Inferring Reporting Biases in Hedge Fund Databases from Hedge Fund Equity Holdings[1]

Vikas Agarwal[2] Vyacheslav Fos[3] Wei Jiang[4]

First Version: March, 2009
This Draft: January, 2010

ABSTRACT

This paper is a first study that formally analyzes the degree of the self-reporting bias in the hedge funds databases by exploring the quarterly equity holdings of a complete list of hedge fund companies that file the Form 13F to the SEC between 1980 and 2008 and self-report their performance to a union of five major hedge fund databases. We find that the propensity to self-report is consistent with the trade-offs between the benefits (access to prospective investors) and costs (revealing trading secrecy and losing flexibility in selective marketing) of self-reporting. Though self-reporting and non-reporting funds do not differ significantly in return performance, reporting funds experience substantial deterioration in performance after both the reporting initiation and termination dates.

[1] Comments and suggestions from Gerald Gay, Bob Hodrick, and seminar participants at Columbia University have contributed to this draft. The authors thank Linlin Ma for excellent research assistance.
[2] J. Mack Robinson College of Business, Georgia State University, 35 Broad Street, Atlanta, GA 30303. Tel: 404 413 7326; Email: vagarwal@gsu.edu.
[3] Graduate School of Business, Columbia University, 3022 Broadway, New York, NY 10027. Tel: 212 854 8057; Email: vfos14@gsb.columbia.edu.
[4] Graduate School of Business, Columbia University, 3022 Broadway, New York, NY 10027. Tel: 212 854 9002; Email: wj2006@columbia.edu.

Hedge funds are pooled private investment vehicles. Unlike other financial institutions such as banks and mutual funds, they have largely escaped the regulations by raising capital via private placement (under the Securities Act of 1933) and from a limited number of "qualified investors," i.e., accredited institutions and high-net worth individuals (under the Investment Company Act of 1940). Due to their lightly regulated nature, hedge funds are not required to report information about their characteristics, strategies, and performance to any authority or database. As a result, hedge funds are among the least transparent major market participants though according to some estimates by Credit Suisse / Tremont, they managed 1.5 to 2.0 trillion dollars of assets and accounted for about one-third of the equity trading volume in the U.S. during 2007.

The importance of hedge funds has attracted a growing volume of research; and due to the lack of mandatory disclosure, the burgeoning research on hedge funds has mostly relied on commercial hedge fund databases to which hedge funds report voluntarily. Prior research has documented several biases in hedge fund databases including the survivorship bias, backfilling bias, and smoothing bias (e.g., Ackermann, McEnally, and Ravenscraft (1999), Brown, Goetzmann, and Ibbotson (1999), Fung and Hsieh (2000), Liang (2000), Getmanksy, Lo, and Makarov (2004), and Bollen and Pool (2008)). However, the extant literature has not formally addressed the degree of self-reporting bias, arguably one of the most important biases in hedge fund databases. Self-reporting bias is a type of selection bias that results from hedge funds' choices to not report to any database, to initiate reporting at some time, or to discontinue reporting for various reasons, the common ones being liquidation and closed for new investment. Such a bias can potentially affect any study on the performance and risk characteristics of hedge funds but the magnitude or even the direction of the bias is yet unknown. Our paper fills this gap in the hedge fund literature by being 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 their self-reporting.

A hedge fund's choice to voluntarily report to a commercial database is likely to be non-random. Like all other economic activities, the reporting behavior of hedge funds should be determined by the cost-benefit trade-offs. On the benefit side, listing in a database enhances a fund's exposure to potential

investors, which is likely to be more significant for small and medium sized fund companies that desire more publicity but lack the resources for aggressive direct marketing.[5] The main cost of reporting is a partial loss of secrecy and privacy that many hedge funds value.[6] Moreover, keeping the reporting status constitutes a commitment to revealing a fixed set of information at fixed time intervals, depriving a hedge fund of the flexibility in publicizing selective information (such as return performance of a particular period of time) that is most favorable to the fund. Finally, investors attracted to hedge funds through database subscription tend to be more "retail" and short-term. Hedge funds usually value institutional 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.

Even after a fund decides to report to a commercial database, it exercises the discretion on the reporting initiation date and later may choose to exit from the database for both positive and negative reasons. On the positive side, if a hedge fund is closed to new investors due to its success and lack of scalable investment opportunities, then there would be no incentive to attract more capital. On the negative side, embarrassing losses or even the prospect of liquidation could be the reason for a hedge fund to stop reporting.

These scenarios related to the choice of reporting as well as initiation and discontinuation of reporting point towards 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 is a first attempt at quantifying the degree of the self-reporting bias in the hedge funds databases by analyzing the quarterly equity holdings of a complete list of hedge fund companies that file the Form 13F to the Securities and Exchange Commission (SEC) between 1980 and 2008. Because of the mandatory

[5] In order to be exempt from the regulations of the 1934 Securities Exchange Act and the 1940 Investment Company Act (and their amendments), a hedge fund cannot advertise to the general public through mass media such as newspapers and TV channels. Moreover, the investors the fund approaches directly must satisfy the requirement of "qualified investors." Therefore, reporting to a commercial database is often viewed as a cheap way to reach the target investor groups, where the database vendors bear the responsibility of ensuring that the only qualified investors have access to their databases.

[6] Though self-reporting hedge funds in general do not reveal holdings information to hedge fund databases, the reported information, such as descriptions of style classification, asset allocation, monthly returns, and leverage/hedging ratios, is often revealing of the funds' investment strategies.

nature of the 13F filings,[7] this sample is largely free from the selection bias due to 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 – CISDM, HFR, Eureka, MSCI, and TASS. This represents 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.

Upon classifying hedge funds' self-reporting status, we compare the performance and other characteristics of the self-reporting hedge funds to those of the non-reporting ones using returns and other portfolio statistics imputed from the quarter-end holdings of all hedge funds that file 13F forms. We further analyze the return dynamics around the initial and last reporting dates, and the impact of reporting on fund flows.

We find that during the period of 1980-2008, young and medium-sized fund companies that employ more diversified and higher-frequency trading strategies (using portfolio turnover rates as proxy) have a stronger incentive to self-report to databases, presumably to publicize their funds and attract potential investors. Given the characteristics of these funds, trading secrecy is less likely to be revealed through voluntary disclosure because of the diversified nature and the high portfolio turnover rates, both of which reduce the costs of reporting. Interestingly, the difference in the return performance, though slightly in favor of the non-reporting funds, is small. Presumably the positive and negative reasons prompting reporting initiation and termination offset one another. This is good news for the large body of research on hedge fund performance because the self-reporting bias may not have a material impact when it comes to performance evaluation especially if researchers use a multitude of commercial databases to exhaustively cover the universe of self-reporting hedge funds.

Importantly, we find, among the self-reporting funds, significant evidence that performance deteriorates both after the initial self-reporting date and after the reporting termination date. The first evidence supports the hypothesis that hedge funds strategically initiate self-reporting after a run of

[7] All institutions that have investment discretion over $100 million or more in Section 13(f) securities (mostly publicly traded equity; but also include convertible bonds, and some options) are required to disclose their quarter-end holdings in these securities.

superior performance; while the second evidence indicates that reporting termination is usually a sign of deterioration. The latter is further supported by the fact that net flows to funds tend to decrease after reporting termination, even after controlling for performance. Good performance prior to initiation of reporting offsets the poor performance subsequent to termination of report, which offers one explanation why we do not find significant difference between the performance of reporting and non-reporting funds.

The findings of our paper have implications for the growing research on hedge funds which examines their risk-return characteristics and persistence in their performance.[8] Our study contributes to the earlier work on hedge fund data biases by Ackermann, McEnally, and Ravenscraft (1999), Brown, Goetzmann, and Ibbotson (1999), Fung and Hsieh (2000, 2009), Liang (2000), Malkiel and Saha (2005), and Posthuma and Jelle van der Sluis (2003) among others. Interestingly, Fung and Hsieh (2000) argue that the returns of funds of hedge funds (FOFs) contain the aggregate investment experience of hedge fund investors and should also include the performance of non-reporting hedge funds. However, as they acknowledge themselves, their approach does not allow them to measure and assess the direction of self-reporting bias, which is the focus of our study. It is also important to note that computing the self-reporting bias from FOFs is difficult as they attempt to select underlying hedge funds based on qualitative and quantitative information, and therefore the sample of the funds they invest in is not going to be random, introducing another form of selection bias.

Our paper also determines the performance of funds both before they initiate reporting and after they cease reporting. The latter part is closely related with the work by Hodder, Jackwerth, and Kolokolova (2008), who estimate the returns of hedge funds after their disappearance from the databases using data on FOFs that invest in a portfolio of hedge funds, assuming some independence between the component funds' self-reporting status and the FOFs' investment decision. One limitation of their

[8] An incomplete list of hedge fund performance studies includes Amin and Kat (2003), Agarwal and Naik (2004), Agarwal, Bakshi, and Huij (2009), Avramov, Kosowski, Naik, and Teo (2007), Bollen and Whaley (2009), Fung and Hsieh (1997, 2001, 2004), Fung, Hsieh, Naik, and Ramadorai (2008), Getmansky, Lo, and Makarov (2004), Hasanhodzic and Lo (2007), Mitchell and Pulvino (2001), Patton (2009) and persistence in their performance (Kosowski, Naik, and Teo (2007), Jagannathan, Malakhov, and Novikov (2009)). For a survey of the hedge fund literature, see Agarwal and Naik (2005).

approach is that they need to estimate the holdings of FOFs since this data is not commercially available. Moreover, the validity of their assumption is questionable if the FOFs are more likely to pull out the money from funds before or shortly after they disappear from the databases due to the funds' bad performance. Our approach, in contrast, offers a more direct and comprehensive test on the propensity to report, as well as the effects of reporting initiation and termination.

In terms of using the hedge fund companies' 13F quarterly equity holdings, our paper is related to Griffin and Xu (2009) who use returns imputed from holdings of hedge funds to infer their overall performance. In addition to having a different sample (1,199 funds versus 306 funds in Griffin and Xu (2009)), the focus of our paper is also different as we relate the analysis of performance to the propensity and effects of voluntary reporting by hedge funds.

Our research contributes to the literature in several ways. At the specific level, it is the first study that formally analyzes the self-reporting bias in hedge fund databases. Our results will offer important benchmarks and references for hedge fund researchers and investment managers who use such data sources. More generally, the study provides insights into the motivation and consequences of voluntary disclosure by hedge funds, and in the same spirit, by other financial institutions. Finally, it raises interesting questions about the role of hedge fund regulation if voluntary disclosures are deemed inadequate. This is particularly pertinent in view of the on-going debate regarding the mandatory registration of hedge fund managers and more stringent disclosure rules.

The rest of the paper is organized as follows: Section I details data collection and classification, and provides an overview of the complete sample of 13F filing hedge fund companies. Section II compares the characteristics and return performance of self-reporting and non-reporting hedge fund companies. Section III analyzes the change in performance of self-reporting funds before and after their initial and last reporting dates, as well as the effects of reporting initiation and termination on fund flows. Finally, Section IV concludes.

I. Data and Overview

A. 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 (formerly the CDA/Spectrum database), available through Wharton Research Data Services (WRDS). The Form 13F filing, which discloses quarter-end holdings of an institution with a maximum of 45-day delay, is mandatory for all institutions that exercise investment discretion over $100 million of assets in equity and some other publicly traded securities.[9] The second source is a comprehensive self-reported hedge fund database created by the union of five major commercial hedge fund databases: CISDM, Eureka, HFR, MSCI, and TASS (henceforth, the "Union Hedge Fund Database" or simply 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 hedge 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 usually 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 level or at the level of pooled portfolio.[10] Hence, pairing a 13F filing institution to funds in the Union Hedge Fund Databases is often a one-to-multiple match (if a match exists). The matching between the two data sources is facilitated by the fact that the latter database reports the sponsor or management companies of individual funds in most cases.

The Thomson Reuters Ownership database consists of a list of 5,188 unique 13F-filing institutions for the 1980 -2008 period. We go through the list manually in order to identify whether each

[9] More accurately, institutions are required to disclose all securities that appear on the official list of "Section 13(f) Securities," published by the SEC periodically. This list includes almost all publicly traded equity, some preferred stocks, bonds with convertible features, warrant, and publicly traded call and put options. The Thomson Reuters Ownership database contains only holdings of equity, and does not include other securities (Aragon and Martin (2009)).

[10] A fund is usually defined at the level where participating clients combine their investment dollars and purchase/sell pooled portfolio units, rather than individual securities. The unit price is determined by dividing the market value of the pooled portfolio by the number of outstanding units.

filing institution has major hedge fund management business. We classify a 13F-filing institution as a "hedge fund company" if it satisfies one of the following: (i) It matches the name of one or multiple funds from the Union Hedge Fund Database. (ii) It is listed by industry publications (Hedge Fund Group (HFG), Barron's, Alpha Magazine, and Institutional Investors) as one of the top hedge funds. (iii) The company's own website claims itself as a hedge fund management company or lists hedge fund management as a major line of business.[11] (iv) The company is featured by news articles in Factiva as a hedge fund manager/sponsor. (v) Some 13F filer names are those of individuals. In such cases we search the full individual names over the internet (mostly through the filer and co-filer identity information on various types of SEC filings) and classify the name as a hedge fund if the person is the founder, partner, chairman, or other leading personnel of a hedge fund company. Notable examples in this category include Carl Icahn (founder and chairman 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).

