

UC Irvine

UC Irvine Previously Published Works

Title

Revisiting Mutual Fund Portfolio Disclosure

Permalink

<https://escholarship.org/uc/item/9dr6x6gm>

Journal

Review of Financial Studies, 29(12)

ISSN

0893-9454

Authors

Schwarz, Christopher G
Potter, Mark E

Publication Date

2016-12-01

DOI

10.1093/rfs/hhw057

Peer reviewed

Revisiting Mutual Fund Portfolio Disclosure

Christopher G. Schwarz
Paul Merage School of Business, University of California Irvine

Mark E. Potter
Finance Division, Babson College

We document that CRSP and Thomson contain many voluntarily reported mutual fund portfolios that are not in SEC filings while, additionally, CRSP and Thomson are missing many SEC mandated portfolios available in SEC filings. We document that the voluntary disclosures are likely driven by convenience rather than duplicity. Although mandated portfolios contain securities with more return momentum, we find use of SEC or Thomson data lead to similar empirical findings. CRSP, however, contains inaccurate position information prior to 2008. Our findings have important implications, such as highlighting a 35% increase in observed manager trading by combining data sources. (*JEL* G11, G23)

The authors thank the Editor, Laura Starks, and an anonymous referee for extensive, helpful comments and Meerae Park, Teri Wang, Chengdong Yin, and Thomas Howland for outstanding research assistance. The authors acknowledge and thank seminar participants at Babson College, UC Irvine, and the 2012 NFA Annual Meeting for helpful comments. All errors are the responsibility of the authors. Send correspondence to Christopher Schwarz, University of California – Irvine, Paul Merage School of Business, SB2 329, Irvine, CA 92697; telephone: 949-824-0936. E-mail: cschwarz@uci.edu.

Financial research has greatly benefited from data availability. Instead of hand-collecting data, researchers have largely relied on database vendors, such as the Center for Research in Security Prices (CRSP) and Thomson Financial. Over time, researchers have examined these commercial databases for both their coverage and their reliability. For example, the accuracy of the CRSP Stock Databases has been improved through the documentation of various data issues (e.g., Rosenberg and Houglet 1974; Bennin 1980; Shumway 1997; Canina et al. 1998; Shumway and Warther 1999). Elton, Gruber, and Blake (2001) compare the coverage of CRSP Mutual Fund Database to that of Morningstar, and Evans (2010) finds some mutual funds are incubated, leading to backfilled returns. Ljungqvist, Malloy, and Marston's (2009) discovery of changes in the I/B/E/S database has led to more reliable data for researchers. Agarwal, Daniel, and Naik (2009) document an increase in data from combining hedge fund databases due to the voluntary nature of hedge fund reporting. These studies are important since financial research findings heavily depend on the availability and accuracy of the underlying data.

In this paper, we document that the Thomson Mutual Fund Holdings (Thomson) database contains many mutual fund portfolios reported as of dates not mandated by the 1940 Investment Company Act. At the same time, we find that many mandated mutual fund portfolio disclosures made in Securities and Exchange Commission (SEC) filings are not in Thomson.¹ Since Thomson contains portfolios not reported in SEC filings, the non-SEC portfolios must be voluntarily disclosed by funds. Given the primary source for Thomson is often believed to be SEC filings and prior research documents the potential costs of more frequent portfolio

¹ Throughout the manuscript, "SEC filings" refers to the Forms N-30D, N-CSR, and N-Q filings, which contain mutual fund portfolios as required by the 1940 Investment Company Act. "Thomson" refers to the Thomson Financial Mutual Fund Holdings data, also known as the s12 data file, which contains mutual fund portfolios. Thomson was previously known as the CDA Spectrum database (e.g., Wermers 1999).

disclosure (e.g., Wermers 2001; Coval and Stafford 2007; Shive and Yun 2013; Verbeek and Wang 2013), this finding is surprising.²

The differences in report dates across SEC filings and Thomson are significant. For the same universe of mutual funds, we find a total of 77,555 unique portfolios across both sources. Only 48% of those portfolios appear, or overlap, in both SEC filings and Thomson. Thirty-two percent of portfolios are contained in Thomson, but not in SEC filings, and 20% of the portfolios are reported in SEC filings, but not in Thomson. A majority of the reporting differences are driven by the funds whose fiscal year-ends do not align with calendar quarter-ends. Many of these funds' portfolios are reported in Thomson as of non-SEC-mandated calendar quarter-end months. At the same time, their fiscal year-end-aligned portfolios required by the 1940 Act are included in SEC filings but are not included in Thomson.

We find a similar lack of overlap between the SEC's portfolios and the CRSP Mutual Fund Database's (CRSP) portfolios, which become available starting in the second half of 2003. When we merge CRSP with SEC and Thomson data, we find only 39% of portfolios overlap in all three sources for the same universe of funds. More importantly, we document CRSP portfolios' positions are inaccurate prior to the fourth quarter of 2007. During that period, one in five CRSP fund portfolios has 25% or more of their positions reported inaccurately.

Convenience is the likely motivation for voluntarily reporting portfolios to vendors. Portfolios are required to be reported to investors as of dates based on fiscal year-ends. Because funds within the same management company can have distinct fiscal-year ends, many management companies would need to provide some of their funds' portfolios to Thomson

² At the 2007 WRDS Users Meeting, SEC N-30D, N-CSR, and N-Q forms were specifically listed as the underlying data source for Thomson (Moussawi 2007). Additionally, Wermers (1999) notes that some of the portfolios are obtained through informal reports made to CDA Spectrum (which was bought by Thomson), but notes the main source for Thomson is SEC filings.

almost every month. These companies likely find it more efficient to simply report all of their funds' portfolios as of calendar quarter-ends when they also report other data to Thomson (e.g., Form 13F).³ The convenience motivation is supported by the fact that 98% of portfolios reported in Thomson are either as of SEC-mandated months or voluntary disclosures as of calendar quarter-end months.

Next, given the size of the mismatch, we compare the consistency of empirical findings using either Thomson or SEC data. To provide a complete analytical comparison, we hand-collect the approximately 15,000 SEC filed portfolios not contained in Thomson. While we do find that portfolios reported to the SEC have securities with relatively higher return momentum as compared with voluntarily reported portfolios (e.g., window dressing), we find using either database leads to similar empirical conclusions. Specifically, both databases lead to similar estimates of manager skill and the cost of mandatory disclosure. Thus, we conclude that prior empirical findings that rely on Thomson's mixture of mandatory and voluntary portfolios are likely unbiased compared with results using only the mandatorily disclosed portfolios in SEC filings.

Finally, as with hedge fund data that are voluntarily reported, we document that combining mutual fund portfolio datasets leads to significantly more data to study. Adding SEC data to Thomson leads to a 25% increase in the observed number of positions and a 16% increase in observed position changes, or trading. The median trade size in the combined dataset is 33% smaller due to enhanced precision from access to more frequent portfolios, and the average

³ Mutual fund companies directly provide data to database vendors. While Thomson did not clearly confirm this, Fidelity stated they directly provide data to Thomson. WRDS is also under the impression that Thomson receives data directly from fund companies. Additionally, Morningstar stated mutual fund companies directly provide data. To confirm Thomson is not simply collecting data from fund companies' Web sites, we examined twenty company Web sites, for which Thomson has voluntarily reported portfolios. We find that a large majority of these companies do not provide non-SEC portfolios via the Web.

maximum delay from when a trade occurs until when it is observed is reduced by three weeks. In 2008, when CRSP data become reliable, a CRSP, SEC, and Thomson combined dataset leads to an increase in observed trades of 35%, as well as a 50% reduction in median trade size and a one-month reduction in average trade delay as compared with Thomson only. These trade observation improvements are significant since prior studies (e.g., Puckett and Yan 2011; Elton, Gruber, and Blake 2011, 2012) find that increased disclosure frequency is important when examining managers' trading performance.

Our findings make several contributions to the commercial database coverage and reliability research and asset management literatures. First, we document that commercial mutual fund portfolio databases are not replicas of the mandatory SEC-filed portfolios; rather, they are combinations of voluntary and mandatory disclosures. More importantly, we find these databases do not contain all mandatory disclosures available through SEC filings for funds they cover. Second, we document that incentives such as convenience can lead to financial firms providing more disclosure information than is required by regulatory agencies. Third, the use of CRSP portfolio data prior to the fourth quarter of 2007 should be avoided. Fourth, we demonstrate a significant increase in observed manager trading by combining multiple mutual fund portfolio databases.

Finally, although we find no significant empirical differences, researchers should perform their analyses using various portfolio datasets as a robustness check because some findings could be sensitive to database selection. For example, researchers should carefully consider the public availability of portfolios when examining how market participants utilize mutual funds' portfolio disclosures (e.g., Shive and Yun 2013; Solomon, Soltes, and Sosyura 2014). Voluntarily reported CRSP and Thomson portfolios are not readily available to all market participants, especially

noninstitutions. Moreover, due to differences in reporting delays between datasets, a fund's SEC portfolio as of a more recent report date may be publicly available before an older, voluntarily disclosed Thomson portfolio.⁴

1. Data and Mutual Fund Holdings Disclosure Rules

Mutual funds are heavily regulated by the SEC and are therefore required to report large amounts of data to regulators and fund shareholders. In this section, we specifically discuss the rules for mutual fund portfolio disclosure, as well as our data sources.

