


REVISITING 'STYLIZED FACTS' ABOUT HEDGE FUNDS

JUHA JOENVÄÄRÄ, ROBERT KOSOWSKI AND PEKKA TOLONEN

Centre for Hedge Fund Research, Risk Management Laboratory
Imperial College Business School
Working Paper 10
Date: 3/2012

Revisiting 'stylized facts' about hedge funds*

Juha Joenväärä[a], Robert Kosowski[b], and Pekka Tolonen[c]

[a] *University of Oulu*
[b] *Imperial College Business School, Imperial College, London*
[c] *University of Oulu and GSF*

The First Draft; May 2011
This Draft; February 2012

Abstract

This paper presents new stylized facts about hedge fund performance and data biases based on a novel database aggregation. By highlighting the effect of data base differences on previously documented results we aim to improve the ability of researchers in this literature to compare results across different studies. We document economically important positive risk-adjusted performance of the average fund while differences in its magnitude are due to differences in fund size, domicile and data biases, but not differences in fund risk exposures. Measures of misreporting and return smoothing by funds are similar across different data bases. Performance persistence is sensitive to share restrictions, rebalancing frequency, fund size and weighting scheme as well as more pronounced biases in certain databases. Hedge funds with greater managerial incentives, smaller funds and younger funds outperform while hedge funds with strict share restrictions are not associated with higher risk-adjusted returns. Since several stylized facts are sensitive to the choice of the database it is important to use a high quality consolidated database such as the one used in this paper.

JEL Classifications: G11, G12, G23
Keywords: hedge fund performance, persistence, sample selection bias, managerial skill

---

*We would like to thank participants at the CREST Paris 4th Annual Hedge Fund conference 2012 and GSF summer workshop in Helsinki, in particular, Arjen Siegmann, the discussant, and Petri Jylhä. We would like to thank Pertrac for providing us access to with the Pertrac Analytical Platform. We are grateful for the support of OP-Pohjola Group Research Foundation. Contact address: Juha Joenväärä, University of Oulu, juha.joenvaara@oulu.fi. Robert Kosowski, Imperial College Business School, Imperial College, London, UK, r.kosowski@imperial.ac.uk. Pekka Tolonen, University of Oulu and GSF, pekka.tolonen@oulu.fi. Online appendix is available at the web page: https://workspace.imperial.ac.uk/business-school/Public/research/f_agroup/JKT_Appendix.pdf. The usual disclaimer applies.

1. Introduction

This paper presents new stylized facts about hedge fund performance and data biases based on a novel database aggregation and a comprehensive analysis of differences between the main commercial hedge fund databases. By highlighting the effect of data base differences on previously documented results we aim to improve the ability of researchers in this literature to compare results across different studies. While several hedge fund studies build and use a large consolidated database containing multiple databases, there is no standard merging methodology described in the literature that could be used as a benchmark to help gauge the sensitivity of findings to the databases employed. Another contribution of this paper is that our results can be viewed as laying the foundations for an *industry standard* for matching hedge fund databases so that the consolidated data is designed to be as close as possible to the true unobserved population. We demonstrate the properties of our consolidated database by revisiting stylized facts about average hedge fund performance, performance persistence, data biases, management company domicile and fund-specific characteristics explaining cross-sectional differences in hedge fund performance. First, we find that the results inferred from the aggregate database qualitatively differ from those found in the previous literature. Importantly, the stylized facts obtained using one database are often in contrast to results inferred from the consolidated database. Our aim is not to produce 'back-tests' of earlier studies and our results should not be interpreted as questioning earlier findings. The reason is that differences in our findings compared to previous studies may also be due to revisions the same database over time, an issue recently documented by Patton, Ramadorai and Streatfield (2011). Second, our overall findings show the importance of using an aggregate database in hedge fund research and also when allocating capital to hedge funds in practice, since the results based on a single database are often not representative and may even be misleading compared to findings based on the aggregate database.

