

Data Observer

Jan Philipp Harries*

A Dataset for Blockholders in US-Listed Firms

<https://doi.org/10.1515/jbnst-2021-0033>

Received August 23, 2021; accepted August 24, 2021

Abstract: I introduce a new dataset for research on institutional and individual blockholders in US-listed firms. As of June 2021, the dataset contains 758,666 parsed Form 13D and 13G filings from November 1993 to May 2021, sourced from SEC EDGAR. Due to the semi-structured nature of these filings, a comparable dataset was not available for research before.

Keywords: blockholdership, long-term ownership, SEC filings, dataset

JEL Classification: C81, G23, G32

1 Introduction

Regulatory guidelines in the US command that every investor needs to submit a filing, specifically a Form 13D or Form 13G filing, to the SEC when his stake in a publicly listed company exceeds 5%. These filings are published on the SEC Edgar website and are provided free of charge and publicly accessible. However, while single filings can easily be found and accessed via the EDGAR website, the SEC does not provide a database with aggregated information from these filings and the creation of a database by third parties is hindered by the many different shapes of these filings with a varying layout and wording, despite general template supplied by the SEC.

This paper is an abbreviated version of my earlier working paper “Determinants of Blockholdership – A new Dataset for Blockholder Analysis”. The dataset is available for download at <https://doi.org/10.7910/DVN/61Z64Q>.

*Corresponding author: Jan Philipp Harries, University of Wuppertal, Gaußstraße 20, 42119 Wuppertal, Germany, E-mail: harries@wiwi.uni-wuppertal.de. <https://orcid.org/0000-0001-5926-5888>

While the SEC demands many forms to be filed in a unified, machine-readable XML or XBRL format, this is not the case for Form 13D and Form 13G filings. This led to a situation where information from a variety of filings is easily available for researchers (e.g. Form 13F filings, which report holdings of institutional money managers or Form 3, 4, or 5 filings which report insider transactions) while the blockholder information contained in the Form 13D and 13G filings is very difficult to access and thus rarely used in research. Some commercial providers (e.g. Factset) offer proprietary databases claiming to contain information from Form 13D and Form 13G filings; however, samples taken from these databases indicate incomplete data. Dlugosz et al. (2006) for example stated that “*despite this important role, there is no standardized data set for these blocks, and the best available data source, Compact Disclosure, has many mistakes and biases*”, referring to a different commercial blockholder database.

With this paper, I introduce and release a new dataset for the information contained in Form 13D and Form 13G filings. Manually scanning hundreds of different filing formats, I developed a parser that is sufficiently accurate and robust to parse the relevant information out of hundreds of thousands of blockholder filings. As of June 2021, the resulting database contains 758,666 blockholder filings from November 1993 to May 2021, each with 76 fields containing various information, e.g. the reported ownership percentage or the addresses of the filing entity and the subject company.

The remainder of this paper is structured as follows: Section 2 gives a brief overview of the regulatory framework for blockholder filings. In Section 3, I introduce the dataset and provide descriptive statistics and figures. Section 4 contains information about data availability and I conclude with closing remarks in Section 5.¹

2 Regulatory Framework

The Securities Exchange Act of 1934 (SEA) contains the filing rules and legal definitions under which blockholders have to submit forms to the SEC. Relevant for blockholders is first and foremost section 13(d) of the SEA.²

Section 13(d)-1(a) of the SEA commands that “any person who, after acquiring directly or indirectly the beneficial ownership of any equity security of a class which is specified in paragraph (i) of this section, is directly or indirectly the beneficial owner of more than 5% of the class shall, within 10 days after the

¹ A description of all fields in the database can be found in the Supplementary Material.

² See also *43 FR 18,495* or *17 CFR 240.13d-1 ff.* in the US Code of Federal Regulations.

acquisition, file with the Commission, a statement containing the information required by Schedule 13D”. As noted by Dlugosz et al. (2006), “this rule has been interpreted to include shares that may be obtained through the exercising of options, warrants, or rights in the next 60 days a part of the beneficial ownership calculation”.³ These filings are known as Form 13D filing.

Some types of investors (broker-dealers, banks, insurance companies, investment companies, investment advisors, employee benefit plans, parent companies, savings associations, and churches) are allowed to file an abbreviated filing, the Form 13G, which waives some required fields (e.g. the source of funds). This is only possible if the investor has “has acquired such securities in the ordinary course of his business and not with the purpose nor with the effect of changing or influencing the control of the issuer, nor in connection with or as a participant in any transaction having such purpose or effect”. The exact conditions are laid out in section 13(d)-1(b) of the SEA and these forms were also parsed for the database.

In the case of a change in a previously reported ownership share in a company, the (former) blockholder has to file an amendment to the original filing. These filings are called Form 13D/A or Form 13G/A, respectively.

