Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Inferring Reporting-Related Biases in Hedge Fund Databases from Hedge Fund Equity Holdings

Vikas Agarwal, Vyacheslav Fos, Wei Jiang,

To cite this article:
Vikas Agarwal, Vyacheslav Fos, Wei Jiang, (2013) Inferring Reporting-Related Biases in Hedge Fund Databases from Hedge Fund Equity Holdings. Management Science 59(6):1271-1289. http://dx.doi.org/10.1287/mnsc.1120.1647

MANAGEMENT SCIENCE
Vol. 59, No. 6, June 2013, pp. 1271–1289
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1120.1647

Inferring Reporting-Related Biases in Hedge Fund Databases from Hedge Fund Equity Holdings

Vikas Agarwal
J. Mack Robinson College of Business, Georgia State University, Atlanta, Georgia 30303; and
Centre for Financial Research, University of Cologne, 50923 Cologne, Germany, vagarwal@gsu.edu

Vyacheslav Fos
College of Business, University of Illinois at Urbana–Champaign, Champaign, Illinois 61820, vfos@illinois.edu

Wei Jiang
Graduate School of Business, Columbia University, New York, New York 10027, wj2006@columbia.edu

This paper formally analyzes the biases related to self-reporting in hedge fund databases by matching the quarterly equity holdings of a complete list of 13F-filing hedge fund companies to the union of five major commercial databases of self-reporting hedge funds between 1980 and 2008. We find that funds initiate self-reporting after positive abnormal returns that do not persist into the reporting period. Termination of self-reporting is followed by both return deterioration and outflows from the funds. The propensity to self-report is consistent with the trade-offs between the benefits (e.g., access to prospective investors) and costs (e.g., partial loss of trading secrecy and flexibility in selective marketing). Finally, returns of self-reporting funds are higher than that of nonreporting funds using characteristic-based benchmarks. However, the difference is not significant using alternative choices of performance measures.

Key words: hedge funds; mandatory and voluntary disclosure; reporting and selection biases
History: Received May 23, 2011; accepted July 20, 2012, by Brad Barber, finance. Published online in *Articles in Advance* February 1, 2013.

1. Introduction

Hedge funds, unlike other financial institutions such as banks and mutual funds, have largely escaped regulation under the Securities Act of 1933 (Pub. L. 112-106) and the Investment Company Act of 1940 (Pub. L. 112-90) by raising capital via private placement from accredited institutions and high net worth individuals. Because of their lightly regulated nature, hedge funds are not required to report information about their characteristics, strategies, and performance to any authority or database. For this reason, the growing volume of research on hedge funds has mostly relied on commercial hedge fund databases to which hedge funds report voluntarily. Although the extant research has documented several biases in hedge fund databases, including the survivorship bias, backfilling bias, and smoothing bias (see §2.1 for a review), it has not formally addressed the degree of self-reporting bias, i.e., a selection bias that results from hedge funds' choices to not report to any database, to initiate reporting at some time, or to discontinue reporting. Our paper is the first to assess the extent of self-reporting bias in a comprehensive sample of hedge funds as well as to analyze the determinants of self-reporting.

A hedge fund's choice to voluntarily report to a commercial database should be determined by the cost-benefit trade-offs. In terms of benefits, listing in a database enhances a fund's exposure to potential investors, which is likely to be more significant for smaller fund companies that desire more publicity but lack the resources for aggressive direct marketing.[1] The main cost of reporting is a partial loss of secrecy and privacy that many funds value. Moreover, keeping the reporting status current constitutes a commitment to revealing a fixed set of information at fixed time intervals, depriving a fund of the flexibility in publicizing selective information (such as return performance over a particular period of time) that is most favorable to the fund. Conditional on a decision to report, a fund exercises its discretion on the reporting initiation date and later may choose to exit the database. Termination can be for positive as well as negative reasons. On the positive side, if a fund is closed to new investors because of its success and lack of scalable investment opportunities, then there would be no incentive to attract more capital.

[1] To be exempt from the regulations of the Securities Exchange Act of 1934 (Pub. L. 112-158) and the Investment Company Act of 1940, a hedge fund can only directly approach accredited and qualified investors. By reporting to a commercial database, a hedge fund shifts the burden of ensuring that only such investors have access to hedge fund information to the database vendors.

On the negative side, embarrassing losses or even the prospect of liquidation can also trigger exit from databases.

These scenarios indicate a potential selection bias among self-reporting databases. However, the magnitude, or even the direction of the bias, is hard to assess a priori (Fung and Hsieh 2000). This paper attempts to quantify the degree of the self-reporting bias in the hedge fund databases by analyzing the quarterly equity holdings of a complete list of hedge fund companies that filed Form 13F with the Securities and Exchange Commission (SEC) between 1980 and 2008. Due to the mandatory nature of the 13F filings, this sample is largely free from the selection bias caused by hedge funds' reporting incentives. Among all 13F-filing hedge fund companies, we determine their self-reporting status by matching them to the union of five major hedge fund databases: Center for International Securities and Derivatives Markets (CISDM), Hedge Fund Research Inc. (HFR), Eureka, Morgan Stanley Capital International (MSCI), and Trading Advisor Selection System (TASS). These union data represent the most comprehensive database of self-reporting hedge funds that has been used in the literature and hence minimizes the inaccuracy in the classification of funds' self-reporting status.

Our analyses consist of two steps. First, we analyze the return dynamics around the initial and last reporting dates and the impact of reporting on fund flows for the subsample of self-reporting funds. Conditional on self-reporting, we find that performance deteriorates significantly after both the reporting initiation and termination dates. The deterioration amounts to 73 and 28 basis points, respectively, using monthly market-adjusted returns imputed from quarter-end portfolio holdings. The first difference represents a bias that funds strategically initiate self-reporting after a run of superior performance; the second indicates that reporting termination or "delisting" is usually a sign of performance deterioration. The latter point is further supported by the fact that net flows to funds tend to decrease after reporting termination, after controlling for performance. Moreover, the termination-related timing bias is likely to be an underestimate, as extremely underperforming funds that disappear from commercial databases may also drop out of the 13F database because of substantial drops in the value of their U.S. equity portfolios.

Second, we compare the performance and other characteristics of the self-reporting hedge funds to those of the nonreporting ones. Unconditionally, we find that young and medium-sized fund companies that employ more diversified and higher-turnover trading strategies have a stronger incentive to self-report, presumably to publicize their funds and attract potential investors. Trading secrecy is less likely to be revealed through voluntary disclosure for these funds because of their diversified nature and the high portfolio turnover rates, reducing the costs of reporting. Returns of self-reporting funds are higher than those of nonreporting funds using characteristic-based benchmarks. However, the difference is not significant using alternative choices of performance measures. A lack of consistent difference could be because the positive and negative reasons prompting reporting initiation and termination largely offset one another.

In a closely related work, Aiken et al. (2013) assess self-reporting bias using a relatively small sample of funds of hedge funds based on the premise that their returns and holdings contain information of nonreporting hedge funds and of hedge funds that terminate reporting, and that selection by these funds of the underlying funds is random. Our study of the "timing bias" is also related to prior work by ter Horst and Verbeek (2007) and Hodder et al. (2008); both studies support a positive correlation between negative performance and reporting termination. Ter Horst and Verbeek (2007) uncover the survivorship bias indirectly applying a selection model on a commercial database. Hodder et al. (2008) resort to data from funds of hedge funds, assuming some independence between the component funds' self-reporting status and the umbrella funds' investment decision.

Our approach avoids these limitations using a comprehensive sample of hedge funds that are required to report their positions in the 13F securities. Needless to say, this approach has its own limitations, as it relies on the quarter-end long-equity positions at the hedge fund company (rather than at the individual fund) level; ignores intraquarter trading, hedge fund fees, and trading costs; and can be potentially affected by strategic behavior of hedge funds to reduce the information content of their holdings to "front runners" and "copycats" (Brown and Schwarz 2012). Without knowledge about the returns for the short equity or nonequity positions of hedge funds, our study provides a direct estimate for the reporting bias in the equity component of the hedge fund returns. Findings from alternative approaches can be viewed as complementary to obtain a more complete picture of the self-reporting biases.

Our results offer benchmarks and references for hedge fund researchers and investment managers who use self-reported data sources. More generally, the study provides insights into the motivation and consequences of voluntary disclosure by hedge funds. Mandatory disclosure may help investors and regulators complement the information from voluntary disclosure to assess the performance and operational

risk involved in hedge fund investing. This is particularly pertinent in view of the ongoing debate regarding more stringent disclosure rules for hedge funds, including their mandatory registration.

2. Literature Review and Data Description

2.1. Review of Related Literature

Most of the academic research on hedge funds relies on self-reported data that are subject to several biases. The first of these is the survivorship bias that arises from databases containing information only about surviving funds. However, unlike mutual funds, where such bias originates from poorly performing funds dropping out of the databases (e.g., Brown and Goetzmann 1995, Malkiel 1995, Elton et al. 1996),[2] both well-performing and poorly performing hedge funds may stop reporting, as the former have fewer incentives to attract capital by disseminating information about their performance (Brown et al. 1999; Fung and Hsieh 2000, 2002, 2009; Liang 2000). The second type of bias, closely related to survivorship bias, is the multiperiod sampling bias or look-ahead bias that results from researchers conditioning on funds' survival for econometric analysis of performance persistence and other issues (see, e.g., Brown et al. 1992 and ter Horst et al. 2001 for mutual funds, and Fung and Hsieh 2000 and Baquero et al. 2005 for hedge funds). The third bias—instant history or backfill bias—is somewhat unique to hedge funds. It is a consequence of hedge funds choosing to "backfill" their historical performance when initiating reporting, usually subsequent to good performance (e.g., Fung and Hsieh 2000, Edwards and Caglayan 2001). The fourth bias, closely related to the backfill bias, is the incubation bias, which results from hedge funds using internal capital to build up their track records and reporting better performance over the incubation period. This has been studied by Evans (2010) for mutual funds and Avramov et al. (2011) for hedge funds. Fifth, there is stale price or return smoothing bias that can arise from hedge funds investing in illiquid securities or intentionally smoothing their returns to reduce their volatility (e.g., Asness et al. 2001, Getmansky et al. 2004, Bollen and Pool 2008, Cassar and Gerakos 2011). Finally, self-reporting bias, the subject of our study, arises from hedge funds self-selecting to report to commercial databases. Our paper contributes to this burgeoning literature on hedge fund biases, especially the last one.

[2] More recently, Linnainmaa (2013) documents a reverse survivorship bias in mutual funds when the poor performance is driven by negative idiosyncratic shocks.

2.2. Collection of Hedge Funds

The key inputs to our analyses are data from two sources. The first is the 13F quarter-end equity holdings data from the Thomson Reuters Ownership Data. The Form 13F filing, which discloses quarter-end holdings of an institution with a maximum 45-day delay, is mandatory for all institutions that exercise investment discretion over $100 million of assets in equity and other publicly traded securities.[3] The second source is a comprehensive self-reported hedge fund database created by merging five major commercial hedge fund databases: CISDM, Eureka, HFR, MSCI, and TASS (henceforth, the "Union Database"). Throughout the paper, we call a hedge fund company that appears in the first database a "13F-filing hedge fund company" and a fund that appears in the second data source a "self-reporting hedge fund." It is worth noting that the level of reporting is often different between the two data sources. The 13F filings are aggregated at the institution level, comparable to the level of management companies or sponsors of hedge funds. The reporting unit in the self-reporting databases is usually at the fund or pooled portfolio level. Hence, pairing a 13F-filing institution to funds in the Union Database is often a one-to-multiple match (if a match exists), based on the information about the management companies of individual funds in the Union Database.