While the hedge fund literature is continuing to grow significantly, to the best of our knowledge there is no comprehensive study that would compare the stylized facts in the literature based on individual and consolidated databases. Such a comparison would be very useful for academic researchers and practitioners; for mutual funds, for example, Elton, Gruber and Blake (2001) find systematic differences in returns between the popular Morningstar and CRSP mutual fund databases. They show that these differences are important, since they may change the conclusion about individual mutual funds or a group of mutual funds. Our paper fills the equivalent research gap for hedge funds and aims to assist hedge fund researchers in evaluating their database choice and the understanding of differences in results between databases. We argue that the database selection is even more important in hedge fund studies than in the mutual fund literature for three reasons. First there are 5-10 commercially available hedge fund databases while there are only two main databases used in the majority of mutual fund studies (CRSP and Morningstar). Second, hedge fund databases are also non-overlapping - we find that almost 70% of funds in our consolidated database report only to one of the used major databases. Third, existing research documents a larger number of data biases in hedge fund databases than in mutual funds which highlights the importance of comparing the quality of individual databases.

To understand why the performance results differ across databases, we start by highlighting how the return, the asset under management (AuM) time-series, and the attrition rates differ between commercial and consolidated databases based for a data download that we carried out in Q3 of 2011. Overall, we find that the aggregate data set consists of 24,749 unique hedge funds that report at least 12 return observations. For these hedge funds, 8,512 are active, while 16,237 are not providing any data to vendors, thus, we classify them as defunct. The number of hedge funds in our database is close to that reported by the UBS' proprietary AIS database consisting of about 20,000 hedge funds and 45,000 share classes[1], while the PerTrac 2010 hedge fund database study finds that hedge fund industry contains about 23,600 funds. Therefore, we believe that our aggregate database containing the union of five major databases is close to the true unobservable population of hedge funds.

We document that the number of hedge funds ranges across data vendors from 6,772 for Morningstar to 9,719 for BarclayHedge. EurekaHedge is the largest data vendor in terms of 4,452 active hedge funds. Importantly, the proportion of alive and defunct funds varies significantly across databases. BarclayHedge, HFR and TASS (EurekaHedge and Morningstar) contain relatively more (fewer) defunct funds than alive funds. The attrition rates are also remarkably different across data vendors and time periods. We find that EurekaHedge and Morningstar have very limited information about defunct funds before 2004. In contrast, BarclayHedge, HFR and TASS do not seem to suffer from the same lack of data. Hence, the survivorship and backfilling biases are much higher in EurekaHedge and Morningstar than in the BarclayHedge, HFR and TASS databases. Therefore, we test the hypotheses that EurekaHedge and Morningstar should have higher average returns, but weaker performance persistence than the other databases due to

[1] See, Güner, Rachev, Edelman, and Fabozzi (2010). The AIS database includes funds that UBS allocated capital to, but that do not report to commercial databases.

these biases. The rationale is that the EurekaHedge's and Morningstar's bottom deciles may not contain a significant number of liquidated funds that deliver poor performance suggesting the spread portfolio between the top and bottom deciles may be indistinguishable from zero.

Overall, we find that about 28 percent of AuM observations are missing from our aggregate database. The amount of missing AuM observations varies significantly across data vendors, being lowest for BarclayHedge (11%) and HFR (19%), while significantly higher for EurekaHedge (37%), TASS (34%), and Morningstar (32%). Since the performance of the overall hedge-fund industry is evaluated using value-weighted portfolios, we test the hypothesis that these differences in the AuM time-series properties lead to different average performance results based on different databases. To evaluate hedge-fund industry's average performance, we examine equal-weighted (EW) and value-weighted (VW) portfolios formed using the aggregate and single databases. We find clear evidence that hedge funds deliver, on average, economically and statistically significant abnormal performance on an equal- and value-weighted basis, as well as across investment strategies, domiciles, size categories, and time-periods, a finding that is consistent with previous studies such as Kosowski, Naik and Teo (2007).

Specifically, for the aggregate EW portfolio, we find an annualized average excess return of 8.5 percent and an annualized alpha of 5.9 percent. The average excess returns (alpha) is lower for the aggregate VW portfolio, at 7.4 (4.8) percent per year. The conclusion that the average hedge fund (EW or VW) delivers risk-adjusted returns for investors differs significantly across databases. For example, the annualized equal-weighted average return ranges from 7.9 percent for TASS to 10.3 percent for EurekaHedge. In contrast, when returns are measured on a value-weighted basis, TASS shows the highest value-weighted average returns of 8.6 percent, being almost a quarter higher than the lowest respective return for BarclayHedge.