Applying the above procedure yields 1,199 unique hedge fund companies among all 13F filing institutions. They include 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. In our analysis, we do not include full-service banks whose investment arms engage in hedge funds business (even if their hedge funds made to the top fund lists compiled by industry publications), 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, Boyson, and Naik (2009), Cici, Gibson, and Moussawi (2009), and Nohel, Wang, and Zheng (2009)). The reason for the exclusion is that the equity holdings of these full-service institutions in their 13F filings may not be informative about the investments of their hedge funds. Our results are qualitatively similar if we include the institutions with major hedge fund business in

[11] There is no official definition of a hedge fund. Even if a company's website does not formally mention hedge fund management as part of their business, we still classify the company as a hedge fund manager or sponsor if it manages pooled, privately organized investment vehicles that adopt performance-based compensation, after excluding private equity and venture capital businesses that also have performance-based compensation.

the list of hedge funds, except their presence will skew the statistics related to portfolio size because they tend to be much larger than the other hedge funds on the list.

Due to our top-down approach, our list of 13F filing hedge funds companies is considerably longer than those 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 hedge fund companies. In both papers, authors use a one-sided match from published hedge fund lists to the 13F database for the purpose of their research and did not classify hedge funds that fail to make to a major published list or choose not to report to any database. Given that the focus of this paper is to analyze the selection bias, it is particularly important that we adopt the top-down approach to compile a complete list of 13F-filing hedge funds.

Equally important for our research is the composition of 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. For example, Fung and Hsieh (1997) use monthly data from TASS Management and Paradigm LDC, Ackermann et al. (1999) use a combination of HFR and MAR databases, Liang (1999) uses HFR data and Liang (2000) compares the HFR and TASS databases for different data biases. More recently, Agarwal, Daniel, and Naik (2009) show that there is limited overlap among four commercial databases, and using one or two of them may result in exclusion of a large number of self-reporting hedge funds. We extend the approach of Agarwal, Daniel, and Naik (2009) by adding one more database (Eureka) to their list of four and use the union of five major databases to minimize the under-classification of self-reporting funds. Using multiple databases also enables us to resolve occasional discrepancies among different databases. Fung and Hsieh (2009) also emphasize the importance of using multiple databases by documenting that some funds that stop reporting to a database and are classified as defunct/graveyard funds by that database may be active and reporting to another database. We address

these concerns and account for the differing performance histories of a fund reporting to more than one database.

The Union Hedge Fund Database contains a sample of 11,417 hedge funds, which includes 6,245 equity-oriented hedge funds, over our sample period.[12] Figure 1 plots a Venn diagram that shows the percentages of funds report to each database individually and to all possible combinations of multiple databases. One of the most striking observations from Figure 1 is that 71% of the funds are covered exclusively by only one database with CISDM and MSCI having the maximum (25.8%) and minimum (5.8%) fraction of unique funds.[13] This underscores the importance of using multiple databases in order to achieve a comprehensive coverage of the hedge fund universe.

[Insert Figure 1 here.]

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

We next classify the self-reporting status of all the 1,199 hedge fund companies that file 13F by matching them to the Union Database. The classification entails two steps. In the first step, we match by name allowing minor variations. For example, "DKR Capital" from the 13F list is matched to "DKR Capital Inc." in the Union Database. The name-matching produces 645 self-reporting fund companies, or 53.8% of all 13F filing fund companies.

Second, 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 average monthly returns of all funds reported in the Union Database that belong to the same fund management company, weighted

[12] We take advantage of using multiple databases to fill the missing strategy information if the fund is covered by more than one database. However, despite this exercise, we still have strategy field missing for 483 out of the 11,417 funds in our sample and therefore we cannot determine if these funds are equity-oriented.

[13] 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). Most of the funds choose not to report to multiple databases because of the additional cost due to the different requirement imposed by different data vendors on reporting funds, such as the types of data fields, availability of audited financial statements, etc.

by their assets under management. 60 pairs (or 9.3% of the 645 self-reporting fund companies) turn out to have negative correlations[14], and for 219 pairs, the correlation is not defined due to a lack overlapping periods of data from both data sources. The self-reporting status of these funds is not convincingly established and therefore we exclude them from our main analysis (that is, they are considered neither self-reporting nor non-reporting). As a result, we end up with 366 self-reporting funds and 554 non-reporting funds. Our results are qualitatively similar, but noisier, if we leave the 279 (60+219) funds in the self-reporting subsample based on the name matching.

Figure 2 plots the distribution of all 13F-filings and the subset of self-reporting hedge fund companies over the years. Also plotted is the average portfolio size imputed from the 13F quarter-end holdings for both groups of fund companies, expressed in 2008 constant dollars using the Consumer Price Index (CPI) deflator. Figure 2 shows that both the number of 13F filing hedge fund companies and that of the subset of self-reporting fund companies have steadily increased over our sample period from 1980 to 2008, with a marked jump in the number of 13F filing hedge fund companies since 2001. Interestingly, the average portfolio size of self-reporting funds was higher than that of the non-reporting funds before 1988, but has been consistently lower than the latter since 1988.

Several forces underlie the changes in the relative size of the reporting and non-reporting funds. First, macro funds, which tend to be large in size, dominated the hedge fund industry prior to 1988. The trading strategies of these funds are hard to reverse engineer, implying lower costs of reporting to databases. Since 1988, in contrast, smaller long-equity short funds have become more popular. These funds are more sensitive about trading secrecy and hence are less willing to report to databases. Finally and maybe most importantly, there has been a structural change in the hedge fund investor profile in the 1990s. While high net-worth individuals were the predominant investors in the earlier period, institutional investors became the mainstay in the more recent time. This shift can potentially explain why large funds chose to report to commercial databases prior to 1988 to reach out to prospective retail investors but switched to alternative channels afterwards to target institutions.

[14] The same percentage number in the Griffin and Xu (2009) sample is 8.5%.

[Insert Figure 2 here.]

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 information from 13F filings. As a result, the unit of observation is at the hedge fund management company level, which we will term interchangeably as "hedge funds" for the rest of the paper when there is no danger of confusion.

The main advantage of relying on the 13F data source is that there is little selection bias associated with self-reporting as long as they meet the minimum hurdle of assets under management ($100 million). Therefore, comparing the portfolio composition and return performance of self-reporting with non-reporting funds could offer an unbiased view of hedge fund performance and shed light on the selection bias introduced by self-reporting.

The limitation of the Thomson Reuters Ownership database is that it only captures the long-equity portfolios of hedge fund companies and masks intra-quarter trading. Needless to say, our research methodology hinges on the proposition that long-equity positions are a substantive portion of the portfolios of equity-oriented hedge funds and that the returns imputed from quarter-end equity long positions are informative about the total returns of these hedge funds. This proposition is also the premise that underlies the earlier work by Brunnermeier and Nagel (2004) and Griffin and Xu (2009).

We believe that this proposition is valid for several reasons. First, among the self-reporting fund 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 fund company level and including returns from short positions and non-equity securities, and are net of fees) is 0.54; and the median number is slightly higher at 0.57. Both numbers are comparable to the correlation of 0.55 (mean) and 0.64 (median) reported in Griffin and Xu's (2009) sample. By resorting to the original 13F filings (rather than the data processed by Thomson Reuters) that include additional 13F securities such as some types of debt, call/put options and warrants, Aragon and Martin (2009) find that common stock account for, on average, 54.7% of the portfolio value of a random sample of 250 hedge fund companies.

Second, the importance of equity positions in hedge fund total returns is evident from the equity market betas of hedge funds. Using the monthly Credit Suisse/Tremont hedge fund indices from January 1993 to May 2009,[15] we find that the market beta of the index of all hedge funds is 0.33, and the four-factor equity return model explains 49.8% of the total variation. For the sub-index of equity long-short hedge funds, the two figures are 0.48 and 77.1% respectively.

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 condemning the 13F rule for taking the fund's "property without just compensation in violation of the Fifth Amendment to the Constitution."[16] 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, which eventually contributed to the "death spiral" in the summer of 2007 when many funds employing similar strategies attempted to cut their risks simultaneously in response to their losses (Khandani and Lo (2007)). A recent paper by Agarwal, Jiang, Tang, and Yang (2009) presents large-sample evidence of strategic delays by asset management companies (including hedge funds) in their 13F disclosure.

C. Overview of Hedge Fund using Quarter-End Equity Holdings Data

Before we compare self-reporting hedge fund companies to non-reporting 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. Further, we compare their statistics with those of other categories of 13F-filing institutional investors. Such an analysis represents the most complete overview of the long equity positions of hedge funds in the literature.

[15] Available from: http://www.hedgeindex.com/hedgeindex/en/default.aspx?cy=USD.
[16] For a more detailed discussion, see Philip Goldstein's interview in September 12, 2006 issue of *Business Week*: http://www.businessweek.com/print/investor/content/sep2006/pi20060913_356291.htm.

The other categories that we compare hedge funds to include: (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 (the type 5 institutions by the Thomson classification, excluding hedge funds classified by us). The Thomson Reuters type code 5 since 1998 is known to be problematic in that the category could include many misclassified institutions that should be assigned with the other type codes (mostly, type code 4). Therefore, we reassign an institution which has type code 5 after 1998 to an earlier code if available and different from 5. The comparison is reported in Table 1.

[Insert Table 1 here.]

Table 1 shows that hedge fund companies are much smaller in size compared to institutions of other categories where size is calculated as the total value of the quarter-end equity portfolio using reported shares and corresponding quarter-end stock prices reported in CRSP. In particular, the average size of a hedge fund company's long equity portfolio is 16.5% of that of a mutual fund management company; though the difference in the total assets under management is likely to be smaller because the former may have other asset exposures while mutual funds are more or less constrained to hold long positions in publically traded securities.

Hedge funds also tend to be younger. Because age changes mechanically with the reporting year for the same institution in a panel data, we simply consider the inception year of a filing institution as a proxy for age. The inception year is left-censored at 1980 which is the earliest year that Thomson Reuters has data coverage. The median hedge fund company started 13F filing 19 years after the median bank/insurance company; and the same differences with mutual fund companies and investment advisors are 17 and 7 years, respectively. These differences are all statistically significant at the 1% level.

Three measures point uniformly to the more active nature of hedge funds in portfolio management. First, they are significantly (at the 1% level) less diversified than all other categories as measured by median portfolio Herfindahl index, and the biggest difference is with respect to the mutual

funds (0.047 vs. 0.018). The same relation holds using the mean statistic except for the comparison with the "Other" category. Second, hedge funds' portfolio volatility is higher than all other categories using both mean (5.53%) and median (4.93%) standard deviation of monthly returns imputed from quarter-end holdings, and the differences are all significant at the 1% level.

Third, hedge funds' inter-quarter portfolio turnover rates, average (median) of 91.6% (81.5%) annually, is 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, with all differences being significant at the 1% level. Here, the portfolio turnover rate is compounded from the inter-quarter turnover rates, calculated as the lesser of purchases and sales, divided by the average portfolio size of the last and the current quarter.[17] 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 to the current quarter-end and the average of the stocks prices at the two quarter-ends. The logic of using the *lesser* (rather than the average) of purchases and sales is to free the measure from the impact of net flows. 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 hedge funds' more active management bring about superior returns? The answer is not obvious from Table 1. We compute monthly excess return for each institution as the difference between the imputed portfolio return and the value-weighted CRSP all 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[18] and calculate the buy-and-hold return for the month. It turns out that all categories have average

[17] 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 does not account for intra-quarter trading (including that of non-equity securities), which may significantly contribute to the funds' turnover.

[18] 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.

and median excess returns close to zero.[19] Moreover, hedge funds outperform all the other institutions on average, though only the differences between the average excess returns of hedge funds and those of investment advisors and other institutions are statistically significant. If we use median excess return as the metric, hedge funds outperform all other institutions significantly except "Other" institution type. When we use one-factor and four-factor alphas as the performance metric, hedge funds seem to underperform other institutions on average, with all pair-wise differences being significant except the difference in one factor alphas of hedge funds and "Other" institutions.[20] However, the magnitude of the differences is small. The overall evidence suggests that hedge funds do not command superior returns from their long equity positions on average. We will analyze the performance within the hedge fund group in more detail in the following sections.

II. Comparing Self-Reporting and Non-Reporting Hedge Funds

A. The Economics of Self-Reporting

After characterizing the sample of all 13F filing hedge fund companies, the first natural question to ask is whether hedge funds which choose to report to commercial hedge fund databases are systematically different from those that choose not to report. Answer to this question is the key to understanding the selection bias in hedge fund databases that has important implications for the current research on hedge funds.

Like other economic activities, the reporting behavior of hedge funds are outcomes of cost-benefit trade-offs. The benefit that is 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

[19] Given that institutions as a whole hold a majority stake in public equities (the 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.

[20] Since we examine the performance of long equity portfolios of institutions, we do not need to use multifactor models augmented by option factors as in Agarwal and Naik (2004) and Fung and Hsieh (2001, 2004).

benefits are likely to be more significant for small and medium sized 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,[22] indicating the industry's general reluctance for or even strong opposition to more transparency. Though self-reporting hedge funds in general do not reveal holdings information to hedge fund databases, the reported information, such as general descriptions of style classification, asset allocation, monthly returns, and leverage/hedging ratios, is often revealing of the funds' investment strategy. For example, proposed "hedge fund replication" strategies that promise to provide 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. Such a 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 "retail" based and shorter-term, 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 (James and Karceski (2006), Chen, Goldstein, and Jiang (2009)). 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 databases subscribers.