The Thomson Reuters Ownership database consists of 5,188 unique 13F-filing institutions for the 1980–2008 period, which we classify into the following four categories in addition to hedge funds: (1) banks and insurance companies (a combination of type 1 and type 2 institutions by the Thomson classification); (2) mutual fund management companies (type 3 institutions by the Thomson classification); (3) independent investment advisors (type 4 institutions by the Thomson classification, excluding hedge funds classified by us); and (4) others (type 5 institutions by the Thomson classification, excluding hedge funds classified by us).[4]

There is no official definition of a hedge fund. We adopt the generally accepted notion of hedge

[3] Institutions are required to disclose all securities that appear on the official list of "Section 13(f) Securities," published by the SEC periodically and including almost all publicly traded equity, some preferred stocks, bonds with convertible features, warrant, and exchange-traded call and put options. The Thomson Reuters database contains only holdings of equity. Some institutions can request to defer the disclosure of certain long positions by providing justification for each of those positions. Agarwal et al. (2013) find that these amendments constitute about 3% of the filings that are immediately disclosed.

[4] Since 1998, the Thomson Reuters type 5 is known to be problematic. Therefore, we reassign an institution that has type 5 after 1998 to an earlier code if available that is different from 5 and manually reclassify the remaining institutions.

funds as pooled private investment vehicles that adopt performance-based compensation and are operated outside of the securities regulation and registration requirements. As such, we manually classify a 13F-filing institution as a "hedge fund company" if it satisfies one of the following: (1) it matches the name of one or multiple funds from the Union Database; (2) it is listed by industry publications (e.g., Hedge Fund Group, Barron's, Alpha Magazine, and Institutional Investors) as one of the top hedge funds; (3) the company's website claims it to be a hedge fund management company or lists hedge fund management as a major line of business;[5] (4) the company is featured by news articles in Factiva as a hedge fund manager/sponsor; or (5) some 13F filer names are those of individuals. In such cases, we search the individuals' names over the Internet (mostly through the filer and cofiler identity information on various types of SEC filings) and classify the name as a hedge fund if the person is the founder, partner, chair, or other leading personnel of a hedge fund company.[6]

Applying the above procedure yields 1,199 unique hedge fund companies among all 13F-filing institutions. Although this number is low relative to the universe of hedge fund companies (our Union Database consists of 5,342 companies), the total value of equity positions held by 13F hedge funds is $1.25 trillion, which is 83% of the size of the hedge fund industry as of 2008, according to Credit Suisse/Tremont. The difference between the number of hedge fund companies in 13F and commercial databases is mostly a result of the minimum requirement of $100 million in 13(f) securities (mainly equities) for 13F-filing institutions, which rules out smaller hedge fund companies and most of the companies specializing in nonequity strategies. Because we use the long-equity holdings for our analysis, it is comforting to notice that the largest percentage of our sample funds belongs to the "equity" or "equity long/short" category (38.4%). Other major categories include event driven (10.2%), sector (5.4%), and multistrategy (5.7%), which are also likely to have substantial equity exposure.

We restrict our sample to relatively "pure-play" hedge funds (such as Renaissance Technologies and Pershing Square, and investment companies where hedge funds represent their core business, such as D. E. Shaw and the Blackstone Group/Kailix Advisors) and do not include full-service banks whose investment arms engage in hedge fund business (such as Goldman Sachs Asset Management and UBS Dillon Read); nor do we include mutual fund management companies that enter the hedge fund business, a new phenomenon in recent years (Agarwal et al. 2009a, Cici et al. 2010, Nohel et al. 2010). The reason for the exclusion is to ensure that the 13F filings are informative about the investments of the institutions' hedge funds. Our results are qualitatively similar if we include the full-service institutions with major hedge fund business in the sample, except that their presence will skew the statistics related to portfolio size because they tend to be much larger than other hedge funds in the list.

Because of our top-down approach, our list of 13F-filing hedge funds companies is comprehensive and is considerably longer than that used in prior literature. For example, Brunnermeier and Nagel (2004) analyze the role of hedge funds during the late 1990s technology bubble with a sample of 53 hedge fund companies, and Griffin and Xu (2009) examine the portfolio characteristics and performance of 306 companies. In both papers, the authors used a one-sided match from published hedge fund lists to the 13F database and did not classify hedge funds that failed to make it to a major published list or that chose not to report to any database. Given that the focus of this paper is the selection bias, it is particularly important that we have a complete list of 13F-filing hedge funds. Equally important is a comprehensive sample of self-reporting hedge funds, given that a key variable of our analysis is the self-reporting status of a hedge fund. Most of the research in the area of hedge funds has been conducted using one or more of the self-reported databases.[7] More recently, Agarwal et al. (2009b) show that there is limited overlap among four commercial databases, which is confirmed by a Venn diagram for the five components of our Union Database of 11,417 hedge funds (including 6,245 equity-oriented funds) (see Figure 1). One of the most striking observations from Figure 1 is that 71% of the funds are covered exclusively by only one database.[8] This underscores

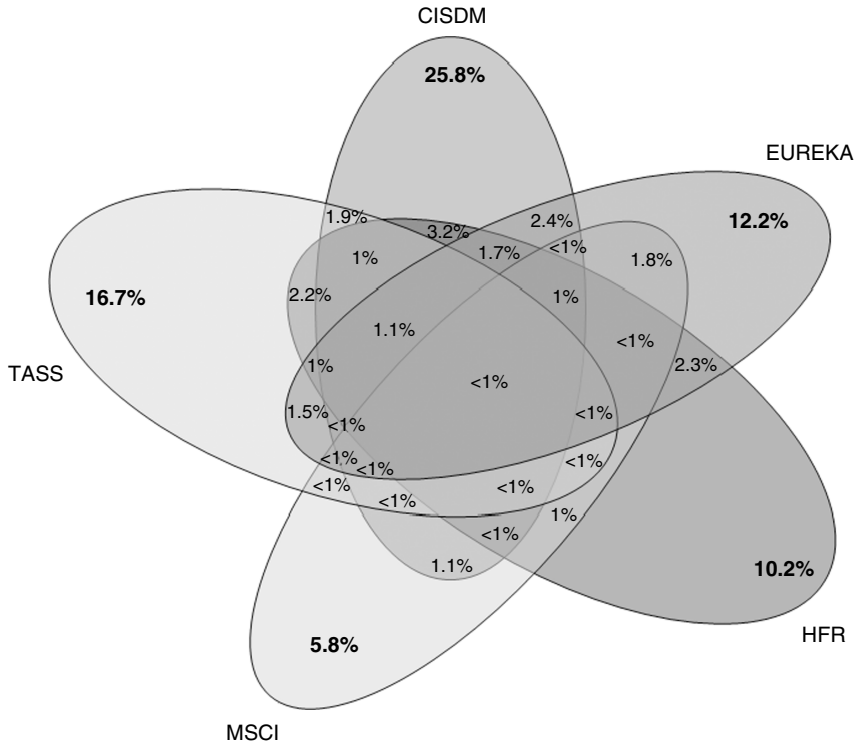
[5] Even if a company's website does not formally mention hedge fund management as part of its business, we still classify the company as a hedge fund manager or sponsor if it manages investment vehicles whose descriptions fit our definition of hedge funds. We exclude private equity and venture capital businesses that also have performance-based compensation.

[6] Notable examples in this category include Carl Icahn (founder and chair of the hedge funds, Icahn Capital, L.P. and Icahn Partners) and George Soros (founder and chairman of Soros Fund Management, a hedge fund management company).

[7] Fung and Hsieh (1997) used monthly data from TASS Management and Paradigm LDC; Ackermann et al. (1999) used a combination of HFR and Managed Account Reports, Inc., databases; Liang (1999) used HFR data; and Liang (2000) compared the HFR and TASS databases for different biases in the data.

[8] A major determinant in the choice of databases to which funds report is the subscriber clientele of the databases (in terms of both characteristics and geography). The additional cost of multiple reporting includes the different requirements imposed by different data vendors, such as the types of data fields, availability of audited financial statements, etc. Jorion and Schwarz (2013) find that funds report to multiple databases after good performance and when they need capital.

Figure 1 Venn Diagram of the Union Database



Notes. The Union Database contains a sample of 11,417 hedge funds by merging the following databases: CISDM, Eureka, HFR, MSCI, and TASS. This figure shows the percentage of funds covered by each database individually and by all possible combinations of multiple databases.

the importance of using multiple databases to accurately classify reporting status and record the timing of reporting initiation and termination. Using multiple databases also enables us to resolve occasional discrepancies among different databases.

2.3. Classification of the Self-Reporting Status of Hedge Funds

We classify the self-reporting status of all 1,199 13F-filing hedge fund companies by matching them to the Union Database. The classification entails two steps. In the first step, we match by name allowing minor variations,[9] producing 645 self-reporting companies, or 53.8% of all 13F-filing companies. Self-reporting fund companies may report for a period shorter than our sample span of 1980–2008. A lower percentage—33%—of our sample hedge fund companies are reporting, on average, at any given month during our sample period.

In the second step, we compute the correlation between returns imputed from the 13F quarterly holdings (henceforth, "13F portfolio returns") and returns reported in the Union Database (henceforth, "self-reported returns"). For the former, we compute the monthly returns of a fund company, assuming it holds the most recently disclosed quarter-end holdings. For the latter, we compute the value-weighted average monthly returns of all funds reported in the Union Database that belong to the same fund management company. We find that 60 pairs (or 9.3% of the 645 self-reporting fund companies) have negative correlations,[10] and for 219 pairs, the correlation is not defined because of the lack of overlapping periods of data from both data sources.[11] The self-reporting status of these funds is not convincingly established; therefore, we exclude them from our main analysis (i.e., they are considered neither self-reporting nor nonreporting). As a result, we end up with 366 self-reporting funds and 554 nonreporting funds. We note that the 554 nonreporting funds hold long-equity positions totaling \$570 billion, or 45.6% of the long-equity positions held by all the 1,199 13F hedge fund companies. Hence, commercial databases

[9] For example, "DKR Capital" from the 13F list is matched to "DKR Capital Inc." in the Union Database.

[10] Griffin and Xu (2009) reported the same percentage number in their sample as 8.5% and discussed various reasons for correlation being less than one.

[11] Several factors contribute to the lack of data overlap between the two databases: hedge funds might initiate reporting toward the end of our sample period; there are occasional missing return data in the Union Database; and some reporting hedge funds do not file 13F in all periods because of their small size.

do not provide information about hedge fund companies that hold almost half of the aggregate equity positions, underscoring the importance of using the 13F source as a supplement to examine hedge fund performance.

