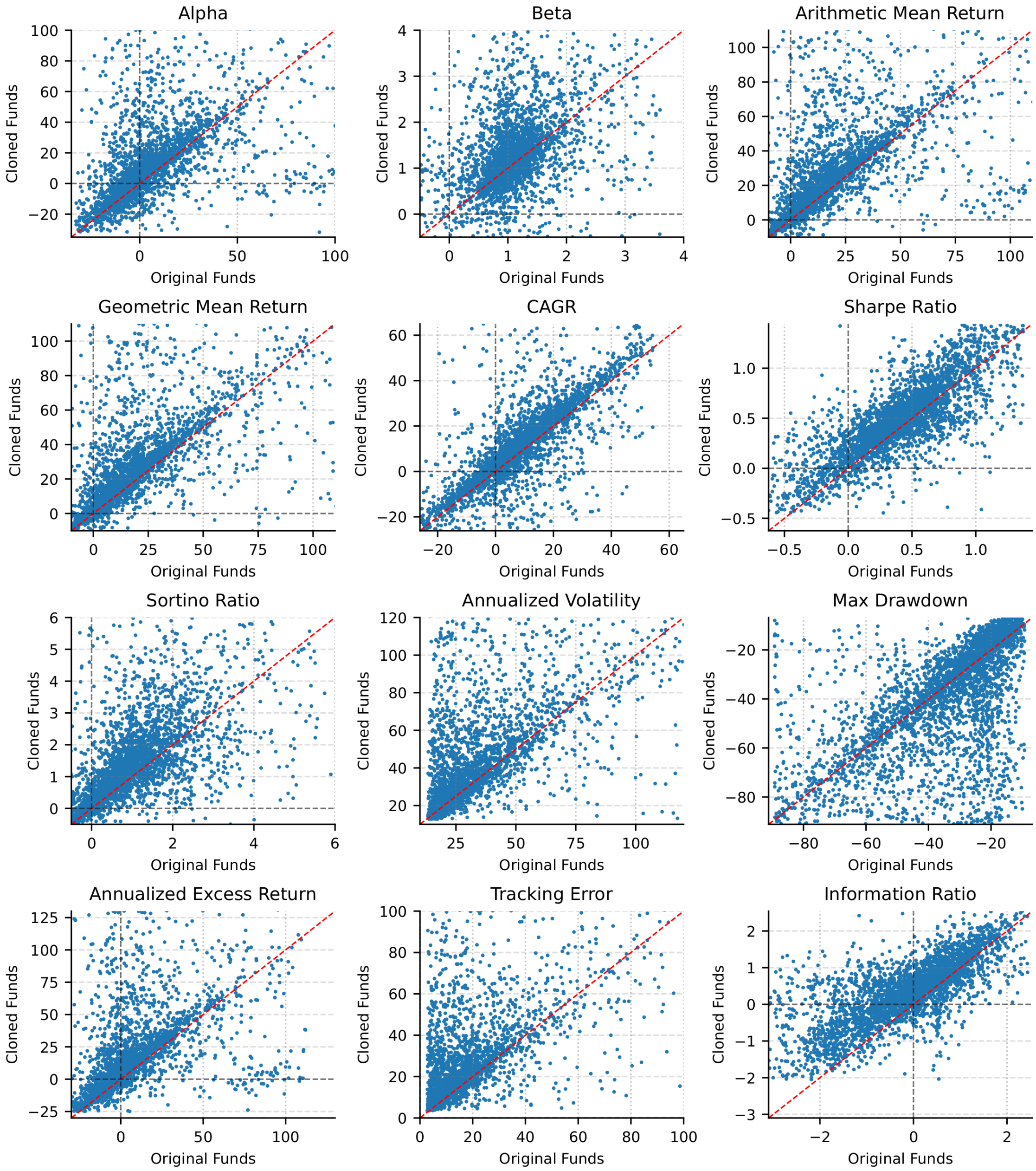


Figure 4.9: Correlation of Performance Metrics between Original and Cloned Funds across Various Factors.

Twelve scatterplots display the relationship between 3,643 *original and cloned funds across a series of performance metrics, such as Alpha, Beta, and various types of returns, ratios, volatility, drawdown, and tracking error. Each plot on the x-axis represents a factor's value for the original funds, while the y-axis corresponds to the values for the cloned funds. The nearly linear patterns, highlighted by the red line of equality, illustrate a strong positive correlation between the original and cloned funds across most metrics. Notably, factors like Alpha, arithmetic mean return, Compound Annual Growth Rate and information ratio showcase a tight clustering around the line of equality, indicating that cloned funds often mirror the performance of their originals closely. The plots for Sharpe Ratio and Sortino Ratio also display close alignment, suggesting a consistent risk-adjusted return profile. Meanwhile, the graphs for annualized volatility and maximum drawdown reflect a greater variance, pointing to potential differences in risk behaviors. Overall, these visual comparisons affirm the cloning approach's ability to replicate fund performance effectively, with variations in certain risk-related factors.*

Factor	t-stat	Samples	Mean	Std. Dev.	Min	25%	50%	75%	Max
Alpha	1,317,518.5***	3,643	9.40	30.57	−123.80	0.00	3.70	10.40	202.00
Beta	1,996,741.0***	3,643	0.20	0.96	−5.02	−0.15	0.09	0.46	6.56
Arithmetic Mean Return	1,054,798.0***	3,643	10.54	32.40	−111.70	0.90	3.60	10.30	213.80
Geometric Mean Return	1,063,496.5***	3,643	16.44	57.72	−179.10	1.00	4.00	12.10	458.00
CAGR	1,392,466.5***	3,643	2.65	10.33	−53.60	0.30	2.90	5.10	71.40
Sharpe Ratio	1,519,189.0***	3,643	0.08	0.24	−1.46	−0.02	0.08	0.20	1.49
Sortino Ratio	1,270,866.5***	3,643	0.49	1.28	−4.83	0.00	0.29	0.77	8.87
Annualized Volatility	1,492,135.0***	3,643	20.31	66.70	−279.20	−1.90	3.00	21.65	422.90
Max Drawdown	2,358,882.0***	3,643	−4.40	19.00	−74.86	−10.28	0.27	4.99	81.23
Annualized Excess Return	914,614.5***	3,643	11.45	32.55	−110.60	1.90	4.50	11.10	215.80
Tracking Error	1,056,415.0***	3,643	10.96	33.25	−142.60	0.00	2.10	11.60	216.80
Information Ratio	829,797.0***	3,643	0.46	0.74	−2.58	0.07	0.34	0.77	4.08

Table 4.7: Statistical Analysis of Performance and Risk Metrics Delta Distributions (Cloned vs. Original Funds).

This table delineates the delta distributions for various performance and risk metrics, where each metric for cloned funds is subtracted from that of the original funds. It presents the t-statistics from the Wilcoxon signed-rank test alongside highly significant p-values, indicating the strength of the evidence against the null hypothesis. The metrics where higher values indicate better performance (e.g., Alpha, CAGR, Sharpe Ratio) show positive median values, suggesting that cloned funds generally match or outperform the original funds. Conversely, for metrics where lower values are preferred (e.g., Volatility, Max Drawdown), the median values are less than or equal to zero, aligning with the desired outcomes. The results overall support the rejection of the null hypothesis, affirming that cloned portfolios – rebalanced on the disclosure date and modeled after institutionally successful strategies – do replicate or exceed the historical market performance of the original funds. Significance levels are denoted as follows: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

In analyzing the delta distribution of alpha, we observe that 50% of cloned portfolios have an alpha exceeding their source portfolios by at least 3.7%. Specifically, the top 25% of cloned funds, corresponding to the fourth quartile, demonstrate a remarkable alpha difference of 10.04% compared to their source portfolios, signifying that these clones achieve superior risk-adjusted returns.

The beta distribution shows a consistent median difference of 0.09, indicating stability across the cloned portfolios in terms of market sensitivity.

Further supporting the alpha findings, both the arithmetic and geometric mean returns in the third and fourth quartiles of cloned portfolios surpass those of the source portfolios. The arithmetic mean return at the 75th percentile is 10.3% higher, while the geometric mean return at the same percentile reaches 12.1% higher, indicating significant annualized return improvements in the top quartile clones. These observations are complemented by higher CAGR values, Sharpe, and Sortino ratios within the top quartile. Specifically, CAGR exceeds 5.1%, the Sharpe ratio surpasses 0.2, and the Sortino ratio exceeds 0.77, underscoring the clones' enhanced performance metrics.

However, it's noteworthy that cloned portfolios exhibit higher annualized volatility, with a median increase of 3%, and a greater maximum quarterly drawdown, with a median increase of 0.27%, suggesting increased risk in cloned portfolio strategies. Moreover, cloned portfolios demonstrate larger tracking errors with the median value of 2.1 indicating negative deviation from their source portfolios.

The outcomes of the Wilcoxon signed-rank test, indicated by the t-values and corresponding *p*-values in Figure 4.7, establish that all evaluated delta distributions significantly diverge from zero. This suggests that the performance of cloned funds is markedly different from that of the source funds. Specifically, nine out of the twelve examined performance and risk factors show that the cloned funds not only match the performance of the source funds but in several cases exceed it. This is evidenced by the positive delta values across the median and top quartile for these factors. However, for three factors—annualized volatility, maximum drawdown, and tracking error—the cloned funds present lower performance metrics.

Given these results, we confidently reject the null hypothesis. Instead, we accept the alternative hypothesis, affirming the premise that replicating the buy and sell activities of institutional investment managers based on their 13F filings on the date of disclosure can effectively mirror and surpass the performance of the original funds, thereby outperforming the S&P 500 benchmark.

4.4 Conclusion

We cloned 161,408 hedge fund portfolios from the period of 2013 to 2023, revealing that replicating the strategies of institutional investment managers from SEC 13F filings, rebalanced on the filing disclosure date, aligns with or exceeds the performance of the original funds. Performance and risk characteristics of cloned and original funds were evaluated across twelve dimensions. Our results confirm that cloned portfolios tend to replicate the superior performance of top-performing original funds, suggesting the efficacy of a straightforward cloning strategy in surpassing the market benchmark, particularly in terms of risk-adjusted returns. The upper quartile of original funds shows a minimum alpha of 12.8%, while the upper quartile of cloned counterparts exhibits a minimum alpha of 24.3%. This finding signifies that the top 25% of cloned funds do not merely track the performance of their source funds but also surpass the S&P 500 benchmark by a margin exceeding 24%, indicating substantial outperformance. However, despite the overall strong performance, the cloned funds exhibit increased annualized volatility and maximum drawdowns, indicating higher risk levels compared to their source funds. Additionally, the higher tracking error suggests a discrepancy in the performance consistency of cloned funds against the market benchmark. Our results on fund survival rates show a median operational lifespan of 5 years, with 3.4% of funds exceeding a 26-year lifespan, providing context for the durability of investment strategies over time. Furthermore, 75% of all funds manage portfolios smaller than $2 billion, and the portfolio value distribution is heavily skewed by a small number of funds, with less than 2.5% of funds managing upwards of $45 billion.