1.1 Disclosure rules for mutual funds

In this paper, we focus on the required portfolio reporting by mutual funds. Under the Investment Company Act of 1940, a mutual fund is required to report its holdings (or "portfolio") to the SEC, as well as to each individual fund shareholder, as of dates ("report dates") that coincide with its fiscal year-end. These disclosures must include all of the fund's portfolio positions, including, but not limited to common equities, preferred equities, options, bonds, and short positions. The required portfolio disclosures are found in the fund's SEC Forms N-30D, N-Q, and N-CSR filings.⁵ These disclosures must be mailed to shareholders within sixty days of the report date and filed with the SEC's EDGAR database within ten days of the mailing. The date this portfolio information is filed with EDGAR is known as the "file date," and the difference between the file date and report date is known as the "reporting delay."

⁴ Some voluntarily reported portfolios in Thomson may be available on fund company Web sites. However, based on our twenty fund company sample, this constitutes only a small fraction of the total voluntarily reported data.

⁵ Form N-CSR replaced Form N-30D starting in 2003. Form N-Q was added for quarterly holdings disclosures starting in mid-2004. A handful of funds disclose what appear to be voluntary portfolio snapshots via an additional form, N-30B-2. However, this is rare and declines over our sample period.

Although not examined in this paper, under the 1934 Securities Act, all mutual fund management companies with more than \$100 million in 13(f) securities, such as publicly traded equities and bonds, must file SEC Form 13F. Form 13F is much different than the previously discussed mutual fund portfolio disclosure mandate. First, Form 13F is targeted at all institutional money managers, including hedge funds (e.g., Agarwal et al. 2013; Aragon, Hertz, and Shi 2013; Brown and Schwarz 2015). Second, Form 13F discloses holdings at the management company level. In other words, a management company (e.g., Fidelity) would report one company-level aggregated portfolio for all funds in a 13F, whereas each individual fund's (e.g., Magellan) portfolio is disclosed in the previously described SEC Forms N-30D, N-Q, and N-CSR. Third, Form 13F reports holdings information as of each calendar quarter-end and Form 13F must be filed within forty-five days of the report date.

1.2 Data sources

We gather mutual fund holdings data from three sources starting in 1996, when the SEC's EDGAR database came fully online, to 2008. The first source is the Thomson mutual fund holdings database (i.e., Thomson s12 data file), which is used in almost all studies that employ mutual fund portfolio holdings (e.g., Daniel et al. 1997; Wermers 2000; Coval and Stafford 2007; Kacperczyk, Sialm, and Zheng 2008; Cremers et al. 2009; Huang, Sialm, and Zhang 2011; Wermers et al. 2012; Shive and Yun 2013). Prior to its ownership by Thomson, this database was known as CDA Investment Technologies data (Daniel et al. 1997), as well as the CDA Spectrum holdings database (Wermers 1999). We restrict our sample to portfolios that have between 50% and 102% of their assets in common equity securities. It is possible for equity funds to have more than 102% of their assets invested in equities, but we often find that

portfolios with over 100% of net assets in equity positions contain errors related to stock splits. This is especially true in the third quarter of 1999.

We link these portfolios to the CRSP Survivorship Bias-Free Mutual Fund database (CRSP) using the MFLinks file available on WRDS. We filter our sample, removing portfolios belonging to funds with non-equity-related style codes or that are missing prior year return, asset, turnover, or expense data in CRSP. We then combine the Thomson and CRSP data with the portfolios contained in the previously described SEC Forms N-30D, N-CSR, and N-Q filings. This process took a number of steps, which are outlined in Section A of the Online Appendix. We only merge SEC portfolios for funds that are in our CRSP and Thomson data to keep the set of mutual funds consistent across all data sources. SEC portfolios' positions are contained in nonstandardized free text files. Later, we discuss the collection of these portfolios' positions in a tabulated format.

As of July 2003, the CRSP database also contains mutual fund portfolios. Although not used as frequently as Thomson, more studies are using CRSP's portfolios (e.g., Shive and Yun 2013; Cici and Palacios 2015). Unlike Thomson, CRSP does not directly collect its data. Prior to the fourth quarter of 2007, CRSP obtained its portfolios from Morningstar, and, subsequently, CRSP received its data from Lipper (Cici and Palacios 2015). Hence, CRSP mutual fund portfolios could exhibit different characteristics during these two periods.

Thus, our final dataset consists of mutual fund portfolios sourced from CRSP, Thomson, and SEC filings. One difference between the three data sources is the portfolios' position coverage. Thomson portfolios contain only long common equity positions. CRSP portfolios have some coverage of nonequity and short equity positions. However, this coverage appears to be rather sparse in the pre-2008 period. On the other hand, as discussed earlier, portfolios in SEC

filings contain all position information, regardless of security type. Thus, in terms of a portfolio's positions, SEC data are the most extensive.

2. Comparison of Portfolios across Data Sources

In total, for the same set of mutual funds, we have three different sources for mutual fund portfolios (CRSP, SEC, and Thomson). In this section, we investigate the consistency of the portfolio report dates across databases, as well as the accuracy of position information in the CRSP and Thomson databases.

2.1 Consistency of portfolio report dates

While combining the CRSP and Thomson data with the SEC data, we discovered that a large number of CRSP and Thomson portfolios are not reported in the SEC filings. Additionally, a large number of portfolios contained in the SEC filings do not appear in CRSP and Thomson. In other words, these three sources contain portfolios for the same funds, but as of different report dates. To illustrate this issue, in Table 1, we list portfolios from 2007 to 2008 for the Heritage Capital Appreciation Fund, which has an October fiscal year-end. We include the portfolio's report date, the data source, and whether the portfolio is required by SEC regulations, as well as the number of shares held by the fund for two representative securities.

<Insert Table 1 about here>

For this fund, the three data sources have few portfolios with the same report date (i.e., “overlapping” portfolios). In fact, CRSP and Thomson have no portfolios that overlap with the

SEC portfolios. The share counts across all sources are similar but are not identical, indicating the CRSP and Thomson portfolios are for the same mutual fund as the SEC portfolios, but as of different dates. Thomson and CRSP have portfolios reported as of the end of calendar quarters, and the SEC filings contain portfolios aligned with the fund's fiscal year. Thus, the portfolios unique to CRSP and Thomson must be voluntarily reported by the fund since they are not contained in SEC filings. Given the expected costs of portfolio disclosure (e.g. Wermers 2001), it is surprising funds voluntarily disclose additional non-SEC portfolios to data vendors. That all SEC portfolios are not contained in CRSP and Thomson is also surprising.

The Heritage example is not an isolated incident. In Table 2, we report a comparison of the SEC and Thomson portfolios for all of our funds. We exclude CRSP at this time as it is not available over our entire same period. For each year, we report the number of portfolios in Thomson (*Thomson*). We then report the number of Thomson portfolios that overlap with the SEC portfolios (*SEC Overlap*), the overlap percentage (*Overlap %*), and the number of SEC portfolios for these same funds that do not overlap with Thomson (*SEC only*). We also report the percent increase in portfolios by combining SEC and Thomson (*% increase*). We report results over our entire sample period (*Total*), as well from 1996 through the first half (FH) of 2004, when the SEC required semiannual disclosures (*Semiannual*), and from the second half (SH) of 2004 to 2008, when the SEC required quarterly disclosures (*Quarterly*).

<Insert Table 2 about here>

Overall, we find a low degree of overlap between the two data sources. For our entire sample period, we find a total of 77,555 unique reported portfolios. Of these, 48% (37,393

portfolios) overlap between SEC and Thomson data, 20% (15,635) are reported in the SEC filings, but not in Thomson, and 32% (24,527) are reported in Thomson, but not to the SEC. Moreover, a combined dataset contains 25% more portfolios than does Thomson alone. To our knowledge, we are the first to document the mismatch between portfolio disclosures in the primary data source (e.g., SEC filings) and Thomson.⁶

The overlap is significantly different during the two disclosure periods. In the SEC-mandated semiannual period (1996 to 2004 FH), the match rate is less than 50%. This increases to 76% during the SEC-mandated quarterly disclosure period (2004 SH to 2008). While the overlap rate is higher during this later period, the increase in portfolios achieved by combining sources is also larger. These changes are expected since twice as many portfolios are disclosed through SEC filings per year during the SEC-mandated quarterly disclosure period. Thus, even though the overlap is higher more recently, the advantage of combining datasets continues into the higher frequency disclosure regime.

In addition to the Thomson portfolios, we also have CRSP portfolios from 2003 SH to 2008. In Table 3, we report results for a three-way comparison of CRSP, SEC, and Thomson portfolios for this period. We report the number of portfolios contained in only one of the data sources, the number contained in only two data sources, and, finally, the number of portfolios contained in all three databases. We also report the total number of portfolios available by combining all of the datasets (*All combined*) and the total number of portfolios available in each individual dataset. As an example, the Heritage Capital Appreciation Fund data from 2007 to 2008 in Table 1 contain a total of 15 combined portfolios. Four of those portfolios are Thomson

⁶ Ge and Zheng (2006) examine the frequency of mutual fund portfolio disclosure. However, they only use the data from Thomson and make no comparison to SEC filings.

only, eight are SEC only, three are CRSP and Thomson only, and zero are reported to all three sources.