[22] Such resistance culminated in *Goldstein vs. Securities and Exchange Commission* (details in http://www.seclaw.com/docs/ref/GoldsteinSEC04-1434.pdf) where Phillip Goldstein, the manager of hedge fund Bulldog, challenged an SEC 2004 rule that required most hedge fund advisors to register with the SEC by early 2006. The decision of the Court, made in June 2006, was mostly in favor of Goldstein.

While it is understandable that the worst performing funds are unlikely to report to commercial databases—either because they do not wish to publicize the embarrassment or because they do not survive long enough to satisfy the requirement for track records by most data vendors—Lhabitant (2004) offers one explanation to the general absence of the largest and most successful hedge funds in the commercial databases. Not only these funds have low needs for enhanced visibility, but they are also concerned that communicating performance to a data vendor may lead to inclusion in that data vendor's index, which automatically raises the performance of that index. As a result, these hedge funds' individual performance will appear less differentiated. If this reason is valid, then the sample of self-reporting funds will be biased toward funds with average performance.

B. Comparison of Self-Reporting and Non-Reporting Hedge Fund Companies

A first step toward answering the question "who report" is to compare the pooled sample of 13F-filing hedge fund companies that never appear in the Union Database (there are 554 such non-reporting companies) and those that appear in the database for some time during our sample period (there are 366 such self-reporting companies). We do not include the 279 fund companies whose reporting status cannot be accurately verified. Table 2 reports the comparison of fund companies and portfolio characteristics: portfolio size, portfolio concentration, returns volatility, portfolio turnover rate, and fund company inception year.

[Insert Table 2 here.]

Table 2 Panel A reveals several patterns regarding the characteristics of self-reporting funds. First, the portfolio size of self-reporting hedge funds are more or less comparable to the non-reporting ones, though the latter has much higher standard deviation. The self-reporting funds are slightly smaller by the mean statistic but somewhat larger by median comparison, indicating that the largest fund companies are under-represented in the set of self-reporting funds. This finding is intuitive as the larger funds possibly facing decreasing returns to scale and capacity constraints may have weaker incentives to report to commercial databases for attracting more capital. Second, the self-reporting hedge funds have

lower portfolio concentration than that of the non-reporting funds as measured by the average portfolio Herfindahl index (average of 0.08 versus 0.11, significant at 1% level). The average monthly return volatilities of the two categories are almost identical, but the self-reporting funds have considerably higher portfolio annualized turnover rates (106%) than that of the non-reporting funds (79%) and the difference is significant at the 1% level. Again these findings conform to the economics of reporting as less concentrated (or more diversified) and higher turnover funds need to worry less about their trading strategies being revealed through self-reporting. Finally, the average inception year (defined as a fund company's first appearance in the Thomson Reuters database) is very similar for both groups, though the median self-reporting fund is two years younger than its non-reporting counterpart.

Table 2 Panel A further compares the loadings on common risk factors by self-reporting and non-reporting funds. Interestingly, the equity positions of self-reporting funds have significantly higher exposure to the size (SMB) and book-to-market (HML) factors where the differences in both mean and median are significant at the 1% level. The difference in the loadings on the market factor follows 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 exposure to common risk factors hardly constitutes trading secrecy, these results support the hypothesis that fund with less conventional trading strategies (i.e., lower factor loadings) are more reluctant to reveal their information to databases.

Moving on to the differences in return performance reported in the Table 2 Panel B, we observe that average (median) raw returns of self-reporting funds are significantly higher, at the 1% (5%) level, than those of the non-reporting funds. However, both the magnitude and significance of the differences drop precipitously when the returns are adjusted by the market benchmark (i.e., return in excess of the market), or by the CAPM one-factor or using the four factors (market, size, book-to-market, and momentum) developed in Carhart (1997).

The two pooled samples compared in Table 2 are not necessarily directly comparable in that self-reporting and non-reporting fund companies may exist in the Thomson database for different periods and different lengths of time. To refine the comparison, we make the following adjustments: For each self-

reporting fund, we crop out the period for which it appears in the Thomson Reuters Ownership database (which may contain periods before, during, and after its reporting to the Union Database). We then find non-reporting fund companies that have 13F data over the same period (or with the maximum overlap). If there are ties in matches, we choose the one that is closest in portfolio size as the self-reporting fund to be the "matching fund." Table 3 reports the results from such refined comparison.

[Insert Table 3 here.]

The comparisons between the two groups regarding the differences in mean and median of turnover rates (reported in Table 3 Panel A) are qualitatively similar to those shown in Table 2, but the magnitudes of the differences are strengthened. However, the differences in the median portfolio concentration are now positive and significant. The portfolio sizes of the paired funds are almost identical, only due to the matching algorithm that is based on this variable. Finally, the matching non-reporting funds are now much older, which is again an artifact of the algorithm which favors matching funds with longer periods of 13F filings.

The return differences between the mean and median return measures over the matched time period (reported in Table 3 Panel B) indicate that self-reporting funds underperform non-reporting funds by 2-8 basis points monthly using the various performance measures, but none of the differences are statistically significant. Interestingly, the differences by percentile values indicate that for lower percentiles (e.g., the 5^{th} percentile), self-reporting funds perform significantly worse (at the 5% and 10% levels) using two of the three benchmark-adjusted return measures, while the pattern does not hold at percentiles above median. Combined evidence indicates that a small fraction of reporting funds has poor performance and may be struggling; while the most successful ones are no more prone to self-reporting.

The only conflicting difference between the results from Table 2 Panel B and Table 3 Panel B is the relative ranking of raw performance between the two groups of funds: it is significantly positive in favor of the self-reporting funds in Table 2 while negative (but short of significance) in Table 3. But such an inconsistency is not observed using any of the benchmark-adjusted returns. Taken together, these figures indicate timing of hedge fund reporting based on the market condition: hedge funds that were

active during years when the overall market performed well were more likely to report to hedge fund databases. Section III will discuss the issue of timing of reporting in more detail.

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 (using portfolio Herfindahl index and turnover rates as proxies), and have higher loadings on common factors—presumably trading secrecy is less likely to be revealed through voluntary disclosure or is less important when portfolio involves more stocks, evolves more quickly, and have more exposure to common risk factors. This pattern echoes Agarwal, Jiang, Tang, and Yang's (2009) finding that hedge funds adopting less conventional investment strategies are more likely to resort to confidential 13F filing in order to delay revealing their quarter-end positions. In both cases, funds who value privacy more are more likely to refrain from voluntary disclosures or to seek exemptions from mandatory ones.

Finally, the difference in the return performance, though slightly in favor of the non-reporting funds, is small.[23] This is good news for the existing and ongoing studies on hedge fund performance because the self-reporting bias may not have a material impact when it comes to performance evaluation. Section II.A. hypothesizes that the sample of self-reporting funds might be over-represented by funds with average performance. Therefore the selection bias due to self-reporting could be offset between the absence of both the most and least successful funds. Fung and Hsieh (2000) conjectured, with the support of some anecdotal evidence, that the selection bias due to self-reporting is limited because on the one hand "only funds with good performance want to be included in a database," while on the other hand "managers with superior performance did not necessarily participate in vendors' databases." Our results are supportive of their conjecture.

[23] This result is consistent with Brav, Jiang, Partnoy, and Thomas (2008) who find that hedge funds reporting to two commercial databases perform worse than the non-reporting ones among the sample of activist hedge funds, but the difference is not statistically significant. They measure performance of these funds mainly by the abnormal returns of the companies targeted by the activist funds during the event window.

III. Comparison of Self-Reporting Funds Before and After Reporting Initiation and Termination

This section analyzes the second type of selection bias conditional on the subsample 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 also contribute to a selection bias. Until now, the extant literature has not been able to quantify this type of selection bias as the performance of funds before reporting initiation and after termination of reporting is not observable from the commercial databases. Our study explores the 13F filings, which are not constrained by funds' reporting status to the commercial databases.

A. Comparison of Fund Companies Before and After the Initial Reporting Date

The Union Hedge Fund Database provides information on the dates when the hedge funds enter the databases. If a fund company reports to multiple constituent sources in the Union Database, we use the earliest date. Among all 13F-filing hedge fund companies, 103 out 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. For rest of the funds, such exact information is not available and all we can observe is the first date of the performance data recorded in the database. Following the practice of the literature (e.g., Ackermann, McEnally, and Ravenscraft (1999)), for such funds we add 24 months to the first performance dates to form the approximate first reporting dates, effectively assuming a typical practice of 24 months' back-filling by reporting funds. We know this assumption is problematic as Fung and Hsieh (2009) document periods longer than 24 months between the inception and first reporting dates. Hence, for robustness, we conduct our analyses using both the entire sample and the subsample with accurate information on initial reporting date, and focus more on the latter results for our discussion that follows.

For each fund whose reporting date falls within the 1980-2008 period, 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). Results are reported in Table 4.

[Insert Table 4 here.]

Panel A Table 4 shows no significant difference in the performance of funds before and after the initial reporting date when we use the full sample, though performance is overall worse after initial reporting compared to the period before. The difference in the raw monthly return is 52 basis points, or 6 percent at an annualized rate.

However, when we use the subsample of funds for which we have accurate initial reporting dates, we observe from Panel B of Table 4 that the performance after initial reporting is significantly lower than that before reporting. The average raw returns and measures of risk-adjusted performance (excess returns, CAPM alpha, and four-factor alpha) are lower by 90, 73, 58, and 24 basis points per month respectively, and all except four-factor alpha differences being statistically significant at the 1% level in addition to being economically meaningful. We obtain similar results using the median performance with the corresponding figures being 49, 32, 33, and 19 basis points per month respectively. The results in Panel B are much more significant and coherent, compared to the full-sample results in Panel A, albeit with a smaller sample, indicating that accurate reporting dates are essential to identify the selection bias around reporting initiation for the sample of self-reporting funds, providing support to the arguments in Fung and Hsieh (2009).

The interpretation of this difference is further facilitated by Figure 3. Panels A and B plot the time series of the monthly raw returns and excess returns averaged across the 77 hedge funds (with accurate initial self-reporting dates) from 24 months before the reporting month, to 24 months afterwards. The two dotted horizontal lines marked the time-series averages of the two sub-periods. The figure indicates 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.

[Insert Figure 3 here.]

The subsequent normal performance after a run of superior one supports the hypothesis that hedge funds strategically time their initiation of reporting to hedge fund databases, if they decide to report. Given the customary back-filling practice (that is, hedge funds usually send retrospective return data to commercial databases), our analysis shows that the early periods of reported returns contain an upward bias for inferring the reported funds' normal performance. Hence, the trimming of early-period returns in return analysis as adopted by the literature is justified. However, the different results between Panels A and B of Table 4 also points to the limit of the simple 24-month trimming practice as it does not seem to identify the true initial reporting dates, and hence does not completely clear the selection bias due to strategic timing in reporting initiation.

B. Hazard Analysis for reporting initiation

To provide more detailed evidence about the reporting initiation, we present the hazard analysis of reporting initiation for the subsample of fund companies with accurate initial reporting date information. In the language of hazard analysis, 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. Once the hedge fund has initiated reporting, it exits the sample because the spell has "failed". We estimate our instant hazard model with respect to a set of time-varying explanatory variables (*X*), such as fund characteristics. That is, the 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 Hedge Fund Database (observations of completed spells) or to the end of our sample period (observations of censored spells).

We adopt the semi-parametric 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:

$$h(t) = h(0)e^{X_t'\beta} \quad (1)$$

where t is the number of periods since the fund company's first appearance in the Thomson Reuters database. In this setting, a positive coefficient β_k indicates that an increase in the covariate X_k is associated with an increase in the instantaneous probability of hedge funds' initiating reporting to a database during period t. We conduct the analysis at the quarterly frequency and results are reported in Table 5. Following the norm adopted in hazard analyses and to facilitate interpretation, Table 5 reports the hazard ratio (also called "exponentiated coefficient" associated with each covariate rather than the 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 testifies the significance of raw coefficient (β_k) being difference from zero, or of the hazard ratio (e^{β_k}) being different from unit.

[Insert Table 5 here]

According to Table 5, hedge funds after better performing periods have higher probability of reporting initiation during the current period: hazard ratios associated with performance (lagged) are significantly higher than one. This result supports evidence in Figure 3: hedge fund's performance tends to be abnormally high before reporting initiation. When risk-adjusted measures of performance are considered (columns (2)-(4) in Table 5) and market returns are controlled for, the evidence suggests that hedge funds have higher probability of reporting initiation after a period of good market performance. This result is consistent with the ease in marketing funds when overall market performs well. The coefficient of the market return is insignificant when performance is measured by raw returns because the latter already contains information about market returns.

Table 5 highlights additional elements in hedge funds' strategic reporting initiation. First, when the proxy for the aggregate flow to hedge fund industry is high, hedge funds have 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 due to stock price appreciation. This evidence suggests that booming of the hedge fund industry provides enough capital to many funds, leading to their lowered needs to enhance exposure to potential investors by reporting initiation.