Given that cloned funds matched or outperformed their source funds in nine out of twelve factors, our study rejects the null hypothesis and demonstrates that replicating the holdings of institutional investment managers can indeed result in performance that is comparable to or even better than the original funds, thereby providing a potential avenue to outperform the S&P 500 benchmark.

Bibliography

Ackermann, C., McEnally, R., and Ravenscraft, D. (1999). The performance of hedge funds: risk, return, and incentives. *The Journal of Finance*, 54(3):833–874.

Agarwal, V., Boyson, N. M., and Naik, N. Y. (2009). Hedge funds for retail investors? An examination of hedged mutual funds. *Journal of Financial and Quantitative Analysis*, 44(2):273–305.

Agarwal, V., Jiang, W., Tang, Y., and Yang, B. (2013). Uncovering Hedge Fund Skill from the Portfolio Holdings They Hide. *The Journal of Finance*, 68(2):739–783.

Agarwal, V. and Naik, N. Y. (2003). Risks and portfolio decisions involving hedge funds. *Review of Financial Studies*, 17(1):63–98.

Amin, G. S. and Kat, H. M. (2003). Hedge fund performance 1990–2000: Do the "Money machines" really add value? *Journal of Financial and Quantitative Analysis*, 38(2):251.

Amir-Ghassemi, F., Papanicolaou, A., and Perlow, M. (2022). Aggregate alpha in the hedge fund industry: A further look at best ideas. *The Journal of Portfolio Management*, 48(3):220–239.

Anderson, A. and Brockman, P. (2018). An Examination of 13F Filings. *Journal of Financial Research*, 41(3):295–324.

Antón, M., Cohen, R. B., and Polk, C. (2010). Best Ideas. *Social Science Research Network*.

Aragon, G. O., Hertzel, M. G., and Shi, Z. (2013). Why Do Hedge Funds Avoid Disclosure? Evidence from Confidential 13F Filings. *Journal of Financial and Quantitative Analysis*, 48(5):1499–1518.

Blume, M. E. (1975). Betas and their regression tendencies. *The Journal of Finance*, 30(3):785.

Bollen, N. P. B. and Fisher, G. S. (2013). Send in the clones? Hedge fund replication using futures contracts. *The Journal of Alternative Investments*, 16(2):80–95.

Brinson, G. P., Hood, L. R., and Beebower, G. L. (1986). Determinants of portfolio performance. *Financial Analysts Journal*, 42(4):39–44.

Brown, G. W., Howard, P., and Lundblad, C. (2021). Crowded trades and tail risk. *The Review of Financial Studies*, 35(7):3231–3271.

Cao, S., Da, Z., Jiang, D., and Yang, B. (2021a). The strategic use of 13F restatement by hedge funds. *Social Science Research Network*.

Cao, S. S., Du, K., Yang, B., and Zhang, A. L. (2021b). Copycat skills and disclosure costs: Evidence from peer companies' digital footprints. *Journal of Accounting Research*, 59(4):1261–1302.

Chaudhry, A. and Johnson, H. L. (2008). The efficacy of the Sortino ratio and other benchmarked performance measures under skewed return distributions. *Australian Journal of Management*, 32(3):485–502.

DiPietro, D. M. (2019). Alpha Cloning: using quantitative techniques and SEC 13F data for equity portfolio optimization and generation. *The Journal of Financial Data Science*, 1(4):159–171.

Fama, E. F. and French, K. R. (2004). The capital asset pricing model: Theory and evidence. *The Journal of Economic Perspectives*, 18(3):25–46.

Fleiss, A., Cui, H., Stoikov, S., and DiPietro, D. M. (2019). Constructing Equity Portfolios from SEC 13F Data Using Feature Extraction and Machine Learning. *The Journal of Financial Data Science*, 2(1):45–60.

Fleiss, A., Kumaar, A., Rida, A., Shin, J., Lai, X., Fang, V., Chen, J., and Li, A. (2021). Deep reinforcement learning & feature extraction for constructing alpha generating equity portfolios. Available at SSRN: https://ssrn.com/abstract=3958478 or http://dx.doi.org/10.2139/ssrn.3958478.

Frank, M. M., Poterba, J. M., Shackelford, D. A., and Shoven, J. B. (2004). Copycat Funds: Information Disclosure Regulation and the Returns to Active Management in the Mutual Fund Industry. *The Journal of Law and Economics*, 47(2):515–541.

Fung, W. and Hsieh, D. A. (2000). Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases. *Journal of Financial and Quantitative Analysis*, 35(3):291.

Fung, W. and Hsieh, D. A. (2001). The risk in hedge fund strategies: Theory and evidence from trend followers. *Review of Financial Studies*, 14(2):313–341.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267–270.

Griffin, J. M. and Xu, J. (2009). How smart are the smart guys? A unique view from hedge fund stock holdings. *Review of Financial Studies*, 22(7):2531–2570.

Groggel, D. J. and Conover, W. (2000). Practical nonparametric statistics. *Technometrics*, 42(3):317.

Hasanhodzic, J. and Lo, A. W. (2007). Can hedge funds be replicated? The linear case. *Journal of Investment Management*, 5(2):5–45.

Hayes, B. T. (2012). On the Market-Timing Ability of Factor-Based Hedge Fund Clones. *The Journal of Alternative Investments*, 15(1):8–42.

Hayes, B. T. and Ba, Y. A. (2014). Beta Regime-Switching Hedge Funds and their Clones. *The Journal of Alternative Investments*, 17(3):87–110.

Hubbard, J. W. (2008). Can Hedge-Fund Returns Be Replicated?: The Linear Case. *The C.F.A. Digest*, 38(1):6–7.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, 23(2):389–416.

Kacperczyk, M., Sialm, C., and Zheng, L. (2006). Unobserved actions of mutual funds. *Review of Financial Studies*, 21(6):2379–2416.

Kaizoji, T. and Miyano, M. (2016). Why does the power law for stock price hold? *Chaos, Solitons & Fractals*, 88:19–23.

Lazo-Paz, R., Moneta, F., and Chincarini, L. B. (2023). Crowded spaces and anomalies. *Social Science Research Network*.

Lewellen, J. (2011). Institutional investors and the limits of arbitrage. *Journal of Financial Economics*, 102(1):62–80.

Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4):36–52.

Malkiel, B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance*, 50(2):549–572.

Markowitz, H. M. and Blay, K. A. (2013). *Risk-Return Analysis: The Theory and Practice of Rational Investing (Volume One)*. McGraw-Hill Professional.

Martin, G. S. and Puthenpurackal, J. (2008). Imitation Is the Sincerest Form of Flattery: Warren Buffett and Berkshire Hathaway. *Social Science Research Network*.

Mitchell, M. and Pulvino, T. (2001). Characteristics of risk and return in risk arbitrage. *The Journal of Finance*, 56(6):2135–2175.

O'Shaughnessy, J. P. (2005). *What Works on Wall Street*. McGraw Hill LLC.

Paleologo, G. A. (2021). *Advanced Portfolio Management: A Quant's Guide for Fundamental Investors*. John Wiley & Sons.

Pedersen, C. S. and Satchell, S. E. (2002). On the foundation of performance measures under asymmetric returns. *Quantitative Finance*, 2(3):217–223.

Sandvik, S. H., Frydenberg, S., Westgaard, S., and Heitmann, R. K. (2011). Hedge fund performance in bull and bear markets: Alpha creation and risk exposure. *The Journal of Investing*, 20(1):52–77.

Schroeder, J. (2025a). Anomalies, Trends and Patterns in Disclosure Activities: Understanding EDGAR. Working Paper.

Schroeder, J. (2025b). Outperforming the Market: Portfolio Strategy Cloning from SEC 13F Filings. Working Paper.

Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1):49–58.

Sias, R. W., Turtle, H. J., and Zykaj, B. B. (2016). Hedge fund crowds and mispricing. *Management Science*, 62(3):764–784.

Sortino, F. A. and Price, L. N. (1994). Performance measurement in a downside risk framework. *The Journal of Investing*, 3(3):59–64.

Verbeek, M. and Wang, Y. (2013). Better than the Original? The Relative Success of Copycat Funds. *Journal of Banking and Finance*, 37(9):3454–3471.

Wallerstein, E., Tuchschmid, N. S., and Zaker, S. (2009). How do hedge fund clones manage the real world? *The Journal of Alternative Investments*, 12(3):37–50.

Weber, A. (2017). Annual risk measures and related statistics.

Weber, V. and Peres, F. (2013). Hedge fund replication: putting the pieces together. *The Journal of Investment Strategies*, 3(1):61–119.

Chapter 5: Effects of Non-Reliance Disclosures in Form 8-K Filings on Stock Prices

Author: Jan L. Schroeder (jan@sec-api.io)*

Abstract
We analyze market reactions to 8,006 non-reliance disclosures filed under Item 4.02 in SEC Form 8-K filings (2004–2023). These disclosures result in cumulative abnormal returns of -2.6% to -5.4%, with revenue recognition errors triggering 114% more severe reactions than net income errors. Disclosures expressing uncertainty about the impact of identified errors face three times harsher market responses than those with clearly stated effects, while SEC-identified issues do not amplify negative price reactions. As Item 4.02 disclosures precede financial restatements, they serve as a significant and robust early risk signal for investors, analysts and stakeholders.

Publication Details: Working Paper. Submitted to the Journal of Corporate Accounting & Finance.

Keywords: Non-Reliance Disclosures, SEC Form 8-K Filings, Item 4.02, Event Study, Market Reactions

Acknowledgements: We thank the participants of the GLOBAFA 2025 Conference (30–31 May, Budapest) for their invaluable comments during the presentation of this paper. We are especially grateful to session chair Conrado Diego García Gómez (University of Valladolid, Spain), as well as Seda Bilyay-Erdogan (Kadir Has University, Türkiye) and Pradip Banerjee (Indian Institute of Management Indore), for their feedback and insights. We also sincerely thank Christiane Pott, Professor of International Accounting and Auditing at TU Dortmund, Germany, and Celine Daute, TU Dortmund, Germany, for their feedback and comments.

The following is based on Schroeder (2025).

*Jan L. Schroeder is Research Fellow at the Department of Finance, TU Dortmund, Germany, and the founder and CEO of SEC-API.io. The authors have no competing interests to declare. All errors are our own.

Highlights

- Non-reliance disclosures in SEC Form 8-K, Item 4.02 trigger significant negative stock price reactions, with effects robust from 2007 to 2023.
- Revenue recognition errors lead to 114% more severe market reactions compared to net income errors.
- Disclosures expressing uncertainty about the impact of identified errors experience three times harsher market reactions than those with clearly stated effects.
- Auditor reputation does not mitigate the negative impact of non-reliance disclosures, with Big 4 involvement showing no protective effect.
- SEC-identified issues do not amplify negative stock price reactions, suggesting regulatory involvement does not inherently worsen investor sentiment.

5.1 Introduction

The reliability of financial statements is fundamental to investment decisions and risk management. When companies file Item 4.02 disclosures in SEC Form 8-K, they publicly acknowledge that previously issued financial statements can no longer be relied upon. These non-reliance disclosures act as early warning signals, typically preceding formal financial restatements, yet their market impact has received limited empirical attention due to the unstructured nature of the disclosure text.