Since we find that Sharpe ratios, risk loadings and R^2's of the Fung and Hsieh (2004) seven-factor model are very similar across individual and the aggregate databases, this suggests that return differences cannot be explained by risk exposures. Consistent with our hypothesis we find that the equal-weighted return differences across databases are driven by the different levels of survivorship and backfilling biases in the commercial databases. The equal-weight returns are the highest (lowest) for EurekaHedge and Morningstar (BarclayHedge, HFR and TASS), which has the lowest (highest) amount of defunct funds. The missing AuM observations and a relatively high performance of large funds drive the differences in value-weighted portfolios. We find that high value-weighted returns for TASS are due to the outstanding performance of large funds and non-randomly missing AuM observations. Importantly, only in the TASS data, a significant number of funds have a missing AuM observation exactly at the same time when a fund exhibits a poor return. A low value-weighted performance in the BarclayHedge data is associated with the fact that it is able to gather significantly more AuM observations for poor performing funds compared to other databases. We conclude that the return differences between databases are smaller after 2004 when the survivorship and backfilling biases decrease in the EurekaHedge and Morningstar databases suggesting that the properties of databases seems to converge over time. In the paper, we classify hedge funds into twelve main strategies as shown in Online Appendix[2] and show that there are significant differences in the average performance across strategies. Our findings show that the fund size is strongly related to the levels of average performance. In the aggregate database, we find that for ten of the twelve style groups small funds outperform large funds. We also find the onshore based management firms and hedge funds to outperform offshore vehicles suggesting that domicile also explains differences in the average performance.

Next, we address the economically important question of whether hedge funds are able to add value after fees by delivering superior performance persistence consistently through time. We focus on an annual horizon, a realistic frequency given practical redemption restrictions. We tackle this issue by sorting funds on the *t*-statistic of alpha obtained from the Fung and Hsieh (2004) model. The main results obtained using the aggregate database show that there is economically and statistically significant performance persistence at annual horizon for equal-weight portfolios. However, for value-weight portfolios, we document significant performance persistence only at monthly horizons, which implies that performance persistence vanishes rather quickly as the holding period increases. This suggests that performance persistence may be driven by small funds.

We demonstrate that the conclusions about performance persistence are sensitive to the choice of the data vendor. To highlight the issue, we find that BarclayHedge, HFR and TASS show economically significant performance persistence for the equal-weighted portfolios at semi-annual horizons, whereas using EurekaHedge and Morningstar databases we do not find any evidence about performance persistence. Consistent with our hypothesis, we find that the average risk-adjusted returns of EurekaHedge's and Morningstar's bottom deciles are significantly higher

[2] Online appendix is available at the web page: https://workspace.imperial.ac.uk/businessschool/Public/research/f_agroup/JKT_Appendix.pdf.

than the respective returns in the BarclayHedge, HFR and TASS databases. In addition, the dropout rates show that bottom deciles in BarclayHedge, HFR, and TASS databases contain a significantly higher number of defunct funds compared to the respective deciles for EurekaHedge and Morningstar. Hence, one plausible reason driving the results is a larger survivorship and backfilling bias in EurekaHedge and Morningstar. An alternative explanation would be that BarclayHedge, HFR and TASS contain a set of 'high quality' funds that do not report to EurekaHedge and Morningstar. However, the returns of top decile portfolios are very similar across data vendors ruling out the possibility that high quality funds are only present in the BarclayHedge, HFR, and TASS databases. Finally, since none of the databases show significant persistence at annual horizon as the results for the aggregate database suggest, we interpret this as evidence supporting the use of a consolidated database in making the conclusion about hedge fund performance persistence.

We perform also a performance persistence test for the aggregate database in order to measure whether a real-time investor is able to exploit the short-term performance persistence of hedge funds. We implement feasible rebalancing strategies by taking into account fund-specific share restrictions within each rebalancing horizon. Specifically, we exclude funds having lockup, notice, and redemption periods longer than the rebalancing horizon in question. Based on the length of notice period, we use lagged information in persistence tests to mitigate look-ahead bias. For the aggregate database, the results suggest persistence only at quarterly horizon. These results indicate that one should take share restrictions into account when implementing persistence tests.