Once we identify the self-reporting status of hedge fund companies and the periods during which they report to the Union Database, our analyses almost exclusively rely on the information from 13F filings where the unit of observation is at the hedge fund management company level, which we henceforth interchangeably call "hedge funds" when there is no confusion. Comparing the 13F portfolio composition and return performance of self-reporting with nonreporting funds sheds light on the selection bias introduced by self-reporting among the equity-oriented hedge funds that meet the hurdle of managing $100 million in 13(f) securities.

Needless to say, the 13F database captures only the quarter-end long-equity portfolios of hedge fund companies. Therefore, our results inform about the reporting bias associated with total portfolio returns of individual funds only if the long-equity positions are a substantive portion of the portfolios of equity-oriented hedge funds and the returns imputed from quarter-end equity long positions are informative about the total returns of these hedge funds. Earlier work by Brunnermeier and Nagel (2004) and Griffin and Xu (2009) relies on the same premise, which we believe is valid, on average, for several reasons.

First, among the self-reporting companies, we find that the average return correlation between their 13F holdings (equity-long positions only, and before fees) and their fund returns reported to hedge fund databases (aggregated at the company level and including returns from short positions and nonequity securities, and net of fees) is 0.54; the median number is slightly higher at 0.57, and the interquartile range is 0.34–0.77.[12] The median slope in a regression of 13F returns on fund returns is 0.91, close to unity. We calculate the correlation using an average duration of data overlap of 12 years between a fund's appearance in the Union Database and in the 13F database. Both figures are comparable to the correlation of 0.55 (mean) and 0.64 (median) reported by Griffin and Xu (2009). Moreover, using a proprietary data set of funds of hedge funds, Ang et al. (2010) reported that funds following equity- and event-driven strategies (which constitute a great majority of our sample) mainly invest in equity and distressed corporate debt and hence have relatively low leverage.

Second, the contribution of equity positions to the total returns of funds is evident from the equity market betas of funds. Using the monthly Credit Suisse/Tremont hedge fund indices from January 1993 to May 2009,[13] we find that the market beta of the index of all equity-oriented funds is 0.48. Similarly, the average market beta from the Carhart (1997) four-factor model of the return index of all the self-reporting funds in our sample is 0.40.

Finally, the constant resistance of hedge funds against ownership disclosure, including the 13F filings, implies that the equity positions are critically informative of their investment strategies. Philip Goldstein, an activist hedge fund manager at Bulldog Investors, likens his stock holdings to "trade secrets" as much as the protected formula used to make Coke, and condemns the 13F rule for taking the fund's "property without just compensation in violation of the Fifth Amendment to the Constitution" (McCormack 2006). In the wake of the "quant meltdown" in August 2007, 13F filings that publicize equity positions of major quant hedge funds took much of the blame for inviting "copycats" into the increasingly correlated and crowded strategy space. It led to many funds employing similar strategies attempting to cut their risks simultaneously in response to their losses (Khandani and Lo 2007). Agarwal et al. (2013) presents comprehensive evidence of strategic delays by hedge funds in their 13F disclosure.

2.4. Overview of Hedge Funds Using Quarter-End Equity Holdings Data

Before we compare self-reporting hedge fund companies to nonreporting ones, we take advantage of the complete list of 13F-filing hedge funds to report the summary statistics of their equity-portfolio characteristics and the return performance of their long-equity positions. Furthermore, we compare their statistics with those of other categories of 13F-filing institutional investors (especially mutual funds) without tabulation. All the differences discussed in Table 1 are significant at the 1% level.

The average (median) equity position size of the hedge funds in our sample is $1,041 ($368) million, much smaller than institutions of other categories (about 16.5% of that of a mutual fund management company). Hedge funds also tend to be younger. The median hedge fund started 13F filing in 2002, compared to 1985 for mutual funds and 1995 for other investment advisors. Three measures point uniformly to the active nature of hedge funds in portfolio management. With a median portfolio Herfindahl index of 0.047, hedge funds are much less diversified than

[12] A further investigation reveals that the 10 hedge fund companies that exhibit the highest return correlations (ranging from 0.96 to 0.99) all have funds in equity-oriented strategies, including long/short equity, equity hedge, event driven, and sector.

[13] Available from http://www.hedgeindex.com/hedgeindex/en/default.aspx?cy=USD (accessed January 12, 2013).

Table 1 Summary Statistics of Hedge Funds' Equity-Portfolio Characteristics and Performance

	Mean	Median
Portfolio size ($, million)	1,041	368
Portfolio Herfindahl Index	0.0953	0.0465
Monthly return volatility	0.0553	0.0493
Annualized portfolio turnover rate	0.9162	0.8149
Inception year	1999	2002
Market-adjusted return	0.0008	0.0011
One-factor alpha	−0.0006	−0.0002
Four-factor alpha	−0.0020	−0.0011
DGTW abnormal return	0.0067	0.0057
Market factor	1.0917	1.0553
SMB (*small minus big*) *factor*	0.3344	0.2861
HML (*high minus low*) *factor*	0.0781	0.0706
Momentum factor	−0.0126	−0.0047
R-square from four-factor model	0.8538	0.8775
Number of institutions	1,199	

Notes. This table reports the mean and median of the characteristics and return performance of the long-equity portfolios of hedge funds. The "Hedge fund" category is manually classified (see §2.1). The portfolio size is calculated as the total value of quarter-end equity portfolio using reported shares and corresponding quarter-end stock prices reported in CRSP. The *Portfolio Herfindahl index* is the Herfindahl index of the disclosed quarter-end equity holdings. The *Monthly return volatility* is the volatility of the imputed portfolio return, which is same as defined in Figure 2. The *Annualized portfolio turnover rate* is compounded from the quarterly turnover rates, calculated as the lesser of purchases and sales, divided by the average portfolio size of the last and the current quarters. The *Inception year* is the year of the institution's first appearance in Thomson Reuters (censored at 1980). The *Market-adjusted return* is the same as defined in Figure 2. *One-factor alpha* and *Four-factor alpha* are the monthly intercepts from CAPM one-factor and Carhart (1997) four-factor models using all available data. *DGTW abnormal return* is the Daniel et al. (1997) characteristic-based benchmark-adjusted return. *Market factor*, *SMB factor*, *HML factor*, *Momentum factor*, and *R*-squares are estimated factor loadings from and explanatory power of four-factor model. The sample period is 1980–2008.

mutual funds (median = 0.018). Second, hedge funds' median monthly portfolio return volatility (4.93%) is higher than that of mutual funds (4.48%). Third, the average (median) of hedge funds' interquarter portfolio turnover rate is 91.6% (81.5%) annually, about twice as high as that of mutual funds, investment advisors, and other institutions, and more than three times that of bank and insurance companies. Here, the portfolio turnover rate is compounded from the interquarter turnover rates,[14] calculated as the lesser of purchases and sales, divided by the average portfolio size of the last and the current quarter.[15] The comparison between hedge funds and mutual funds in terms of portfolio concentration and turnover rates is consistent with Griffin and Xu's (2009) findings using similar measures.

Does the more active management of hedge funds bring about superior returns for their long-equity positions? To address this question, we compute the monthly market-adjusted return for each institution as the difference between the imputed portfolio return and CRSP value-weighted equity market return. For the former, we assume that in each month, the institution holds the portfolio disclosed at the most recent past quarter-end[16] and calculate the buy-and-hold return for the month. It turns out that all categories, including hedge funds, have average and median market-adjusted returns close to zero.[17] Finally, both the mean and median *R*-squares from the four-factor model are significantly lower for hedge funds, compared with each of the other institutional categories. This evidence, combined with greater overall portfolio volatility, suggests that hedge funds exhibit higher idiosyncratic volatility. We analyze the performance within hedge fund group in more detail in the following sections.

3. The Economics of Self-Reporting: Hypothesis Development

Like other economic activities, the reporting behavior of hedge funds is an outcome of cost-benefit tradeoffs. The benefit most cited by hedge fund data vendors in marketing their services to hedge funds is that listing in a database enhances a fund's exposure to potential investors, including fund of funds, foundations, banks, endowments, pensions, consultants, and high net worth individuals. Such benefits are likely to be more significant for smaller fund companies that desire more publicity but lack the resources for aggressive direct marketing. The main cost of reporting is a partial loss of secrecy and privacy that some hedge funds value. The SEC's efforts to push for more disclosure by hedge fund companies have faced strong resistance, indicating the industry's

[14] It is possible that some hedge funds may be very high-frequency traders by actively trading within the quarter and therefore may not report any long-equity positions at the end of a quarter. However, this will only result in our underestimating the actual portfolio turnover rates of such hedge funds.

[15] We follow the practice of Morningstar, the leading mutual fund research company, in defining portfolio turnover rates. It is worth pointing out that our turnover figures for mutual funds are lower than those reported in the Morningstar database because the 13F data do not account for intraquarter trading, which may significantly contribute to the funds' turnover. Purchases (sales) are calculated as the sum of the products of positive (negative) changes in the number of shares in the holdings from the previous quarter-end to the current quarter-end and the average of the stock prices at the two quarter-ends. The logic of using the *lesser* (rather than average) of purchases and sales is to free the measure from the impact of net flows.

[16] We code the monthly return as missing if the lag between the current month and the last quarter-end when the portfolio information is available exceeds six months.

[17] Given that institutions as a whole hold a majority stake in public equities (percentage increased from 32% in the beginning to 66% to the end of our sample period), it is not surprising that on average they simply perform at par with the market.

general reluctance for or even strong opposition to more transparency. Although self-reporting hedge funds in general do not reveal holdings information to hedge fund databases, the reported information, such as style classification, asset allocation, monthly returns, and leverage/hedging ratios, often reveals the funds' investment strategy. For example, proposed "hedge fund replication" strategies that promise low-cost hedge fund exposure are mostly built on the self-reported information (Kat and Palaro 2006). Moreover, keeping the reporting status constitutes a commitment to revealing a fixed set of information at fixed time intervals. The rigid schedule reduces a hedge fund company's flexibility in marketing, such as featuring a subset of information or a chosen period of return performance that is most favorable to the fund.

An additional cost is related to the clientele of database subscribers. Potential long-term investors targeted directly by hedge funds (mostly large institutions, fiduciaries, and some funds of funds) are different from those attracted to hedge funds through database subscription, which tend to be more "shorter-term" driven, consisting disproportionately of small institutions and individuals. Stulz (2007) mentions that retail investors may require more "hand-holding" subsequent to poor performance. Mutual fund literature also provides some evidence on institutional money being more "sticky" than retail in that the former does not chase short-term performance as much as the latter (Chen et al. 2010). Hedge funds usually value long-term investors whose investing or divesting decisions are not sensitive to short-term performance. Hence, some hedge funds may not want to be exposed to the clientele that are typical of database subscribers.

Although it is understandable that funds may not desire to appear in commercial databases during periods of poor performance because they do not wish to publicize the embarrassment, it is much less clear whether reporting funds are overall better or worse performers than nonreporting ones. On one hand, the extreme poor performers may be unlikely to appear in a database simply because they do not survive long enough to satisfy the requirement for track records by most data vendors. In addition, some successful hedge funds may prefer to voluntarily report as it serves as a strong signal for better transparency and institutional quality. On the other hand, very successful funds can also shun reporting, given their low need for enhanced visibility and possibly full capacity. Furthermore, Lhabitant (2006) offers an explanation for the general absence of the largest and most successful hedge funds in the commercial databases: these funds might be concerned that communicating performance to a data vendor may lead to inclusion in that data vendor's index (often averaged over member funds weighted by assets), which automatically raises the performance of index. As a result, these hedge funds' individual performances will appear less differentiated.