Our study examines the stock price impact of Item 4.02 filings, leveraging a newly available, structured dataset of 8,006 disclosures from 2004 to 2023. In contrast to prior research, which has focused on the market effects of formal restatements disclosed in Forms 10-K/A and 10-Q/A (Kravet and Shevlin, 2009; Files et al., 2009; Umar et al., 2023), we isolate the market response to the initial disclosure of unreliability. While it is well established that restatements can trigger abnormal returns of -4% to -12% (Dechow et al., 1996; Anderson and Yohn, 2002; Kinney and McDaniel, 1989; Scholz, 2008; Robbani et al., 2006), there is little understanding of the investor response to the earlier 8-K filings that first signal the need for such restatements (Palmrose et al., 2004; Scholz, 2008). Feldman et al. (2008) remain the only study to analyze Item 4.02 disclosures directly, based on a limited sample of 182 filings.

Item 4.02 disclosures are legally required to be published within four business days of identifying a material misstatement, as mandated by Section 13(a) of the Securities Exchange Act and introduced by the Sarbanes–Oxley Act (SOX)[1] in 2004. These filings provide qualitative information on the nature of the error, affected line items and periods, and whether the issue was identified by the company, its auditor, or the SEC. Unlike amended filings, Item 4.02 disclosures do not contain all corrected figures, but serve as the first formal acknowledgment of a reliability breach, often weeks or months before a restatement is filed.

Despite the absence of revised financials, these disclosures frequently trigger immediate stock price adjustments. We find that Item 4.02 filings are associated with cumulative abnormal returns (CARs) ranging from -2.6% to -5.4%, reflecting the materiality and informational content investors extract from them.

We contribute to the literature on financial restatements and disclosure-driven price reactions in several ways. First, we provide the most comprehensive empirical analysis of Item 4.02 disclosures to date, based on a large-scale dataset spanning two decades.

[1] Section 13(a) of the Securities Exchange Act of 1934 (15 U.S.C. § 78m(a))

Second, we shift the focus from restatement outcomes to the signaling effects of the initial disclosure event. Third, we analyze how specific content features–such as stated uncertainty, bundling with other material events, and disclosure timing–shape investor responses. Finally, our findings offer actionable insights for investors, regulators, and corporate decision-makers seeking to understand the informational value and market impact of reliability-related disclosures.

The remainder of this paper is structured as follows: Section 5.2 describes the dataset and methodology. Section 5.3 presents the empirical findings and discussion. Section 5.4 concludes with key takeaways, practical implications, and directions for future research.

5.2 Data & Methods

Item 4.02 is one of 33 material events that require disclosure in SEC Form 8-K filings within four business days of occurrence. It provides a narrative summary of material financial reporting errors, identifying instances in which a company, its auditor, or the SEC has determined that previously issued financial statements can no longer be relied upon. These disclosures often detail the nature and discovery of the error, its financial implications and the presence of material weaknesses in internal controls, among others.

Item 4.02 is one of only two Form 8-K items (alongside Item 4.01) that the SEC prohibits from being embedded in periodic filings such as Forms 10-K or 10-Q. It must be filed as a standalone Form 8-K within four business days of a determination by the board, audit committee, or authorized officer. Failure to do so constitutes a technical violation of Exchange Act Rule 13a-11 and Section 13(a), potentially triggering SEC enforcement actions (e.g., Accounting and Auditing Enforcement Releases), civil penalties, and shareholder litigation. While a single violation may not lead to delisting, repeated violations may threaten listing eligibility.

Illustrative examples include Squarespace (SQSP), which disclosed on November 15, 2021, that earnings per share had been overstated by 100% due to a miscalculation in weighted average shares (Figure 5.1). The stock declined by -37.4% within 20 trading days. Similarly, Exela Technologies (XELA) reported long-term debt misclassification and going concern uncertainties (Figure 5.2), resulting in a -68% decline and eventual delisting from NASDAQ.

Item 4.02. Non-Reliance on Previously Issued Financial Statements or a Related Audit Report or Completed Interim Review.

On November 10, 2021, the Audit Committee of the Board of Directors of Squarespace, Inc. (the "Company"), in consultation with management and Ernst & Young LLP, the Company's independent registered public accounting firm, concluded that the Company's previously issued unaudited interim consolidated financial statements for the interim period ended September 30, 2021 included in its quarterly report on Form 10-Q for the quarter ended September 30, 2021, as originally filed with the Securities and Exchange Commission on November 8, 2021, should no longer be relied upon due to the identification of an error in the calculation of its weighted-average shares used in computing net income/(loss) per share attributable to Class A, Class B, and Class C stockholders, basic and diluted ("WASO") for the three months ended September 30, 2021. Due to this error, the Company's WASO, the net income/(loss) per share attributable to Class A, Class B, and Class C stockholders, basic and dilutive for the three months ended September 30, 2021 was also incorrectly calculated as $0.04 per share instead of $0.02 per share.

As a result of the error, the Company determined that it must restate the unaudited interim consolidated financial statements for the interim period ended September 30, 2021 and will file an amendment to the Company's Form 10-Q for the three months ended September 30, 2021, on or around November 15, 2021.

Figure 5.1: Squarespace (ticker: SQSP) announced on November 15, 2021 (since taken private) under Item 4.02 in a Form 8-K filing[2] *that its earnings per share were overstated by* 100% *due to an error in calculation of its weighted average shares and experienced a* −37.4% *stock price collapse 20-days post disclosure.*

Historically, the unstructured nature of Item 4.02 disclosures has made large-scale empirical analysis difficult. To address this, we utilize a new structured dataset of 8,006 Item 4.02 filings from 2004 to 2023, sourced from the commercially available SEC-API.io database.[4] The structured data allow us to extract content features for systematic analysis. The variables used in this study are described in Table 5.1.

We conduct an event study to measure cumulative abnormal returns (CARs) following disclosure. CARs are calculated using the market model, with the S&P 500 ETF (SPY) as the benchmark. Event windows span 1 to 20 trading days post-disclosure to capture immediate and delayed investor responses. Stock price data are sourced from AlgoSeek, while quarterly fundamentals are obtained via Capital IQ on WRDS. The sample is survivorship-bias free, including both listed and delisted firms.

To examine how specific disclosure features influence market reactions, we apply three empirical strategies: Binary Variable Comparison: Disclosures are split by the True/False outcome of each variable (e.g., "Impact yet to be determined"), and the mean and median CARs are compared across groups. Quintile-Based Analysis: Disclosures are sorted into quintiles based on 20-day CARs (Group A: most negative; Group E: most positive). For binary variables, we compare the share marked "True" in the top vs. bottom quintiles. For continuous variables, we compare mean/median values between Groups A and E. OLS Regression: We estimate six ordinary least squares models using 5-day CAR as the

[2] Squarespace' Form 8-K, Item 4.02 disclosure: `https://www.sec.gov/Archives/edgar/data/1496963/000110465921139150/tm2132787d1_8k.htm`

[3] Exela Tech. Form 8-K, Item 4.02 disclosure: `https://www.sec.gov/Archives/edgar/data/1620179/000110465922118697/tm2230425d1_8k.htm`

[4] See: `https://sec-api.io/docs/form-8k-data-search-api`

dependent variable. Each model includes different combinations of fixed effects (firm, auditor, sector, and year) to control for unobserved heterogeneity.

To test whether average CARs significantly differ from zero, we apply Wilcoxon signed-rank and one-sample t-tests, depending on normality (checked via Shapiro–Wilk tests). To evaluate whether price reactions evolve over time, we apply Mann–Kendall tests to assess the presence of monotonic trends in CARs across event windows and quintiles.

Of the 8,006 disclosures, 2,390 were included in the event study, covering the period from 2007 to 2023 for firms listed on NYSE, NASDAQ, or AMEX. Disclosures were excluded if the firm was not publicly traded, was listed only on OTC markets, or the event occurred prior to 2007, outside the period for which pricing data was available to us. In total, our sample spans 5,028 unique filers and 856 distinct auditors.

Item 4.02. Non-Reliance on Previously Issued Financial Statements or a Related Audit Report or Completed Interim Review.

Subsequent to the filing of its Annual Report on Form 10-K for the fiscal year ended December 31, 2021 (the "Original 10-K"), Exela Technologies, Inc. (the "Company") re-evaluated its application of ASC Subtopic 205-40, *Presentation of Financial Statements—Going Concern* ("ASC 205-40") as of March 16, 2022, the date of the Original 10-K. Under ASC 205-40, the Company has the responsibility to evaluate whether conditions and/or events raise substantial doubt about its ability to meet its obligations as they become due within one year after the date that the financial statements are issued. In re-performing this evaluation as of the date of the Original 10-K, the Company determined the need to take into account the potential impact of certain true-up guaranties that the Company had issued in connection with the Revolving Loan Exchange and Prepayment Agreement, dated March 7, 2022, that had not previously been taken into account in its assessment. If the Company had taken the true-up guaranties into account in addition to other existing factors, the Company may not have had sufficient liquidity under its financial model to fund payment of this true-up obligation in addition to its other commitments for the twelve months following the date of the Original 10-K.

Based on this evaluation, management has determined that if that contingent liability created by the true up guaranty were to settle within one year from March 16, 2022, there was substantial doubt about the Company's ability to continue as a going concern for the twelve months following the date of the Original 10-K, which determination should have been disclosed in the Company's previously issued audited financial statements included in the Original 10-K (the "audited financial statements").

As a result of the foregoing, on November 9, 2022, the audit committee of the Company's board of directors concluded, after discussion with the Company's management, that the audited financial statements included within the Original 10-K should be restated and should no longer be relied upon. As such, the Company intends to restate the audited financial statements and related notes in Amendment No. 2 to the Original 10-K, to be filed with the SEC (the "Amendment") to restate management's conclusion that substantial doubt exists and to include appropriate related disclosures and to reclassify certain long-term debt from noncurrent to current, as discussed below (the "Restatement").

The Amendment includes reissued audit reports from KPMG LLP ("KPMG"), the Company's independent registered public accounting firm, due to the Restatement and other limited related changes to Part II, Item 7. Management's Discussion and Analysis of Financial Condition and Results of Operations, and Part II, Item 9A. Controls and Procedures. As result of issuance of an amended audit report including a going concern explanatory paragraph, the Company determined that the amount owed under a credit facility which is no longer existing as of the date of this report would become current, and accordingly restated it from noncurrent to current classification in the restated balance sheet as of December 31, 2021. The above changes did not have an effect on retained earnings, or other components of equity or net assets of the Company and result from management's conclusion that substantial doubt exists and the amended audit reports issued by KPMG to reflect such conclusion.

The Company has discussed with KPMG the matters described herein.

The Amendment is being filed concurrently with this report.

Figure 5.2: Item 4.02 disclosure[3] by Exela Technologies (ticker: XELA) on November 14, 2022 (since moved from NASDAQ to OTC), announcing that the company misclassified long-term debt and concerns about going concern assumptions, resulting in a −68% stock price decline over the following 20 days.