Finally, we examine the cross-sectional relationship between fund characteristics and hedge fund performance. The existing literature has documented that managerial incentives, share restrictions and capacity constraints are associated with cross-sectional differences in hedge fund performance. Using portfolio sorts and the Fama and MacBeth (1973) regressions, we demonstrate that smaller and younger funds, and funds having greater capital flows deliver better future returns than their peers. The result is very robust and homogenous across databases. Our conclusion is in line with the previous literature (e.g., Teo, (2010) and Aggarwal and Jorion (2010)). In contrast to the existing literature, we find, however, that fund characteristics related to managerial discretion or illiquidity do not consistently explain hedge fund cross-sectional returns. In fact, we find very little evidence that share restrictions in the form of lockup, notice and redemption periods are related to higher risk-adjusted returns when we control for the role of other characteristics in multivariate regression. Like Aragon (2007), we find the strongest evidence using the TASS database, whereas using other commercial and aggregate databases the conclusion is much weaker. For example, we find a significantly positive relation between notice period and performance only using TASS. Using our aggregate database, we find a significantly positive cross-sectional relationship between managerial incentives proxied by incentive fees and high-water mark provision and hedge fund return. Importantly, the signs and significance of high-mark provision change wildly across databases suggesting that the conclusion about the impact of managerial incentives on hedge fund performance varies based on the used data vendor. However, the overall results that are obtained using the aggregate database are consistent with Agarwal, Daniel and Naik (2009) showing that hedge funds with greater managerial incentives deliver higher performance than their peers. We stress that these results should not be interpreted as ‘back-tests’ of earlier studies since differences between our findings and previous studies may also be due to revisions of the same data base over time (Patton, Ramadorai and Streatfield (2011)).

Our paper is related to four streams of performance evaluation literature. First, Elton, Gruber and Blake (2001) document systematic return differences in CRSP and Morningstar mutual fund databases. Harris, Jenkinson and Kaplan (2012) show that there are economically important performance differences among private equity fund databases. Liang (2000) compares hedge fund survivorship rates between HFR and TASS databases. We add to this literature by showing that the hedge fund performance results are sensitive to the choice of database or sample. Importantly, the stylized facts do not only differ between relatively young databases such as EurekaHedge and Morningstar, but we also document significant heterogeneity among mature databases such as BarclayHedge, HFR and TASS.

Second, our paper contributes to the hedge fund literature providing an economically and statistically well-motivated merging approach that other researchers can easily follow. Indeed, it is not a trivial task to remove duplicate funds from the aggregate database due to the fact that there is no common identifier for the same hedge fund in the different vendors’ databases. Therefore, few of the existing papers provide transparent and detailed explanations of how their database is constructed. Notable exceptions are Ramadorai and Patton (2011) and Aggarwal and Jorion (2010). We add to this literature by using additional variables such domicile and legal structure to match share classes, and by providing a formal statistical algorithm in identifying unique hedge funds.

Third, the paper relates to the literature examining hedge fund data biases, misreporting, and strategic reporting behaviour. Due to the voluntary reporting, it is well known that hedge fund databases are associated with many data biases (e.g., Fung and Hsieh (2000, 2009), Liang (2000), and Getmansky, Lo, and Makarov (2004)), while the recent studies (e.g., Bollen and Pool (2008, 2009), Patton, Ramadorai, and Streatfield (2011), and Aragon and Nanda (2011)) show that hedge funds misreport, revisit, and strategically delay their returns when reporting in commercial databases. We add to this literature by showing that a database selection bias may arise when a study relies only on one of the hedge fund databases making a conclusion about hedge fund performance.

Fourth, we contribute to the literature documenting new stylized facts about hedge fund performance. The recent literature (e.g., Kosowski, Naik, and Teo (2007) and Jagannathan, Malakhov, and Novikov (2010)) has shown using the sophisticated econometric methods that hedge fund performance persists at annual horizons, while earlier studies (e.g., Brown, Goetzmann, and Ibotson (1999), Agarwal and Naik (2000), and Liang (2000)) document that hedge fund performance persists only at quarterly horizons. Using the consolidated database, we confirm that hedge fund performance persists at annual horizon even when just using unsophisticated econometric methods. However, there is persistence only at monthly horizon when it is evaluated using value-weighted portfolios. When the fund-level share restrictions and the effect of the look-ahead bias are controlled in persistence tests, we document equal-weight performance persistence at quarterly horizon.