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, Getmansky, Liang, and Wermers (2009), Huang, Wei, and Yan (2007)), suggesting investors' general aversion to risk. Moreover, some commercial databases adopt Sharpe Ratio as a performance measure which is unfavorable to funds with volatile returns. As a result, funds are reluctant to publicize themselves to commercial databases when their returns are volatile.

Finally, hedge funds have higher probability of reporting initiation in their young stage if they decide to report: the hazard ratios associated with fund age are significantly lower than one. This result is expected 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 the costs of revealing trading secrecy when funds report to databases.

Please note that the pair-wise comparison analyses (reported in Tables 2 and 3) and the hazard analysis (reported in Table 5) do not necessarily yield coefficients of the same sign or of similar significance levels. While 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. For example, the hazard analysis indicates that funds are less likely to initiate reporting

during the period of volatile returns; but reporting funds as a whole do not have less return volatility as compared to non-reporting funds.

C. Comparison of Fund Companies Before and After the Last Reporting Date

There are 187 funds in our sample that terminated reporting to the Union Database at some point during the 1980-2007 period. For these funds, we are able to analyze the determinants of reporting termination using the same method as we used in Table 4 for report initiation. Moreover, for these funds we have more information about their termination decision due to their reporting status when the decision is made. Results are reported in Table 6.

[Insert Table 6 here.]

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 when their performance starts deteriorating (Ackermann, McEnally, and Ravenscraft (1999), Liang (2000), and Fung and Hsieh (2000, 2002) among others). What is interesting and unique about our analysis here is that we are able to determine the performance of funds after they disappear from the commercial databases. Hence, our analysis is analogous to computing the delisting returns for stocks in Shumway (1997) and Shumway and Warther (1999). Table 6 shows that the average monthly raw returns and the three measures of risk-adjusted performance: excess returns, CAPM alpha, and four-factor alpha, are lower by 1.9%, 0.3%, 0.1%, and 0.2% on a monthly basis after the termination of reporting (the first two being significant at the 1% and 5% levels). We obtain similar results for median performance differences with the corresponding figures being 1.5%, 0.2%, 0.03%, and 0.2% per month, with the first and last differences being significant at the 1% level. A graphical illustration similar to the returns around reporting initiation date is provided in Panels C and D of Figure 3. The message is also similar to what is conveyed by the table.

About 64% of the funds (119 funds) that terminate reporting in our sample provide reasons for termination to the commercial databases. In 112 out of the 119 cases, the given reasons indicate distress (such as liquidation, fund being dormant or data vendor being unable to contact the fund). Other given

reasons could be positive (such as being closed to new investors) or unclear (such as being merged to another fund) but such cases are rare. When we focus on the subsamples partitioned by stated reasons, we do not find significant changes in performance after reporting termination due to the much reduced sample size. 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 report termination—manifested by the higher before-after return gap in raw returns than benchmark-adjusted returns as shown in Table 6. These findings about hedge fund report termination are highly consistent with the patterns associated with stock delisting. Finally, the combination of good performance prior to reporting initiation (results in the previous section) and poor performance following reporting termination could offer one reason for similar average performance of reporting and non-reporting funds as analyzed in Section IIB as the two forces might be offsetting each other.

D. Effects of Self-Reporting on Hedge Fund Flows

D1. Reporting initiation

If the primary benefit of reporting to hedge fund databases (enhancing a hedge fund company's exposure to potential clients) is justified, then a hedge fund should experience, on average, an increase in flows after the initiation of reporting compared 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 afterwards. We then conduct the following regression at the fund (indexed by *i*)-quarter (indexed by *t*) level:

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

In (2), $Flow_{i,t}$ 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 due to changes in investment (and not due to appreciation of the stock prices), and is a proxy for the fund flows. The all-sample average (median) percentage flow to hedge funds companies is 3.6% (1.4%). D_{t-j} are the dummy variables for four quarters before and after the initial reporting date. $Performance_{t-3:t}$ is the monthly average of the

performance measure during the past four quarters that end in the current quarter, and $Control_{i,t-1}$ are lagged control variables including portfolio size (in log), fund age (numbers of quarters since first appearance on Thomson Reuters, in log), portfolio turnover rates, and portfolio volatility. Based on the lessons learnt from Table 4 (discussed in Section III.A), we focus on the subsample of funds with accurate initial reporting dates only. Results are reported in Panel A of Table 7.

[Insert Table 7 here.]

The three columns in Table 7 Panel A estimate equation (2) using three benchmark-adjusted return performance measures: return in excess of the market, CAPM one-factor alpha, and four-factor alpha. The coefficients on *Performance* tell us that flows are highly responsive (significant at the 1% level) to risk-adjusted returns, regardless of whether we use a simple market benchmark (return in excess of the market) or alphas from one-factor or four-factor models. Our findings are economically significant too. For example, for a one percentage point increase in monthly return in excess of the market (or 12 percentage points during the four quarters when performance is measured), net flows to a fund increase by 2.5% of the total portfolio value (see column 1 of Table 7 Panel A). This flow pattern is similar to what the literature has documented for mutual funds (e.g., Chevalier and Ellison (1997)).

Table 7 shows a small increase in flows during quarters *t+1* and *t+2* using four-factor alphas, where *t* is the initial reporting quarter. However, this increase is transient and does not persist into future quarters, possibly due to a deterioration in performance after reporting initiation, as we show earlier in our paper. When we test for changes in flows over the full window through a formal F-test: $\sum_{j=0}^{4} \lambda_j - \sum_{j=-4}^{-1} \lambda_j = 0$, we are unable to reject the null of equality. Therefore, reporting to databases does not lead to higher flows over a longer window comparing flows during in the year following initiation to those during the year preceding reporting initiation. It is worth pointing out 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 would choose to report to databases,

and such a decision process biases down the estimate for the incremental flows from exposure through the databases.

D2. Reporting Termination

Lastly, we repeat the analysis used in regression (2) on reporting termination. Results reported in Panel B of Table 7 show that funds encounter significantly lower net flows (or more outflows) after reporting termination. An F-test of $\sum_{j=0}^{4} \lambda_j - \sum_{j=-4}^{-1} \lambda_j = 0$ is strongly rejected (at the 1% level) in favor of a negative change in net flows across all regression specifications. More specifically, the cumulative net outflows during the reporting termination quarter and the four quarter afterwards amount to 29-34 percent of the lagged portfolio size. This evidence adds further support to the hypothesis that delisting from hedge fund databases is in general a sign of deterioration.

[Insert Table 7 here.]

IV. Conclusion

We document several interesting results from the first comprehensive study that formally analyzes the self-reporting bias in hedge fund databases. First, we find that using multiple commercial databases can mitigate the self-reporting bias in performance, and that the self-reporting and non-reporting funds do not differ significantly in return performance on average. However, we observe significant differences in the portfolio concentration, turnover, portfolio size, and age of the self-reporting and those of non-reporting funds. Specifically, we document 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 employ more diversified and higher-frequency trading strategies, indicating that trading secrecy is less likely to be revealed through voluntary disclosure when it involves more stocks and evolves more quickly. Second, we show that reporting funds time their reporting initiation and termination strategically. Initiation tends to happen after a run of superior performance both in terms of high returns and low volatility, and termination usually occurs after early signs of return

deterioration. These funds experience substantial deterioration in performance after both the reporting initiation and termination dates. Related to this finding, we provide evidence on transitory impact on inflows after reporting initiation and greater outflows from funds after they cease reporting to commercial databases.

Finally, our analyses indicate non-trivial impacts of market-wide returns on fund reporting initiation/termination and fund flows in that both variables are more sensitive to raw returns than to risk-adjusted returns. Such evidence suggests that hedge funds investors chase absolute as well as excess returns, even though market-wide conditions cannot be attributed to skills of fund managers. As a result, hedge funds time their reporting or termination of reporting in response to their own performance as well as to the market-wide conditions.

Taken together, 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 shed light on the motivation and consequences of voluntary disclosure by hedge funds. Finally, by comparing databases from mandatory and voluntary sources, our research also contributes to the on-going debate regarding more stringent disclosure rules for hedge funds.

References

Ackermann, Carl, Richard. McEnally, and David. Ravenscraft, 1999, The performance of hedge funds: Risk, return and incentives, *Journal of Finance* 54, 833-874.

Agarwal, Vikas, and Narayan Y. Naik, 2004, Risks and Portfolio Decisions Involving Hedge Funds, *Review of Financial Studies*, 17, 63–98.

Agarwal, Vikas, and Narayan Y. Naik, 2005, Hedge Funds, *Foundations and Trends in Finance*, 1, 103-170.

Agarwal, Vikas, Nicole M. Boyson, and Narayan Y. Naik, 2009, Hedge Funds for Retail Investors? An Examination of Hedged Mutual Funds, *Journal of Financial and Quantitative Analysis*, 44, 273-305.

Agarwal, Vikas, Naveen D. Daniel, and Narayan Y. Naik, 2009, Role of Managerial Incentives and Discretion in Hedge Fund Performance, *Journal of Finance*, 64, 2221-2256.

Agarwal, Vikas, Wei Jiang, Yuehua Tang, and Baozhong Yang, 2009, Do Institutional Investors Have an Ace up Their Sleeves? –Evidence from Confidential Filings of Portfolio Holdings, Working paper, Columbia University and Georgia State University.

Amin, Gaurav, and Harry Kat, 2003, Hedge Fund Performance 1990-2000: Do the Money Machines Really Add Value?, *Journal of Financial and Quantitative Analysis*, 38, 251–274.

Aragon, George and Spencer Martin, 2009, A Unique View of Hedge Fund Derivative Usage: Safeguard or Speculation? Working paper, Arizona State University.

Avramov, Doron, Robert Kosowski, Narayan Y. Naik, and Melvyn Teo, 2007, Investing in hedge funds when returns are predictable, Working paper, London Business School, Singapore Management University, and University of Maryland.

Bollen, Nicholas P., and Veronika K. Pool, 2008, Conditional Return Smoothing in the Hedge Fund Industry, *Journal of Financial and Quantitative Analysis*, 43, 267–98.

Bollen, Nicholas P., and Robert Whaley, 2009, Hedge Fund Risk Dynamics: Implications for Performance Appraisal, *Journal of Finance*, 64, 985-1035.

Brav, Alon, Wei Jiang, Frank Partnoy, and Randall Thomas, 2008, Returns to Hedge Fund Activism, *Financial Analyst Journal* 64, 45-61.

Brown, Stephen J., William N. Goetzmann, and Roger G. Ibbotson, 1999, Offshore Hedge Funds: Survival and Performance 1989-95, *Journal of Business*, 72, 91–117.

Brunnermeier, Markus and Stefan Nagel, 2004, Hedge Funds and the Technology Bubble, *Journal of Finance*, 59, 2013–2040.

Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance*, 52, 57-82.

Chen, Qi, Itay Goldstein, and Wei Jiang, 2009, Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Outflows, *Journal of Financial Economics* forthcoming.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167-1200.

Cici, Gjergji, Scott Gibson, and Rabih Moussawi, 2009, For Better or Worse? Mutual Funds in Side-by-Side Management Relationships with Hedge Funds, *Journal of Financial Intermediation* forthcoming.

Cox, D.R., 1972, Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society* 34, 187–220.

Ding, Bill, Mila Getmansky, Bing Liang, and Russ Wermers, 2009, Share Restrictions and Investor Flows in the Hedge Fund Industry, Working paper, University of Maryland and University of Massachusetts.

Fung, William, and David A. Hsieh, 1997, Empirical characteristics of dynamic trading strategies: the case of hedge funds, *Review of Financial Studies*, 10, 275–302.

Fung, William, and David A. Hsieh, 2000, Performance Characteristics of Hedge Funds and CTA funds: Natural versus Spurious Biases, *Journal of Financial and Quantitative Analysis*, 35, 291–307.

Fung, William, and David A. Hsieh, 2001, The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers, *Review of Financial Studies*, 14, 313–341.

Fung, William, and David A. Hsieh, 2002, Hedge-fund benchmarks: Information content and biases, *Financial Analysts Journal* 58, 22-34.

Fung, William, and David A. Hsieh, 2004, Hedge Fund Benchmarks: A Risk-Based Approach, *Financial Analyst Journal*, 60, 65–81.

Fung, William, and David A. Hsieh, 2009, Measurement biases in hedge fund performance data: An update, *Financial Analysts Journal*, 65, 36-38.

Fung, William, David A. Hsieh, Narayan Y. Naik, and Tarun Ramadorai, 2008, Hedge Funds: Performance, Risk and Capital Formation, *Journal of Finance*, 63, 1777–1803.

Getmansky, Mila, Andrew Lo, and Igor Makarov, 2004, "An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns," *Journal of Financial Economics*, 74, 529–609.

Griffin, John and Jin Xu, 2009, How Smart Are the Smart Guys? A Unique View from Hedge Fund Stock Holdings, *Review of Financial Studies*, 22, 2531-2570.

Hasanhodzic, Jasmina, and Andrew Lo, 2007, Can Hedge Fund Returns be Replicated? The Linear Case, *Journal of Investment Management*, 5, 5–45.

Hodder, James E., Jens C. Jackwerth, and Olga Kolokolova, 2008, Recovering Delisting Returns of Hedge Funds, Working paper, University of Konstanz and University of Wisconsin-Madison.