Variable Name	Description
`identifiedIssues`	List of all identified issues as disclosed in an Item 4.02 filings, e.g. `"Unexplained reconciling differences in inventory balances"`, `"Misstated inventory by a former employee"`
`identifiedBy`	The entity or entities that discovered the issue. Possible values: `Company`, `Auditor`, `SEC`. For example, the audit committee of the company might have identified a material accounting error (Company), or the company's external auditors might have discovered the issue (Auditor). The SEC might have also identified the error during a routine review and requested the company to disclose the issue using a SEC comment letter or other means.
`restatementIsNecessary`	`true` if the Item 4.02 disclosure informs stakeholders that a restatement of previously disclosed financial statements is necessary, `false` otherwise.
`reasonsForRestatement`	Reasons for the restatement, e.g. `"Misstated inventory"`, `"Inaccurate and unsupported manual journal entries"`
`affectedReportingPeriods`	List of affected reporting periods that potentially require restatement, including quarters or financial years, e.g. `"Q1 2023"`, `"FY 2022"`.
`affectedLineItems`	The line items of the respective income statement, balance sheet, or cash flow statement that are affected by the identified errors , e.g. `"Inventory"`, `"COGS"`, `"Income Before Income Taxes"`
`impactYetToBeDetermined`	`true` if the company explicitly states that the impact of the error is yet to be determined, `false` otherwise.
`impactOfError`	Impact of the error, e.g. "Income before income taxes was overstated by approximately $5 to $7 million for FY 2023. The total overstatement is approximately $26 to $29 million."
`impactIsMaterial`	`true` if the company explicitly states that the impact of the error is material, `false` otherwise.
`materialWeaknessIdentified`	`true` if the company discloses a material weakness in its internal financial controls, `false` otherwise.
`auditors`	List of auditors involved in the restatement process. Typically the auditor employed by the company at the time of disclosure.
`netIncomeDecreased`	`true` if the company explicitly states that net income has been overstated and will require a downward adjustment in a subsequent restatement; `false` otherwise.
`netIncomeIncreased`	`true` if the company explicitly states that net income has been understated and will require an upward adjustment in a subsequent restatement, `false` otherwise. Note, this field is not mutually exclusive with `netIncomeDecreased`as both can be `true` if the restatement affects different periods.
`netIncomeAdjustment`	If explicitly disclosed, the magnitude of the net income adjustment, e.g. "$348.0 million".
`revenueDecreased`	`true` if the company explicitly states that revenue has been overstated and will require a downward adjustment as a result of a restatement, `false`.
`revenueIncreased`	`true` if the company explicitly states that revenue has been understated and it will have to be increased as a result of a restatement, `false` otherwise. Note, this field is not mutually exclusive with `revenueDecreased` as both can be `true` if the restatement affected different periods.
`revenueAdjustment`	If explicitly disclosed, the magnitude of the revenue adjustment, e.g. "$1.2 billion".
`reportedWithOtherItems`	`true` if the filing reports additional material events other than Item 4.02, `false` otherwise. Source: SEC filing metadata.
`reportedWithEarnings`	`true` if the filing includes quarterly or annual operating results disclosed under Item 2.02 or 9.01, `false` otherwise. Source: SEC filing metadata.

Table 5.1: Variables Describing Content Features Capturing Different Parts of Information Disclosed in Item 4.02 Filings.

5.3 Results & Discussion

Our analysis of Item 4.02 disclosures from 2004 to 2023 reveals a clear downward trend in the number of non-reliance filings over time (Figure 5.3, left, Table 5.2). Annual volumes peaked in 2005 (962 disclosures) and 2006 (1,004 disclosures), but have declined steadily since, reaching lows of 98 and 96 disclosures in 2019 and 2020, respectively. A sharp anomaly occurred in 2021, with 864 disclosures–largely driven by equity and warrant misclassification errors involving Marcum and WithumSmith+Brown (Table 5.2, Table 5.3).

This decline suggests a sustained improvement in financial reporting quality following the implementation of SOX. Enhanced regulatory oversight and strengthened internal controls appear to have contributed to the reduced frequency of non-reliance events.

We also observe a consistent seasonal pattern in disclosure timing (Figure 5.3, right; Table 5.2). March and November show elevated filing activity, averaging 55 and 52 disclosures per month, respectively, compared to a baseline monthly average of 34. These peaks align with key reporting deadlines–10-K filings in February and March, and final 10-Qs in November–indicating that most errors are identified during year-end audits and quarterly reviews.

In contrast, July and September have the lowest disclosure activity, averaging just 21 filings per month. The July-October period generally sees fewer disclosures, with the exception of August, which aligns with the Q2 filing cycle. The clustering of Item 4.02 filings around key reporting periods underscores the role of audit procedures in uncovering financial misstatements.

Auditor involvement in Item 4.02 disclosures reveals distinct patterns over time (Figure 5.4, Table 5.3). The surge in filings during 2005 and 2006 was largely driven by companies audited by Big 4 firms, reflecting the heightened regulatory scrutiny following SOX. In 2005 alone, PwC was associated with 159 disclosures, KPMG with 116, Deloitte with 103, and EY with 87 (Figure 5.4, Table 5.3). By contrast, the 2021 spike in non-reliance filings marks a sharp departure from this pattern. That year, two non-Big 4 firms–Marcum and WithumSmith+Brown–accounted for 340 and 237 disclosures, respectively, comprising 74% of all Item 4.02 events. This concentration is notable given their historically minimal involvement (typically under 1% annually) and highlights firm-specific issues related to widespread misclassification errors of equity and warrants[5] [6].

[5]https://pcaobus.org/news-events/news-releases/news-release-detail/imposing-3-million-fine-and-requiring-first-ever-changes-to-supervisory-structure-pcaob-sanctions-marcum-llp-for-significant-quality-control-violations

[6]https://pcaobus.org/news-events/news-releases/news-release-detail/imposing-2-million-fines-pcaob-sanctions-withumsmith-brown-pc-pervasive-quality-control-violations-involving-spac-audits

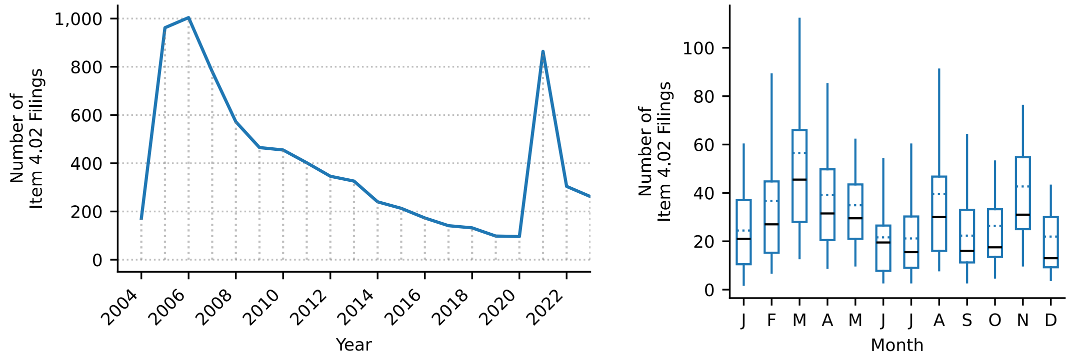


Figure 5.3: Item 4.02 Disclosures per Year from 2004 to 2023 (Left) and Box Plot of Item 4.02 Disclosures per Month for January through December over the Same Period, 2004 to 2023, Excluding 2021 (Right).

	Months												Statistics			
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total	Mean	Std. Dev.	Median
2005	37	89	175	85	91	45	46	91	53	63	123	64	962	80.2	39.0	74.5
2006	45	84	152	81	78	63	60	115	64	76	109	77	1,004	83.7	29.1	77.5
2007	60	67	112	85	62	54	41	91	36	53	76	43	780	65.0	22.6	61.0
2008	40	74	73	63	45	24	32	62	36	31	60	32	572	47.7	17.7	42.5
2009	33	40	45	50	33	22	31	46	47	36	58	24	465	38.8	10.8	38.0
2010	37	41	54	49	48	34	23	47	24	34	32	32	455	37.9	10.0	35.5
2011	24	26	70	49	39	27	28	31	17	30	45	16	402	33.5	15.2	29.0
2012	26	28	45	39	31	25	26	35	18	26	38	9	346	28.8	9.7	27.0
2013	18	40	48	41	36	20	11	29	14	15	39	15	326	27.2	13.1	24.5
2014	16	25	31	22	26	19	11	24	12	17	26	11	240	20.0	6.7	20.5
2015	10	19	28	24	21	4	20	19	15	18	25	10	213	17.8	7.0	19.0
2016	17	14	28	15	21	6	9	11	11	5	25	11	173	14.4	7.2	12.5
2017	10	9	15	20	11	10	7	15	6	10	21	7	141	11.8	5.0	10.0
2018	12	14	15	14	13	7	3	11	6	16	15	6	132	11.0	4.4	12.5
2019	6	7	14	12	10	6	9	8	3	8	10	5	98	8.2	3.1	8.0
2020	2	13	13	9	12	3	8	10	4	7	11	4	96	8.0	4.0	8.5
2021	4	10	15	54	264	52	20	11	8	6	271	149	864	72.0	99.9	17.5
2022	38	46	52	23	23	10	6	38	15	13	30	10	304	25.3	15.4	23.0
2023	9	25	46	24	28	10	10	28	21	17	25	19	262	21.8	10.3	22.5
Statistics																
Total	444	671	1,031	759	892	441	401	722	410	481	1,039	544	7,835	653.1	330.2	563.5
Mean	23.4	35.3	54.3	39.9	46.9	23.2	21.1	38.0	21.6	25.3	54.7	28.6	412.4	34.4	17.4	29.7
Std. Dev.	16.2	25.9	46.1	25.0	57.1	18.5	15.6	31.1	17.6	20.0	60.9	35.4	293.0	24.4	22.0	21.0
Median	18.0	26.0	45.0	39.0	31.0	20.0	20.0	29.0	15.0	17.0	32.0	15.0	326.0	27.2	10.3	23.0

Table 5.2: Item 4.02 Disclosures per Year and Month with Descriptive Statistics per Year and Month from 2005 to 2023.

To avoid distortion of statistics, 2004 is excluded because of SOX reporting requirements taking effect mid-year through.

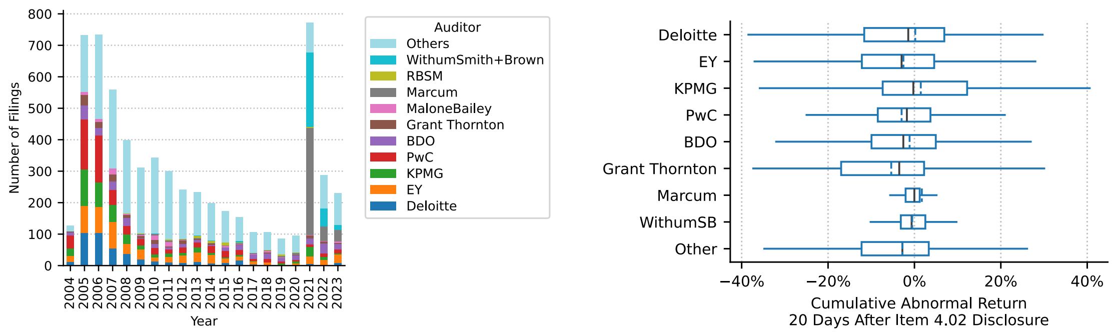


Figure 5.4: Auditor Involvement in Item 4.02 Disclosures from 2004 to 2023 (Left) and Corresponding Stock Price Reactions (Right) from 2007–2023.