Finally, the hedge fund literature has documented cross-sectional performance differences among hedge funds by showing that funds with greater managerial incentives (e.g. Agarwal, Daniel, and Naik (2009), Aggarwal and Jorion (2010)), strict share restrictions (e.g., Aragon (2007)), and less binding capacity constraints (e.g., Teo (2010)), on average, outperform their peers on a risk-adjusted basis. We contribute to this literature by confirming that smaller and younger funds and funds having greater capital flows and managerial incentives deliver higher future returns than their peers, while our results suggest that strict share restrictions are not associated with the higher risk-adjusted returns after we control the role of other fund characteristics using multivariate regressions.

The paper is structured as follows. Section 2 describes the data and methodology. Section 3 summarizes the stylized facts about average fund performance based on different databases.

Section 4 reports stylized facts about performance persistence. Section 5 describes stylized facts about hedge fund performance and cross-sectional characteristics. Section 6 concludes.

2. Data and merging approach

In this section, we propose an *industry standard* in constructing an aggregate hedge fund database by merging multiple commercial databases. Our merging approach is based on the two transparent main steps that can be easily replicated on a regular basis. Due to the fact that these steps are nearly automatic, we update our database each month making data real time applicable. Easily replicable steps also imply that other researchers can follow them in constructing their own data set or even use the same aggregate data.

2.1. Merging procedure of hedge fund databases

We combine five major hedge fund databases (BarclayHedge, EurekaHedge, Hedge Fund Research (HFR), Morningstar and TASS) to the aggregate data set. It is not a trivial task to merge several commercial hedge fund databases and to separate unique hedge funds from the share classes. The main reason is that all the commercial data vendors only provide an identifier to unique share classes, but there are no identifiers for unique hedge funds. The problem is serious even for the studies that are conducted using only one of the commercial databases, since the individual databases contain significant numbers of multiple share classes that cannot be captured only by excluding different currency classes. Thus, it is important to remove duplicate share classes even if a study is conducted using only one of the databases. Figure 1 highlights the issue by showing the histograms of pairwise correlation coefficients of share classes estimated within management companies for each database. All share class pairs are required to have at least 12 monthly return observations. Figure 1 shows a significant spike at the 0.99 correlation level for each database. The results suggest that multiple share class structures exist within management companies. The histogram of the aggregate database suggests that highly correlated share classes exist within management companies, which can be due to (i) multiple share class structures within management companies or (ii) duplicate share classes between databases.

We address these issues by combining five major hedge fund databases to form the aggregate data set using a novel merging approach that is based on two main steps. First, we develop a matching algorithm of hedge fund share classes that aims to identify exactly the same share classes with each other across databases. Second, we propose a formal statistical algorithm that combines sufficiently similar share classes within management companies as a group allowing us to obtain the longest possible time-series for each unique hedge fund. In contrast to other studies that often very loosely describe that they 'carefully' remove duplicate hedge funds without describing in detail their merging process, we report and rationalize clearly the variables that are used in share class matching and open up the details of the statistical algorithm that identifies the unique hedge funds.[3] Specifically, along with the share class and company names, we opt to use information about the legal structure, domicile and currency base. We do not use compensation structure and share restriction information due to different reporting standards across databases, and the changing nature of these variables.[4] Furthermore, applying the statistical algorithm, we can select the individual share class along with the longest track record. This also allows us to define a unique fund by taking either the equal-weighted or the value-weighted cross-sectional average return of share classes belonging to the same group within the management company. Hence, compared to other studies, our innovation is aimed at combining share class information so that data biases are mitigated. Finally, we emphasize that there is no reason to believe that previous papers do not use rigorous merging procedures. Due to the fact that the merging procedures are not discussed in these papers, there is a high demand for the proposed standardized merging approach that could be exploited by researchers, investment professionals and even regulators. Appendix discusses the details of the merging algorithm.

The baseline version of our database contains monthly net-of-fees returns, assets under management (AUM), and other characteristics, such as manager compensation (management fee, performance-based fee, and high-watermark provision), share restrictions (lockup period, advance notification period, and redemption frequency), domicile, currency code, style category, and inception. The database can be easily extended to contain other interesting fund characteristics that are provided only by some of the databases. For example, EurekaHedge provides detailed information about hedge fund soft and hard lockups as well as the locations of hedge funds' head and research offices. This feature makes the merged aggregate database more attractive for

[3] Notable exceptions are Ramadorai and Patton (2011) and Aggarwal and Jorion (2010) who provide the details about the used merging or cleaning procedures. However, we add to this literature by using additional variables such as domicile and legal structure to match share classes and by providing a formal algorithm in cleaning databases from the multiple share classes.