All forms are to be published on the SEC’s EDGAR platform.⁴ Besides Form 13D and G filings, other filings of interest for company ownership are for example the Forms 3, 4, and 5, which are used to disclose insiders’ transactions and Form DEF 14A proxy statements. Further information regarding other filings and forms can be found, e.g. at Meredith (2007).

3 Dataset and Descriptive Analysis

The database contains 758,666 unique blockholdership filings, thereof 205,063 Form 13G filings, 374,531 Form 13G/A filings, 57,122 Form 13D filings, and 121,950 Form 13D/A filings. Since 2008, the number of filings, especially Form 13 D filings, has declined slowly but steadily, as can be seen in Figure 1.

There are 28,246 unique CIK’s among the filing subjects and 59,628 unique CIK’s among the filers themselves in the database. The list of subjects with the most filings is dominated by several ETFs. *iShares ETFs* were most reported on with 2.206

³ For more information and interpretation of these rules, see also “Exchange Act Sections 13(d) and 13(g) and Regulation 13D-G Beneficial Ownership Reporting” at <https://www.sec.gov/divisions/corpfin/guidance/reg13d-interp.htm>.

⁴ The platform can be accessed at <https://www.sec.gov/edgar/searchedgar/companysearch.html>.

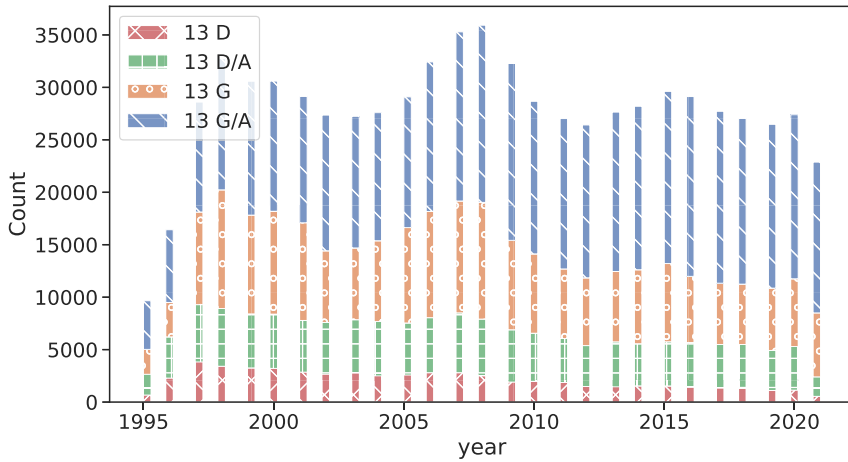


Figure 1: Filing count per year and filing type.

filings, followed by *WM Advisors* with 667 filings, and *IndexIQ ETF Trust* with 649 filings.

Among Filers, data is more concentrated. Leading the pack is *BlackRock Inc.* with 34,253 filings (which amounts to almost 5% of all filings) followed by *VANGUARD* with 21,059 filings and two different entities belonging to *Fidelity* which, if taken together, even top *BlackRock* with a combined total of 37,769 filings. All big filers report a median blocksize around 7 or 8%.

Geographically, most subject companies are located in New York, followed by Chicago, San Diego, and San Francisco. While New York also takes the top spot for the location of the filing entity (with a total of 161,266 filings over all Zip codes), some cities like Boston and Baltimore have significantly more local filers than subjects, probably indicating a bigger financial industry based in these locations.

Looking at the distribution of filing entities and subject companies, this picture is confirmed. The largest US states all fall in the highest quintile of filings. California is the state where most subject companies are based with a total of 109,550 filings. For filers, New York takes unsurprisingly the top spot followed by California (85,971 filings) and Massachusetts (81,645 filings). States that don't contain financial hubs like e.g. Florida or Colorado will be found more often as the location of the subject company than the location of filing entity.

The histogram of reported ownership (measured in % of outstanding shares) in Figure 2 shows that only very few filings report ownership stakes larger than 20%. The overwhelming majority of filings report a share of 10% or less of a company.

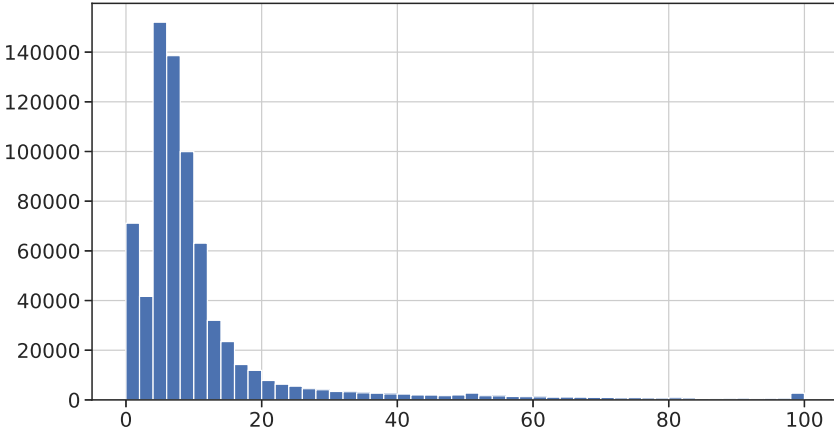


Figure 2: Histogram of reported ownership percentages.