In addition to the entire CRSP availability period, we also report results from 2003 to 2007 Q3 (2007 Q4 to 2008), when CRSP receives its data from Morningstar (Lipper).⁷ As a baseline, we report results for our combined Thomson and SEC dataset over the entire sample period.

<Insert Table 3 about here>

When adding CRSP, we find an even larger mismatch. Only 39% of portfolios overlap in all three data sources. As a comparison, Jorion and Schwarz (2014) find that the overlap between two hedge fund databases, where there is no SEC-mandated reporting and thus all data are voluntarily reported, is 32%. Interestingly, we see a significant change in CRSP when it migrates from Morningstar to Lipper data. In the Morningstar period, CRSP is essentially a subset of Thomson with the lowest coverage of any individual data source. However, in the fourth quarter of 2007, it becomes the most thorough individual data source.

Overall, we find that mutual fund portfolio databases have low degrees of overlap. CRSP and Thomson contain a large number of portfolios voluntarily reported by funds not in SEC filings, while, perhaps more importantly, they are also missing a large number of SEC-mandated portfolios available in SEC filings. In other words, while the SEC database contains only mandatory portfolio disclosures, both Thomson and CRSP contain a mix of mandatory and voluntarily disclosed portfolios. Similar to hedge fund databases, combining multiple mutual fund holdings databases leads to much greater data availability.

⁷ During our analyses, it becomes clear that Lipper portfolio data begin in the fourth quarter of 2007.

2.2 Accuracy of portfolio position information

In the prior section, we find considerable differences in the report dates of the portfolios contained in our three data sources. In this section, we examine overlapping Thomson and CRSP portfolios to determine whether they contain identical equity position information. For each overlapping portfolio position, if the reported number of shares held is different or one of the sources does not list the position, we designate our *Difference* variable one, and zero otherwise. We then compute the proportion of all positions across all portfolios that are different (*Positions pct. diff.*). We also report the average difference size (*Position diff. size %*), which is the absolute value of the share count difference, divided by the larger share count. Difference size is only computed if a difference exists. Finally, we count the number of portfolios in which 10%, 25%, and 50% or more of their positions differ across the two data sources. We report results by year, except for 2007, when CRSP changed data sources. We also aggregate results over the entire CRSP data period and the Morningstar (2003–2007 Q3) and Lipper (2007 Q4–2008) periods. Results are reported in Table 4, panel A.

<Insert Table 4 about here>

We find surprising differences. Over the entire period, 14% of all positions disagree and 15% of portfolios have 25% or more of their equity positions different across the two sources. These differences are significant, averaging more than 60% of the position size. There is a clear demarcation in 2007 Q4, when CRSP receives its data from Lipper. During the Morningstar period, 20% of CRSP and Thomson portfolios disagree by 25% or more, whereas in the Lipper

period only 2% of portfolios disagree by the same magnitude. Thus, it seems likely the disagreement stems from inaccurate position information in CRSP during the Morningstar data period. To confirm this hypothesis, we hand-collect SEC portfolio positions when Thomson and CRSP disagree by 10% or more and the SEC has an overlapping portfolio. We then compare the SEC portfolio positions to those in CRSP and Thomson. We report these results in panel B of Table 4.⁸

Based on the SEC comparison, CRSP clearly has inaccurate position information during the Morningstar data period. Prior to 2008, almost all CRSP portfolios disagree with the overlapping SEC portfolio by 10% or more, whereas this level of disagreement rarely occurs with overlapping SEC and Thomson portfolios. Starting in 2008, both CRSP and Thomson have a rather large disagreement rate, although this is for a small proportion of the total number of portfolios in the datasets.⁹

Overall, our results show CRSP portfolios prior to 2008 contain inaccurate position information and use of CRSP should be avoided during this period. Because of these inaccuracies, we largely focus on Thomson and SEC data in the rest of our analyses.

3. Motivation for Voluntary Reporting

Since many portfolios in Thomson do not overlap with portfolios in SEC filings, mutual fund companies are voluntarily reporting portfolio data. Prior research (e.g., Wermers 2001) has

⁸ More details about how we hand-collect SEC portfolios are contained in Section B of the Online Appendix.

⁹ In untabulated results, we find that some large discrepancies between overlapping SEC and Thomson portfolios are due to incorrect links between the CRSP mutual fund database and Thomson. For example, the MFLinks file matches the Vanguard Windsor fund in CRSP with the Vanguard Windsor II fund in Thomson and vice versa.

noted the potential costs associated with more frequent portfolio reporting, such as funds having an increased susceptibility to front-running and copycatting. Thus, voluntary reporting of additional portfolios to Thomson is somewhat surprising. In this section, we investigate why funds make these disclosures.

Our previous Heritage Capital Appreciation Fund example suggests that funds are systematically shifting portfolio reporting to calendar quarter-end months in Thomson, rather than randomly reporting their portfolios across months. If funds are systematically voluntarily reporting as of the same months, this suggests the motivation for voluntarily reporting is not nefarious. To examine the consistency of voluntary portfolio reporting, we place the portfolios reported in Thomson into three groups. *SEC required* contains portfolios included in the SEC filings. *CQE* contains portfolios voluntarily reported as of calendar quarter-end months. *Non-CQE* contains portfolios voluntarily reported as of noncalendar quarter-end months.

We report this breakdown of portfolios in Table 5. We report results for all funds, funds whose calendar and fiscal quarters align (*Aligned funds*), and funds whose calendar and fiscal quarters do not align (*Unaligned funds*). We also report results with and without index funds by removing funds with “index” or “idx” in their names.¹⁰

<Insert Table 5 about here>

We find almost all voluntarily disclosed portfolios are reported as of calendar quarter-end months. Ninety-eight percent of all Thomson portfolios are either SEC-mandated disclosures or voluntary disclosures coinciding with calendar quarter ends, regardless of funds’ fiscal year-ends. Not surprisingly, *Aligned funds* exhibit a greater propensity to report portfolios to both the

¹⁰ Except for the motivation analyses, empirical tests are run with index funds removed.

SEC and Thomson as of the same months. Given all funds are voluntarily reporting as of the same months, funds likely have the same motivation to voluntarily report.

One motivation for voluntarily reporting portfolios as of calendar quarter-ends is simply convenience. Many mutual fund companies have funds with different fiscal year-end months. For example, Fidelity has funds with at least eight different fiscal year-ends. If Fidelity reported mandatory SEC portfolios to Thomson, they would be providing some of their funds' portfolios to Thomson every month. Fund companies may therefore find it administratively efficient to simply report all of their funds' portfolios as of calendar quarter-ends, when they also report other institutional data to Thomson (e.g., Form 13F). Alternatively, Thomson could have simply asked fund companies to provide funds' calendar quarter-end portfolios to make it easier for investors to compare funds.

If administrative convenience is the motivation, we would expect that all funds under the same management company would have the same disclosure pattern in each filing period. For example, say a fund company has one fund that is required to report an April portfolio to the SEC and another fund that is required to report a May portfolio. We would expect this fund company to either report both portfolios as of June to consolidate reporting to Thomson or report both as of the original SEC-mandated months. Selectively changing only one fund's disclosure to June would require as much effort as defaulting to the required disclosure dates. We therefore examine reporting consistency of fund management companies. Because the convenience motivation is likely much stronger for *Unaligned funds*, we only perform this analysis on these funds.¹¹

¹¹ *Aligned funds* are already reporting as of calendar quarter-ends. Thus, the uploading of extra calendar quarter-end portfolios could be more random based on the administrator. Additionally, *Aligned funds* can only report extra calendar quarter-end portfolios during the SEC-mandated semiannual disclosure period, which is also likely to

For *Unaligned funds*, each period a mandatory filing is due, a management company is designated one if all the company's funds report portfolios to Thomson as of calendar quarter-ends or all as of fiscal quarter-ends. A mixture of reporting policies results in a label of zero. After eliminating periods in which the management company only has one fund, we average these values for each company across reporting periods and report the distribution in Figure 1. Companies with an average of 100% (0%) never (always) selectively disclose voluntary portfolios for their funds.

<Insert Figure 1 about here>

The distribution is largely consistent with the convenience motivation. For example, only about 5% of fund companies selectively disclose all the time, whereas roughly 55% of fund companies almost always report all their funds' portfolios to Thomson as of either fiscal- or calendar quarter-ends. We generally find averages near 100%, suggesting companies never selectively report voluntary portfolios for their funds.

Overall, our results suggest the primary motivation for voluntary reporting is convenience. Fund companies almost universally report portfolios as of calendar quarter-end months. We also find voluntary reporting is similar for all funds within the same fund company.¹²

4. Consistency of Empirical Results

impact these funds' consistency results. Indeed, in untabulated results, we find the voluntary reporting of their portfolios is less consistent than for *Unaligned funds*.

¹² In untabulated results, we examine the motivation for voluntary reporting using a probit model with voluntary reporting as the dependent variable and various fund and portfolio characteristics as the independent variables. These results do not suggest any other primary motivation for voluntary reporting.

Given Thomson and SEC data both contain large numbers of unique portfolios, an open question is whether empirical results are invariant for database selection. In this section, we investigate this question by comparing empirical estimates from both databases for three topics: manager skill, performance characteristics of portfolio holdings (e.g., window dressing), and the cost of mandatory disclosure. The latter two topics are more likely to be sensitive to database selection given the inclusion of voluntarily reported portfolios in Thomson. The former topic is selected due to its prominence in the literature.