The right panel presents a box plot of 20-day cumulative abnormal returns (CARs) following disclosure, grouped by auditor.

	Years																				Statistics				
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	Total	(%)	Mean	Std.	Median
Auditors																									
Deloitte	11	103	104	54	38	20	13	10	8	11	7	9	16	5	3	0	1	6	2	9	430	6.9	22	31	10
EY	20	87	83	85	31	31	12	18	25	31	28	14	14	8	7	3	7	24	16	26	570	9.2	28	26	22
KPMG	24	116	78	54	30	13	11	10	7	15	5	3	4	3	0	4	5	29	10	4	425	6.9	21	30	10
PwC	42	159	149	48	28	21	20	14	18	17	22	20	16	5	11	6	6	9	10	13	634	10.2	32	43	18
BDO	8	44	23	26	24	2	12	6	11	7	4	11	15	15	17	11	13	18	33	17	317	5.1	16	10	14
Grant Thornton	4	34	20	23	10	7	14	3	11	3	4	1	3	0	3	1	3	10	6	3	163	2.6	8	9	4
MaloneBailey	0	8	9	19	3	4	14	15	2	1	0	5	1	2	6	1	0	2	2	5	99	1.6	5	6	2
Marcum	0	0	0	0	0	3	5	5	1	4	3	1	3	2	1	5	5	340	45	36	459	7.4	23	76	3
WithumSmith+Brown	0	0	0	0	0	0	1	0	0	2	2	1	4	0	0	0	0	237	58	17	322	5.2	16	54	0
Other	19	181	267	251	234	211	240	219	158	143	124	109	77	67	58	54	56	97	105	101	2,771	44.8	139	77	116
Statistics																									
Total	128	732	733	560	398	312	342	300	241	234	199	174	153	107	106	85	96	772	287	231	6,190	100.0	310	360	199
Mean	13	73	73	56	40	31	34	30	24	23	20	17	15	11	11	8	10	77	29	23	619		31	36	20
Std.	14	66	85	74	70	64	72	67	48	43	38	33	22	20	18	16	17	117	33	29	774		39	26	35
Median	10	66	50	37	26	10	12	10	10	9	4	7	9	4	4	4	5	21	13	15	428		21	30	10

Table 5.3: Number of Item 4.02 Disclosures by Auditor per Year and Descriptive Statistics (2004 to 2023).

Auditor	Samples	Mean	Std. Dev.	Min	25th	Median	75th	Max
Deloitte	129	0.19%	30.35	−55.36%	−11.63%	−1.43%	6.91%	248.35%
EY	267	−2.55%***	20.08	−54.58%	−12.17%	−2.98%	4.57%	116.50%
KPMG	126	1.47%	21.99	−68.31%	−7.35%	−0.25%	12.20%	110.61%
PwC	179	−2.97%**	15.98	−61.87%	−8.49%	−1.75%	3.71%	41.31%
BDO	157	−1.12%*	21.24	−45.08%	−9.94%	−2.57%	4.94%	144.06%
Grant Thornton	59	−5.37%**	17.64	−37.55%	−16.97%	−3.52%	2.24%	57.99%
Marcum	334	1.72%	36.18	−37.21%	−2.05%	−0.02%	1.22%	631.01%
WithumSmith+Brown	263	−0.51%	7.56	−45.26%	−3.16%	−0.65%	2.52%	35.53%
Other	955	−2.89%***	23.33	−63.97%	−12.24%	−2.80%	3.32%	353.09%

Table 5.4: Statistics of Cumulative Abnormal Returns 20 Days Post Item 4.02 Disclosure by Auditor (2007–2023).

*Significance levels are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).*

Our event study reveals a significant negative market reaction to Item 4.02 disclosures, with cumulative abnormal returns (CARs) declining steadily over the 20-day post-disclosure window (Tables 5.5, 5.6 and 5.7). On day +1, the mean CAR is −0.87% for the full sample and −1.10% when excluding 2021, indicating an immediate adverse investor response. This effect deepens over time, reaching −1.54% and −2.00%, respectively, by day +20. The persistent decline suggests that the market continues to process and reassess the implications of these disclosures well beyond the initial announcement.

Median CARs are statistically significant across all event windows ($p < 0.001$), underscoring the robustness of the effect. Excluding 2021 strengthens the observed impact: by day +20, the median CAR is −2.64%, compared to −1.31% in the full sample. This suggests that the 2021 spike–dominated by Marcum and WithumSmith+Brown related disclosures–was perceived as less material by investors.

Importantly, the dispersion in CARs widens over time. While some firms face sharp declines (as much as −90%), others experience gains exceeding +320%, pointing to a highly heterogeneous market response. Positive outliers likely reflect disclosures involving upward adjustments or effective corrective messaging that reassures investors.

The pattern of returns also supports the notion of gradual information diffusion. Rather than fully pricing in the implications on day +1, investors appear to adjust their valuations over time, consistent with prior studies on restatement-related disclosures (Gleason et al., 2008; He et al., 2018; Myers et al., 2011).

Further insights come from our quintile-based analysis (Figure 5.5; Tables 5.9). Disclosures in the bottom three quintiles generate negative price reactions, while those in the top two yield positive returns. The contrast is stark: the top quintile posts a median 20-day CAR of +13.4%, while the bottom quintile sees a decline of −21.4%. This variation reinforces the idea that market responses are shaped not just by the existence of an error, but by its nature, framing, and surrounding context.

Disclosures stating that the identified issue is material are nearly universal across all quintiles, ranging from 78.3% in Group A to 81.6% in Group E (Table 5.10). Similarly, 97% of disclosures indicate that a restatement of prior financial statements will be required. These minimal differences suggest that while these attributes are essential for compliance, they do little to differentiate investor responses. Instead, factors related to uncertainty and context appear far more influential.

CAR:	$t_0 + 1$ Day	+2	+3	+4	+5	+10	+20
Statistics:							
Samples	2,398	2,395	2,397	2,397	2,396	2,378	2,341
Mean	−0.87%	−1.05%	−1.12%	−1.12%	−1.15%	−1.17%	−1.54%
Median	−0.26%***	−0.38%***	−0.34%***	−0.38%***	−0.46%***	−0.71%***	−1.31%***
Std. Dev.	10.93	11.49	12.52	13.77	14.06	17.34	23.89
Min	−81.45%	−71.77%	−78.38%	−78.38%	−93.13%	−94.19%	−68.31%
Max	320.62%	335.13%	335.13%	333.76%	341.79%	363.04%	631.01%

Table 5.5: Statistics of Cumulative Abnormal Returns (CAR) for 1 to 20 Business Days Post Item 4.02 Disclosures Between 2007 and 2023.

*Significance levels are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).*

CAR:	$t_0 + 1$ Day	+2	+3	+4	+5	+10	+20
Statistics:							
Samples	1,712	1,709	1,712	1,712	1,712	1,694	1,664
Mean	−1.10%	−1.34%	−1.45%	−1.41%	−1.54%	−1.58%	−2.00%
Median	−0.50%***	−0.73%***	−0.78%***	−0.81%***	−1.07%***	−1.40%***	−2.64%***
Std. Dev.	12.82	13.39	14.58	16.01	16.23	19.69	27.36
Min	−81.45%	−71.77%	−78.38%	−78.38%	−93.13%	−94.19%	−68.31%
Max	320.62%	335.13%	335.13%	333.76%	341.79%	363.04%	631.01%

Table 5.6: Statistics of Cumulative Abnormal Returns (CAR) for 1 to 20 Business Days Post Item 4.02 Disclosures Between 2007 and 2023, Excluding 2021.

*Significance levels are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).*

Year	Samples	Mean	Median	Std. Dev.	Min	25th	75th	Max
2007	135	−2.65%	−2.32%**	10.07	−41.88%	−7.64%	2.25%	24.29%
2008	176	−1.37%	−1.21%	18.35	−54.58%	−12.55%	8.09%	57.60%
2009	140	2.13%	−4.76%*	45.28	−61.87%	−16.42%	5.20%	353.09%
2010	120	0.02%	−1.70%	22.37	−60.23%	−11.01%	5.97%	110.61%
2011	96	−4.15%	−5.12%***	19.71	−52.69%	−11.42%	1.80%	118.12%
2012	98	−2.85%	−1.80%*	13.89	−37.42%	−11.02%	4.28%	34.93%
2013	109	−1.32%	−2.57%	18.21	−47.25%	−7.90%	4.60%	98.84%
2014	68	−4.22%	−2.47%**	12.61	−38.16%	−9.53%	1.67%	26.31%
2015	81	−2.77%	−3.64%*	21.84	−54.93%	−13.24%	4.63%	109.39%
2016	61	0.43%	0.39%	18.23	−41.66%	−7.39%	8.53%	59.67%
2017	49	13.35%	−1.70%	92.93	−43.21%	−8.62%	6.68%	631.01%
2018	53	−7.82%	−6.26%***	18.7	−53.52%	−16.24%	0.27%	53.23%
2019	47	−1.34%	−2.79%	22.46	−53.84%	−13.16%	8.51%	58.85%
2020	48	−2.53%	−4.24%	26.75	−52.00%	−19.80%	5.76%	90.19%
2021	677	−0.40%	−0.57%***	11.49	−45.26%	−2.73%	1.13%	112.30%
2022	224	−3.70%	−0.79%*	18.7	−68.31%	−8.80%	4.87%	144.06%
2023	159	−5.97%	−3.05%***	22.42	−59.96%	−17.52%	1.25%	108.10%

Table 5.7: Statistics of Cumulative Abnormal Returns (CARs) 20 Days Post Item 4.02 Disclosures Per Year From 2007 to 2023.

*Significance levels are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).*

Uncertainty surrounding the financial impact of the error significantly amplifies negative market reactions. Disclosures indicating that the impact is "yet to be determined" are associated with a mean CAR of −3.19%, compared to −1.05% when the impact is known (Table 5.8). This uncertainty is present in 33.7% of Group A disclosures but only 25.0% in Group E, a statistically significant gap ($p < 0.01$) (Table 5.10). OLS regressions further confirm this effect, with consistently negative and significant coefficients ranging from −1.8 to −4.0 percentage points (Table 5.12).

Disclosures bundled with other material events also trigger stronger negative reactions. Filings accompanied by additional 8-K items, such as executive departures or impairments, show a mean CAR of −2.76%, compared to −0.66% when disclosed alone (Table 5.8). These bundled disclosures appear in 58.6% of Group A cases but only 47.6% in Group E ($p < 0.001$) (Table 5.10). Regression models reinforce this finding, with significant negative coefficients between −5.5 and −6.4 percentage points (Table 5.12).