Huang, Jennifer, Hong Yan, and Kelsey Wei, 2009, Volatility of Performance and Mutual Fund Flows Working paper, University of Texas Austin and University of Texas Dallas.

Jagannathan, Ravi, Alexey Malakhov, and Dmitry Novikov, 2009, Do Hot Hands Exist Among Hedge Fund Managers? An Empirical Evaluation, *Journal of Finance* forthcoming.

James, Christopher and Jason Karceski, 2006, Investor monitoring and differences in mutual fund performance, *Journal of Banking and Finance*, 30, 2787–2808.

Kat, Harry, and Helder Palaro, 2006, Who Needs Hedge Funds? A Copula-Based Approach to Hedge Fund Return Replication, Working paper, City University London.

Khandani, Amir, and Andrew Lo, 2007, What Happened to the Quants in August 2007?, *Journal of Investment Management*, 5, 5-54.

Kosowski, Robert, Narayan Y. Naik, and Melvyn Teo, 2007, Do Hedge Funds Deliver Alpha? A Bayesian and Bootstrap Analysis, *Journal of Financial Economics*, 84, 229–264.

Liang, Bing, 2000, Hedge Funds: The Living and the Dead, *Journal of Financial and Quantitative Analysis*, 35, 309-326.

Malkiel, Burton G., and Atanu Saha, 2005, Hedge Funds: Risk and Return, *Financial Analysts Journal*, 61, 80–88.

Lhabitant, Francois-Serge, 2004, *Handbook of Hedge Funds,* John Wiley & Sons Ltd, West Sussex, England.

Mitchell, Mark, and Todd Pulvino, 2001, Characteristics of Risk in Risk Arbitrage, *Journal of Finance*, 56, 2135–2175.

Nohel, Tom, Zhi Jay Wang, and Lu Zheng, 2009, Side-By-Side Management of Hedge Funds and Mutual Funds, *Review of Financial Studies* forthcoming.

Patton, Andrew, 2009, Are "Market Neutral" Hedge Funds Really Market Neutral?, *Review of Financial Studies*, 22, 2495-2530.

Posthuma, Nolke, and Pieter Jelle van der Sluis, 2003, A Reality Check on Hedge Funds Returns, Working paper, ABP Investments, APG Investments, Free University of Amsterdam, and GTAA Fund.

Shumway, Tyler, 1997, The Delisting Bias in CRSP Data, *Journal of Finance*, 52, 327–340.

Shumway, Tyler, and Vincent A. Warther, 1999, The Delisting Bias in CRSP's Nasdaq Data and Its Implications for the Size Effect, *Journal of Finance*, 54, 2361–2379.

Stulz, René, 2007, Hedge Funds: Past, Present, and Future, *Journal of Economic Perspectives*, 21, 175–194.

Figure 1
Venn Diagram of the Union Hedge Fund Database

The Union Hedge Fund 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.

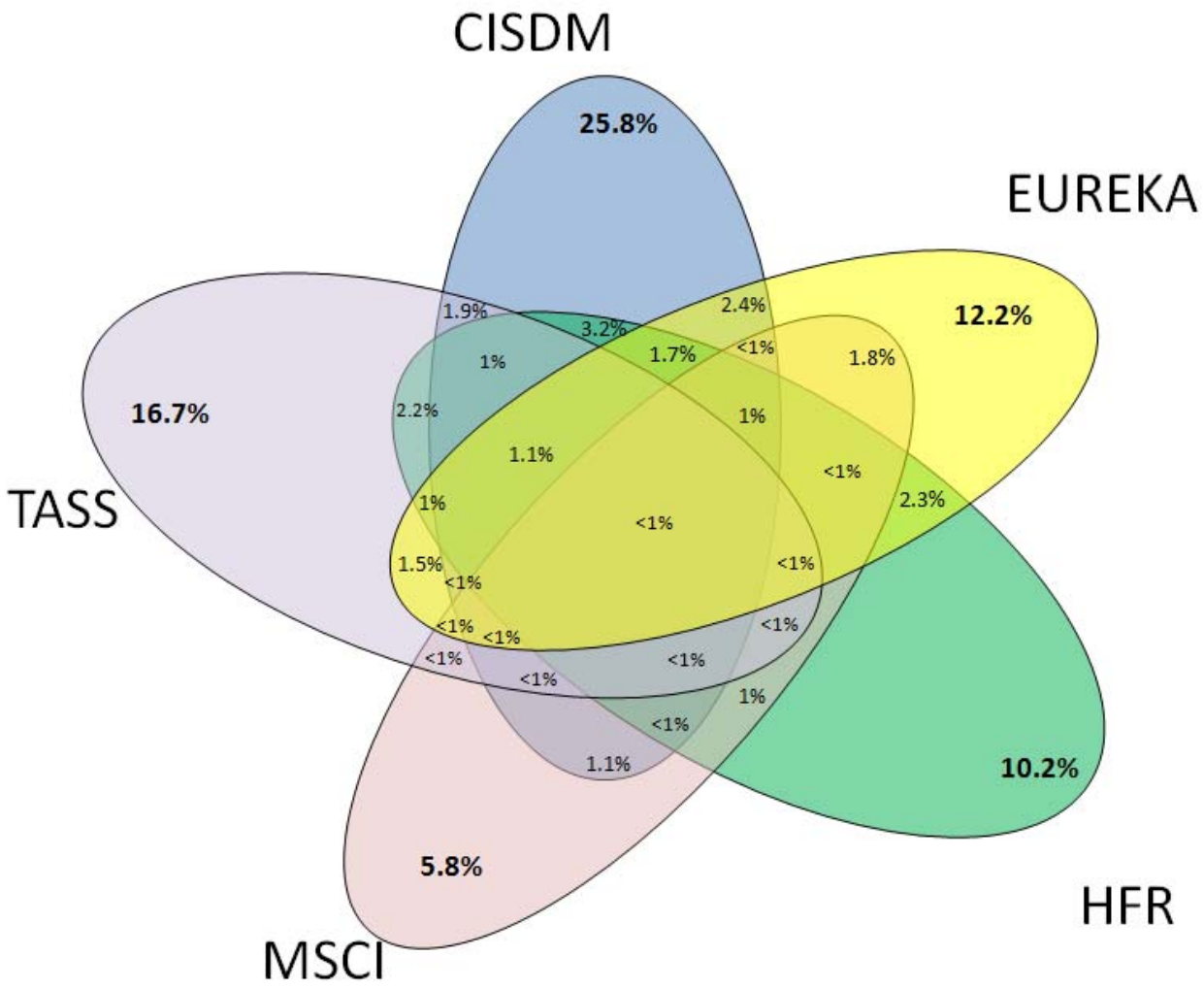


Figure 2

Number of Hedge Funds and Average Portfolio Size

The two solid lines (scale to the left axis) plot the number of 13F-filing hedge funds and the number of self-reporting hedge funds over the period 1980-2008. The two dotted lines (scale to the right axis) plot the average equity portfolio size of self-reporting hedge funds and non-reporting ones. The portfolio size is calculated using the quarter-end holdings disclosed in 13F filings, and is expressed in 2008 constant dollars using the CPI deflator.

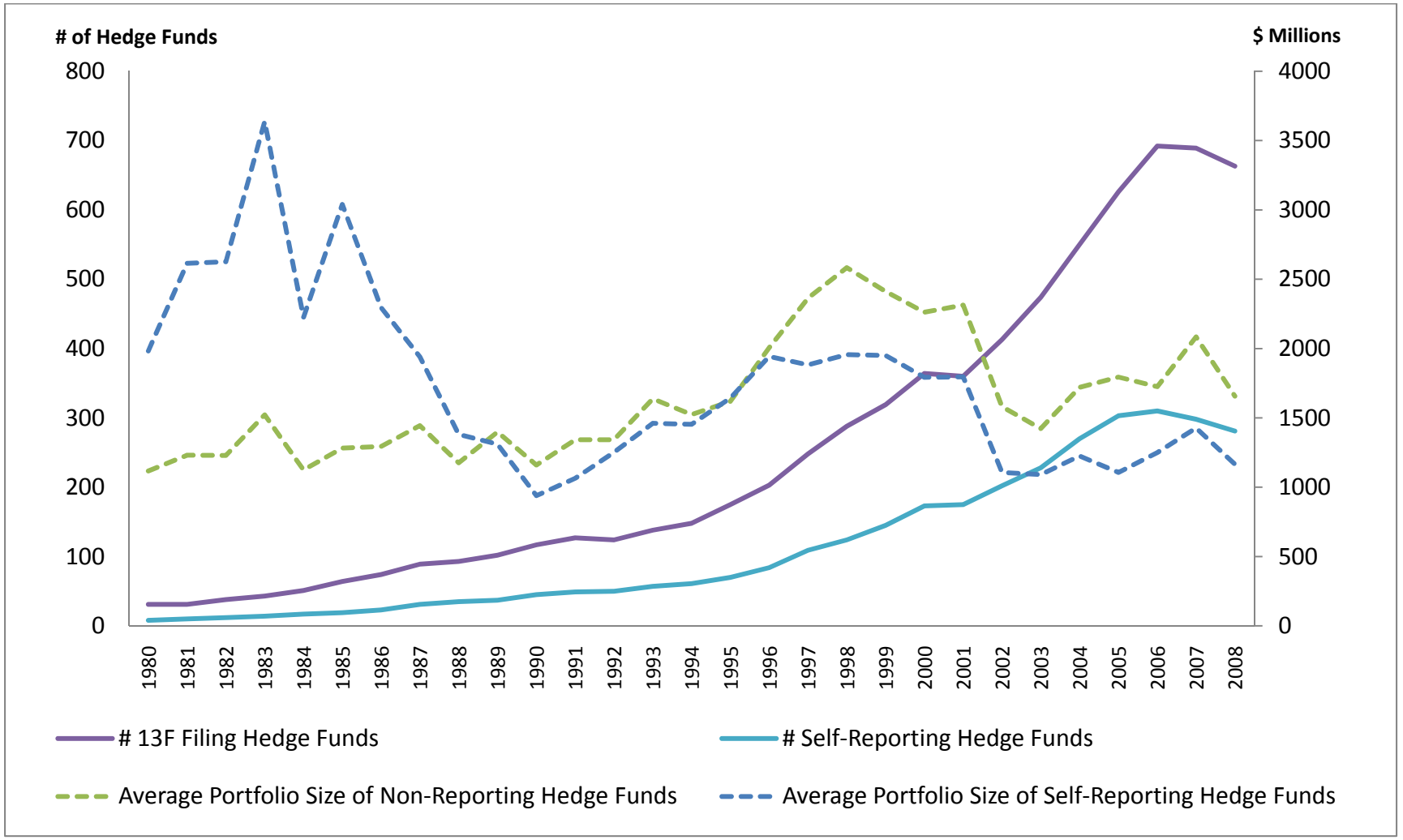
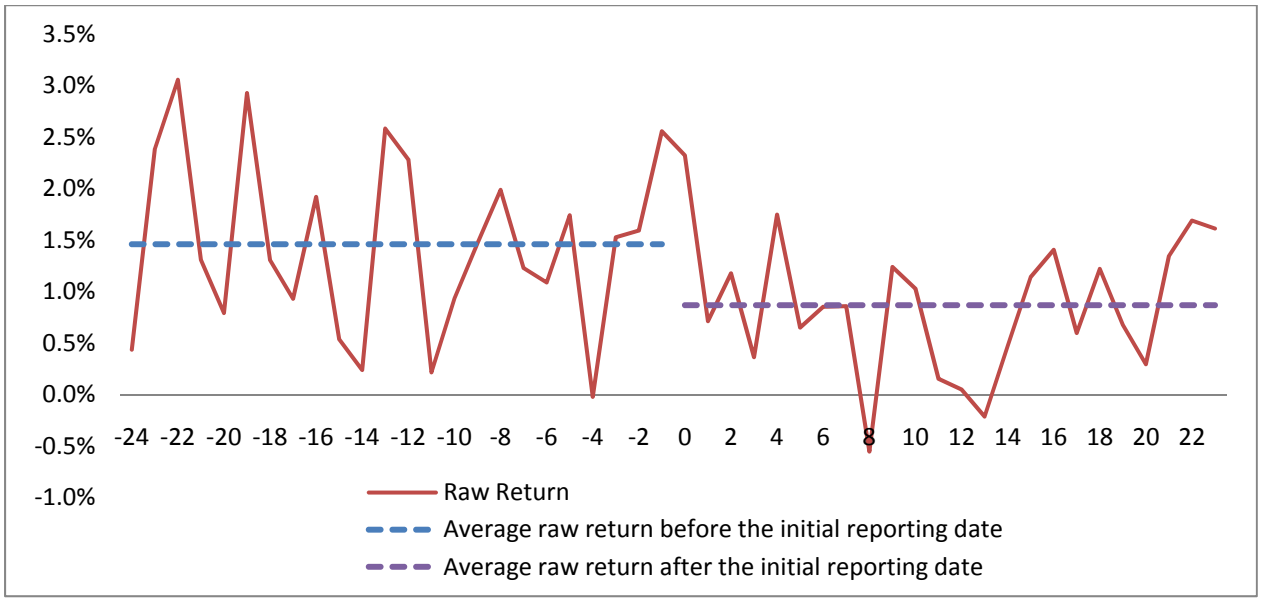