4. Biases Conditional on Self-Reporting: Reporting Initiation and Termination

We start with the first type of selection bias concerning the subsample of self-reporting funds: When do fund companies initiate reporting and when do they terminate? If funds tend to choose reporting initiation after a run of superior performance or to terminate reporting following subpar returns, examining the performance of funds while they appear in the database can contribute to a "timing bias." Until now, the literature has not been able to quantify these two forms of timing bias, as the performance of funds "before birth" and "after death" with respect to the databases is not observable. Because 13F filings are not constrained by funds' reporting status to commercial databases, we are able to assess these two biases.

4.1. First Form of Timing Bias: Comparison of Fund Companies Before and After the Reporting Initiation

The Union Database provides the dates when the hedge funds enter the commercial databases. If a fund company reports to multiple constituent sources in the Union Database, we use the earliest date. Because the 13F reporting is at the company level and reporting to the commercial databases is at the fund level, we assume that the first (last) reporting date of a company is the earliest and latest of the first (last) reporting dates of all funds in a company. Among all 13F-filing hedge fund companies, 103 of the 366 self-reporting funds afford the before–after analysis if we require a minimum of 12 months of return information around the initial reporting date and the existence of such information on both sides of the date. For 77 funds, there is accurate information on the initial reporting dates provided by one commercial database. This last subsample is the key in assessing the initiation-related timing bias. For each of the 77 funds, we compare the return measures (imputed from the 13F holdings) during the 24-month period before reporting to the Union Database and the 24-month period thereafter (or as many months as possible, subject to a minimum of 12 months in total on both sides of the reporting initiation month). We report the results in Table 2.

We observe from the numbers in Table 2 that the performance after initial reporting is significantly lower than that before reporting. The average

Table 2 Comparison of Return Performance Before and After the Initial Reporting Date

	(1) *Raw return*	(2) *Market-adjusted return*	(3) *One-factor alpha*	(4) *Four-factor alpha*	(5) *Difference-in-difference*	(6) *DGTW abnormal return*
Before initial reporting						
Mean	0.0160	0.0059	0.0034	0.0007	0.0024	0.0093
Median	0.0161	0.0033	0.0018	0.0011	0.0012	0.0088
No. of funds	77	77	76	76	76	75
After initial reporting						
Mean	0.0070	−0.0014	−0.0024	−0.0017	−0.0033	0.0071
Median	0.0112	0.0001	−0.0014	−0.0008	−0.0014	0.0058
No. of funds	76	76	76	76	76	76
Differences (*t*-statistics)						
Mean	−0.0090***	−0.0073***	−0.0058***	−0.0024	−0.0057***	−0.0022*
	[−3.09]	[−3.32]	[−2.85]	[−1.42]	[−2.82]	[−1.83]
Median	−0.0049***	−0.0032**	−0.0033**	−0.0019	−0.0026**	−0.0030**
	[−2.88]	[−2.36]	[−2.51]	[−1.33]	[−1.97]	[−2.53]

Notes. This table compares the return measures for fund companies during the 24-month periods before and after the initial reporting date for a subsample of funds where there is accurate information on the initial reporting date. *Raw return* and *Market-adjusted return* are the portfolio returns without adjustment, and in excess of the CRSP value-weighted return, respectively. *One-factor alpha* and *Four-factor alpha* are the intercepts from CAPM and Carhart (1997) four-factor model, using pooled 48-month period for estimating the betas. Alphas are coded as missing if there are fewer than 12 observations for estimation. The *Difference-in-difference* is the difference around the initial reporting date between raw returns of reporting and nonreporting funds. *DGTW abnormal return* is the Daniel et al. (1997) characteristic-based benchmark-adjusted return. The *t*-statistics for the differences between the two samples are based on bootstrapped standard errors and are reported below difference estimates in brackets.

***, **, *Significant at the 1%, 5%, and 10% levels, respectively.

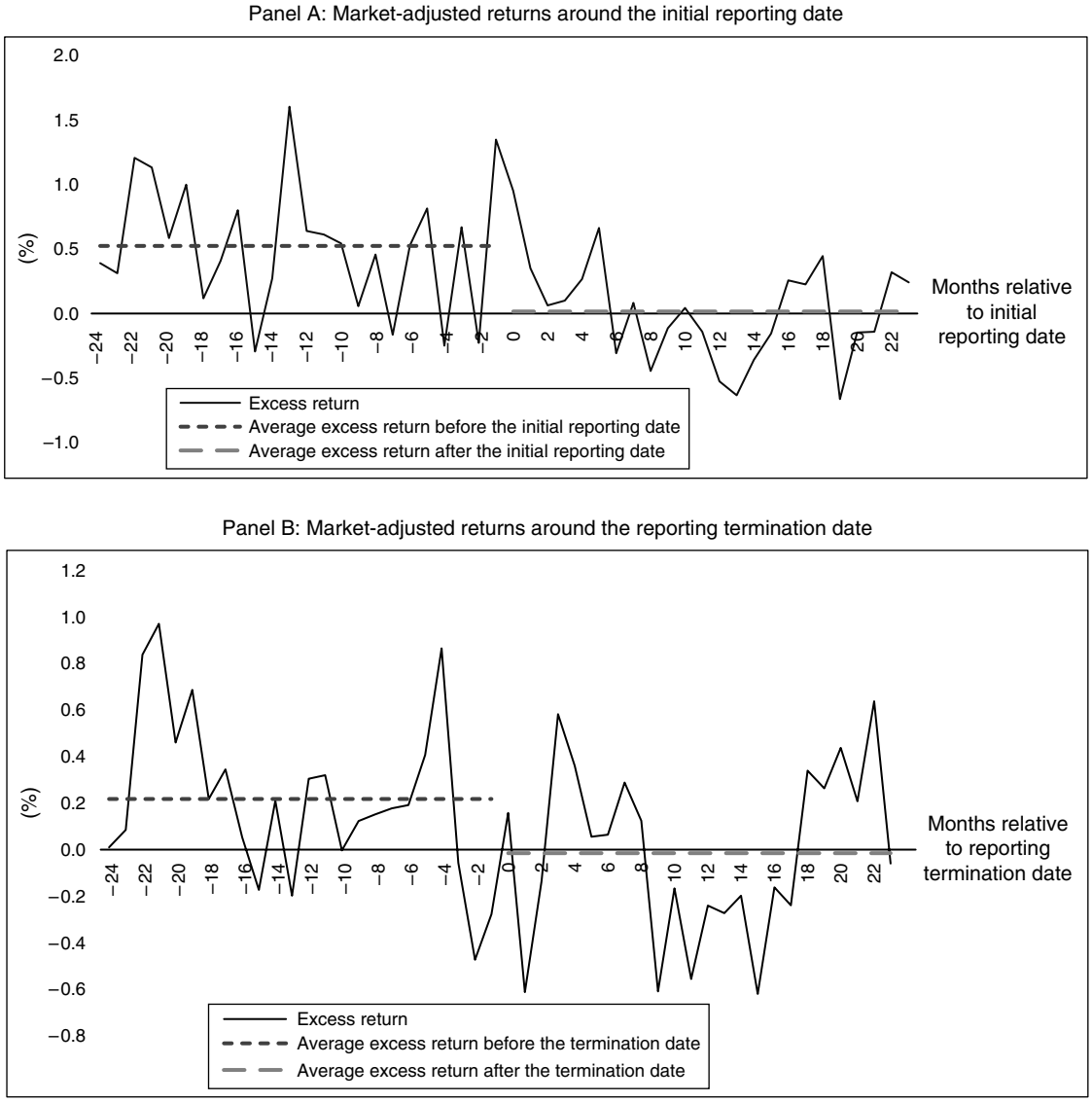
(median) raw monthly return is lowered by 90 (49) basis points, and measures of risk-adjusted performance—market-adjusted returns, capital asset pricing model (CAPM) alpha, four-factor alpha, and the Daniel et al. (1997) characteristic-based benchmark-adjusted return (henceforth, the "DGTW abnormal return")—deteriorate by 22–73 basis points per month on average, or 19–33 basis points per month using median values. The magnitude is economically significant, and all the four average differences except four-factor alpha are also statistically significant (mostly at the 5% level). Finally, a difference-in-difference approach, which computes the difference around the initial reporting date between raw returns of reporting and nonreporting hedge funds, also indicates significant deterioration using both the median and mean values. The results reported in Table 2 use bootstrapped standard errors but do not explicitly control for cross-sectional dependence of returns. For robustness, when we estimate ordinary least squares regressions of the different performance measures on an indicator variable for reporting and cluster the standard errors at the monthly level, the *t*-statistics do shrink by 30%–35%, but none of our general conclusions change.

Panel A of Figure 2 plots the time series of the average monthly market-adjusted returns of the 77 hedge funds from 24 months before the reporting month to 24 months afterward. The two dotted horizontal lines mark the time-series averages of the two subperiods. The figure confirms that funds choose to initiate self-reporting after a run of superior performance, but such performance does not persist in that it mean-reverts to levels at par with the market after reporting initiation.

Applying the same method on all hedge funds that report to the Union Database during our sample period, including the additional 26 funds for which we observe the first date of performance data but not the accurately recorded first reporting date, we continue to find that performance is overall worse (by 52 basis points monthly) after initial reporting compared with the period before, though the difference is not statistically significant. Following the practice in the literature (e.g., Ackermann et al. 1999), we add 24 months to the first performance dates to form the approximate first reporting dates for the 26 funds, effectively assuming a typical practice of 24 months' backfilling by reporting funds. The weaker results indicate that applying uniform backfilling duration is far from accurate (this echoes the point made by Fung and Hsieh 2009).

The subsequent normal performance after a run of superior performance supports the hypothesis of strategic timing in initiating self-reporting by hedge funds, if they decide to report at some time. A related bias, the backfilling or instant history bias, has been analyzed in the prior literature. It refers to the practice by hedge fund managers of not reporting their funds' performance to databases from inception, but instead

Figure 2 Return Performance Around the Initial Reporting Date and the Reporting Termination Date



Notes. Panel A shows the time series of monthly market-adjusted return for the self-reporting hedge funds from 24 months before the initial reporting date to 24 months afterward. The market-adjusted return is the difference between the imputed portfolio return and the CRSP value-weighted equity market return. The imputed portfolio return is constructed by calculating the buy-and-hold return for the month using the most recent past disclosed quarter-end holdings. Panel B repeats the analyses in panel A for the reporting termination date.

"backfilling" the database later after a successful track record. The initial returns of a hedge fund that appear in a commercial database tend to be higher than the expected returns from the same fund. The timing bias that we analyze compares the returns (proxied by the 13F-imputed returns) of hedge funds before and after reporting. These two biases are likely to be correlated but are not the same.