Revenue recognition errors are particularly damaging. Firms disclosing a need to decrease previously reported revenue face mean CARs of −5.47% (Table 5.8). Similarly, when net income must be adjusted downward, mean CARs drop to −2.55%, versus −1.23% when no adjustment is required.

Firm size also matters. While Group A firms have a lower median market cap ($195M) than those in Group E ($328M) (Table 5.11), regression results offer clearer evidence: market capitalization is significantly negatively associated with 5-day CARs in five of six models, with coefficients from −2.4 to −4.4 percentage points (Table 5.12). This suggests that larger firms face steeper penalties, possibly due to heightened expectations of reporting reliability.

Timing plays a role as well. Disclosures released during trading hours are more frequent in Group E (11.5%) than Group A (7.5%), and Monday disclosures–often viewed as opportunistic–are more common in Group A (25.6%) than Group E (20.1%). By contrast, Tuesday disclosures are more prevalent in Group E (23.3%) than in Group A (17.3%) (Table 5.10). These patterns may reflect strategic timing or investor sensitivity to perceived disclosure transparency.

The "Reported with earnings" variable presents a more complex picture. Descriptively, these filings yield a mean CAR of −2.49%, compared to −0.96% when not reported with earnings (Table 5.8). They also appear more frequently in Group A (Table 5.10). However, regression analysis suggests the opposite: the variable is significantly positive in five models, with coefficients ranging from +4.2% to +6.6% (Table 5.12), indicating that concurrent earnings releases may help offset investor concern by providing broader financial context.

Interestingly, SEC involvement does not lead to more negative reactions. Disclosures identified by the SEC are slightly more common in Group E (12.8%) than in Group A (10.7%) (Table 5.10). The regression results support this, with significantly positive coefficients in four of six models (Table 5.12), suggesting that regulatory oversight may reduce uncertainty and reassure investors.

Auditor reputation also shows limited influence. Mean CARs are nearly identical between firms audited by Big 4 firms (−1.51%) and those without (−1.55%) (Table 5.7). A more granular look reveals variation: disclosures involving EY, PwC, Grant Thornton, and "Other" auditors yield significantly negative returns (−2.55% to −5.37%), while Marcum and KPMG are associated with neutral or positive returns (Table 5.4). These results imply that audit firm affiliation may shape investor perceptions, though not uniformly.

Overall, the empirical findings align with signaling theory (Akerlof, 1970; Ross, 1977): non-reliance disclosures act as negative signals that challenge prior financial credibility. Investors interpret these signals as indicators of governance quality, internal control weaknesses, and potential future risk. Market responses reflect not just the existence of an error but what that error signals about the firm's underlying financial integrity.

While our study focuses on short-term cumulative abnormal returns (1–20 trading days) to isolate immediate investor reactions to Item 4.02 disclosures, it does not capture longer-term consequences. Future research could explore reputational effects, litigation risk, and changes in financing costs. Additionally, broader firm responses–such as SEC enforcement actions (e.g., AAERs), executive turnover, shifts in auditor opinion language, and institutional ownership behavior–remain outside the scope of this analysis but offer promising directions for deeper investigation.

We also do not examine the tone or linguistic framing of the disclosures, though these factors likely influence how investors interpret and respond to such filings. Applying natural language processing (NLP) techniques could help quantify sentiment, specificity, and perceived uncertainty. Another important avenue for future research is identifying firms that restate financials without a prior Item 4.02 disclosure, which could reveal potential gaps in compliance and regulatory enforcement.

In summary, our findings show that investor reactions to non-reliance disclosures are most strongly shaped by uncertainty around financial impact, the presence of additional material events, and specific firm characteristics. In contrast, SEC involvement does not appear to amplify negative reactions and may instead reduce perceived risk. The near-universal presence of disclosures stating that the issue is material or will result in a formal restatement offers little explanatory power in distinguishing investor responses. Likewise, Big 4 auditor affiliation offers limited protection against adverse market responses.

Variable	Value	Samples	Pct.	Mean	Std. Dev.	25th	Median	75th
Impact is Material	True	1,865	79.7	−1.64%***	19.78	−8.12%	−1.16%	2.98%
	False	476	20.3	−1.15%***	35.72	−9.13%	−1.85%	3.06%
Restatement Necessary	True	2,298	98.2	−1.56%***	23.99	−8.34%	−1.31%	2.86%
	False	43	1.8	−0.20%	18.31	−11.92%	−1.00%	7.27%
Impact yet to be determined	True	531	22.7	−3.19%***	24.16	−13.62%	−3.37%	3.28%
	False	1,810	77.3	−1.05%***	23.80	−6.84%	−0.95%	2.97%
Material Weakness Identified	True	1,057	45.2	−2.06%***	26.18	−7.62%	−0.76%	2.64%
	False	1,284	54.8	−1.11%***	21.83	−8.70%	−1.74%	3.34%
Reported with Other Items	True	976	41.7	−2.76%***	28.88	−13.27%	−3.14%	3.93%
	False	1,365	58.3	−0.66%***	19.53	−5.18%	−0.69%	2.60%
Reported with Earnings	True	886	37.8	−2.49%***	29.51	−12.76%	−3.04%	3.87%
	False	1,455	62.2	−0.96%***	19.65	−5.44%	−0.76%	2.67%
Net Income Decreased	True	542	23.2	−2.55%***	18.70	−11.14%	−2.90%	5.05%
	False	1,799	76.8	−1.23%***	25.24	−7.36%	−1.02%	2.54%
Net Income Increased	True	213	9.1	−1.79%***	18.99	−10.42%	−0.96%	6.14%
	False	2,128	90.9	−1.51%***	24.33	−8.11%	−1.32%	2.96%
Revenue Decreased	True	166	7.1	−5.47%***	19.51	−16.31%	−4.74%	4.55%
	False	2,175	92.9	−1.24%***	24.17	−7.90%	−1.19%	2.77%
Revenue Increased	True	64	2.7	0.36%	22.29	−8.78%	−0.51%	4.62%
	False	2,277	97.3	−1.59%***	23.94	−8.35%	−1.32%	2.92%
Has Big 4 Auditor	True	685	29.3	−1.51%***	21.43	−10.26%	−1.90%	5.51%
	False	1,656	70.7	−1.55%***	24.84	−7.64%	−1.17%	2.09%
Revenue Adj. contains ‘million’	True	71	3.0	−4.40%**	23.84	−14.72%	−5.40%	2.95%
	False	2,270	97.0	−1.45%***	23.89	−8.10%	−1.21%	2.95%
Net Income Adj. contains ‘million’	True	177	7.6	−2.63%***	15.81	−10.95%	−3.12%	3.31%
	False	2,164	92.4	−1.45%***	24.43	−8.11%	−1.19%	2.95%
Identified by Company	True	2,150	91.8	−1.84%***	20.15	−8.14%	−1.24%	2.86%
	False	191	8.2	−1.88%**	20.49	−9.87%	−1.26%	2.98%
Identified by Auditor	True	548	23.4	0.18%	32.24	−7.44%	−0.91%	4.01%
	False	1,793	76.6	−2.06%***	20.57	−8.54%	−1.45%	2.58%
Identified by SEC	True	427	18.2	1.08%	25.17	−8.19%	−1.32%	3.99%
	False	1,914	81.8	−1.90%***	24.87	−9.53%	−1.33%	3.59%

Table 5.8: Statistics of Cumulative Abnormal Returns 20 Days Post Item 4.02 Disclosure by Content Feature (2007–2023).

*Significance levels are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).*

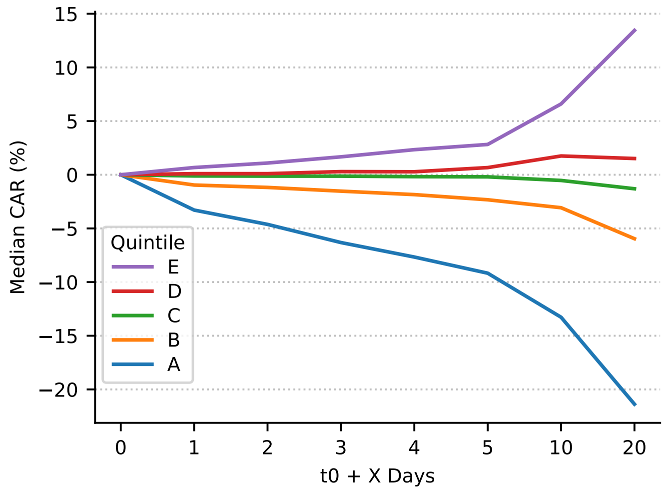


Figure 5.5: Median Cumulative Abnormal Returns (CARs) Over 20 Days Post-Disclosure, Segmented by Market Reaction Quintile for Item 4.02 Disclosures (2007–2023)

Presented is the evolution of median CARs across different quintiles, where Group A (− − −) represents the most negative market reactions and Group E (++) represents the most positive reactions.

Quintile		Samples	Mean	Std. Dev.	Min	25th	Median	75th	Max
A	(− − −)	469	−24.6%	11.4	−68.3%	−30.5%	−21.4%	−15.7%	−11.2%
B	(−−)	468	−6.5%	2.3	−11.2%	−8.3%	−6.0%	−4.6%	−3.3%
C	(−)	468	−1.4%	1.0	−3.3%	−2.3%	−1.3%	−0.6%	0.0%
D	(+)	468	1.9%	1.4	0.1%	0.7%	1.5%	3.0%	5.0%
E	(++)	468	23.0%	39.3	5.0%	8.1%	13.4%	25.3%	631.0%

Table 5.9: Summary Statistics of 20-Day Cumulative Abnormal Returns by Market Reaction Quintile for Item 4.02 Disclosures (2007–2023).

	A (− − −)	B (−−)	C (−)	D (+)	E (++)	Diff(E, A)
Impact is Material	78.3	79.3	79.9	79.3	81.6	−4.0
Impact yet to be determined	33.7	23.5	16.0	15.2	25.0	34.8**
Revenue Decreased	11.9	6.8	3.0	4.9	8.8	35.2
Revenue Increased	3.0	2.4	1.7	3.4	3.2	−6.3
Net Income Increased	11.1	9.2	5.3	7.5	12.4	−10.5
Net Income Decreased	28.6	27.8	14.3	15.4	29.7	−3.7
Restatement Necessary	97.7	98.1	99.6	98.3	97.2	0.5
Material Weakness Identified	45.4	43.6	39.1	55.1	42.5	6.8
Identified by SEC	10.7	16.7	33.3	17.7	12.8	−16.4
Identified by Auditor	21.7	20.7	24.8	23.5	26.3	−17.5
Identified by Company	90.6	90.8	93.4	93.8	90.6	0.0
Has Big 4 Auditor	33.5	31.2	19.9	22.9	38.9	−13.9
Revenue Adj. contains 'million'	4.9	3.8	1.7	1.5	3.2	53.1
Net Income Adj. contains 'million'	8.7	10.0	5.1	5.6	8.3	4.8
Reported with Other Items	58.6	44.7	29.9	27.6	47.6	23.1***
Reported with Earnings	52.2	41.0	27.6	25.4	42.9	21.7**
Published: Pre-Market (6:00 am – 9:30 am)	21.1	16.2	13.5	13.0	19.2	9.9
Published: Regular Hours (9:30 am – 4:00 pm)	7.5	10.0	11.5	10.0	11.5	−34.8*
Published: After Hours (4:00 pm – 10:00 pm)	71.4	73.7	75.0	76.9	69.2	3.2
Published: Monday	25.6	20.1	23.9	23.1	20.1	27.4*
Published: Tuesday	17.3	19.0	16.9	22.9	23.3	−25.8*
Published: Wednesday	19.6	17.7	15.6	16.7	15.8	24.1
Published: Thursday	19.0	22.9	20.9	15.0	21.4	−11.2
Published: Friday	18.6	20.3	22.6	22.4	19.4	−4.1

Table 5.10: Proportions of Item 4.02 Disclosures (in %) That Include Specific Content Features, Segmented Into Quintile Groups Based on the Cumulative Abnormal Return (CAR) Impact.