Figure 3

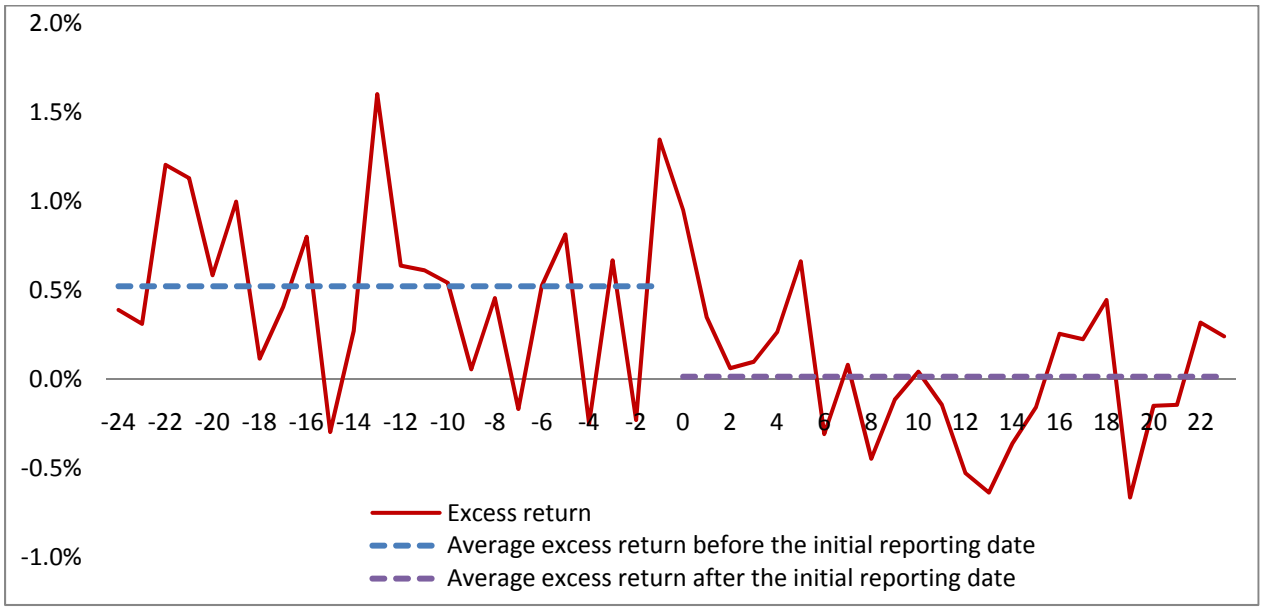
Return Performance around the Initial Reporting Date and the Reporting Termination Date

Panel A shows the time series of monthly raw return for the self-reporting hedge funds from 24 months before the initial reporting date to 24 months afterwards. 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 shows the time series of monthly excess return for the self-reporting hedge funds from 24 months before the initial reporting date to 24 months afterwards. The excess return is the difference between the imputed portfolio return and the CRSP value-weighted equity market return. Panel C repeats the analyses in Panel A for the reporting termination date. Panel D repeats the analyses in Panel B for the reporting termination date.

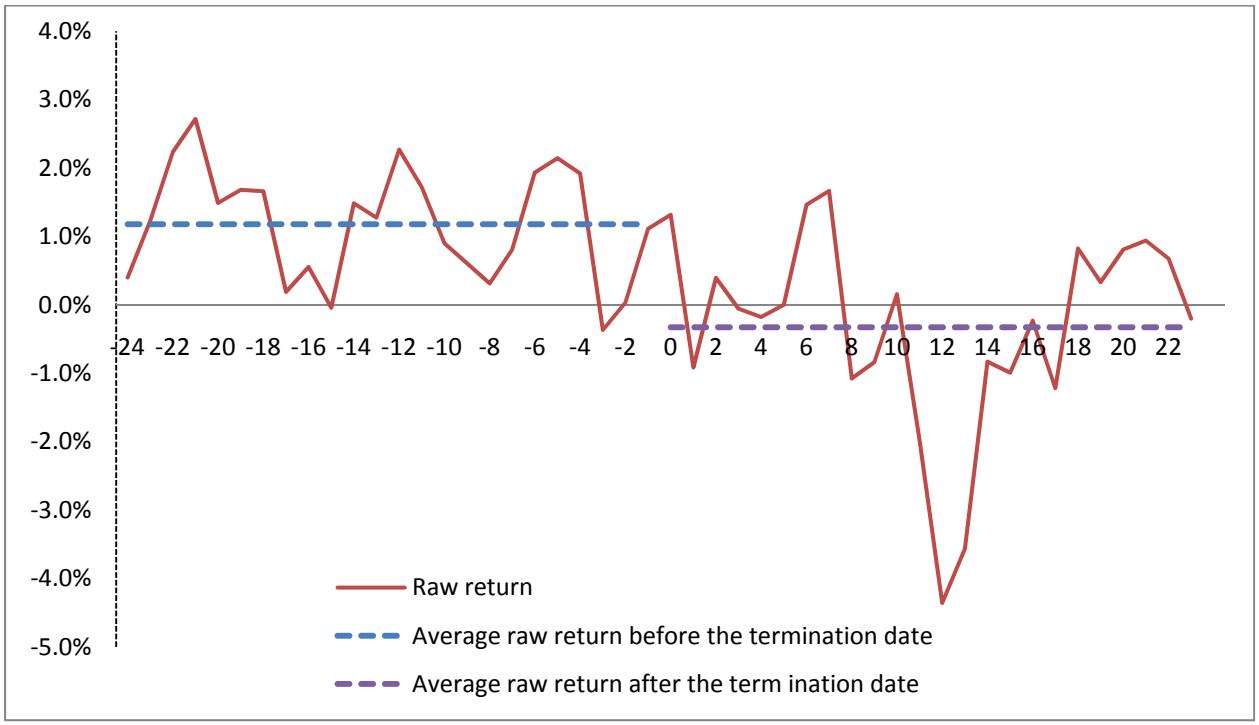
Panel A: Raw Returns around the Initial Reporting Date



Panel B: Excess Returns around the Initial Reporting Date



Panel C: Raw Returns around the Reporting Termination Date



Panel D: Excess Returns around the Reporting Termination Date

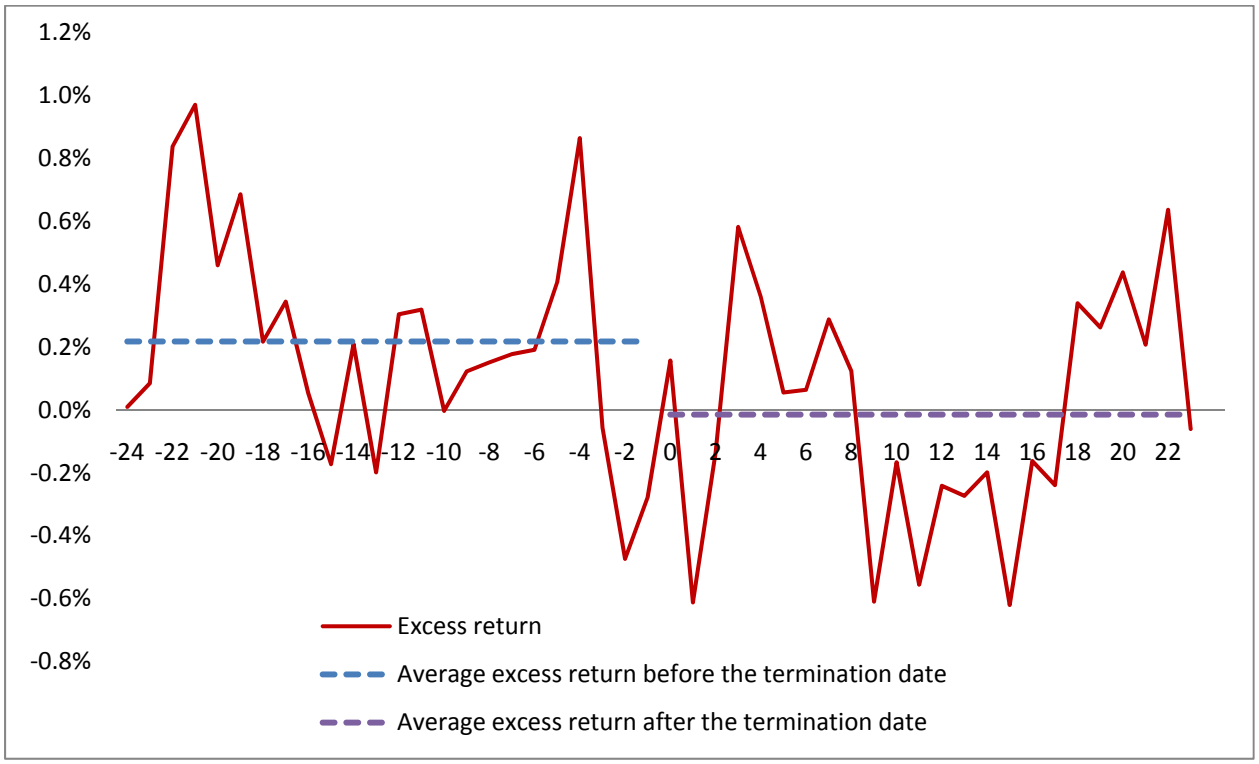


Table 1
Comparison of Hedge Funds with Other Categories of 13F-Filling Institutional Investors

The "Hedge fund" category is manually classified (see section I.A.). The "Bank/insurance" category is a combination of type 1 and type 2 institutions by the classification of Thomson Reuters Ownership Database for 13F filings. The "Mutual fund" category consists of type 3 institutions by Thomson Reuters. The "Investment advisor" category consists of type 4 institutions by Thomson Reuters. The "Other" category includes type 5 institutions by Thomson Reuters (with corrections for coding after 1998). All non-hedge-fund categories exclude classified hedge funds. 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. The imputed portfolio return is same as defined in Figure 3. The annual 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 quarter. The inception year is the year of the institution's first appearance in Thomson Reuter (censored at 1980). The return in excess of the market is the same as defined in Figure 3. One-Factor Alpha and Four-Factor Alpha are the intercepts from CAPM one-factor and Carhart (1997) four-factor models using all available data. Market Factor, SMB Factor, HML Factor, and Momentum Factor are estimated factor loadings from Carhart (1997) four-factor model. The t-statistics correspond to the difference between the "Hedge fund" category and other categories. The sample period is 1980-2008. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level respectively.

	(1) Hedge fund	(2) Bank/insurance	(3) Mutual fund	(4) Investment advisor	(5) Other
Portfolio size ($, million)					
Mean	1041	2609***	6305***	1809***	2431***
t-statistic of the difference	−	−6.58	−5.38	−5.37	−6.76
Median	368	600***	1036***	371	304***
t-statistic of the difference	−	−5.71	−3.71	−0.13	2.97
Portfolio Herfindahl index					
Mean	0.0953	0.0664***	0.0549***	0.0693***	0.1059*
t−statistic of the difference	−	5.23	3.48	4.70	−1.84
Median	0.0465	0.0285***	0.0175***	0.0277***	0.0341***
t-statistic of the difference	−	9.51	14.19	10.77	6.72
Monthly return volatility					
Mean	0.0553	0.0420***	0.0499***	0.0535*	0.0533*
t-statistic of the difference	−	14.28	3.23	1.94	1.96
Median	0.0493	0.0406***	0.0448***	0.0466***	0.0453***
t-statistic of the difference	−	10.02	4.35	2.99	4.02
Annual portfolio turnover rate					
Mean	0.9162	0.2683***	0.4901***	0.5217***	0.6026***
t-statistic of the difference	−	29.72	13.42	18.40	13.23
Median	0.8149	0.2313***	0.4258***	0.3948***	0.4044***
t-statistic of the difference	−	27.02	11.30	20.19	16.55

	(1) Hedge fund	(2) Bank/insurance	(3) Mutual fund	(4) Investment advisor	(5) Other
Inception year					
Mean	1999	1986***	1987***	1994***	2000**
t-statistic of the difference	−	36.56	25.04	14.07	−2.20
Median	2002	1983***	1985***	1995***	2003*
t-statistic of the difference	−	25.98	23.25	16.65	−1.69
Return in excess of the market					
Mean	0.0008	0.0005	0.0007	0.0000**	0.0001*
t-statistic of the difference	−	0.83	0.04	2.26	1.85
Median	0.0011	0.0007***	0.0007**	0.0008**	0.0008*
t-statistic of the difference	−	2.73	2.17	2.39	1.78
One-Factor Alpha					
Mean	−0.0006	0.0002*	0.0016***	0.0006**	−0.0003
t-statistic of the difference	−	−1.66	−3.44	−2.49	−0.62
Median	−0.0002	0.0002	0.0002	0.0001	−0.0002
t-statistic of the difference		−1.42	−1.18	−0.96	0.10
Four-Factor Alpha					
Mean	−0.0020	0.0008***	−0.0003**	−0.0003***	0.0003***
t-statistic of the difference	−	−6.49	−2.27	−3.87	−5.17
Median	−0.0011	0.0004***	−0.0003**	−0.0002***	0.0000***
t-statistic of the difference	−	−6.01	−2.36	−3.60	−4.40
Market Factor					
Mean	1.0917	0.9573***	1.0439***	1.0418***	1.0398***
t-statistic of the difference	−	10.59	2.71	3.96	4.03
Median	1.0553	0.9628***	1.0309**	1.0209***	1.0014***
t-statistic of the difference	−	10.72	2.16	3.90	5.74
SMB Factor					
Mean	0.3344	−0.0780***	0.1600***	0.1448***	0.1267***
t-statistic of the difference	−	22.60	5.43	10.28	11.82
Median	0.2861	−0.1038***	0.0724***	0.0560***	0.0278***
t-statistic of the difference	−	19.18	7.95	11.06	13.14
HML Factor					
Mean	0.0781	−0.0356***	−0.0953***	−0.0477***	0.0344*
t-statistic of the difference	−	5.42	4.12	6.18	1.88
Median	0.0706	−0.0311***	−0.0599***	−0.0275***	0.0251***
t-statistic of the difference	−	7.18	4.29	7.71	2.98
Momentum Factor					
Mean	−0.0126	−0.0156	−0.0044	−0.0048	−0.0087
t-statistic of the difference	−	0.26	−0.40	−0.64	−0.29
Median	−0.0047	−0.0147	0.0050	−0.0084	−0.0121
t-statistic of the difference	−	1.52	−0.93	0.68	1.16
Number of institutions					
	1199	804	204	2007	1801

Table 2
Comparison of Self-Reporting and Non-Reporting Fund Companies

Panel A shows the characteristics of the self-reporting and the non-reporting fund companies. The sample of self-reporting fund companies includes all 13F-filing hedge fund companies that report to the Union Hedge Fund Database (as defined in Figure 1) for some period of time. The sample of non-reporting fund companies includes all 13F-filing hedge fund companies that never report to the Union Hedge Fund Database. The portfolio size, the portfolio Herfindahl index, the monthly return volatility, the annual portfolio turnover rate, the inception year, and the factor loadings are the same as defined in Table 1. The t-statistics correspond to the difference between the self-reporting fund companies and the non-reporting fund companies. The sample period is 1980-2008. Panel B shows the performance measures of the self-reporting fund companies and the non-reporting fund companies. 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 returns without adjustment. Excess 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. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level respectively.