As can be seen in Figure 3, reported ownership percentages have stayed remarkably similar over time. However, the 75% quantile of reported ownership has come down since the early 2000s from about 13 to 10% in 2021.

When filing a Form 13D or 13G with the SEC, investors have to self-report their investor category from a choice of presets. Investors from different categories exhibit different characteristics in their number of filings and the average reported block size, as can be seen in Table 1. Most filings are filed by either investment advisors or individuals with 211,711 and 131,474 total filings, respectively. Interestingly, the size of the average reported ownership percentages differs

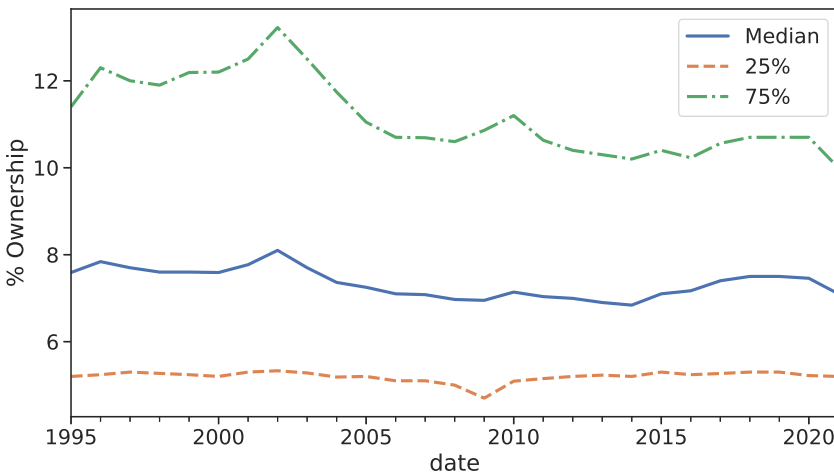


Figure 3: Development of reported ownership percentages.

significantly by investor type. While investment advisors have among the lowest average reported ownership percentage with 7.85%, individuals report average ownership that is two times as high with 15.65% which is only topped by corporations with an average of 20.56%.

Figure 4 shows an interesting trend in the data, which might be explained by the recent trend to passive investments, the increasing market capitalization of listed companies, and more concentrated ownership in general. While the number of filings from individuals and investment advisors was relatively similar until 2005, since then the development is very stable for investment advisors with about 8000 filings so far in 2021, while the number of filings from individuals declined sharply from about 7000 in 1998 to only 2000 in 2021.

For Form 13D and 13D/A filings, the filer has also to declare the source of funds, e.g. how he financed the purchase of the reported blockholding. Descriptive statistics for this category are given in Table 2. The average reported ownership size is the biggest for *Bank* as a funding source with an average of 36.01%. It's not immediately clear why the average size for reported bank-funded blocks is larger than for the other categories. One possible explanation could be that M&A transactions are often the reason for a large reported ownership share and these transactions are often (at least partially) financed by banks.

Among the other fund sources, blocks financed by *Affiliates*, *Personal Funds*, and *Working Capital* exhibit a below-average size. Categories with most filings are *Working Capital* and *Other*.

Table 1: Summary statistics: investor type.

Investor type	n	Mean	Std	Min	Q1	Median	Q3	Max
Bank	11,545	9.54	13.50	0.00	4.31	6.18	9.31	100.00
Broker	11,061	9.80	9.72	0.00	5.30	7.48	11.30	100.00
Church	13	15.83	13.66	0.00	5.80	11.12	19.99	49.23
Corporation	53,254	20.56	23.59	0.00	5.52	9.99	26.60	100.00
Employee benefit	7969	9.89	8.28	0.00	5.92	8.40	11.17	100.00
Holding company	95,178	9.56	12.17	0.00	5.17	6.76	9.99	100.00
Individual	131,474	15.65	17.95	0.00	5.70	9.00	17.60	100.00
Insurance	7145	10.61	13.98	0.00	4.96	6.88	10.80	100.00
Investment advisor	211,711	7.85	7.31	0.00	5.18	6.70	9.49	100.00
Investment company	17,359	8.98	8.29	0.00	5.50	7.42	10.46	100.00
Non-US institution	3807	6.84	9.62	0.00	3.30	4.74	7.20	100.00
Other	56,062	13.37	17.34	0.00	5.00	7.40	13.60	100.00
Partnership	66,248	11.03	13.94	0.00	4.40	6.90	11.30	100.00
Savings association	158	7.56	4.79	0.00	5.03	6.27	8.79	43.60

This table reports summary statistics for the ownership size (as % of floating shares) reported in Form 13D and 13G (/A) filings by self-reported investor category. Period: from November 1993 to May 2021.