4.2. Hazard Analysis for Reporting Initiation

To relate the timing bias to other time-varying fund characteristics in addition to return performance, we present a hazard analysis of reporting initiation for the subsample of fund companies with accurate initial reporting date information. In the language of hazard analysis, in our case, the "failure" event is the hedge fund's first appearance in the hedge fund Union Database. Thus, the hazard rate $h(t)$ is the hedge fund's probability of reporting initiation in a given period t, conditional on the fact that it did not initiate reporting in any of the previous periods. We start with a time-varying sample of nonreporting funds. Once a hedge fund has initiated reporting, it exits the sample because the spell has "failed." We estimate our instantaneous hazard model with respect to a set of time-varying explanatory variables (X),

such as fund characteristics, i.e., values of these variables are tracked dynamically since the fund's first appearance in the Thomson Reuters database until its first reporting date to the Union Database (observations of completed spells) or to the end of our sample period (observations of censored spells).

We adopt the semiparametric Cox proportional hazard model (Cox 1972), which estimates the relation between the instantaneous hazard rates and the covariates by maximizing a partial-likelihood function. In this model, the hazard rate is assumed to be

$$h(t) = h(0)e^{X_i'\beta}, \tag{1}$$

where t is the number of periods since the fund company's first appearance in the Thomson Reuters database. A positive coefficient β_k indicates that an increase in the covariate X_k is associated with an increase in the instantaneous probability of funds' initiating reporting to a database during period t. We conduct our analysis at the quarterly frequency and report the results in Table 3. Following the norm adopted in hazard analyses and to facilitate interpretation, Table 3 reports the hazard ratio (also called "exponentiated coefficient") associated with each covariate rather than raw coefficients β_k, where the ratio is defined as $h(t \mid X_k' = X_k + 1, X_{-k})/h(t \mid X_k) = e^{\beta_k}$. A hazard ratio that is greater (smaller) than unit indicates a positive (negative) contribution of the covariate to the instantaneous probability of reporting initiation. The z-statistics in the table testify to the significance of raw coefficient (β_k) being different from zero, or of the hazard ratio (e^{β_k}) being different from one.

According to Table 3, funds after better-performing periods have a higher probability of reporting initiation during the current period: hazard ratios associated with performance (lagged) are significantly higher than one. This result is consistent with the pattern uncovered in Figure 2. This effect is economically

Table 3 Hazard Analysis of the Reporting Initiation

Performance measure	(1) *Raw return*	(2) *Market-adjusted return*	(3) *One-factor alpha*	(4) *Four-factor alpha*	(5) *DGTW abnormal return*
Performance (%)	1.0558*** [2.90]	1.0539*** [2.76]	1.0452*** [3.12]	1.0377** [2.01]	1.3461*** [2.98]
Aggregate flow to hedge fund industry	0.2509*** [−4.84]	0.2517*** [−4.85]	0.2619*** [−4.67]	0.2659*** [−4.68]	0.2719*** [−4.62]
Portfolio volatility (%)	0.8245*** [−6.05]	0.8242*** [−6.05]	0.8287*** [−5.89]	0.8342*** [−5.70]	0.8235*** [−6.08]
Manager age (log)	0.9243*** [−3.17]	0.9238*** [−3.19]	0.9236*** [−3.19]	0.9216*** [−3.28]	0.9148*** [−3.43]
Portfolio Herfindahl index	0.1280* [−1.74]	0.1307* [−1.73]	0.1200* [−1.83]	0.1331* [−1.78]	0.1391* [−1.71]
Portfolio size (log)	1.0000 [0.62]	1.0000 [0.63]	1.0000 [0.72]	1.0000 [0.66]	1.0083 [0.85]
Turnover	0.6650 [−1.17]	0.6623 [−1.18]	0.6723 [−1.14]	0.6975 [−1.05]	0.6725 [−1.15]
Flow	0.8962 [−1.14]	0.8959 [−1.13]	0.9183 [−0.93]	0.9230 [−0.90]	0.8928 [−1.17]
Market return (%)	1.0077 [0.27]	1.0636** [2.52]	1.0571** [2.25]	1.0560** [2.26]	1.0751*** [2.90]
Observations	23,618	23,618	23,618	23,619	23,363

Notes. This table presents the hazard analysis of reporting initiation for the subsample of fund companies with accurate initial reporting date information using the Cox proportional hazard model. *Raw return* is the portfolio return without adjustment. *Market-adjusted return* is the portfolio return in excess of the CRSP value-weighted return. *One-factor alpha* and *Four-factor alpha* are the intercepts from CAPM one-factor and Carhart (1997) four-factor models using all available data. *DGTW abnormal return* is the Daniel et al. (1997) characteristic-based benchmark-adjusted return. *Performance*, *Flow*, *Aggregate flow to hedge fund industry*, and *Market return* are calculated over [−1,0] quarters relative to the quarter of reporting initiation. *Portfolio size* (in log), *Turnover*, and *Return volatility* are as defined in Table 1. *Manager age* (in log) is the number of years since the fund company's first appearance in Thomson Reuters. *Flow* is the change in total portfolio value during the current quarter net of the asset value appreciation/depreciation due to returns, scaled by the portfolio value at the end of the previous quarter. Reported coefficients are hazard ratios, which are greater (smaller) than unit when the original coefficients are positive (negative). The z-statistics are calculated using the original coefficients (not hazard ratios) and are reported below coefficient estimates in brackets.

***, **, *Significant at the 1%, 5%, and 10% levels, respectively.

significant too, because an interquartile change in the performance when measured by four-factor alpha (column 4 in Table 3) is associated with a 1.34 (imputed from $e^{\beta_k \Delta X_k}$) times higher probability of reporting initiation in the current period, conditional on not reporting through the last period. The coefficients on market returns are overall significantly greater than unit (columns (2)–(5) in Table 3), suggesting that funds have a higher probability of reporting initiation after a period of good market performance, which eases funds' marketing. The market return effect is also economically meaningful: conditional probability of reporting is 2.22 times higher for an interquartile increase in market returns.[18]

Table 3 highlights additional elements in hedge funds' strategic timing in reporting initiation. First, when the proxy for the aggregate flow to hedge fund industry is high, hedge funds have a significantly lower probability of reporting initiation. Here we approximate the aggregate flow by the total increase in the equity portfolio value of all 13F-filing hedge funds, netting out the increase caused by stock price appreciation. This evidence suggests that a boom in the hedge fund industry provides enough capital to many funds, leading to their lowered needs to enhance exposure to potential investors by reporting initiation. We do not find a significant relation between institution size and reporting initiation. This may be a result of two offsetting effects. On one hand, larger institutions may benefit from the economies of scale of reporting. On the other hand, larger institutions need to rely less on attracting capital through reporting, as they can better afford direct marketing and/or may have less need to raise more capital.

Second, hedge funds are less likely to initiate reporting during periods of higher portfolio return volatility. Prior literature shows that flows to hedge funds and mutual funds are dampened by return volatility conditional on performance (Ding et al. 2009, Huang et al. 2012), indicating that investors tend to discount fund returns when the volatility is high. Moreover, Sharpe Ratio is a common performance measure adopted by commercial databases, and this metric is unfavorable to volatile funds. One interquartile increase in portfolio return volatility lowers the probability of reporting initiation by a factor of 0.20.

Finally, hedge funds have higher probability of reporting initiation in their youth stage if they decide to report: the hazard ratios associated with fund age are significantly lower than one. This result is intuitive, as young funds are the most likely to benefit from reporting initiation. The impact of the portfolio concentration (as measured by the average portfolio Herfindahl index) on the reporting initiation is negative and significant at the 10% level. Thus, hedge funds operating more concentrated portfolios are less likely to initiate reporting. This is consistent with greater costs of reporting associated with revealing trading secrecy when investors can use the funds' return and strategy information reported to databases in conjunction with their disclosed 13F holdings.

4.3. Second Form of Timing Bias: Comparison of Fund Companies Before and After Reporting Termination

There are 187 funds in our sample that terminated reporting to the Union Database at some point during the 1980–2008 period. For these funds, we are able to analyze the determinants of reporting termination using the same method as we use in Table 2 for reporting initiation. Moreover, for these funds we have more information about their termination decision because of their reporting status when the decision is made. We report the results for the overall sample of 187 funds in panel A of Table 4. We then examine the subsample of funds (in panel B of Table 4) with assets exceeding \$250 million 24 months after termination of self-reporting[19] in order to avoid any potential survivorship bias caused by some funds' disappearance from the 13F database simply because their assets drop below \$100 million.

We observe that the performance after termination of reporting is significantly lower than that before termination. This is not surprising, given that most funds exit from commercial databases after their performance starts deteriorating (Ackermann et al. 1999; Liang 2000; Fung and Hsieh 2000, 2002; among others). What is interesting and unique about our analysis is that we are able to determine the performance of funds after they disappear from the commercial databases. Our analysis is thus analogous to computing the delisting returns for stocks in Shumway (1997) and Shumway and Warther (1999); hence, this second form of the timing bias is analogous to a "delisting bias."

Panel A of Table 4 shows that the average monthly raw returns deteriorate remarkably (by 1.9%, significant at less than the 1% level) after reporting termination. The magnitude is more modest at 12–28 basis points monthly using risk-adjusted performance measures (market-adjusted returns, CAPM alpha, four-factor alpha, and the DGTW abnormal return). However, these differences are uniformly negative, and the difference in the market-adjusted returns

[18] The insignificance of the coefficient of the market return when raw performance of hedge funds enters the regression is caused by the latter already containing information about market returns.

[19] Applying the \$250 million filter at 24 months before reporting termination produces qualitatively similar results.

Table 4 Comparison of Return Performance Before and After Reporting Termination

	(1) *Raw return*	(2) *Market-adjusted return*	(3) *One-factor alpha*	(4) *Four-factor alpha*	(5) *Difference-in-difference*	(6) *DGTW abnormal return*
Panel A: Overall sample						
Before reporting termination						
Mean	0.0118	0.0028	0.0018	0.0016	0.0001	0.0075
Median	0.0131	0.0032	0.0016	0.0019	0.0003	0.0067
No. of funds	187	187	187	187	187	180
After reporting termination						
Mean	−0.0072	0.0000	0.0006	−0.0001	0.0003	0.0063
Median	−0.0015	0.0014	0.0013	0.0002	0.0007	0.0065
No. of funds	187	187	187	187	187	181
Differences (*t*-statistics)						
Mean	−0.0190***	−0.0028**	−0.0012	−0.0017	0.0002	−0.0012
	[−8.92]	[−2.13]	[−0.98]	[−1.57]	[0.19]	[−1.09]
Median	−0.0146***	−0.0018	−0.0003	−0.0017***	0.0004	−0.0002
	[−4.54]	[−1.55]	[−0.81]	[−2.80]	[0.21]	[−0.23]
Panel B: Subsample of funds with long-equity holdings of more than $250 million						
Before reporting termination						
Mean	0.0143	0.0048	0.0034	0.0027	0.0021	0.0083
Median	0.0141	0.0036	0.0023	0.0020	0.0015	0.0067
No. of funds	135	135	135	135	135	134
After reporting termination						
Mean	−0.0078	−0.0006	0.0006	0.0000	0.0001	0.0075
Median	−0.0026	0.0011	0.0013	0.0001	0.0007	0.0065
No. of funds	135	135	135	135	135	135
Differences (*t*-statistics)						
Mean	−0.0221***	−0.0054***	−0.0028***	−0.0027***	−0.0020	−0.0008
	[−9.54]	[−4.40]	[−2.48]	[−3.14]	[−1.72]	[−0.83]
Median	−0.0167***	−0.0025***	−0.0010	−0.0019***	−0.0008	−0.0002
	[−3.94]	[−2.74]	[−1.80]	[−2.58]	[−0.76]	[−0.06]

Notes. This table presents the same analyses as in Table 2 but replaces the event with reporting termination. Panel A reports the results for the overall sample. Panel B reports the results for the subsample of funds with equity holdings greater than $250 million at the end of the 24-month period post report termination. The *t*-statistics based on bootstrapped standard errors are reported below coefficient estimates in brackets.