[4] For example, Agarwal and Ray (2011) document that hedge fund fees are changing. These changes are not updated to databases; therefore, we believe that fees do not provide reliable information that could be used in merging. Anecdotal evidence suggests that hedge funds' share restrictions are also changed due to investors' higher demand for liquidity.

research and investment purposes, since using it one can exploit all the novel properties of individual databases and while benefiting from the advantages of a large aggregate database.

The two main strengths of our database choice are quality and scope. TASS and HFR databases form a natural cornerstone of our database, since they are widely used among researchers and their properties are well-documented in literature. BarclayHedge, EurekaHedge and Morningstar are not as frequently used; however, we argue that a comprehensive coverage depends crucially on these less used data vendors. BarcleyHedge provides the most comprehensive set of asset under management observations, whereas EurekaHedge covers the largest number of active funds among these databases. The large cross-section of funds that we obtain as a result has the advantage of making it possible to include important variables from other sources such as hedge fund equity holdings and proxies of an operational risk.[5] Specifically, BarclayHedge provides information on hedge fund equity holdings obtained from SEC's 13-F fillings, while Morningstar collected variables related to hedge funds' operational risk from SEC's Form ADV disclosures. These additional unique features of BarclayHedge, EurekaHedge and Morningstar suggest that they provide additional useful information for our aggregate database that the most used databases - HFR and TASS - cannot provide at the time. Finally, we opt not to use the CISDM database, since its data seem not to be of the same commercial quality as that of others. The main reason is that CISDM does not provide frequent updates suggesting that one cannot update the data on a monthly or quarterly basis, which is a very attractive feature of our consolidated database.

2.2. Properties of commercial databases

This section compares the properties of commercial and aggregated databases by focusing on return and AuM time-series as well as misreporting and attrition rates. We find that there are significant differences between commercial databases that may be important in explaining performance differences across commercial databases. Specifically, we find that EurekaHedge and Morningstar have significantly lower proportion of defunct funds compared to BarclayHedge, HFR and TASS. The result is driven by the periods before 2004 when EurekaHedge and Morningstar did not gather information about defunct funds. It is important to note that

[5] Aragon and Martin (2010) show that hedge fund equity holdings can be used to predict future stock returns. Goetzmann, Brown, Liang, and Schwart (2008, 2009) measure operational risk using data obtained from ADV disclosures containing information on the business, ownership, and disciplinary events.

BarclayHedge, HFR and TASS do not suffer from such a problem. In addition, we find that the amount of AuM observations varies significantly across databases. BarclayHedge is superior in the terms of AuM coverage, since it has the lowest number of missing AuM observations and the longest AuM time-series suggesting different behavior when aggregate returns are calculated on a value-weighted basis.

Our sample covers a time period from January 1994 to June 2011. We focus on the post-1994 period to mitigate a potential survivorship bias, as most databases started to collect defunct funds only after 1994. Our sample restricts each fund to have at least 12 monthly non-missing return observations. We find that returns and assets under management of share classes are reported in up to 25 different currencies and approximately 77.2% of share classes are reported in US dollars. We use end-of-month spot exchange rates to convert all non-USD observations of currency share classes to US dollars. After all restrictions, the total number of share classes in sample is 48,121. Obviously, the number of unique share classes and hedge funds is considerably smaller because of (i) multiple share class structures within management companies and (ii) duplicate share classes between databases.

After merging five major databases using our merging procedure, we find 24,749 unique hedge funds. Our estimate of unique hedge funds is consistent with other estimates about the number of funds in hedge fund industry. For example, Güner, Rachev, Edelman, and Fabozzi (2010) report that the UBS' proprietary AIS database has information on over 20,000 hedge funds and 45,000 share classes, while PerTrac 2010 hedge fund database study finds that hedge fund industry contains approximately 23,603 hedge funds. The descriptive statistics presenting our aggregate and commercial databases suggest that an empirical hedge fund study that is conducted using only one of the databases may infer biased results. The sample of a single commercial database may not be representative, since even the largest database solely represents less than half of the hedge funds that exist in our aggregate database. Indeed, the aggregate database has 24,749 hedge funds, while BarclayHedge is the largest database having 9,719 funds (39.27% relative to the aggregate data) and Morningstar being the smallest 6772 funds (27.36% relative to the aggregate data).