To facilitate a complete comparison, we collect the approximately 15,000 portfolios available in SEC filings, but not included in Thomson.

3.1 Manager skill

Manager skill is one of the most researched areas in the mutual fund literature. Thus, a natural question is whether both the Thomson and SEC data provide similar empirical conclusions about this topic. To evaluate skill, we use Daniel et al.'s (1997) (DGTW) excess return measure and the "Return Gap" measure developed by Kacperczyk, Sialm, and Zheng (2008). When calculating DGTW returns, we identify the most recently reported portfolio for each fund at the beginning of each calendar quarter.¹³ We assume the portfolio security weights are the same at the beginning of the quarter as when reported them. We then compute the future one-, three-, and twelve-month weighted average DGTW returns for each fund. We compute averages across all funds each quarter and report results using Fama-Macbeth (1973) for both SEC and Thomson, as well as the differences between the two databases.

¹³ DGTW benchmarks are available via www.smith.umd.edu/faculty/rwermers/ftpsite/Dgtw/coverpage.htm. These benchmarks are also used by Wermers (2004).

In addition to reporting the cross-sectional means, we examine how consistent fund rankings are across datasets. We rank funds each quarter by their one- and three-month DGTW returns. We then place funds into quintiles based on their rankings. To compare rankings, we create a contingency table of SEC and Thomson rankings. If the cross-sectional rankings for funds are the same regardless of dataset, the diagonal terms will be 100% and the off-diagonal terms will be 0%.

We also run similar cross-sectional mean and ranking computations using funds' return gaps. Return gaps are computed as outlined by Kacperczyk, Sialm, and Zheng (2008). In summary, we compute the difference between the actual net return reported in CRSP and a simulated return based on the most recent portfolio of the fund. Results for the DGTW and return gap analyses are reported in Table 6. The cross-sectional average results for DGTW (return gap) are reported in panel A (C), and the DGTW (return gap) contingency table results are reported in panel B (D).

<Insert Table 6 about here>

Our results indicate that manager skill analyses are independent of data source choice. DGTW averages and cross-sectional rankings are similar between the two sources. Skill averages are similar across datasets, and 85% of funds are within the same performance quintile. Although largely consistent, the return gap results show more dataset dependency. The SEC return gap estimates are higher, and although cross-sectional rankings are still strong, they are not as stable. One reason for larger return gap differences is SEC portfolios are relatively more distant, or stale, at the start of a calendar quarter from the portfolio used to calculate the actual net return

of the fund.¹⁴ With that said, either dataset produces largely similar conclusions. Thus, database selection appears not to be an issue when evaluating manager skill.

3.2 Performance characteristics of portfolio holdings

Prior research finds evidence that equity fund managers “window dress” their disclosed portfolios. For example, prior researchers find that institutional equity fund managers window dress their portfolios by overweighting recent high-performing securities (e.g., Lakonishok et al. 1991; O’Neal 2001; Meier and Schaumburg 2004). However, Musto (1999) finds that managers’ window dressing actions depend on whether the portfolio is a voluntary or mandatory disclosure. Thus, mandatory SEC-disclosed portfolios may provide higher estimates of return momentum window dressing by managers when compared to the voluntary disclosed Thomson portfolios.

To examine this possibility, we compute each portfolio’s value-weighted average raw and DGTW security returns for various time frames. For each calendar month, we separate portfolios into SEC portfolios and voluntarily reported Thomson portfolios. For each group, we compute the equal-weighted average return across funds. Finally, we compute the annualized return differences between the two groups, where positive values indicate higher values for the SEC portfolios. We compute differences using both the SEC portfolios included in Thomson (*SEC in Thmn only*) and all the SEC portfolios (*All SEC*). Overall values are computed using Fama-Macbeth (1973), and standard errors are calculated using Newey-West (1987), with the number of lags equaling the number of monthly returns used in the return calculation. Results are reported in Table 7.

¹⁴ For example, for a fund with an October fiscal year-end, in SEC data, the portfolio reported in the first calendar quarter would be in January. However, a voluntarily reported Thomson portfolio would have a report date as of March.

<Insert Table 7 about here>

We find evidence consistent with window dressing results being stronger for mandatorily disclosed portfolios. SEC portfolios have higher prior average returns than do non-SEC portfolios. This is especially true in the month prior to disclosure, with a statistically and economically significant difference of 6%, suggesting managers focus most on short-term performance.¹⁵ Thus, in circumstances in which the research question involves investigating manager actions due to disclosure, researchers should make a distinction between mandatory and voluntary portfolio disclosures as the strength of results may depend on the disclosure type.

3.3 Cost of mandatory disclosure

Requiring mutual funds to disclose portfolios on a regular basis allows hedge fund managers or other investors the opportunity to front-run mutual funds based on the predictability of their investor flows or copycat their portfolios (e.g., Wermers 2001; Coval and Stafford 2007; Shive and Yun 2013; Verbeek and Wang 2013). However, these studies use Thomson and CRSP data, which are mixtures of voluntary and mandatory portfolio disclosures. An important question is therefore whether empirical findings regarding disclosure costs are similar across datasets.

We begin by examining the profitability of front-running predictable mutual fund flows. We run our analysis consistent with Coval and Stafford (2007). However, in this case, rather than use the report date of the portfolio when running the analysis, we choose the most recent portfolio based on file dates as this is the most recent portfolio speculators would have access to.

¹⁵ Alternatively, managers may be “painting the tape” just prior to the portfolio report dates (e.g., Carhart et al. 2002; Hu et al. 2014).

When using SEC data, we use the file dates recorded with EDGAR. However, this information is not available for all Thomson portfolios since voluntarily disclosed portfolios are not filed with EDGAR. Thus, consistent with the literature, we assume all portfolios have a sixty-day filing delay when using Thomson data.¹⁶

We split stocks into decile portfolios based on the percent of shares outstanding mutual funds are expected to buy over the next quarter. We report the average quarterly portfolio raw and DGTW returns for these decile portfolios in Table 8, omitting some deciles for the sake of brevity. We report results for our entire sample period, as well as results for the SEC-mandated semiannual portfolio disclosure period (1996 to 2004 FH) and the quarterly portfolio-mandated SEC disclosure period (2004 SH to 2008). We separate these periods as quarterly portfolio updates could make front-running more profitable since speculators would, on average, have more accurate position information over time (e.g., Wermers 2000). Overall coefficients and t-values are computed using Fama-Macbeth (1973).

<Insert Table 8 about here>

We find similar results using SEC or Thomson data. Similar to Coval and Stafford (2007), stocks with the highest flows significantly outperform stocks with the lowest flows over the entire period. During the more frequent quarterly portfolio-mandated disclosure period, front-running is unprofitable.

¹⁶ In untabulated results, we examine the distribution of EDGAR filing delays. On average, the delay is sixty days and is approximately the same over time. Most funds take at least fifty-seven days to disclose their portfolios to the SEC. These results are consistent with those of Brown and Schwarz (2015), who find that hedge funds wait until the end of their filing delay period to file their Form 13Fs with the SEC.

In analyzing the consistency of disclosure cost results across databases, we also evaluate copycatting results using the framework and assumptions of Verbeek and Wang (2013). We compare copycat net returns to the actual net returns of the underlying funds. Negative differences indicate the copycat fund outperforms. We compute the average difference each month for our sample period and then average the differences across time. We report results in Table 9. In panel A, we report results on the raw and absolute differences. Absolute differences measure the tracking error of the copycat fund. In panel B, we again examine tracking error. For each fund, we calculate the t-value of the difference between the fund's actual net return and the copycat fund's net return.

We report the distribution of these t-values. Significantly negative t-values indicate the copycat fund significantly outperforms the underlying fund. In both panels, we report results for the entire sample period, during the SEC-mandated semiannual portfolio disclosure period (1996 to 2004 FH), and for the quarterly portfolio-mandated SEC disclosure period (2004 SH to 2008). We again separate these periods as quarterly portfolio updates provide speculators more accurate information over time, reducing tracking error.

<Insert Table 9 about here>

We again find similar results using either dataset. The average copycat fund outperforms its underlying fund, similar to Verbeek and Wang (2013). With that said, we do find some minor differences between the two sets of results. Copycat funds using SEC portfolios have higher tracking error than those using the Thomson portfolios. This is likely due to SEC portfolios being more stale, on average, at the beginning of each calendar quarter. Additionally, Thomson has

more portfolios per year during the semi-annual-mandated disclosure period for many funds, which also reduces tracking error. The higher tracking error leads to fewer copycat funds outperforming their underlying funds when using SEC portfolios.

In summary, we find similar estimates for the cost of mandatory disclosure using either Thomson or SEC data. Thus, using either Thomson or SEC data, as with estimates of manager skill, is likely to lead to similar findings.

5. Implications

In the prior section, we find empirical results are largely similar using either Thomson or SEC data. Based on our empirical analysis and comparisons of SEC, Thomson, and CRSP mutual fund portfolios, we have four primary recommendations. First, we avoid CRSP portfolio data before 2007 Q4 due to their inaccurate position information. Second, for robustness, we recommend running empirical tests using only SEC data as some empirical findings may be sensitive to the inclusion of voluntarily reported data. The final two implications we discuss more thoroughly below: the increase in observed trading by combining datasets, as well as the impact of disclosure type on data availability.