***, **, *Significant at the 1%, 5%, and 10% levels, respectively.

is significant at the 5% level.[20] We obtain similar results for median performance differences. A graphic illustration of the performance pattern around the reporting termination date is provided in panel B of Figure 2. Note that these estimates of posttermination performance are underestimated, as the extremely distressed funds that exit from commercial databases will also be exempt from 13F reporting if the value of their equity portfolios drops below $100 million. Indeed, of the 221 companies that file 13F prior to termination of reporting to commercial databases, 61 (71) also disappear from the 13F database two (three) quarters after. The attrition rate of 28% (32%) is much higher than the unconditional average of 10% (12%).

About 64% of the funds (119 funds) that terminate reporting in our sample provide reasons to the commercial databases. In 112 cases, the given reasons indicate distress (such as liquidation, fund being dormant, or data vendor being unable to contact the fund). Other given reasons can be positive (such as being closed to new investors) or unclear (such as being merged with another fund), but such cases are rare. When we focus on the subsamples partitioned by stated reasons, we do not find significant differences across the subsamples in the changes in performance after reporting termination, mostly because of the much reduced sample sizes.

The potential survivorship bias caused by the threshold of $100 million required for filing 13F forms is addressed by the subsample of "large" (greater than $250 million) funds. Panel B shows that the average monthly raw returns and the four risk-adjusted performance measures are lower by 2.21 percentage points and 8–54 basis points, respectively. The differences in the median performance of funds after and before reporting termination reveal a similar pattern. A robustness check using $1,000 million as the cutoff yields numerically close results. Panels A and B

[20] The magnitude of market-adjusted returns is qualitatively similar to but compares favorably with the finding of Hodder et al. (2008) that the average delisted hedge fund held by a sample of fund of hedge funds had a monthly return of −1.86% immediately after it was delisted.

combined suggest performance deterioration around reporting termination is not driven by the potential survivorship bias from the minimum asset-under-management requirement for 13F disclosure.

In summary, exiting from commercial databases by the reporting funds is overall a sign of deterioration. Interestingly, negative market returns also contribute to higher incidences of reporting termination—manifested by the significantly higher before–after return gap in raw returns than benchmark-adjusted returns, as shown in Table 4. Moreover, the effect of the market returns on reporting termination seems to be driven by the episode of the financial crisis in 2007–2008. Given the high average market beta (0.4–0.5) of our sample funds, it is not surprising that a major market downturn constitutes a negative common shock to hedge funds. Finally, the combination of good performance prior to reporting initiation (results in the previous section) and poor performance following reporting termination acts as offsetting forces that bias the performance tracked by the commercial database toward average.

4.4. Effects of Self-Reporting on Hedge Fund Flows

4.4.1. Reporting Initiation. We hypothesize that a primary benefit of reporting to hedge fund databases is enhancing a hedge fund company's exposure to potential clients. If such a motive is justified, then a hedge fund should experience, on average, an increase in flows after the initiation of reporting relative to the counterfactual of not reporting. For all funds that initiate reporting during our sample period, we can isolate the quarterly observations from four quarters before the initial reporting date to four quarters afterward. We then estimate the following regression at the fund (indexed by i)-quarter (indexed by t) level:

$$Flow_{i,t} = \sum_{j=-4}^{4} \lambda_j D_{t+j} + \beta Performance_{t-3:t} + \gamma Control_{i,t-1} + \varepsilon_{i,t}, \quad (2)$$

where $Flow_{i,t}$ is calculated as $(Size_{i,t} - Ret_{i,t} * Size_{i,t-1})/Size_{i,t-1}$, all using disclosed holdings in Form 13F. It measures the change in the value of a fund's equity portfolio caused by changes in investment by the funds' investors (and not by the changes in the stock prices) and is a proxy for the net fund flows. Note that we do not use the total net assets from commercial databases, as they are not available for fund-quarter observations in nonreporting status and are often not updated at the quarterly level. The all-sample average (median) percentage flow to hedge funds companies is 3.6% (1.4%). Dummy variables for four quarters before and after the initial reporting date, restricted to the subsample of funds with accurate initial reporting dates are denoted by D_{t+j}. Monthly average of the performance measure during the past four quarters that end in the current quarter is denoted by $Performance_{t-3:t}$, and $Control_{i,t-1}$ are lagged control variables including portfolio size (in log), fund age (numbers of quarters since first appearance in Thomson Reuters, in log), portfolio turnover rate, and portfolio volatility.

The four columns in panel A of Table 5 correspond to the four benchmark-adjusted return performance measures that appear in earlier tables. The coefficients on *Performance* show that flows are highly responsive (significant at the 1% level) to risk-adjusted returns. The effects are economically significant, too. For example, for a one percentage point increase in monthly market-adjusted return, net fund flows increase by 2.5% of the total portfolio value (see column 1), a magnitude similar to the one documented in the mutual fund literature (e.g., Chevalier and Ellison 1997, Sirri and Tufano 1998). Excluding the initiation quarter ($j = 0$), an F-test for changes in flows over the full window $\sum_{j=1}^{4} \lambda_j - \sum_{j=-4}^{-1} \lambda_j = 0$ fails to reject the null of equality. Therefore, reporting to databases does not lead to higher flows over a longer window comparing flows during the year following initiation to those during the year preceding reporting initiation. Note that we do not observe the counterfactuals—flows that would prevail had the reporting funds chosen not to initiate reporting. It is possible that funds anticipating loss of flows from existing sources choose to report to databases, a decision process that can attenuate the estimated incremental flows from exposure through the databases.

4.4.2. Reporting Termination. Last, we repeat the analysis in regression (2) on reporting termination. Results reported in panel B of Table 5 show that funds encounter significantly lower net flows (or more outflows) after reporting termination. The same F-test $\sum_{j=1}^{4} \lambda_j - \sum_{j=-4}^{-1} \lambda_j = 0$ is strongly rejected (at the 5% level) in favor of a negative change in net flows across all specifications. Specifically, the cumulative net outflows during the quarter of reporting termination and four quarters afterward amount to 23%–29% of the lagged portfolio size. This evidence adds further support to a negative delisting bias; i.e., delisting from hedge fund databases is in general a sign of deterioration.

5. The Unconditional Self-Reporting Bias: Comparing Self-Reporting and Nonreporting Hedge Funds

As a next step, we move up from the subsample of self-reporting funds to the full sample and ask

Table 5 Flow to Fund Companies Before and After the Initial and Final Reporting Dates

Performance measure	(1) *Market-adjusted return*		(2) *One-factor alpha*		(3) *Four-factor alpha*		(4) *DGTW abnormal return*	
			Panel A: Effects of reporting initiation on flows					
$Q-4$	0.0854	[1.20]	0.0960	[1.34]	0.1062	[1.43]	0.0929	[1.30]
$Q-3$	−0.0083	[−0.17]	0.0033	[0.07]	−0.0526*	[−1.71]	−0.0022	[−0.04]
$Q-2$	0.0650	[1.16]	0.0733	[1.29]	0.0209	[0.48]	0.0698	[1.23]
$Q-1$	0.0280	[0.51]	0.0345	[0.63]	0.0363	[0.61]	0.0320	[0.58]
Q	0.0387	[0.97]	0.0470	[1.17]	0.0273	[0.69]	0.0428	[1.06]
$Q+1$	0.1282	[1.56]	0.1345	[1.64]	0.1798**	[2.08]	0.1317	[1.60]
$Q+2$	0.0601	[1.40]	0.0683	[1.60]	0.0760*	[1.75]	0.0641	[1.49]
$Q+3$	−0.0034	[−0.06]	0.0026	[0.04]	0.0047	[0.08]	−0.0019	[−0.03]
$Q+4$	0.0504	[0.65]	0.0515	[0.66]	0.0522	[0.67]	0.0484	[0.62]
Performance	2.4853***	[16.32]	0.8716***	[7.79]	0.8689***	[6.34]	1.1395***	[7.04]
Portfolio size	−0.0281***	[−31.00]	−0.0273***	[−30.10]	−0.0256***	[−27.94]	−0.0272***	[−30.04]
Manager age	−0.0077***	[−4.69]	−0.0082***	[−4.99]	−0.0029	[−1.59]	−0.0087***	[−5.24]
Turnover	0.0098***	[3.10]	0.0114***	[3.64]	0.0119***	[3.63]	0.0083***	[2.58]
Portfolio volatility	0.3101***	[5.23]	0.2984***	[5.03]	0.3904***	[6.27]	0.1645***	[2.60]
Nonreporting funds dummy	−0.0042	[−0.56]	−0.0026	[−0.34]	−0.0018	[−0.23]	−0.0045	[−0.59]
Constant	0.2647***	[25.61]	0.2613***	[25.21]	0.2322***	[21.94]	0.2629***	[25.34]
N	141,090		141,089		131,544		139,928	
R-squared	0.016		0.014		0.012		0.014	
F-test								
Point estimate	0.0652		0.0498		0.2019		0.0498	
F-statistics	0.14		0.08		1.34		0.08	
p-value	0.708		0.776		0.248		0.7761	
			Panel B: Effects of reporting termination on flows					
$Q-4$	−0.0106	[−0.30]	−0.0063	[−0.18]	−0.0354	[−1.20]	−0.0103	[−0.29]
$Q-3$	0.0136	[0.34]	0.0163	[0.41]	0.0268	[0.65]	0.0152	[0.38]
$Q-2$	−0.0079	[−0.34]	−0.0055	[−0.23]	−0.0236	[−1.07]	−0.0085	[−0.36]
$Q-1$	0.0475	[1.05]	0.0526	[1.16]	0.0520	[1.17]	0.0540	[1.18]
Q	−0.0568	[−1.52]	−0.0584	[−1.57]	−0.0654*	[−1.73]	−0.0621*	[−1.65]
$Q+1$	−0.0418	[−1.00]	−0.0427	[−1.02]	−0.0354	[−0.84]	−0.0440	[−1.04]
$Q+2$	−0.0508	[−1.41]	−0.0522	[−1.44]	−0.0470	[−1.29]	−0.0543	[−1.49]
$Q+3$	−0.0272	[−0.69]	−0.0286	[−0.73]	−0.0256	[−0.65]	−0.0324	[−0.82]
$Q+4$	−0.1030***	[−2.78]	−0.1050***	[−2.82]	−0.1003***	[−2.71]	−0.1075***	[−2.88]
Performance	2.9684***	[5.66]	1.4168***	[3.60]	1.4327***	[2.89]	2.5845***	[3.22]
Portfolio size	−0.0602***	[−9.30]	−0.0587***	[−9.13]	−0.0567***	[−8.84]	−0.0591***	[−9.24]
Manager age	−0.0160*	[−1.80]	−0.0188**	[−2.11]	−0.0114	[−1.20]	−0.0179**	[−2.02]
Annualized portfolio turnover rate	0.0026	[0.24]	0.0031	[0.28]	0.0036	[0.32]	−0.0000	[−0.00]
Portfolio volatility	−0.0131	[−0.05]	−0.0453	[−0.19]	0.0454	[0.18]	−0.2420	[−0.94]
Constant	0.5052***	[11.45]	0.5092***	[11.59]	0.4770***	[10.68]	0.5046**	[11.51]
N	6,301		6,301		5,934		6,240	
R-squared	0.048		0.045		0.041		0.041	
F-test								
Point estimate	−0.2654		−0.2856		−0.2281		−0.2886	
F-statistics	6.22		7.13		4.77		7.25	
p-value	0.0126		0.0076		0.0290		0.0071	