Figure 2 shows a Venn diagram describing the percentage amounts of unique and duplicate share classes in our union of five major databases. Figure shows that the total amount of unique funds in the union is 67.7%. BarclayHedge has the largest percentage share (15.6%), and even the

EurekaHedge database, which has the lowest share (10.4%), has a considerable number of unique share classes. Hence, all of the databases have a large number of share classes that are not in other databases suggesting the importance of the use of multiple databases. The numbers shown above indicate that the properties of the databases differ significantly from each other suggesting that the sample based on a single database may not be representative.

Share classes in commercial databases are classified into two groups, namely active and defunct. Active funds report to commercial data vendors, while defunct funds no longer report to the most recent snapshots. Hedge funds stop reporting because they liquidate, merge with other funds, cannot be reached by the vendor, stop accepting new investments, are dormant or stop reporting for unknown reasons. It is important to note that funds belonging to the defunct category are not necessarily liquidated. However, recent studies document that defunct funds typically have lower performance measures relative to alive funds.

Among the first, Ackermann, McEnally, and Ravenscraft, (1999) and Liang (1999), and Fung and Hsieh (2000) note that hedge fund performance measures are affected by the survivorship bias when the database contains only alive funds. Furthermore, hedge funds often report return data prior to their listing date in the database, thereby creating a backfill bias. Clearly, due to the incentives, the backfilled returns are usually higher than the nonbackfilled returns. To confirm that our findings are robust to incubation and backfill biases, we repeat our analysis by excluding the first 12 months of data. We also ensure that serial correlation does not drive our findings. We unsmooth hedge fund returns using the Getmansky, Lo, and Makarov (2004) specification. We then repeat performance tests to confirm our conclusions.[6]

Table 1 presents the summary statistics of hedge fund database returns and AUMs categorized by all, alive, and defunct funds. The sample period extends from January 1994 to June 2011 and contains funds with at least 12 non-missing monthly returns. Single commercial databases are cleaned and merged using the proposed statistical algorithm. The statistics of returns are annualized and reported in percentages. According to Panel A, our aggregate database has 37,592 unique share classes after databases are merged at the share class level using the name matching. The aggregate database has 8,353 unique management firms and 24,749 unique hedge funds (of which 34.39% are alive) after databases are merged using the statistical algorithm. Annualized performance measures are higher for alive funds if compared to defunct funds

[6] The results are available upon request.

(12.15% / year vs 8.35% / year). The aggregate data has 1.36% (27.56%) of missing return (AuM) observations.[7] The results suggest that alive hedge funds are larger than defunct hedge funds. The average AuM for alive (defunct) funds is 160.2 (83.82) million US dollars.

Overall, for each of the databases, a typical alive fund has a larger annualized return than a typical defunct fund. Morningstar database has the largest amount of missing returns, 2.58%. Alive funds are also larger based on a statistics of the average AUM. EurekaHedge and TASS have the largest amounts of missing AUM observations, namely, 37.12% and 34.4%. Based on the coverage of AuM reporting, BarclayHedge is superior to its peers, since it has only 11.49% of missing AuM observations. In each database, the coverage of AUM observations is important if performance is measured on a value-weighted basis.

Table 1 (Panel A) presents the summary statistics of the length of the time series of both the return and assets under management. The results are reported separately for all hedge funds and separately for alive, and defunct funds. The aggregate data shows that the average length of fund-return time series is 62 months. On average, the longest return time series are in the Morningstar database (70 months). Defunct funds seem to have shorter return time series compared to alive funds. BarclayHedge has the lowest amount of missing AUM observations and the longest fund-level AUM time series on average (57.22 months).

Panel A shows that the largest commercial database in term of the number of hedge funds is BarclayHedge having 9,719 funds from which 36.4% are alive. The proportion of alive funds in TASS and HFR are close to our aggregate data (37.83% and 39.61%, respectively). The largest proportions of alive funds are in EurekaHedge (62.14%) and Morningstar (52.54%). These results suggest that a survivorship bias may have the largest effects in the EurekaHedge and Morningstar databases.