5.1 Increase of observed trading by merging datasets

More frequent portfolio disclosures, leading to better trade information, is one major implication of combining datasets. Greater data availability is particularly useful for studies that examine manager trades. as well as reactions to information events such as earnings announcements. For example, using Table 1 as an illustration, we observe the fund's ATVI

position increased in July 2007 in the combined dataset rather than having to wait until September 2007 using only Thomson.

To quantify the benefits of more portfolios, we examine the increase in the number of trades, as well as the quality of trade information observed by combining databases. For example, we calculate the trade size in shares. We also compute the maximum number of days between when the trade could have occurred and when we observe the trade. Both of these characteristics help quantify the reduction in the gap between trading occurrences and trading observations. With more frequent portfolios, we should find trade sizes that are smaller over time, thereby reducing the maximum time delay between trade action and observation, similar to the prior ATVI example. We report results for this analysis in Table 10.

<Insert Table 10 about here>

Using only Thomson data, we observe seven million positions and 5.45 million trades for the entire sample period. The average (median) trade size is 110,000 (9,462) shares, and the average (median) maximum observed trade delay is 114 (105) days. Because of the amount of data voluntarily reported to Thomson, Thomson has better trade coverage than does the SEC, although this advantage disappeared once quarterly disclosure began in the second half of 2004.

Combining databases leads to a significant increase in observed trading. Using Thomson and SEC data together, we observe 8.8 million positions and 6.34 million trades, a 25% and 16% increase, respectively. Trade metrics improve substantially. The average (median) trade size declines to 101,826 (6,453), and the average (median) maximum delay is reduced to 95 (92) days. In other words, combining data reduces observed median trade sizes by 33%, while trade

information is available, on average, nearly three weeks sooner. In 2008, when we can also draw from CRSP data, we find an even larger advantage from using all three data sources. As compared with only using Thomson, we find a 50% increase in observed positions, a 35% increase in observed trades, a 50% reduction in trade size, and a one-month reduction in trade observation delay.

Overall, combining datasets is advantageous to researchers. Not only does this lead to more portfolios but this also leads to a large corresponding increase in the amount and quality of trading information.

5.2 Portfolio availability

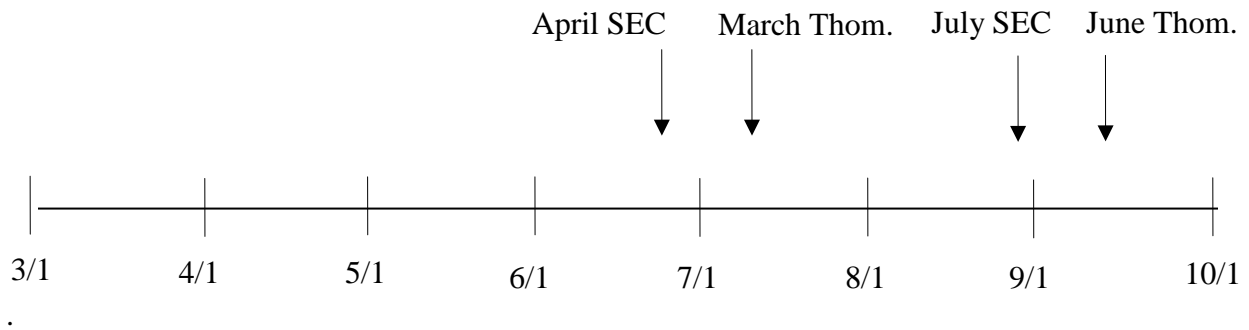
Voluntary reporting to CRSP and Thomson has important implications for studies that rely on the knowledge of portfolio holdings. Presently, studies presume that portfolios in Thomson represent SEC portfolio disclosures and are therefore homogenous in terms of access and availability. It is likely, however, that there are significant differences in the public availability and timing of Thomson's voluntary disclosures vis-à-vis mandatory SEC disclosures.

First, only those investors who subscribe to CRSP or Thomson can observe the voluntary filings. These subscribers are likely to be institutional investors given the cost of these databases. Since a large proportion of mutual fund shareholders are likely noninstitutions, most fund shareholders likely exclusively observe portfolios through mandated SEC disclosures. It is therefore advantageous for studies that examine mutual fund investor reactions (e.g., Solomon, Soltes, and Sosyura 2014) to use only mandatory disclosures available to everyone.¹⁷

¹⁷ Some of the portfolios voluntarily reported to CRSP and Thomson may be available to investors through other data sources, such as fund companies' Web pages. However, when we sampled a number of voluntarily reporting fund companies' Web sites, we found very few voluntarily disclosed portfolios. Thus, many voluntary disclosures in commercial databases will only be available to subscribers.

Second, even with access to CRSP and Thomson, SEC disclosures are likely to be observable prior to voluntary Thomson disclosures, although precise analysis is difficult because Thomson’s file dates do not represent actual upload dates. Thomson updates are provided shortly after the end of calendar quarters. At the same time, most voluntary disclosures are as of the end of a calendar quarters. Given it is highly unlikely that funds upload their portfolios to Thomson within a few days of the end of the quarter, almost all voluntarily disclosed portfolios will be delayed by a minimum of approximately three months.¹⁸ This potentially results in voluntary portfolios being ignored by traders because they are never the most recent portfolios available.

For example, assume a fund discloses its January, April, July, and October portfolios to the SEC and reports its March, June, September, and December portfolios to Thomson. Using the minimum expected Thomson delay (three months and a few weeks) and the average SEC delay (60 days), the disclosure time line for the first two calendar quarters’ portfolios would be approximately



Because of the delays, the March and June Thomson portfolios would be preceded by the April and July SEC portfolios in their availability. Thus, these voluntary disclosures are not

¹⁸ By construction, the only file dates in Thomson are as of the end of March, June, September, and December. For example, consider the data contained in the June 2010 version of the Thomson database we use. In the first, second, and third quarters of 2009, there are a total of 1,598, 1,655, and 1,585 portfolios in our sample. In the fourth quarter of 2009, this drops to only 1,149. Thus, Thomson is still missing close to 500 (33%) portfolios a full five months after the end of the quarter. Although we did not include 2009 in our sample, we are able to locate almost all of the missing 2009 Q4 portfolios in SEC filings. Thus, the omitted portfolios are not due to fund death.

likely to be employed by investors unless investors are specifically looking at historical trading behavior.

For these reasons, researchers should carefully determine whether the use of voluntarily disclosed portfolios is appropriate for their studies. More generally, although most empirical results are likely agnostic about the choice of Thomson or SEC data, we suggest researchers run results using each dataset individually as a robustness check, especially those research hypotheses that rely on the timing of portfolio availability as mixing voluntary and mandatory portfolios could confound any results related to market reactions to portfolio disclosures.

6. Conclusion

In this paper, we document that mutual fund portfolios as of different dates are reported to different data sources. From 1996 to 2008, only 48% of the total number of portfolios available across Thomson and SEC filings for the same set of funds appear in both datasets. Thirty-two percent of the portfolios are only reported in Thomson, and 20% are only reported in SEC filings. When examining CRSP, SEC, and Thomson data together, only 39% of portfolios overlap. Given the primary source for CRSP and Thomson is believed to be SEC filings, these low overlap rates are surprising. It is also curious to observe mutual funds providing more portfolio data than mandated, given the potential for front-running and copycatting.

We document the primary motivation for these voluntary filings to likely be due to administrative convenience. Almost all voluntarily disclosed portfolios are as of calendar quarter-end months. Although SEC-filed portfolios have securities with, on average, higher past performance, the use of either Thomson or SEC filings leads to similar conclusions regarding manager skill and cost of mandatory portfolio disclosure. These results validate prior research

using Thomson. However, we document CRSP portfolio positions are inaccurate prior to the fourth quarter of 2007.

Overall, our findings have several important implications. First, the use of CRSP data prior to 2008 should be avoided. Second, we show combining datasets leads to the availability of significantly more portfolio, position, and trade information. Third, studies examining the use of portfolio disclosures by market participants should consider the heterogeneity in both accessibility and timing of the SEC and non-SEC portfolios. More generally, researchers should assess the robustness of their empirical results about database choice. Finally, our results emphasize the importance of verifying any database data, even those data disclosed through regulatory filings.

References

- Agarwal, V., N. Daniel, and N. Naik. 2009. Role of managerial incentives and discretion in hedge fund performance. *Journal of Finance* 64:2221–56.
- Agarwal, V., W. Jiang, Y. Tang, and B. Yang. 2013. Uncovering hedge fund skill from the portfolios holdings they hide. *Journal of Finance* 68:739–83.
- Aragon, G., M. Hertzel, and Z. Shi. 2013. Why do hedge funds avoid disclosure? Evidence from confidential 13F filings. *Journal of Financial and Quantitative Analysis* 48:1499–518.
- Bennin, R. 1980. Error rates in CRSP and COMPUSTAT: A second look. *Journal of Finance* 35:1267–71.
- Brown, S., and C. Schwarz. 2015. The impact of mandatory hedge fund portfolio disclosure. Working Paper.
- Canina, L., R. Michaely, R. Thaler, and K. Womack. 1998. Caveat compounder: A warning about using the daily CRSP equal-weighted index to compute long-run excess returns. *Journal of Finance* 53:403–416.
- Carhart, M., R. Kaniel, D. Musto, and A. Reed. 2002. Leaning for the tape: Evidence of gaming behavior in equity mutual funds. *Journal of Finance* 57:661–93.
- Cici, G., and L. Palacios. 2015. On the use of options by mutual funds: Do they know what they are doing? *Journal of Banking and Finance* 50:157–68.
- Coval, J., and E. Stafford. 2007. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics* 86:479–512.
- Cremers, M., and A. Petajisto. 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies* 22:3329–65.
- Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristics-based benchmarks. *Journal of Finance* 52:1035–58.
- Elton, E., M. Gruber, and C. Blake. 2001. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases. *Journal of Finance* 56:2415–30.