Panel A: Comparison of Fund Characteristics

	(1) Self-reporting fund companies	(2) Non-reporting fund companies	(3) Difference	(4) t-statistics of the difference
Portfolio size ($, million)				
Mean	927	1029	−102	−0.76
Median	415	341	74**	2.12
Std. Dev.	1517	2394	−877**	−2.34
Portfolio Herfindahl index				
Mean	0.0798	0.1056	−0.0258***	−3.24
Median	0.0458	0.0480	−0.0022	−0.50
Std. Dev.	0.0922	0.1547	−0.0625***	−4.27
Monthly return volatility				
Mean	0.0557	0.0556	0.0002	0.11
Median	0.0509	0.0474	0.0036**	2.40
Std. Dev.	0.0213	0.0295	−0.0081***	−3.25
Annual portfolio turnover rate				
Mean	1.0562	0.7937	0.2626***	7.19
Median	0.9909	0.6243	0.3666***	5.72
Std. Dev.	0.5526	0.5946	−0.0420**	−2.03
Inception year				
Mean	1998.7	1999.0	−0.3	−0.60
Median	2000	2002	−2.0**	−2.48
Std. Dev.	6.6	7.8	−1.2***	−3.08
Market Factor				
Mean	1.0940	1.0900	0.0040	0.18
Median	1.0787	1.0373	0.0414***	2.69
Std. Dev.	0.2652	0.3624	−0.0973***	−3.34

	(1) Self-reporting fund companies	(2) Non-reporting fund companies	(3) Difference	(4) t-statistics of the difference
SMB Factor				
Mean	0.3863	0.2980	0.0883**	2.56
Median	0.3416	0.2383	0.1033***	3.63
Std. Dev.	0.3912	0.5335	−0.1423***	−3.95
HML Factor				
Mean	0.1284	0.0428	0.0855***	2.60
Median	0.1140	0.0407	0.0733***	3.05
Std. Dev.	0.4333	0.5821	−0.1489***	−3.81
Momentum Factor				
Mean	−0.0083	−0.0156	0.0074	0.34
Median	−0.0019	−0.0059	0.0039	0.30
Std. Dev.	0.2740	0.3366	−0.0626*	−1.94
Number of institutions				
	366	554	−	−

Panel B: Comparison of Return Performance

	(1) Raw return	(2) Return in excess of the market	(3) One-Factor Alpha	(4) Four-Factor Alpha
Self-reporting fund companies				
5th Percentile	−0.0178	−0.0139	−0.0096	−0.0105
25th Percentile	−0.0019	−0.0011	−0.0010	−0.0021
Median	0.0047	0.0017	0.0016	0.0009
75th Percentile	0.0095	0.0048	0.0047	0.0038
95th Percentile	0.0164	0.0108	0.0117	0.0086
Mean	0.0025	0.0009	0.0014	0.0005
Std. Dev.	0.0112	0.0082	0.0067	0.0059
# funds	366	366	355	355
Non-reporting fund companies				
5th Percentile	−0.0322	−0.0183	−0.0137	−0.0109
25th Percentile	−0.0073	−0.0025	−0.0018	−0.0021
Median	0.0028	0.0011	0.0009	0.0006
75th Percentile	0.0095	0.0041	0.0039	0.0032
95th Percentile	0.0185	0.0120	0.0124	0.0105
Mean	−0.0006	0.0000	0.0005	0.0003
Std. Dev.	0.0178	0.0107	0.0083	0.0081
# funds	554	554	512	512
Differences (t-statistics)				
5th Percentile	0.0144***	0.0043	0.0041	0.0004
	[3.05]	[1.19]	[1.53]	[0.18]
25th Percentile	0.0054***	0.0015**	0.0008	0.0000
	[4.37]	[2.09]	[1.28]	[−0.03]
Median	0.0019**	0.0007**	0.0007	0.0003
	[2.09]	[2.04]	[1.75] *	[1.20]
75th Percentile	0.0001	0.0007	0.0007	0.0006
	[0.11]	[1.19]	[1.41]	[1.31]
95th Percentile	−0.0021	−0.0011	−0.0007	−0.0018
	[−0.74]	[−0.58]	[−0.42]	[−1.23]
Mean	0.0031***	0.0009	0.0009*	0.0003
	[3.21]	[1.44]	[1.85]	[0.54]

Table 3
Comparison of Self-Reporting and Non-Reporting Matching Fund Companies

This table repeats the analyses in Table 2 except using a sample of non-reporting fund companies that is matched with the sample of reporting fund companies through the following procedure: For each self-reporting fund, we crop out the period for which it appears in the Thomson Reuters 13F database. We then find non-reporting fund companies that have 13F data over the same period (or with the maximum overlap). If there are ties in matches, we choose the one that is closest in portfolio size as the self-reporting fund to be the "matching fund." The t-statistics for the differences are reported. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level respectively.

Panel A: Comparison of Fund Characteristics

	(1) Self-reporting fund companies	(2) Non-reporting "matching fund companies	(3) Difference	(4) t-statistics of the difference
Portfolio Size				
Mean	927	846	81	0.84
Median	415	394	21	0.49
Std. Dev.	1517	1133	384**	2.09
Portfolio Herfindahl Index				
Mean	0.0798	0.0709	0.0089	1.41
Median	0.0458	0.0377	0.0082***	2.88
Std. Dev.	0.0922	0.0926	−0.0004	−0.02
Volatility				
Mean	0.0557	0.0550	0.0007	0.43
Median	0.0509	0.0479	0.0030	1.58
Std. Dev.	0.0213	0.0214	0.0000	−0.02
Portfolio Turnover Rate				
Mean	1.0562	0.6484	0.4079***	11.59
Median	0.9909	0.4389	0.5521***	9.77
Std. Dev.	0.5526	0.5002	0.0524**	2.09
Inception year				
Mean	1998.7	1993.5	5.2***	9.91
Median	2000	1995	5.0***	4.78
Std. Dev.	6.6	7.3	−0.7**	−2.16
Market Factor				
Mean	1.0940	1.0674	0.0267	1.48
Median	1.0787	1.0330	0.0457***	3.12
Std. Dev.	0.2652	0.2250	0.0402	1.54
SMB Factor				
Mean	0.3863	0.2949	0.0913***	3.27
Median	0.3416	0.2153	0.1263***	4.46
Std. Dev.	0.3912	0.3875	0.0037	0.13

	(1) Self-reporting fund companies	(2) Non-reporting "matching fund companies	(3) Difference	(4) t-statistics of the difference
HML Factor				
Mean	0.1284	−0.0221	0.1504***	4.65
Median	0.1140	0.0113	0.1028***	3.72
Std. Dev.	0.4333	0.3926	0.0406	1.42
Momentum Factor				
Mean	−0.0083	−0.0085	0.0003	0.02
Median	−0.0019	−0.0109	0.0089	0.71
Std. Dev.	0.2740	0.1796	0.0944***	3.95
Number of institutions				
	366	366	–	–

Panel B: Comparison of Return Performance

	(1) Raw return	(2) Return in excess of the market	(3) One-factor alpha	(4) Four-factor alpha
Self-reporting fund companies				
5th Percentile	−0.0181	−0.0146	−0.0113	−0.0105
25th Percentile	−0.0020	−0.0009	−0.0007	−0.0020
Median	0.0047	0.0018	0.0017	0.0010
75th Percentile	0.0095	0.0048	0.0047	0.0039
95th Percentile	0.0156	0.0108	0.0113	0.0092
Mean	0.0024	0.0009	0.0015	0.0005
Std. Dev.	0.0113	0.0082	0.0067	0.0059
# funds	366	366	355	355
Non-reporting fund companies				
5th Percentile	−0.0134	−0.0091	−0.0088	−0.0069
25th Percentile	−0.0025	−0.0013	−0.0009	−0.0012
Median	0.0045	0.0013	0.0014	0.0008
75th Percentile	0.0091	0.0040	0.0048	0.0032
95th Percentile	0.0174	0.0118	0.0119	0.0098
Mean	0.0032	0.0013	0.0017	0.0011
Std. Dev.	0.0096	0.0065	0.0061	0.0051
# funds	366	366	357	357
Differences (t-statistics)				
5th Percentile	−0.0047*	−0.0055*	−0.0025	−0.0036**
	[−1.66]	[−1.81]	[−1.11]	[−2.21]
25th Percentile	0.0004	0.0004	0.0002	−0.0008**
	[0.40]	[0.68]	[0.35]	[−2.34]
Median	0.0002	0.0005	0.0003	0.0002
	[0.24]	[1.05]	[0.65]	[0.54]
75th Percentile	0.0005	0.0008	−0.0001	0.0007
	[0.66]	[1.38]	[−0.24]	[1.20]
95th Percentile	−0.0018	−0.0009	−0.0007	−0.0006
	[−1.28]	[−0.64]	[−0.46]	[−0.40]
Mean	−0.0008	−0.0004	−0.0002	−0.0005
	[−1.06]	[−0.79]	[−0.37]	[−1.32]

Table 4

Comparison of Return Performance before and after the Initial Reporting Date

This table compares the return measures (defined as in Table 2) for fund companies during the 24-month period before the initial reporting date, and during the 24-month period afterwards. The pooled 48-month period is used to estimate the beta loadings for the one-factor alpha and four-factor alpha. The one-factor alpha and four-factor alpha are coded as missing if there are fewer than 12 observations during the estimation window. The Diff-in-Diff is the difference (difference-in-difference) around the initial reporting date between raw returns of reporting and non-reporting hedge funds. Panel A includes the full sample of self-reporting fund companies where the initial reporting dates for some companies are imputed from the first performance dates. Panel B uses only the subsample where such information is accurately recorded. The t-statistics for the differences between the two samples are reported below difference estimates in parentheses. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level respectively.