*Group A (− − −) represents disclosures with the strongest negative market reaction, while Group E (++) consists of disclosures associated with the strongest positive stock price movements. Diff(E, A), indicates the percentage difference between the most positive (E) and most negative (A) quintiles. Significance levels are marked with * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$).*

	A (− − −)		B (−−)		C (−)		D (+)		E (++)	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Income Statement										
Revenue	365.4	109.5	784.8	70.8	434.0	132.3	938.7	115.2	408.5	128.2
Cost of Goods Sold	280.3	54.9	463.2	45.6	290.1	77.5	710.7	76.0	229.9	57.2
R&D	14.1	7.4	35.8	2.4	35.8	9.8	17.5	3.5	28.6	9.0
Operating Income After Depreciation	−6.3	−4.3	85.4	−0.2	14.5	2.1	53.6	5.7	15.0	0.8
Net Income	−29.0	−6.3	43.7	−0.1	5.3	−0.8	20.5	3.1	−58.6	−3.7
Balance Sheet										
Cash & Equivalents	74.6	37.8	166.0	19.1	79.7	34.6	92.2	35.6	107.3	32.9
Current Assets	211.3	116.0	542.3	74.7	350.5	113.9	350.3	155.9	347.0	145.8
Total Assets	438.5	190.6	1,337.8	162.6	558.8	205.1	972.9	380.2	986.9	225.4
Current Liabilities	114.0	51.0	287.1	38.5	162.2	46.8	229.8	53.1	185.1	43.6
Long-Term Debt	94.7	12.4	345.9	7.9	73.6	9.9	359.2	11.4	310.8	12.9
Total Liabilities	236.5	80.4	732.6	82.8	262.8	90.2	691.5	121.6	571.3	81.4
Cash Flow Statement										
Free Cash Flow	−15.9	−7.1	50.2	−1.7	9.1	2.2	32.7	3.0	8.8	−0.4
Operating Activities – Net Cash Flow	−0.9	−1.2	84.0	−0.4	22.9	6.6	65.2	6.5	25.3	3.8
Financial Ratios										
Price to Earnings	−257.3	−2.5	36.3	−0.3	−1.8	−2.3	27.0	8.8	−0.9	−5.0
Debt to Equity	−1.5	0.5	−0.1	0.8	25.2	0.7	2.9	1.0	0.0	0.8
Others										
Market Cap	616.6	195.6	1,728.1	280.4	814.0	180.5	702.8	305.7	907.2	328.7
Years Since IPO	14.1	15.0	17.0	17.0	13.3	13.0	12.8	13.5	14.5	13.0

Table 5.11: Summary Statistics of Financial Statement Metrics and Ratios for Companies Disclosing Item 4.02, Segmented by Market Reaction Quintiles.

The table presents the mean and median values of key financial indicators, including income statement, balance sheet, cash flow statement, and financial ratios, for firms in each quintile. Income statement figures represent annual data, with all financial values (e.g., revenue, net income, assets) reported in millions of USD.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	2.060	−2.640	3.333	2.708	3.290	−2.760
	(0.92)	(−0.21)	(0.23)	(0.91)	(1.32)	(−0.42)
Impact is Material	0.407	−0.347	0.679	0.510	0.230	0.679
	(0.53)	(−0.26)	(0.83)	(0.62)	(0.29)	(0.43)
Impact yet to be determined	−2.043***	−4.038***	−1.821**	−1.943**	−1.922***	−3.470**
	(−2.76)	(−2.98)	(−2.33)	(−2.49)	(−2.58)	(−2.17)
Revenue Decreased	−1.178	−0.061	−0.757	−1.165	−0.944	0.009
	(−0.86)	(−0.03)	(−0.51)	(−0.81)	(−0.69)	(0.00)
Revenue Increased	−1.360	−1.970	−1.735	−1.489	−1.114	−0.847
	(−0.72)	(−0.62)	(−0.87)	(−0.76)	(−0.59)	(−0.23)
Net Income Increased	1.364	0.232	1.546	1.338	1.545	0.428
	(1.27)	(0.11)	(1.34)	(1.19)	(1.42)	(0.18)
Net Income Decreased	−0.478	1.428	−0.114	−0.234	−0.513	1.316
	(−0.61)	(0.96)	(−0.13)	(−0.28)	(−0.64)	(0.73)
Restatement Necessary	−2.861	−1.744	−2.901	−3.345	−2.732	−0.547
	(−1.22)	(−0.51)	(−1.14)	(−1.35)	(−1.16)	(−0.12)
Material Weakness Identified	−0.289	−0.812	−0.317	−0.553	0.249	−1.299
	(−0.47)	(−0.69)	(−0.48)	(−0.84)	(0.39)	(−0.93)
Identified by SEC	1.860**	0.546	1.849**	1.663*	1.500*	−0.587
	(2.30)	(0.38)	(2.18)	(1.90)	(1.69)	(−0.32)
Identified by Auditor	0.740	2.330**	1.043	0.778	0.931	2.170
	(1.09)	(2.06)	(1.44)	(1.07)	(1.35)	(1.59)
Identified by Company	−0.548	−0.758	−0.822	−0.677	−0.638	−0.592
	(−0.47)	(−0.37)	(−0.64)	(−0.55)	(−0.54)	(−0.24)
Has Big 4 Auditor	1.297*	−1.025	2.687	1.185	1.042	−11.242
	(1.84)	(−0.51)	(0.90)	(1.58)	(1.44)	(−1.26)
Revenue Adj. contains 'million'	−1.041	−4.290	−0.004	−1.495	−1.123	−3.416
	(−0.52)	(−1.24)	(0.00)	(−0.71)	(−0.56)	(−0.87)
Net Income Adj. contains 'million'	0.096	−0.846	0.627	−0.212	0.298	2.467
	(0.08)	(−0.37)	(0.49)	(−0.17)	(0.25)	(0.90)
Reported with Other Material Events	−5.593***	−4.496	−6.399***	−5.655***	−5.500***	−5.809*
	(−3.81)	(−1.62)	(−4.13)	(−3.64)	(−3.72)	(−1.76)
Reported with Earnings	4.155***	4.436	4.943***	4.349***	4.166***	6.625*
	(2.81)	(1.57)	(3.18)	(2.79)	(2.80)	(1.96)
Affected Years	0.095	0.383	0.084	0.090	0.052	0.027
	(0.43)	(0.88)	(0.36)	(0.38)	(0.22)	(0.05)
Market Cap (log)	−2.626***	−4.424*	−2.821***	−3.785***	−2.365***	−5.274
	(−3.17)	(−1.69)	(−3.16)	(−3.97)	(−2.83)	(−1.56)
Revenue (log)	0.298	0.388	0.152	−0.086	0.246	−1.869
	(1.20)	(0.33)	(0.56)	(−0.30)	(0.98)	(−1.17)
Operating Income After Depreciation (log)	0.516	0.512	0.553	0.561	0.606*	0.528
	(1.45)	(0.54)	(1.46)	(1.51)	(1.69)	(0.46)
Net Income (log)	−1.762**	−0.862	−2.294***	−1.754**	−1.533*	−1.141
	(−2.19)	(−0.53)	(−2.73)	(−2.10)	(−1.90)	(−0.60)
Goodwill (log)	−0.156	0.916	−0.136	−0.065	−0.098	1.481
	(−0.67)	(1.11)	(−0.56)	(−0.26)	(−0.42)	(1.47)
Operating Activities - Net Cash Flow (log)	−1.247***	0.591	−1.108**	−1.128**	−1.269***	0.684
	(−2.62)	(0.53)	(−2.25)	(−2.22)	(−2.66)	(0.50)
Free Cash Flow (log)	1.076**	−0.504	0.848*	0.976**	0.986**	−1.490
	(2.41)	(−0.50)	(1.85)	(2.04)	(2.20)	(−1.18)
Total Debt (log)	1.035	−3.365	1.656	0.916	0.939	−2.209
	(0.75)	(−0.91)	(1.07)	(0.63)	(0.67)	(−0.49)
Total Assets (log)	0.746	1.876	0.824	1.520**	0.692	3.296
	(1.06)	(0.97)	(1.06)	(1.98)	(0.96)	(1.34)
Firm FE	No	Yes	No	No	No	Yes
Auditor FE	No	No	Yes	No	No	Yes
Sector FE	No	No	No	Yes	No	Yes
Year FE	No	No	No	No	Yes	Yes
R-squared	0.033	0.804	0.105	0.051	0.045	0.835
Adjusted R-squared	0.022	0.292	0.022	0.036	0.028	0.298
Number of observations	2,396	2,396	2,396	2,229	2,396	2,229

Table 5.12: OLS Regression Results for 5-Day Cumulative Abnormal Returns (CAR) Following Form 8-K Item 4.02 Disclosures from 2007 to 2023, with CAR as the Dependent Variable.

*Statistical significance is denoted by * ($p < 0.1$), ** ($p < 0.05$), *** ($p < 0.01$).*

5.4 Conclusion

We provide new insights into the market impact of Item 4.02 non-reliance disclosures, analyzing 8,006 filings from 2004 to 2023. Our findings confirm that these disclosures act as robust early risk signals, often triggering significant negative stock price reactions, with cumulative abnormal returns (CARs) ranging from -2.6% to -5.4%. Investors react most severely to revenue recognition errors, which lead to 114% stronger negative returns than net income misstatements. Furthermore, disclosures explicitly stating that the financial impact is yet to be determined experience three times harsher market reactions than those where the full impact is specified, reinforcing the importance of transparency and clarity in financial reporting.

The results carry important implications for stakeholders and financial professionals. Market participants should closely monitor Item 4.02 disclosures as early indicators of financial restatements which follow 97% of such disclosures, incorporating them into risk models and investment decision frameworks. Given that SEC-identified issues do not exacerbate price reactions, investors should focus more on the nature of the reported error rather than regulatory involvement. Additionally, the declining role of Big 4 auditors in these disclosures (from 61% in 2004–2006 to 12% in 2021–2023) suggests that auditor reputation alone does not mitigate financial reporting risks, making due diligence on non-Big 4 auditors increasingly critical. We find that March and November consistently experience the highest volume of Item 4.02 disclosures, aligning with key reporting cycles of audited annual reports (Form 10-K) in February and March and year-end quarterly reports (Form 10-Q) in November.