Notes. This table reports the regression estimates of the flows to fund companies before and after the initial and final reporting dates. Dependent variable is the net percentage flow in a given quarter, where flow is defined as the change in total portfolio value during the current quarter net of the asset value appreciation/depreciation due to returns, scaled by the portfolio value at the end of the prior quarter. Panel A reports the results for the subsample of fund companies with accurate initial reporting date information using four risk-adjusted *Performance* measures: market-adjusted return, CAPM one-factor alpha, Carhart (1997) four-factor alpha, and the characteristic-based Daniel et al. (1997) benchmark-adjusted return. $Q+j$, where $j=-4,\ldots,4$, is the dummy variable for j quarters relative to the quarter of initial reporting. Other variables are as defined in Table 1. *Manager age* (in log) is the number of years since the company's first appearance in Thomson Reuters. All covariates lag the dependent variable by one quarter. The *F*-test reported at the bottom of the table tests the null hypothesis that sum of coefficients on $Q+1$ to $Q+4$ and the sum of coefficients of $Q-4$ to $Q-1$ are equal. Panel B presents the same analyses as in panel A for the full sample, except it examines the flows to fund companies before and after reporting termination. The *t*-statistics are reported next to the coefficient estimates in brackets.

***, **, *Significant at the 1%, 5%, and 10% levels, respectively.

the question "Who reports?" Our answer relies on the comparison of the pooled sample of 13F-filing hedge fund companies that never appear in the Union Database (there are 554 such nonreporting companies) and those that appear in the database for some time during our sample period (there are 366 such self-reporting companies). To reduce noise, we do not include the 279 fund companies whose reporting status cannot be accurately verified.

5.1. Comparison of Fund Characteristics

We adopt the following procedure to match reporting funds to their nonreporting counterparts: For each self-reporting fund, let $[t_1, t_2]$ be the time period for which it appears in the Thomson Reuters database (which may contain periods before, during, and after its reporting to the Union Database). The matched fund is one among all nonreporting funds in existence at t_1 that minimizes the following two-dimensional "distance score:"[21]

$$\text{Distance} = \frac{1}{2}\left[\left|\frac{\text{Nonreporting fund size} - \text{Reporting fund size}}{\text{Reporting fund size}}\right| + \left|\frac{\text{Nonreporting fund age} - \text{Reporting fund age}}{\text{Reporting fund age}}\right|\right].$$

If at some time t_3, where $t_1 < t_3 < t_2$, the matched fund disappears from the 13F database, we continue with a new matched fund using the same procedure based on information available at t_3. By construction, the portfolio size and age of the paired funds are comparable. Unconditionally, the median size and initiation year do not differ across the two samples, but the largest hedge funds are underrepresented in the reporting sample.

Table 6 reveals several characteristics of portfolios. First, the self-reporting hedge funds have lower median portfolio concentration (as measured by the portfolio Herfindahl index) than the nonreporting funds. Second, the average monthly return volatilities of the two categories are almost identical, but the self-reporting funds have considerably higher average portfolio annualized turnover rates (106%) than the nonreporting funds (76%), and the difference is significant at the 1% level. This finding is intuitive, as higher turnover funds need to worry less about their trading strategies being revealed through self-reporting.

Table 6 further compares the loadings on common risk factors for self-reporting and nonreporting funds. Interestingly, the equity positions of self-reporting funds have significantly higher exposure to the size (small minus big, SMB) factor, where the differences in both mean and median are significant at the 1% level. The differences in the loadings on the market and book-to-market (high minus low, HML) factors follow the same pattern using the median statistic only, and the difference in the loadings on the momentum factor is not significant. To the extent that exposures to common risk factors hardly constitute trading secrecy, these results support the hypothesis that funds with less conventional trading strategies (i.e., lower factor loadings) are more reluctant to report to databases.

Note that the pairwise comparison analyses reported in Table 6 and the hazard analysis (reported in Table 3) do not necessarily yield coefficients of the same sign or of similar significance levels. The former relates the fund characteristics (averaged over the time series) to their propensity to ever report; the latter focuses on how the time-variation in fund characteristics prompt report initiation at certain points of time. For example, the hazard analysis indicates that funds are less likely to initiate reporting during periods of volatile returns; but reporting funds as a whole do not have lower return volatility than nonreporting funds.

5.2. Comparison of Return Performance

We next move on to return performance comparison, which underlies the important consequences of the self-reporting-related biases in commercial databases. We report these results in Table 7.

The return differences between the mean (median) return measures over the matched time period indicate that self-reporting funds outperform nonreporting funds by 1–6 (0–11) basis points monthly using the various performance measures. These differences are overall insignificant with the exception of the median difference in DGTW abnormal returns. The combined results suggest that reporting hedge funds exhibit somewhat superior stock picking abilities (which manifest in the DGTW measure) but do not command a convincing edge over the nonreporting funds in return performance. A lack of consistent difference could exist because the positive and negative reasons prompting reporting initiation and termination largely offset one another.

In a robustness check (not tabulated), we examine the subsample of funds with high correlation (in the top two quintiles across all funds); the median self-reporting fund significantly outperforms the median nonreporting fund by 19, 18, and 29 basis points using three of the four risk-adjusted performance measures: market-adjusted returns, four-factor alpha, and DGTW abnormal returns, respectively. However, the mean difference in the risk-adjusted performance of

[21] The average/median score is below 5% and the 90th percentile value is 18%. Hence, the matching is overall close.

Table 6 Comparison of Self-Reporting and Nonreporting Fund Companies

	(1) Self-reporting fund companies	(2) Nonreporting matches	(3) Difference	(4) *t*-statistics of the difference
Portfolio Herfindahl index				
Mean	0.0798	0.0860	−0.0062	−0.91
Median	0.0458	0.0551	−0.0093**	−2.13
Volatility				
Mean	0.0557	0.0541	0.0016	1.15
Median	0.0509	0.0497	0.0013	0.90
Annualized portfolio turnover rate				
Mean	1.0562	0.7596	0.2967***	7.40
Median	0.9909	0.6652	0.3258***	5.49
Market factor				
Mean	1.0940	1.0870	0.0070	0.37
Median	1.0787	1.0429	0.0358**	2.40
SMB (small minus big) factor				
Mean	0.3863	0.2935	0.0928***	3.04
Median	0.3416	0.2489	0.0927***	2.76
HML (high minus low) factor				
Mean	0.1284	0.0782	0.0502	1.48
Median	0.1140	0.0616	0.0524**	2.11
Momentum factor				
Mean	−0.0083	−0.0340	0.0258	1.27
Median	−0.0019	−0.0220	0.0201	1.49
Number of institutions	366	366	—	—

Notes. This table compares the characteristics of the self-reporting with those of the nonreporting fund companies after matching the two samples through the following procedure: For each self-reporting fund, let $[t_1, t_2]$ be the time period for which it appears in the 13F database. The matched fund is one among all nonreporting funds in existence at t_1 that minimizes "distance" in both portfolio size and age. If at some time t_3, where $t_1 < t_3 < t_2$, the matched fund disappears from the 13F database, we continue with a new matched fund using the same procedure based on information available at t_3. The sample of self-reporting fund companies includes all 13F-filing funds that report to the Union Database for some period of time. The sample of nonreporting fund companies includes all 13F-filing hedge fund companies that never report to the Union Database. Other variables are as defined in Table 1. The *t*-statistics correspond to the difference between the self-reporting fund companies and the nonreporting fund companies.

***, **, *Significant at the 1%, 5%, and 10% levels, respectively.

self-reporting funds and nonreporting funds ranging from 8–24 basis points continues to be statistically insignificant, with the exception of DGTW abnormal returns. These findings suggest a statistically significant (but modest) reporting bias for the subsample of funds which are more equity-oriented, but little evidence of bias for the overall sample.

To account for any differences in the liquidity of the underlying stocks held by self-reporting and nonreporting funds, we include lagged market returns, as in Asness et al. (2001) or the Pastor and Stambaugh (2003) value-weighted traded liquidity factor in the multifactor model. Our results (not tabulated) for the differences in performance between self-reporting and nonreporting funds remain unchanged. We further compare the probability of attrition for self-reporting and nonreporting funds by estimating the probability of a fund filing 13F in a quarter conditional on the same fund filing 13F in the prior quarter. In results not tabulated, we observe that the average conditional probabilities for self-reporting and nonreporting funds are 5.2% and 3.8%, respectively, but the difference is not statistically significant. The somewhat higher 13F-attrition probability of the reporting funds is not surprising, given that less established funds are more likely to report.