- . 2011. Monthly Holdings data and the selection of superior mutual funds. *Journal of Financial and Quantitative Analysis* 46:341–67.
- . 2012. An examination of mutual fund timing ability using monthly holding data. *Review of Finance* 16:619–45.
- Evans, R. 2010. The incubation bias. *Journal of Finance* 65:1581–611.
- Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: empirical tests. *Journal of Political Economy* 81:607–636.
- Ge, W., and L. Zheng. 2006. The frequency of mutual fund portfolio disclosure. Working Paper.
- Hu, G., R. D. McLean, J. Pontiff, and Q. Wang. 2014. The year-end trading activity of institutional investors: evidence from daily trades. *Review of Financial Studies* 27:1593–614.
- Huang, J., C. Sialm, and H. Zhang. 2011. Risk shifting and mutual fund performance. *Review of Financial Studies* 24:2575–616.
- Jorion, P., and C. Schwarz. 2014. The strategic listing decisions of hedge funds. *Journal of Financial and Quantitative Analysis* 49:773–96.
- Kacperczyk, M., C. Sialm, and L. Zheng. 2008. Unobserved actions of mutual funds. *Review of Financial Studies* 21:2379–416.
- Lakonishok, J., A. Schleifer, R. H. Thaler, and R. W. Vishny. 1991. Window dressing by pension fund managers. *American Economic Review* 81:227–31.
- Ljungqvist, A., C. Malloy, and F. Marston. 2009. Rewriting history. *Journal of Finance* 64:1935–60.
- Meier, I., and E. Schaumburg. 2004. Do funds window dress? Evidence for U.S. equity mutual funds. Working Paper.
- Moussawi, R. 2007. Mutual fund data sources in WRDS: Overview and empirical issues. WRDS Users Meeting, Philadelphia, PA.
- Musto, D. K. 1999. Investment decisions depend on portfolio disclosures. *Journal of Finance* 54:935–52.
- Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708.

- O'Neal, E. S. 2001. Window dressing and equity mutual funds. Working Paper.
- Puckett, A., and S. Yan. 2011. The interim trading skill of institutional investors. *Journal of Finance* 66:601–633.
- Rosenberg, B., and M. Houglet. 1974. Error rates in CRSP and Compustat data bases and their implications. *Journal of Finance* 29:1303–1310.
- Shive, S. and H. Yun. 2013. Are mutual funds sitting ducks? *Journal of Financial Economics* 107:220–37.
- Shumway, T. 1997. The delisting bias in CRSP data. *Journal of Finance* 52:327–40.
- Shumway, T., and V. A. Warther. 1999. The delisting bias in CRSP's Nasdaq data and its implications for interpretation of the size effect. *Journal of Finance* 54:2361–79.
- Solomon, D., E. Soltes, and D. Sosyura. 2014. Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics* 113:53–72.
- Verbeek, M., and Y. Wang. 2013. Better than the original? The relative success of copycat funds. *Journal of Banking and Finance* 37:3454–71.
- Wermers, R. 1999. Mutual fund herding and the impact on stock prices. *Journal of Finance* 54:581–622.
- . 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses. *Journal of Finance* 55:1655–95.
- . 2001. The potential effects of more frequent portfolio disclosure on mutual fund performance. *Investment Company Institute Perspectives* 7:1–12.
- . 2004. Is money really “smart”? New evidence on the relation between mutual fund flows, manager behavior, and performance persistence. Working Paper.
- Wermers, R., T. Yao, and J. Zhao. 2012. Forecasting stock returns through an efficient aggregation of mutual fund holdings. *Review of Financial Studies* 25:3490–529.

Table 1. Portfolio filings example

This table lists the portfolios reported in the Thomson Mutual Fund Holdings Database (Thomson), CRSP Survivorship-Free Mutual Fund database (CRSP), and SEC filings for the Heritage Capital Appreciation Trust from 2007 to 2008. The *Report date* is the date the portfolio is reported. *Source* is the source for the portfolios, and *SEC required* indicates if the portfolio's disclosure is SEC mandated. The number of shares held by the fund for both Activision Blizzard, Inc. (ATVI) and Genentech, Inc. (DNA) are also reported.

Report date	Source	SEC required	Sample holdings	
			ATVI	DNA
31 Jan 2007	SEC Form N-Q	Yes	0	221,100
31 Mar 2007	Thomson (s12 File)	No	1,320,450	221,100
30 Apr 2007	SEC Form N-CSR	Yes	1,320,450	221,100
30 Jun 2007	Thomson (s12 File)	No	1,320,450	221,100
31 Jul 2007	SEC Form N-Q	Yes	1,568,050	221,100
30 Sep 2007	Thomson (s12 File)	No	1,886,450	221,100
31 Oct 2007	SEC Form N-CSR	Yes	1,886,450	211,300
31 Dec 2007	Thomson (s12 File)	No	1,409,620	211,300
31 Jan 2008	SEC Form N-Q	Yes	1,579,620	227,200
31 Mar 2008	Thomson (s12 File) and CRSP	No	1,579,620	173,800
30 Apr 2008	SEC Form N-CSR	Yes	1,778,225	173,800
31 Jul 2008	SEC Form N-Q	Yes	1,642,620	137,700
30 Sep 2008	Thomson (s12 File) and CRSP	No	1,878,220	137,700
31 Oct 2008	SEC Form N-CSR	Yes	1,828,920	134,100
31 Dec 2008	Thomson (s12 File) and CRSP	No	1,810,220	132,700

Table 2. Comparison of Thomson and SEC portfolios

This table compares the portfolios contained in the Thomson Mutual Fund Holdings database (s12 file) and those in SEC filings for the same set of mutual funds. For each year, we report the number of portfolios in Thomson (*Thomson*). We then report the number of Thomson portfolios that have the same report date (i.e., “overlap”) with SEC-filed portfolios (*SEC overlap*), as well as the percentage that overlap (*Overlap %*). Finally, we report the number of SEC filed portfolios that do not overlap with Thomson (*SEC Only*). *% increase* is the increase in the number of portfolios by combining SEC and Thomson data (*SEC only* divided by *Thomson*). *Total* reports the total values for our entire 1996–2008 sample period. *Semiannual* reports results for the 1996–2004 FH period when the SEC-mandated semiannual portfolio disclosure for mutual funds, and *Quarterly* reports totals over the 2004 SH–2008 period when the SEC mandated quarterly portfolio disclosure for mutual funds.

	Thomson	SEC overlap	Overlap %	SEC only	% increase
1996	2,165	1,355	63	271	13
1997	2,394	1,632	68	152	6
1998	3,087	1,701	55	478	15
1999	3,025	1,401	46	631	21
2000	4,313	1,797	42	880	20
2001	4,169	1,786	43	1,115	27
2002	4,772	1,679	35	1,311	28
2003	5,717	2,562	45	1,067	19
2004 FH	3,011	1,344	45	501	17
2004 SH	3,459	2,724	79	936	27
2005	6,799	5,173	76	2,202	32
2006	6,568	5,167	79	1,842	28
2007	6,185	4,687	76	2,020	33
2008	6,256	4,385	70	2,229	36
Total	61,920	37,393	60	15,635	25
Semiannual	32,653	15,257	47	6,406	20
Quarterly	29,267	22,136	76	9,229	32

Table 3. Comparison of portfolios reported to CRSP, SEC, and Thomson

This table examines the mutual fund portfolios reported to CRSP, SEC, and Thomson for the same set of mutual funds. In each time period, we aggregate the portfolios available across all available databases. *CRSP*, *SEC*, and *Thomson count (%)* are the number (percent) of portfolios only reported to CRSP, SEC, and Thomson, respectively. *CRSP and SEC*, *CRSP and Thomson*, and *SEC and Thomson count (%)* are the count (percent) of portfolios only reported to those two databases, respectively. *All DBs count (%)* is the number (count) of portfolios reported to all sources. *All combined* is the total number of unique portfolios when combining all data sources. *CRSP total*, *SEC total*, and *Thomson total* are the total number of portfolios reported to each of those databases individually, respectively. *1996–2008* excludes CRSP, whose portfolio availability starts in the second half of 2003.