Panel A: Full Sample

	(1) Raw return	(2) Return in excess of the market	(3) One-factor alpha	(4) Four-factor alpha	(5) Diff-in-Diff
Before initial reporting					
5th Percentile	−0.0346	−0.0331	−0.0237	−0.0228	−0.0282
25th Percentile	−0.0013	−0.0041	−0.0038	−0.0043	−0.0063
Median	0.0129	0.0009	0.0018	0.0010	−0.0010
75th Percentile	0.0211	0.0079	0.0073	0.0060	0.0044
95th Percentile	0.0448	0.0290	0.0254	0.0199	0.0275
Mean	0.0115	0.0036	0.0035	0.0021	0.0010
Std. Dev.	0.0299	0.0253	0.0224	0.0175	0.0236
# funds	103	103	102	102	103
After initial reporting					
5th Percentile	−0.0286	−0.0184	−0.0134	−0.0135	−0.0183
25th Percentile	0.0015	−0.0039	−0.0032	−0.0041	−0.0065
Median	0.0084	0.0016	0.0013	0.0005	0.0001
75th Percentile	0.0174	0.0083	0.0072	0.0055	0.0054
95th Percentile	0.0291	0.0173	0.0164	0.0129	0.0170
Mean	0.0063	0.0012	0.0014	0.0003	−0.0010
Std. Dev.	0.0170	0.0118	0.0093	0.0091	0.0114
# funds	103	103	102	102	103

	(1)	(2)	(3)	(4)	(5)
	Raw return	Return in excess of the market	One-factor alpha	Four-factor alpha	Diff-in-Diff
Differences (t-statistics)					
5th Percentile	0.0061	0.0147	0.0103	0.0093	0.0099
	[0.76]	[0.74]	[0.96]	[0.86]	[0.71]
25th Percentile	0.0028	0.0003	0.0007	0.0002	−0.0002
	[0.18]	[0.10]	[0.26]	[−0.06]	[−0.08]
Median	−0.0046	0.0007	−0.0005	−0.0006	0.0010
	[−1.46]	[0.71]	[−0.54]	[−0.37]	[0.57]
75th Percentile	−0.0037*	0.0004	−0.0002	−0.0005	0.0010
	[−1.82]	[0.09]	[−0.27]	[−0.01]	[0.20]
95th Percentile	−0.0157	−0.0118	−0.0090	−0.0070	−0.0105
	[−1.18]	[−0.91]	[−0.85]	[−0.91]	[−1.02]
Mean	−0.0052	−0.0024	−0.0021	−0.0018	−0.0021
	[−1.52]	[−0.88]	[−0.86]	[−0.90]	[−0.80]

Panel B: Subsample of Fund Companies with Accurate Initial Reporting Date Information

	(1) Raw return	(2) Return in excess of the market	(3) One-factor alpha	(4) Four-factor alpha	(5) Diff-in-Diff
Before initial reporting					
5th Percentile	−0.0147	−0.0094	−0.0116	−0.0099	−0.0116
25th Percentile	0.0075	−0.0006	−0.0013	−0.0030	−0.0040
Median	0.0161	0.0033	0.0018	0.0011	0.0012
75th Percentile	0.0238	0.0094	0.0077	0.0048	0.0057
95th Percentile	0.0454	0.0394	0.0255	0.0136	0.0317
Mean	0.0160	0.0059	0.0034	0.0007	0.0024
Std. Dev.	0.0176	0.0141	0.0114	0.0089	0.0118
# funds	77	77	76	76	76
After initial reporting					
5th Percentile	−0.0333	−0.0226	−0.0236	−0.0159	−0.0257
25th Percentile	0.0029	−0.0041	−0.0053	−0.0046	−0.0062
Median	0.0112	0.0001	−0.0014	−0.0008	−0.0014
75th Percentile	0.0174	0.0056	0.0047	0.0035	0.0036
95th Percentile	0.0271	0.0147	0.0119	0.0102	0.0106
Mean	0.0070	−0.0014	−0.0024	−0.0017	−0.0033
Std. Dev.	0.0185	0.0132	0.0135	0.0119	0.0130
# funds	76	76	76	76	76
Differences (t-statistics)					
5th Percentile	−0.0186	−0.0133	−0.0120	−0.0060	−0.0141
	[−1.31]	[−1.21]	[−1.06]	[−0.35]	[−1.27]
25th Percentile	−0.0046	−0.0035**	−0.0040**	−0.0016	−0.0022
	[−1.02]	[−2.36]	[−2.55]	[−1.21]	[−1.20]
Median	−0.0049***	−0.0032**	−0.0033**	−0.0019	−0.0026*
	[−2.88]	[−2.36]	[−2.51]	[−1.33]	[−1.97]
75th Percentile	−0.0064***	−0.0039	−0.0030	−0.0013	−0.0021
	[−3.34]	[−1.56]	[−1.27]	[−1.19]	[−1.36]
95th Percentile	−0.0184*	−0.0248**	−0.0135*	−0.0034	−0.0211**
	[−1.99]	[−2.51]	[−1.74]	[−0.74]	[−2.04]
Mean	−0.0090***	−0.0073***	−0.0058***	−0.0024	−0.0057***
	[−3.09]	[−3.32]	[−2.85]	[−1.42]	[−2.82]

Table 5
Hazard Analysis of the Reporting Initiation

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. 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), portfolio 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 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 previous quarter. Reported coefficients are hazard ratios. The z-statistics are calculated using the original coefficients (not hazard ratios) and are reported below coefficient estimates in parentheses. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level respectively.

	(1)	(3)	(5)	(7)
Performance Measure	Raw Return	Return in excess of the market	One-factor alpha	Four-factor alpha
Performance	228.13***	190.80***	83.09***	40.60**
	[2.90]	[2.76]	[3.12]	[2.01]
Aggregate Flow to Hedge Fund Industry	0.2509***	0.2517***	0.2619***	0.2659***
	[−4.84]	[−4.85]	[−4.67]	[−4.68]
Portfolio volatility (%)	0.8245***	0.8242***	0.8287***	0.8342***
	[−6.05]	[−6.05]	[−5.89]	[−5.70]
Manager age (log)	0.9243***	0.9238***	0.9236***	0.9216***
	[−3.17]	[−3.19]	[−3.19]	[−3.28]
Portfolio Herfindahl Index	0.1280*	0.1307*	0.1200*	0.1331*
	[−1.74]	[−1.73]	[−1.83]	[−1.78]
Portfolio size (log)	1.0000	1.0000	1.0000	1.0000
	[0.62]	[0.63]	[0.72]	[0.66]
Turnover	0.6650	0.6623	0.6723	0.6975
	[−1.17]	[−1.18]	[−1.14]	[−1.05]
Flow	0.8962	0.8959	0.9183	0.9230
	[−1.14]	[−1.13]	[−0.93]	[−0.90]
Market Return	2.16	475.4**	257.0**	233.0**
	[0.27]	[2.52]	[2.25]	[2.26]
Observations	23618	23618	23618	23619

Table 6
Comparison of Return Performance before and after Reporting Termination

This table presents the same analyses as in Table 4 except replacing the event with reporting termination. The t-statistics are reported below coefficient estimates in parentheses. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level respectively.

	(1) Raw return	(2) Return in excess of the market	(3) One-factor alpha	(4) Four-factor alpha	(5) Diff-in-Diff
Before reporting termination					
5th Percentile	−0.0068	−0.0143	−0.0143	−0.0102	−0.0157
25th Percentile	0.0081	−0.0008	−0.0021	−0.0022	−0.0033
Median	0.0131	0.0032	0.0016	0.0019	0.0003
75th Percentile	0.0180	0.0078	0.0069	0.0058	0.0051
95th Percentile	0.0288	0.0178	0.0173	0.0151	0.0137
Mean	0.0118	0.0028	0.0018	0.0016	0.0001
Std. Dev.	0.0128	0.0101	0.0101	0.0090	0.0095
# funds	187	187	187	187	187
After reporting termination					
5th Percentile	−0.0562	−0.0290	−0.0216	−0.0188	−0.0264
25th Percentile	−0.0277	−0.0045	−0.0039	−0.0033	−0.0031
Median	−0.0015	0.0014	0.0013	0.0002	0.0007
75th Percentile	0.0120	0.0060	0.0065	0.0045	0.0059
95th Percentile	0.0296	0.0179	0.0177	0.0148	0.0208
Mean	−0.0072	0.0000	0.0006	−0.0001	0.0003
Std. Dev.	0.0262	0.0147	0.0137	0.0118	0.0140
# funds	187	187	187	187	187
Differences (t-statistics)					
5th Percentile	−0.0494***	−0.0147*	−0.0074	−0.0087	−0.0107
	[−4.07]	[−1.92]	[−1.07]	[−1.20]	[−1.22]
25th Percentile	−0.0358***	−0.0037***	−0.0018	−0.0011	0.0002
	[−10.42]	[−3.15]	[−1.57]	[−1.13]	[0.10]
Median	−0.0146***	−0.0018	−0.0003	−0.0018***	0.0004
	[−4.54]	[−1.55]	[−0.81]	[−2.80]	[0.21]
75th Percentile	−0.0061***	−0.0019	−0.0004	−0.0013	0.0008
	[−3.91]	[−1.58]	[−0.41]	[−1.33]	[0.90]
95th Percentile	0.0007	0.0002	0.0004	−0.0003	0.0070
	[0.15]	[0.33]	[0.44]	[−0.08]	[1.56]
Mean	−0.0190***	−0.0028**	−0.0012	−0.0017	0.0002
	[−8.92]	[−2.13]	[−0.98]	[−1.57]	[0.19]

Table 7
Flow to Fund Companies before and after the Initial Reporting Date

This table reports the results of multivariate regressions that examine the flow to fund companies before and after the initial reporting date. The dependent variable is the net percentage flow to a fund company in a given quarter, where the 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 previous quarter. Panel A reports the estimates of equation (1) for the subsample of fund companies with accurate initial reporting date information using three benchmark-adjusted *Performance* measures: return in excess of the market, CAPM one-factor alpha, and Carhart (1997) four-factor alpha. *Q+j*, where *j=-4,…, 4,* is the dummy variable for *j* quarters relative to the quarter of initial reporting. Portfolio size (in log), portfolio 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. All covariates lag the dependent variable by one quarter. The F-test reported at the bottom of the table test the null hypothesis that sum of coefficients on Q to Q+4 and the sum of coefficients of Q-4 to Q-1 are equal. The F-test reported at the bottom of the table tests the null hypothesis that sum of coefficients on Q 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 examining the flows to fund companies before and after reporting termination. The t-statistics are reported below coefficient estimates in parentheses. Coefficients marked with ***, **, and * are significant at the 1%, 5%, and 10% level respectively.

Panel A: Effects of Reporting Initiation on Flows

	(1)	(2)	(3)
Performance Measure	Return in excess of the market	One-factor alpha	Four-factor alpha
Q-4	0.0854	0.0960	0.1062
	[1.20]	[1.34]	[1.43]
Q-3	−0.0083	0.0033	−0.0526*
	[−0.17]	[0.07]	[−1.71]
Q-2	0.0650	0.0733	0.0209
	[1.16]	[1.29]	[0.48]
Q-1	0.0280	0.0345	0.0363
	[0.51]	[0.63]	[0.61]
Q	0.0387	0.0470	0.0273
	[0.97]	[1.17]	[0.69]
Q+1	0.1282	0.1345	0.1798**
	[1.56]	[1.64]	[2.08]
Q+2	0.0601	0.0683	0.0760*
	[1.40]	[1.60]	[1.75]
Q+3	−0.0034	0.0026	0.0047
	[−0.06]	[0.04]	[0.08]
Q+4	0.0504	0.0515	0.0522
	[0.65]	[0.66]	[0.67]
Performance	2.4853***	0.8716***	0.8689***
	[16.32]	[7.79]	[6.34]

Performance Measure	(1) Return in excess of the market	(2) One-factor alpha	(3) Four-factor alpha
Portfolio size	−0.0281***	−0.0273***	−0.0256***
	[−31.00]	[−30.10]	[−27.94]
Manager age	−0.0077***	−0.0082***	−0.0029
	[−4.69]	[−4.99]	[−1.59]
Turnover	0.0098***	0.0114***	0.0119***
	[3.10]	[3.64]	[3.63]
Portfolio volatility	0.3101***	0.2984***	0.3904***
	[5.23]	[5.03]	[6.27]
Non-Reporting Funds Dummy	−0.0042	−0.0026	−0.0018
	[−0.56]	[−0.34]	[−0.23]
Constant	0.2647***	0.2613***	0.2322***
	[25.61]	[25.21]	[21.94]
N	141090	141089	131544
R-squared	0.016	0.014	0.012
F-test			
Point estimate	0.1002	0.1000	0.1435
F-statistics	1.04	1.04	1.85
p-value	0.3074	0.3075	0.17

Panel B: Effects of Reporting Termination on Flows

	(1)	(2)	(3)
Performance measure	Return in excess of the market	One-factor alpha	Four-factor alpha
Q-4	−0.0106	−0.0063	−0.0354
	[−0.30]	[−0.18]	[−1.20]
Q-3	0.0136	0.0163	0.0268
	[0.34]	[0.41]	[0.65]
Q-2	−0.0079	−0.0055	−0.0236
	[−0.34]	[−0.23]	[−1.07]
Q-1	0.0475	0.0526	0.0520
	[1.05]	[1.16]	[1.17]
Q	−0.0568	−0.0584	−0.0654*
	[−1.52]	[−1.57]	[−1.73]
Q+1	−0.0418	−0.0427	−0.0354
	[−1.00]	[−1.02]	[−0.84]
Q+2	−0.0508	−0.0522	−0.0470
	[−1.41]	[−1.44]	[−1.29]
Q+3	−0.0272	−0.0286	−0.0256
	[−0.69]	[−0.73]	[−0.65]
Q+4	−0.1030***	−0.1050***	−0.1003***
	[−2.78]	[−2.82]	[−2.71]
Performance	2.9684***	1.4168***	1.4327***
	[5.66]	[3.60]	[2.89]
Portfolio size	−0.0602***	−0.0587***	−0.0567***
	[−9.30]	[−9.13]	[−8.84]
Manager age	−0.0160*	−0.0188**	−0.0114
	[−1.80]	[−2.11]	[−1.20]
Turnover	0.0026	0.0031	0.0036
	[0.24]	[0.28]	[0.32]
Portfolio volatility	−0.0131	−0.0453	0.0454
	[−0.05]	[−0.19]	[0.18]
Constant	0.5052***	0.5092***	0.4770***
	[11.45]	[11.59]	[10.68]
N	6301	6301	5934
R-Squared	0.048	0.045	0.041

	(1)	(2)	(3)
Performance measure	Return in excess of the market	One-factor alpha	Four-factor alpha
F-test			
Point estimate	−0.3222	−0.3440	−0.2935
F-statistics	8.48	9.56	7.03
p-value	0.0036	0.002	0.008