The overall evidence is consistent with the hypothesis that young and medium-sized fund companies have a stronger incentive to report to databases to publicize their funds and attract potential investors. Moreover, self-reporting funds are more diversified, employ higher-frequency trading strategies, and have higher loadings on common factors—presumably trading secrecy is less likely to be revealed through voluntary disclosure or is less important when a portfolio involves more stocks, evolves more quickly, and has more exposure to common risk factors. Fung and Hsieh (2000, p. 299) conjectured, with the support of some anecdotal evidence, that the selection bias from self-reporting is limited because on one hand, "only funds with good performance want to be included in a database," whereas on the other hand, "managers

Table 7 Comparison of Self-Reporting and Nonreporting Matching Fund Companies

	(1) *Raw return*	(2) *Market-adjusted return*	(3) *One-factor alpha*	(4) *Four-factor alpha*	(5) *DGTW abnormal return*
Self-reporting fund companies					
5th percentile	−0.0178	−0.0139	−0.0096	−0.0093	−0.0043
25th percentile	−0.0019	−0.0011	−0.0010	−0.0018	0.0048
Median	0.0047	0.0017	0.0016	0.0011	0.0075
75th percentile	0.0095	0.0048	0.0047	0.0040	0.0106
95th percentile	0.0164	0.0108	0.0117	0.0093	0.0168
Mean	0.0025	0.0009	0.0014	0.0008	0.0072
Std. dev.	0.0112	0.0082	0.0067	0.0058	0.0070
Skewness	−1.5759	−1.6448	−0.9769	−0.5679	−1.1196
Kurtosis	7.2111	8.8857	6.2933	5.6275	8.9828
No. of funds	366	366	355	355	366
Nonreporting fund companies					
5th percentile	−0.0179	−0.0139	−0.0095	−0.0104	−0.0014
25th percentile	−0.0026	−0.0016	−0.0013	−0.0022	0.0042
Median	0.0038	0.0011	0.0016	0.0006	0.0064
75th percentile	0.0096	0.0044	0.0047	0.0035	0.0099
95th percentile	0.0176	0.0123	0.0115	0.0098	0.0162
Mean	0.0024	0.0008	0.0014	0.0002	0.0068
Std. dev.	0.0114	0.0078	0.0070	0.0069	0.0056
Skewness	−0.8687	−0.5841	−0.3980	−0.4524	−0.3131
Kurtosis	5.0598	6.9565	6.4825	6.4968	5.4862
No. of funds	366	366	355	355	365
Differences (*t*-statistics)					
5th percentile	0.0002 [0.04]	0.0000 [−0.01]	−0.0001 [−0.02]	0.0012 [0.50]	−0.0029 [−1.37]
25th percentile	0.0006 [0.55]	0.0005 [0.89]	0.0003 [0.45]	0.0005 [0.79]	0.0005 [1.22]
Median	0.0009 [1.00]	0.0006 [1.61]	0.0000 [0.06]	0.0004 [1.28]	0.0011 [2.58]**
75th percentile	−0.0001 [−0.11]	0.0004 [0.71]	−0.0001 [−0.09]	0.0005 [1.13]	0.0007 [1.37]
95th percentile	−0.0012 [−0.75]	−0.0014 [−0.80]	0.0002 [0.08]	−0.0005 [−0.24]	0.0006 [0.31]
Mean	0.0001 [0.10]	0.0001 [0.15]	0.0001 [0.14]	0.0006 [1.22]	0.0005 [0.89]
Std. dev.	−0.0002 [−0.14]	0.0003 [0.27]	−0.0004 [−0.39]	−0.0011 [−0.71]	0.0014 [1.85]*
Skewness	−0.7073 [−0.78]	−1.0608 [−0.86]	−0.5790 [−0.66]	−0.1154 [−0.07]	−0.8065 [−0.46]
Kurtosis	2.1513 [0.43]	1.9292 [0.36]	−0.1892 [−0.07]	−0.8693 [−0.10]	3.4967 [0.19]

Notes. This table compares the performance measures of the self-reporting and nonreporting funds using samples matched through the procedure described in Table 6. All return performance measures are calculated at the monthly frequency assuming the companies hold their most recently disclosed quarter-end holdings. *Raw return* is the portfolio return without adjustment. *Market-adjusted return*, *One-factor alpha*, *Four-factor alpha*, and *DGTW abnormal return* are as defined in Table 1. The *t*-statistics for the differences are reported in brackets.

***, **, *Significant at the 1%, 5%, and 10% levels, respectively.

with superior performance did not necessarily participate in vendors' databases." Our results are supportive of their conjecture.

6. Conclusion

This paper presents a comprehensive study that formally analyzes the self-reporting-related biases in hedge fund databases. Our research provides important references and benchmarks for hedge fund researchers and investment managers who use commercial databases and publicly available information on portfolio holdings of institutions. Our findings also reveal the motivation and consequences of voluntary disclosure by hedge funds. Finally, by comparing databases from mandatory and voluntary sources, our research contributes to the ongoing debate regarding more stringent disclosure rules for hedge funds.

Acknowledgments

Comments and suggestions from Jonathan Berk, Gavin Cassar, Wayne Ferson, Gerald Gay, John Griffin, Bob Hodrick, Marno Verbeek, an anonymous referee, an associate editor, Brad Barber (the department editor), and seminar and conference participants at Columbia University, the Center for Research in Security Prices Forum, Financial Management Association 2010 Annual Meetings, Tel Aviv University, University of California Davis Symposium on Financial Institutions and Intermediaries, Western Finance Association 2011 Annual Meetings, and the 2nd Annual Conference on Hedge Funds have contributed to this draft. The authors thank Linlin Ma and Yuehua Tang for excellent research assistance. The authors gratefully acknowledge the financial support from the Banque Nationale de Paris Paribas Hedge Fund Centre at Singapore Management University.

References

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

Agarwal V, Boyson NM, Naik NY (2009a) Hedge funds for retail investors? An examination of hedged mutual funds. *J. Financial Quant. Anal.* 44:273–305.

Agarwal V, Daniel ND, Naik NY (2009b) Role of managerial incentives and discretion in hedge fund performance. *J. Finance* 64:2221–2256.

Agarwal V, Jiang W, Tang Y, Yang B (2013) Uncovering hedge fund skill from the portfolios they hide. *J. Finance*. Forthcoming.

Aiken AL, Clifford CP, Ellis J (2013) Out of the dark: Hedge fund reporting biases and commercial databases. *Rev. Financial Stud.* 26:208–243.

Ang A, Gorovyy S, van Inwegen GB (2010) Hedge fund leverage. *J. Financial Econom.* 102:102–126.

Asness C, Krail R, Liew J (2001) Do hedge funds hedge? *J. Portfolio Management* 28:6–19.

Avramov D, Kosowski R, Naik NY, Teo M (2011) Hedge funds, managerial skill, and macroeconomic variables. *J. Financial Econom.* 99:672–692.

Baquero G, ter Horst JR, Verbeek M (2005) Survival, look-ahead bias, and persistence in hedge fund performance. *J. Financial Quant. Anal.* 40:493–517.

Bollen NP, Pool VK (2008) Conditional return smoothing in the hedge fund industry. *J. Financial Quant. Anal.* 43:267–298.

Brown SJ, Goetzmann WN (1995) Performance persistence. *J. Finance* 50:679–698.

Brown SJ, Schwarz C (2012) Do market participants care about portfolio disclosure? Evidence from hedge funds' 13Fs. Working paper, New York University, New York.

Brown SJ, Goetzmann WN, Ibbotson RG (1999) Offshore hedge funds: Survival and performance 1989–95. *J. Bus.* 72:91–117.

Brown SJ, Goetzmann WN, Ibbotson RG, Ross SA (1992) Survivorship bias in performance studies. *Rev. Financial Stud.* 5:553–580.

Brunnermeier M, Nagel S (2004) Hedge funds and the technology bubble. *J. Finance* 59:2013–2040.

Carhart M (1997) On persistence in mutual fund performance. *J. Finance* 52:57–82.

Cassar G, Gerakos JJ (2011) Hedge funds: Pricing controls and the smoothing of self-reported returns. *Rev. Financial Stud.* 24:1698–1734.

Chen Q, Goldstein I, Jiang W (2010) Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *J. Financial Econom.* 97:239–262.

Chevalier J, Ellison G (1997) Risk taking by mutual funds as a response to incentives. *J. Political Econom.* 105:1167–1200.

Cici G, Gibson S, Moussawi R (2010) For better or worse? Mutual funds in side-by-side management relationships with hedge funds. *J. Financial Intermediation* 19:169–187.

Cox DR (1972) Regression models and life-tables (with discussion). *J. Royal Statist. Soc.* 34:187–220.

Daniel K, Grinblatt M, Titman S, Wermers R (1997) Measuring mutual fund performance with characteristic-based benchmarks. *J. Finance* 52:1035–1058.

Ding B, Getmansky M, Liang B, Wermers R (2009) Share restrictions and investor flows in the hedge fund industry. Working paper, State University of New York at Albany, Albany.

Edwards FR, Caglayan MO (2001) Hedge fund performance and manager skill. *J. Futures Markets* 21:1003–1028.

Elton EJ, Gruber MJ, Blake CR (1996) Survivorship bias and mutual fund performance. *Rev. Financial Stud.* 9:1097–1120.

Evans RB (2010) Mutual fund incubation. *J. Finance* 65:1581–1611.

Fung W, Hsieh DA (1997) Empirical characteristics of dynamic trading strategies: The case of hedge funds. *Rev. Financial Stud.* 10:275–302.

Fung W, Hsieh DA (2000) Performance characteristics of hedge funds and CTA funds: Natural versus spurious biases. *J. Financial Quant. Anal.* 35:291–307.

Fung W, Hsieh DA (2002) Hedge-fund benchmarks: Information content and biases. *Financial Analysts J.* 58:22–34.

Fung W, Hsieh DA (2009) Measurement biases in hedge fund performance data: An update. *Financial Analysts J.* 65:36–38.

Getmansky M, Lo AW, Makarov I (2004) An econometric model of serial correlation and illiquidity in hedge fund returns. *J. Financial Econom.* 74:529–609.

Griffin J, Xu J (2009) How smart are the smart guys? A unique view from hedge fund stock holdings. *Rev. Financial Stud.* 22:2531–2570.

Hodder JE, Jackwerth JC, Kolokolova O (2008) Recovering delisting returns of hedge funds. Working paper, University of Wisconsin–Madison, Madison.

Huang JC, Wei KD, Yan H (2012) Investor learning and mutual fund flows. Working paper, Cheung Kong Graduate School of Business, Beijing.

Jorion P, Schwarz C (2013) The strategic listing decisions of hedge funds. *J. Financial Quant. Anal.* Forthcoming.

Kat H, Palaro H (2006) Who needs hedge funds? A copula-based approach to hedge fund return replication. Working paper, Cass Business School, London.

Khandani A, Lo A (2007) What happened to the quants in August 2007? *J. Investment Management* 5:5–54.

Lhabitant F-S (2006) *Handbook of Hedge Funds* (John Wiley & Sons, West Sussex, UK).

Liang B (1999) On the performance of hedge funds. *Financial Analysts J.* 55:72–85.

Liang B (2000) Hedge funds: The living and the dead. *J. Financial Quant. Anal.* 35:309–326.

Linnainmaa JT (2013) Reverse survivorship bias. *J. Finance* Forthcoming.

Malkiel BG (1995) Returns from investing in equity mutual funds 1971 to 1991. *J. Finance* 50:549–572.

McCormack K (2006) Do hedge funds hold 'trade secrets'? *Bus. Week* (September 11), http://www.businessweek.com/.

Nohel T, Wang ZJ, Zheng L (2010) Side-by-side management of hedge funds and mutual funds. *Rev. Financial Stud.* 23:2342–2373.

Pastor L, Stambaugh RF (2003) Liquidity risk and expected stock returns. *J. Political Econom.* 111:642–685.

Shumway T (1997) The delisting bias in CRSP data. *J. Finance* 52:327–340.

Shumway T, Warther VA (1999) The delisting bias in CRSP's NASDAQ data and its implications for the size effect. *J. Finance* 54:2361–2379.

Sirri E, Tufano P (1998) Costly search and mutual fund flows. *J. Finance* 53:1589–1622.

Stulz R (2007) Hedge funds: Past, present, and future. *J. Econom. Perspect.* 21:175–194.

ter Horst JR, Verbeek M (2007) Fund liquidation, self-selection, and look-ahead bias in the hedge fund industry. *Rev. Finance* 11:605–632.

ter Horst JR, Nijman T, Verbeek M (2001) Eliminating look-ahead bias in evaluating persistence in mutual fund performance. *J. Empirical Finance* 8:345–373.