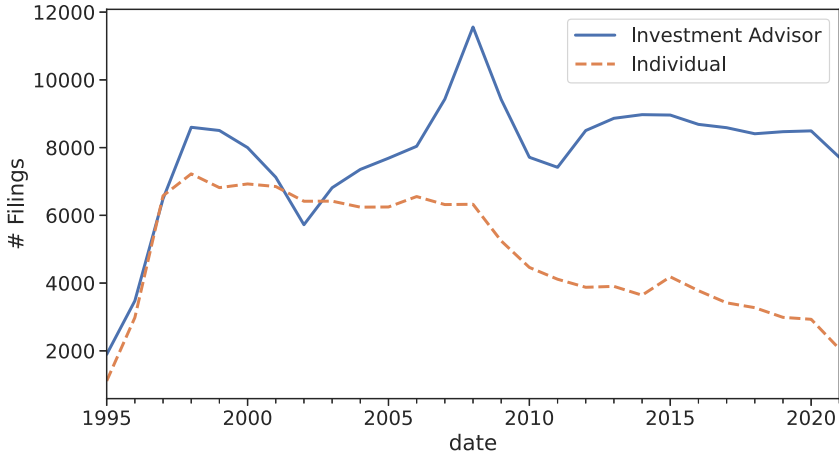


Figure 4: Development of filings filed by investments advisors and individuals.

Table 2: Summary statistics: source of funds.

Source of funds	n	Mean	Std	Min	Q1	Median	Q3	Max
Affiliate	17,922	19.01	20.72	0.00	6.11	10.10	23.20	100.00
Bank	1572	36.01	28.40	0.00	12.14	28.50	52.23	100.00
Other	43,836	22.92	23.32	0.00	6.64	13.68	31.59	100.00
Personal funds	26,430	20.65	21.93	0.00	6.40	11.50	26.60	100.00
Subject company	2090	22.06	20.99	0.00	7.20	13.95	29.76	100.00
Working capital	47,565	19.62	22.04	0.00	5.72	10.30	24.40	100.00

This table reports summary statistics for the ownership size (as % of floating shares) reported in filings by the self-reported source of funds category. Note that the number of observations is limited as the source of funds has only to be reported for Form 13D and 13D/A filings. Period: from November 1993 to May 2021.

4 Data Availability and Field Description

The dataset is made available as a RAR-compressed CSV file, in which each row represents a single filing and the 76 columns contain parsed information for each filing. It is hosted on the Harvard Dataverse repository and can be downloaded under the following permanent url:

<https://doi.org/10.7910/DVN/61Z64Q>.

A description of all 76 included columns can be found in the Supplementary Material.

5 Closing Remarks

In this paper, I introduce a publicly available and easily usable dataset consisting of data extracted from SEC Form 13D, D/A, G, and G/A filings. In contrast to Form 13F filings that contain all shareholdings of institutional investment managers, these filings are not required to be filed in a computer-readable format, which makes accessing them difficult and cumbersome.

While some commercial suppliers offer data products containing parts of the Form 13D/G(A) data, this database is to the best of my knowledge the only publicly available, noncommercial database that contains up-to-date, accurate, and complete filing data compiled from these forms.

I hope that using this data enables researchers to get a better understanding of e.g. the reception of new ownership data and information assimilation in financial markets, especially with regards to nonfinancial blockholders. Combined with other data sources, this data could lead to new insights into the effects of blockholdership and hint at ways to reduce information asymmetry and ultimately improve market efficiency (e.g. through improved, targeted regulation, or additional incentives for long-term ownership).

Additionally, the different characteristics of short-term- and long-term-oriented shareholders and their effects on the long-term health of the economy are an interesting and relevant area of research. I plan to use the blockholder data to conduct an empirical study on whether companies with large long-term blockholders outperform companies with a high public float in the long-term using an asset pricing approach.

This paper is an abbreviated version of my working paper “Determinants of Blockholdership – A new Dataset for Blockholder Analysis” (Harries 2021).

References

- Dlugosz, J., Fahlenbrach, R., Gomers, P., and Metrick, A. (2006). Large blocks of stock: prevalence, size, and measurement. *J. Corp. Finance* 12: 594–618.
- Harries, J.P. (2021). Determinants of blockholdership - a new dataset for blockholder analysis. SSRN Scholarly Paper ID 3899185. Social Science Research Network, Rochester, NY.
- Meredith, M. (2007). A librarian’s guide to the securities and exchange commission’s filings. *Ref. Libr.* 48: 35–55.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/jbnst-2021-0033>).