Regulators, such as the SEC, can view the results as evidence that Item 4.02 disclosure requirements enhance market efficiency by providing timely and material information to stakeholders. Corporate executives should recognize the potential stock price impact of non-reliance disclosures and prioritize clear, immediate communication of financial errors to minimize uncertainty-driven volatility.

Future research could examine insider trading, institutional ownership shifts, and enforcement actions following Item 4.02 disclosures. Reputational effects, litigation risk, and financing costs also warrant exploration. Additionally, analyzing disclosure tone using NLP and identifying restatements without prior 4.02 filings could uncover important patterns in investor response and regulatory compliance.

Bibliography

Akerlof, G. A. (1970). The market for “Lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Anderson, K. L. and Yohn, T. L. (2002). The effect of 10K restatements on firm value, information asymmetries, and investors’ reliance on earnings. *SSRN Electronic Journal*.

Dechow, P. M., Sloan, R. G., and Sweeney, A. P. (1996). Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research*, 13(1):1–36.

Feldman, R., Livnat, J., and Segal, B. (2008). Shorting companies that restate previously issued financial statements. *The Journal of Investing*, 17(3):6–15.

Files, R., Swanson, E. P., and Tse, S. (2009). Stealth disclosure of accounting restatements. *The Accounting Review*, 84(5):1495–1520.

Gleason, C. A., Jenkins, N. T., and Johnson, W. B. (2008). The contagion effects of accounting restatements. *The Accounting Review*, 83(1):83–110.

He, L., Sarath, B., and Wans, N. (2018). Material weakness disclosures and restatement announcements: The joint and order effects. *Journal of Business Finance & Accounting*, 46(1–2):68–104.

Kinney, W. R. and McDaniel, L. S. (1989). Characteristics of firms correcting previously reported quarterly earnings. *Journal of Accounting and Economics*, 11(1):71–93.

Kravet, T. and Shevlin, T. (2009). Accounting restatements and information risk. *Review of Accounting Studies*, 15(2):264–294.

Myers, L. A., Scholz, S., and Sharp, N. Y. (2011). Restating under the radar? Determinants of restatement disclosure choices and the related market reactions. *SSRN Electronic Journal*.

Palmrose, Z.-V., Richardson, V. J., and Scholz, S. (2004). Determinants of market reactions to restatement announcements. *Journal of Accounting and Economics*, 37(1):59–89.

Robbani, M. G., Anantharaman, S., Bhuyan, R., and University-Sacramento, C. S. (2006). Financial Restatements and Their Impact on Stock Prices: Evidence from the US Financial Markets. Working paper, Cameron University.

Ross, S. A. (1977). The Determination of Financial Structure: The Incentive-Signalling Approach. *The Bell Journal of Economics*, 8(1):23–40.

Scholz, S. (2008). The changing nature and consequences of public company financial restatements 1997–2006. Technical report, University of Kansas School of Business, & The Department of the Treasury.

Schroeder, J. (2025). Effects of Non-Reliance Disclosures in Form 8-K Filings on Stock Prices. Working Paper.

Umar, M., Mirza, N., and Ribeiro-Navarrete, S. (2023). The impact of financial restatements on sell-side recommendation accuracy. *Finance Research Letters*, 55:103868.

Chapter 6: Following Insiders to Outperform the Market

Authors: Jan L. Schroeder (jan@sec-api.io), Alexander Krause (alexanderelias.krause@tu-dortmund.de)*

Abstract
We present a novel framework for developing and backtesting a profitable quantitative investment strategy based on insider purchases disclosed in SEC Form 4 filings. Our approach involves constructing long-equity portfolios by systematically selecting positions using filter criteria identified through Monte Carlo simulations. To evaluate the effectiveness of insider trading as a position selection signal, we analyze 1.8 million insider transactions reported between 2011 and 2023. Monte Carlo processes are applied across 1,685 features to generate and test a wide range of selection criteria. From this process, we identify a top-performing strategy that achieves a 5-day cumulative abnormal return (CAR) of 3.4%. Backtesting the resulting portfolios by using a rule-based investment algorithm over the 13-year period reveals consistent outperformance relative to the S&P 500 index across 11 out of 14 performance metrics. Notably, the best-performing strategy delivers a Sharpe ratio of 1.88, a Sortino ratio of 2.92, and an annualized excess return of 41.4%, after accounting for transaction costs.

Publication Details: Accepted for Publication in the Journal of Investing.

Keywords: Quantitative Portfolio Construction, Monte Carlo Simulations, Insider Trading, SEC Form 4 Filings

Acknowledgements: We gratefully acknowledge the feedback received during the 51st EBES Conference (April 11–13, 2025, Rome, Italy). We thank all session participants for their insightful comments, particularly Hasan Fehmi Baklaci (Yaşar University, Türkiye), Szabolcs Blazsek (Mercer University, U.S.A.), Wei Gao (Widener University, U.S.A.), and Stefano Gubellini (San Diego State University, U.S.A.). We also presented this research at the 14th International Conference of the Financial Engineering and Banking Society (FEBS), held in France from June 11–13, 2025, during the "Stock Market Investments" session chaired by George Chalamandaris. We thank all session participants for their valuable feedback and suggestions.

The following is based on Schroeder and Krause (2025).

*Jan L. Schroeder is the founder and CEO of SEC-API.io and Research Fellow at the Department of Finance, TU Dortmund. The authors have no competing interests to declare. All errors are our own. Corresponding author: Jan L. Schroeder.

Bibliography

Aboody, D. and Lev, B. (2000). Information asymmetry, R&D, and insider gains. *The Journal of Finance*, 55(6):2747–2766.

Alldredge, D. M. and Cicero, D. C. (2014). Attentive insider trading. *Journal of Financial Economics*, 115(1):84–101.

Bajo, E. and Petracci, B. (2006). Do what insiders do: Abnormal performances after the release of insiders' relevant transactions. *Studies in Economics and Finance*, 23(2):94–118.

Biesta, M. A., Doeswik, R. Q., and Donker, H. A. (2003). The Profitability of Insider Trades in the Dutch Stock Market. Available at SSRN: http://ssrn.com/abstract=498042.

Biggerstaff, L., Cicero, D., and Wintoki, M. B. (2020). Insider trading patterns. *Journal of Corporate Finance*, 64:101654.

Brochet, F. (2010). Information Content of Insider Trades before and after the Sarbanes-Oxley Act. *The Accounting Review*, 85(2):419–446.

Cheuk, M., Fan, D. K., and So, R. W. (2005). Insider trading in Hong Kong: Some stylized facts. *Pacific-Basin Finance Journal*, 14(1):73–90.

Cohen, L., Malloy, C., and Pomorski, L. (2012). Decoding inside information. *The Journal of Finance*, 67(3):1009–1043.

Collin-Dufresne, P. and Fos, V. (2016). Insider trading, stochastic liquidity, and equilibrium prices. *Econometrica*, 84(4):1441–1475.

Covel, M. W. (2017). *Trend Following: How to Make a Fortune in Bull, Bear, and Black Swan Markets.* Wiley, Hoboken, NJ, 5 edition.

Del Brio, E. B., Miguel, A., and Perote, J. (2002). An investigation of insider trading profits in the Spanish stock market. *The Quarterly Review of Economics and Finance*, 42(1):73–94.

DeVault, L., Cederburg, S., and Wang, K. (2022). Is "Not Trading" Informative? Evidence from Corporate Insiders' Portfolios. *Financial Analysts Journal*, 78(1):79–100.

Dymke, B. M. and Walter, A. (2008). Insider trading in Germany – Do corporate insiders exploit inside information? *BuR – Business Research*, 1(2):188–205.

Etebari, A., Tourani-Rad, A., and Gilbert, A. (2004). Disclosure regulation and the profitability of insider trading: Evidence from New Zealand. *Pacific-Basin Finance Journal*, 12(5):479–502.

Friederich, S., Gregory, A., Matatko, J., and Tonks, I. (2002). Short-run Returns around the Trades of Corporate Insiders on the London Stock Exchange. *European Financial Management*, 8(1):7–30.

Gao, G., Ma, Q., Ng, D. T., and Wu, Y. (2021). The sound of silence: What do we know when insiders do not trade? *Management Science*.

Hsieh, J., Ng, L., and Wang, Q. (2023). How informative are insider trades and analyst recommendations? *Journal of Banking & Finance*, 149:106787.

Inci, A. C., Lu, B., and Seyhun, H. N. (2010). Intraday Behavior of Stock Prices and Trades around Insider Trading. *Financial Management*, 39(1):323–363.

Jaffe, J. F. (1974). Special Information and Insider Trading. *Journal of Business*, 47(3):410–428.

Jeng, L. A., Metrick, A., and Zeckhauser, R. (1999). The Profits to Insider Trading: A Performance-Evaluation Perspective. NBER Working Paper 6913, National Bureau of Economic Research, Cambridge, MA.

Jeng, L. A., Metrick, A., and Zeckhauser, R. (2003). Estimating the returns to insider trading: A performance-evaluation perspective. *The Review of Economics and Statistics*, 85(2):453–467.

Lei, Q., Rajan, M., and Wang, X. (2014). Can traders beat the market? Evidence from insider trades. *China Finance Review International*, 4(3):243–270.

Lorie, J. H. and Niederhoffer, V. (1968). Predictive and Statistical Properties of Insider Trading. *The Journal of Law & Economics*, 11(1):35–53.

Neupane, B., Thapa, C., Marshall, A., and Neupane, S. (2021). Mimicking insider trades. *Journal of Corporate Finance*, 68:101940.

O'Shaughnessy, J. P. (2012). *What Works on Wall Street: The Classic Guide to the Best-Performing Investment Strategies of All Time*. McGraw-Hill, New York, NY, 4 edition.

Schroeder, J. (2025). Anomalies, Trends and Patterns in Disclosure Activities: Understanding EDGAR. Working Paper.

Schroeder, J. and Krause, A. (2025). Following Insiders to Outperform the Market. *Journal of Investing*.

Seyhun, H. (1986). Insiders' profits, costs of trading, and market efficiency. *Journal of Financial Economics*, 16(2):189–212.

Siegel, J. J. and Schwartz, J. (2023). *Stocks for the Long Run: The Definitive Guide to Financial Market Returns & Long-Term Investment Strategies*. McGraw Hill, New York, NY, 6 edition.

Sundling, C. and Verschuur, C. (2014). How to yield abnormal return by replicating insider trades - A study on the Swedish stock market. mastersthesis, mastersthesis.

Veenman, D. (2012). Disclosures of insider purchases and the valuation implications of past earnings signals. *The Accounting Review*, 87(1):313–342.

Wang, W., Shin, Y., and Francis, B. B. (2012). Are CFOs' trades more informative than CEOs' trades? *Journal of Financial and Quantitative Analysis*, 47(4):743–762.

Wisniewski, T. P. and Bohl, M. T. (2005). The information content of registered insider trading under LAX law enforcement. *International Review of Law and Economics*, 25(2):169–185.

Zingg, A., Lang, S., and Wyttenbach, D. (2007). Insider trading in the Swiss stock market. *Swiss Journal of Economics and Statistics*, 143(3):331–362.