	1996–2008		2003SH–2008		2003SH–2007Q4		2007Q4–2008	
	Count	%	Count	%	Count	%	Count	%
CRSP			2,513	5	869	2	1,644	13
SEC	15,635	20	8,748	18	7,126	20	1,622	13
Thomson	24,527	32	3,367	7	2,985	8	382	3
CRSP and SEC			1,462	3	345	1	1,117	9
CRSP and Thomson			6,993	15	5,088	14	1,905	16
SEC and Thomson			6,189	13	5,364	15	825	7
All DBs	37,393	48	18,640	39	13,943	39	4,697	39
Totals								
All combined	77,555	100	47,912	100	35,720	100	12,192	100
CRSP total			29,608	62	20,245	57	9,363	77
SEC total	53,028	68	35,039	73	26,778	77	8,261	68
Thomson total	61,920	80	35,189	73	27,380	75	7,809	64

Table 4. Position consistency and accuracy of overlapping Thomson and CRSP portfolios

This table examines the position consistency and accuracy of overlapping CRSP and Thomson portfolios. In panel A, during each period, we identify portfolios with the same report dates (i.e., “overlapping portfolios”) in both the Thomson and CRSP databases. We then compare the positions across the two databases. If the share counts are different, or one of the sources does not have the position listed at all, we designate our *Difference* variable one, and zero otherwise. *Positions pct. diff.* is the proportion of the total number of positions across all overlapping portfolios that are different. *Positions diff. size %* is the average size of the difference if one exists. It is computed as the absolute value of the share count difference divided by the larger share count across the two databases. *Total* is the number of portfolios used in the analysis. *10%+, 25%+, and 50%+ diff.* are the number (*count*) and percent (%) of portfolios that have 10%, 25%, and 50% or more of their positions different across the two databases. In panel B, we compare the CRSP (*CRSP vs. SEC*) and Thomson (*Thomson vs. SEC*) portfolios that disagree by 10% or more to overlapping SEC portfolios.

A. CRSP and Thomson portfolio comparison

Year	Portfolios						Positions		
	Total	10%+ diff.		25%+ diff.		50%+ diff.		Pct. diff.	Diff. size %
		Count	%	Count	%	Count	%		
2003	1,879	322	17	293	16	137	7	11	86
2004	4,273	803	19	756	18	513	12	17	64
2005	4,623	1,157	25	1,068	23	732	16	20	63
2006	4,800	941	20	891	19	589	12	15	66
2007 Q1–Q3	3,456	755	22	716	21	492	14	18	64
2007 Q4	1,168	18	2	15	1	15	1	4	86
2008	5,434	113	2	92	2	85	2	6	89
2003–2008	25,633	4,109	16	3,831	15	2,563	10	14	68
2003–2007								17	95
Q3	19,031	3,978	21	3,724	20	2,463	13		
2007 Q4–2008	6,602	131	2	107	2	100	2	5	91

B. SEC portfolio comparison

Year	Total	CRSP vs. SEC			Thomson vs. SEC		
		10%+ diff. pct.	25%+ diff. pct.	50%+ diff. pct.	10%+ diff. pct.	25%+ diff. pct.	50%+ diff. pct.
2003	233	94	94	89	7	5	4
2004	671	95	95	90	7	5	4
2005	1,089	98	98	91	6	3	3
2006	882	99	99	93	4	2	2
2007Q1–Q3	696	98	98	93	7	4	4
2007Q4	15	53	53	33	53	53	53
2008	90	57	57	46	49	42	41
2003–2008	3,676	96	96	90	7	5	4

2003–2007Q3	3,571	97	97	91	6	4	3
2007Q4–2008	105	56	56	44	50	44	43

Table 5. Examination of portfolios reported to Thomson

This table describes the breakdown of portfolios reported to Thomson. *Aligned funds* are funds whose calendar and fiscal quarters align, and *Unaligned funds* are funds whose calendar and fiscal quarters do not align. *Total* is the total number of portfolios reported in Thomson in our sample period from 1996 to 2008. *SEC req.* is the number of Thomson portfolios that overlap with SEC disclosed portfolios. *CQE* is the number of portfolios voluntarily reported to Thomson as of calendar quarter-end months. *Non-CQE* is the number of portfolios voluntarily reported to Thomson in noncalendar quarter-end months.

All funds				
	Total	SEC required	CQE	Non-CQE
All	61,920	37,201 60%	23,235 38%	1,484 2%
Aligned funds	31,726	24,024 76%	6,837 22%	865 1%
Unaligned funds	30,194	13,177 44%	16,398 54%	619 2%
Index funds removed				
	Total	SEC required	CQE	Non-CQE
All	58,623	35,251 60%	21,944 37%	1,428 2%
Aligned funds	30,191	22,826 76%	6,534 22%	831 3%
Unaligned funds	28,432	12,425 44%	15,410 54%	597 2%

Figure 1. Portfolio shifting consistency by management company

This figure reports how consistently management companies voluntarily report their noncalendar fiscal quarter-end funds' portfolios to Thomson. In each mandatory filing period, we label a management company one if all of its funds have fiscal quarter-end portfolios in Thomson or all its funds have calendar quarter-end portfolios in Thomson. A mixture of fiscal and calendar quarter-end reporting results in a label of zero. We then average this variable across filing periods by management company and report the distribution. The y-axis represents the percentage of fund companies. The x-axis is the average consistency value. Companies with an average of 100% (0%) never (always) selectively disclose voluntary portfolios for their funds.

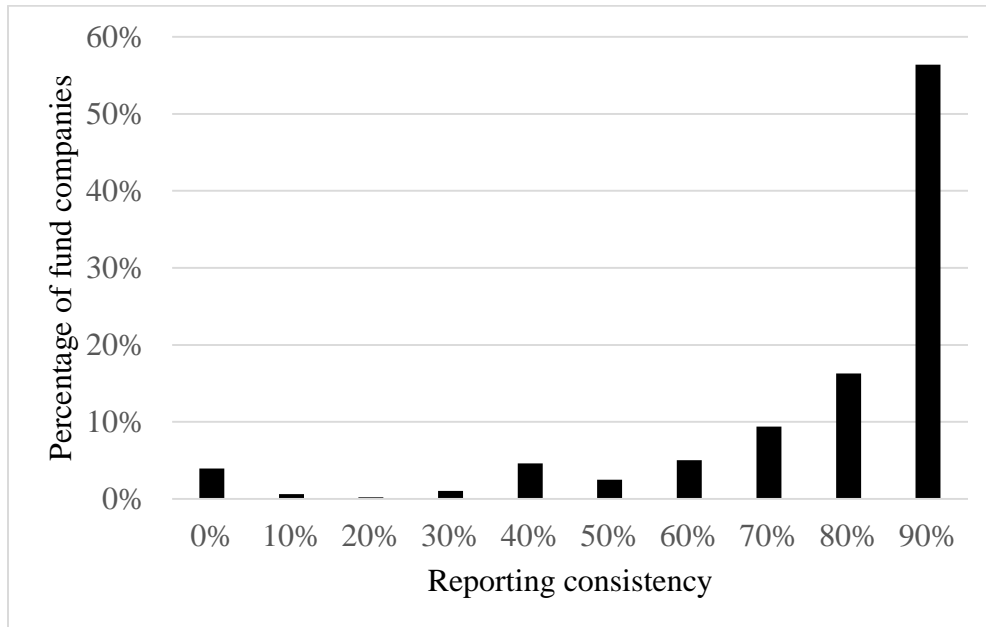


Table 6. Evaluation of manager skill

In this table, we report results comparing estimates of manager skill using SEC and Thomson data. In panels A and B, we report results related to future DGTW returns of the funds' portfolios. At the beginning of each quarter, we find the most recent portfolio disclosed for each fund and then compute its future one-, three-, and twelve-month raw and DGTW returns. In panel A, we then compute overall time-series averages and t-values using Fama-Macbeth (1973). In panel B, we place funds in quintiles each quarter based on their SEC and Thomson computed one- and three-month DGTW ranks. We then report a contingency table comparing quintiles in which funds are placed. In panels C and D, we report similar results, except for the return gap measure. ** and * denote significance at the 1% and 5% level, respectively.

A. Future returns

	Raw returns			DGTW returns		
	1 mo.	3 mo.	12 mo.	1 mo.	3 mo.	12 mo.
SEC	0.21	1.38	7.55	0.07	0.08	0.28
Thomson	0.16	1.43	7.59	0.02	0.09	0.27
Difference	-0.19	0.01	0.18	0.00	-0.01	0.00

B. Comparison of return quintile rankings

Future 1-month DGTW return						Future 3-month DGTW return					
Th/SEC	Low	2	3	4	High	Th/SE C	Low	2	3	4	High
Low	91	7	1	0	0	Low	91	7	1	0	0
2	7	80	10	1	1	2	7	81	9	2	1
3	1	10	78	10	1	3	1	10	80	9	1
4	1	2	9	81	8	4	1	2	9	82	7
High	0	1	1	7	90	High	0	0	1	7	91

C. Return gap

	1 mo.	2 mo.	3 mo.
SEC	0.00	0.03	0.11**
Thomson	0.00	0.03	0.08**
Difference	0.00	0.00	0.03**

D. Comparison of return gap quintile rankings

1-month return gap						3-month return gap					
Th/SEC	Low	2	3	4	High	Th/SE C	Low	2	3	4	High
Low	79	11	6	3	2	Low	85	9	3	2	1
2	13	69	10	5	2	2	10	76	10	4	1
3	3	11	70	12	3	3	2	9	75	11	2
4	2	5	9	70	13	4	2	4	9	74	11
High	2	4	5	10	79	High	1	2	3	9	85

Table 7. Performance characteristics of portfolio holdings

In this table, we present results on the performance characteristics of portfolios' holdings. For each month, we compute the weighted average return of the securities contained in each portfolio over the prior twelve months ($[-12, -1]$), the prior six months ($[-6, -1]$), the prior three months ($[-3, -1]$), and the prior month ($[-1, -1]$). We then compute the average return for the first month ($[0, 1]$), three months ($[0, 3]$), six months ($[0, 6]$), and twelve months ($[0, 12]$) after the portfolio's report date. We compute both raw returns and DGTW-adjusted returns. For each month, we then compute the average values for those portfolios disclosed through SEC filings and those portfolios only in Thomson and compute the differences. Positive values indicate the SEC filings' performance is higher. We report results using only the SEC-disclosed portfolios in Thomson (*In Thmn SEC Only*), as well as all SEC portfolios in our sample (*All SEC*). We use Fama-Macbeth (1973) to compute the average differences with standard errors computed using Newey-West (1987) with lags equal to the number of months used to calculate the return values. ** and * denote significance at the 1% and 5% level, respectively.

	Raw return difference		DGTW-adjusted return difference	
	All SEC	In Thmn SEC only	All SEC	In Thmn SEC only
$[-12, -1]$	1.07 (0.68)	1.27 (0.78)	0.71 (0.73)	0.88 (0.90)
$[-6, -1]$	1.00 (0.77)	1.24 (0.92)	0.88 (1.22)	1.02 (1.40)
$[-3, -1]$	1.72 (1.25)	1.70 (1.26)	1.05 (1.59)	1.12 (1.81)
$[-1, -1]$	5.88** (2.77)	6.18** (2.89)	1.86* (2.50)	2.11** (2.73)
$[0, 1]$	0.21 (0.09)	0.22 (0.09)	-0.08 (-0.11)	0.04 (-0.05)
$[0, 3]$	0.77 (0.90)	0.55 (0.64)	0.63 (1.69)	0.53 (1.38)
$[0, 6]$	0.77 (1.23)	0.66 (1.03)	0.28 (0.99)	0.26 (0.94)
$[0, 12]$	0.65 (1.31)	0.47 (0.92)	0.22 (0.75)	0.16 (0.59)

Table 8. Returns from front-running mutual fund flows

This table reports returns from front-running flow-induced mutual fund trading. For each quarter, we estimate the expected flow into each stock as the total number of shares mutual funds are expected to buy (sell), scaled by the stock's average daily volume over the prior six months. We then place stocks in decile portfolios based on their expected flow, where decile one has the lowest expected flow. For each quarter, we compute the value-weighted raw and DGTW returns for each portfolio for the next quarter. We compute returns for the entire sample period, as well as for the SEC-mandated semiannual disclosure period (1996–2004FH) and the SEC-mandated quarterly disclosure period (2004 SH–2008). We compute returns using only portfolios in Thomson and those in SEC filings. Public disclosure dates for the SEC dataset are the file dates with EDGAR, and the public disclosure dates for the Thomson dataset are sixty days after the portfolio's report date. Overall coefficients and t-values are computed using Fama-Macbeth (1973). ** and * denote significance at the 1% and 5% level, respectively.

All periods				
Decile	Raw returns		DGTW returns	
	SEC	Thomson	SEC	Thomson
1	0.38	0.66	-0.59	-0.47
2	0.03	0.08	-0.77	-0.37
5	1.18	1.04	0.48	0.37
9	1.93	2.23	0.49	0.50
10	3.03*	2.96	0.77*	0.51
10-1	2.65*	2.30*	1.36**	0.98*
Difference		0.35		0.38

SEC-mandated semiannual disclosure period				
Decile	Raw returns		DGTW returns	
	SEC	Thomson	SEC	Thomson
1	1.31	1.66	-0.67	-0.56
2	0.84	0.86	-0.96	-0.42
5	2.31	1.90	0.61	0.29
9	3.73	3.94*	0.97	0.83
10	5.02**	4.98**	1.40**	0.94*
10-1	3.71*	3.32*	2.07**	1.50*
Difference		0.39		0.57

SEC-mandated quarterly disclosure period				
Decile	Raw returns		DGTW returns	
	SEC	Thomson	SEC	Thomson
1	-1.32	-1.16	-0.43	-0.30
2	-1.46	-1.36	-0.43	-0.28
5	-0.88	-0.55	0.23	0.52
9	-1.37	-0.92	-0.40	-0.12
10	-0.62	-0.75	-0.38	-0.27
10-1	0.70	0.41	0.06	0.03
Difference		0.29		0.03

Table 9. Copycatting returns

This table reports results using disclosed portfolios to “copycat” mutual funds. For each, month we compute a fund’s equity portfolio return using the value-weighted returns of the securities held on its most recent publicly disclosed portfolio. Public disclosure dates for the SEC dataset are the file dates with EDGAR, and the public disclosure dates for the Thomson dataset are sixty days after the portfolio’s report date. We then combine this return with the cash and fixed income returns to compute the copycat gross return based on the fund’s portfolio weights in CRSP. This return is reduced by the monthly average of the copycat fund’s transaction costs, as well as by one-twelfth of 20 bps, which is an administrative fee. We then compute the difference between the actual fund return and the copycat return with a negative value equating to copycat fund outperformance. In panel A, we report the monthly averages, as well as the t-values computed using the Fama-Macbeth (1973) approach. We also report the average absolute differences, which represents tracking error. We additionally report the t-values for the difference between the Thomson and SEC values. In panel B, we compute the t-values for each fund’s difference between the copycat fund returns and actual fund returns and report the distribution of those t-values. In both panels, *SEC* represents copycat returns computed using SEC data, and *Thomson* represents copycat returns computed using Thomson data. *Semi* reports results during the SEC’s mandated semiannual disclosure period (1996–2004 FH), and *Qtr* reports results during the SEC’s mandated quarterly disclosure period (2004 SH–2008).

A. Copycat returns

	Difference					Absolute difference		
	Thomson		SEC		Diff	Thomson	SEC	Diff
	Avg.	t-value	Avg.	t-value	t-value	Avg.	Avg.	t-value
Entire period	-0.078	-6.84	-0.065	-4.28	1.29	0.808	0.851	-7.93
Semi.	-0.072	-4.21	-0.054	-2.28	1.17	0.938	0.995	-7.42
Qtr.	-0.088	-8.30	-0.085	-7.99	-1.18	0.590	0.608	-3.91

B. Distribution of t-values for individual fund return differences

	Thomson			SEC		
	All	Semi.	Qtr.	All	Semi.	Qtr.
Max.	3.97	3.62	4.11	3.87	3.87	4.02
99%	2.39	2.50	2.14	2.16	2.40	2.26
95%	1.36	1.66	1.17	1.30	1.58	1.16
90%	0.87	1.08	0.76	0.89	1.08	0.73
75%	0.04	0.26	-0.11	0.04	0.24	-0.03
50%	-0.94	-0.66	-0.91	-0.85	-0.59	-0.84
25%	-1.81	-1.55	-1.80	-1.69	-1.38	-1.67
10%	-2.71	-2.35	-2.64	-2.46	-2.11	-2.53
5%	-3.26	-2.86	-3.19	-3.02	-2.66	-3.02
1%	-4.44	-4.02	-4.50	-3.94	-3.81	-4.05
Min.	-6.53	-5.57	-7.97	-7.37	-6.41	-6.06

Table 10. Position and trading information

This table reports the amount of trading information available across datasets. *Positions* reports the total number of positions disclosed. *Trades* is the number of trades observed. We define a trade occurring whenever the number of shares held changes. *Average (median) trade size* is the average (median) number of shares in each observed trade. *Average (median) maximum trade delay* is average (median) maximum number of days between when a trade could have been executed and when it is observed in the data, which is the difference between the report dates of the portfolios in which the share change occurs. *Thomson* and *SEC* report the number of trades observed using only Thomson and SEC holdings data, respectively. *SEC + Thomson* combines the SEC and Thomson portfolios, and *CRSP + SEC + Thomson* also includes CRSP's portfolios. *All* reports the results for our 1996–2008 sample period. *1996–2004 FH* reports the results during the SEC's mandated semiannual disclosure period, and *2004 SH–2008* reports the results during the SEC's mandated quarterly disclosure period. CRSP data are only used in 2008, at which time they become more accurate.

Thomson						
	Positions	Trades	Trade size		Maximum trade delay	
			Average	Median	Average	Median
All	7,090,563	5,449,235	111,508	10,579	114	105
1996–2004 FH	3,256,972	2,503,730	114,788	12,289	128	120
2004 SH–2008	3,833,591	2,945,505	105,605	7,500	101	92
2008	858,118	666,574	104,385	7,100	98	92

SEC						
	Positions	Trades	Trade size		Maximum trade delay	
			Average	Median	Average	Median
All	6,096,870	4,821,097	127,045	16,141	128	126
1996–2004 FH	2,142,587	1,814,883	141,493	20,961	180	183
2004 SH–2008	3,954,283	3,006,214	101,038	7,466	97	92
2008	868,894	670,829	103,448	7,428	93	92

SEC + Thomson						
	Positions	Trades	Trade size		Maximum trade delay	
			Average	Median	Average	Median
All	8,772,355	6,335,042	102,232	7,566	95	98
1996–2004 FH	3,938,602	2,829,985	103,245	8,944	112	107
2004 SH–2008	4,833,753	3,505,057	100,408	5,086	81	91
2008	1,098,699	810,126	95,094	5,000	78	91

CRSP + SEC + Thomson						
	Positions	Trades	Trade size		Maximum trade delay	
			Average	Median	Average	Median
2008	1,283,154	899,512	89,967	3,557	66	62