Panel B in Table 1 describes the statistical properties of hedge fund returns categorized by alive and defunct funds. In the aggregate database, over 51.36% of funds have non-normal returns based on the Jarque-Bera test of normality (5% level of significance) and 21.44% of funds exhibit significant serial correlation in returns for the first six lags based on the Ljung-Box test (5% level

[7] We calculate missing AuM observations conditional on fund-level return time series that have at least 12 monthly returns for each fund.

of significance. Results are similar across databases. It is well-known from the prior literature (e.g., Bollen and Pool (2008, 2009)) that hedge funds misreport or 'manage' their returns. Panel B summarizes hedge fund misreporting using a measure proposed by Jylhä (2011). The measure formalizes the fact that the number of small gains exceeds the number of small losses documented by Bollen and Pool (2009). Table 1 presents the measure of discontinuity (DC(%)) and z statistic (z(DC)) for testing the existence of discontinuity across databases. We follow Jylhä (2011) and use only USD-based funds. Our findings suggest that misreporting is quite similar across databases. Alive hedge funds seem to misreport less than defunct ones. However, the difference is small. To examine the number of funds showing higher serial correlation when returns are low, we document the estimates of conditional smoothing using the methodology proposed by Bollen and Pool (2006). Consistent with Bollen and Pool (2006), we find that 5.77% of funds exhibit conditional smoothing in our aggregate database. Alive funds in the Morningstar database show the highest amount of funds showing conditional smoothing (7.07%). Agarwal, Daniel and Naik (2011) show that hedge funds with high performance-based incentives, low lockup and restriction periods, high volatility, and low liquidity exhibit significant December spikes in returns. We test the existence of a December spike and apply the Fung and Hsieh (2004) residual returns approach to measure the difference between the average December values and the average of January-November values. After correcting for clustering at the fund-level, consistent with Agarwal, Daniel and Naik (2011), we find significant differences between the average December and January-November values across all databases. To summary, our measures of statistical properties of hedge fund returns are consistent with recent literature.

Table 2 presents the attrition rates across databases and for the aggregate data. Attrition rates are calculated as the ratio of the number of dissolved funds to the number that existed at the start of the year. Attrition rates are remarkably different across commercial databases. Attrition rates for EurekaHedge and Morningstar show an interesting pattern. Specifically, Table 2 shows that these two databases have started to gather information on defunct funds after 2004, since their attrition rates are close to zero before 2004. For EurekaHedge, the results is in a line with Teo (2009), who examines using the EurekaHedge database whether hedge funds' access to local information is associated with superior performance. In contrast, during these early years attrition rates are considerably higher for BarclayHedge, HFR and TASS as well as for the aggregate database. The results of attrition rates suggest that EurekaHedge and Morningstar databases are subject to higher survivorship bias if compared to TASS, HFR, and Barclay. According to recent literature, survivorship bias imparts an upward bias to performance measures (Liang (2000)).

3. Average hedge fund performance

In this section, we investigate hedge funds' average performance on an equal- and value weight basis as well as across time-periods, investment strategies and fund domiciles. An important question in performance evaluation is whether active managers add value on average. A common approach to evaluate hedge fund abnormal performance is to estimate the alpha, which is the return that cannot be explained by exposure to common risk factors. For investors, a positive abnormal performance is crucial since it tells how much an average hedge fund adds positive value, if any, after management and incentive fees. The major interest is in the results of our novel aggregate database that serves a proxy of the unobservable whole hedge-fund industry. Also, we identify possible differences in performance estimates between databases that may suggest existence of sampling biases in hedge funds databases occurring from the fact that all hedge fund databases are not drawn from the same population of hedge funds. Consistent with previous studies (e.g., Kosowski, Naik, and Teo (2007)), we find that hedge funds add value even after fees. Specifically, our results show positive and significant aggregate abnormal returns for the aggregate and commercial databases. The databases that are associated with more pronounce biases deliver the highest average performance.

3.1. Equally-weighted and value-weighted portfolios

We construct equal-weight (EW) and value-weight (VW) portfolios of hedge funds. EW portfolio measures how an average hedge fund performs, while VW portfolio summarizes the performance of aggregate wealth invested in hedge-fund industry as a whole. We expect to find that the aggregate EW portfolio has a higher average return than the VW portfolio, since the evidence suggests (e.g., Teo (2010)) that smaller hedge funds, on average, outperform the larger ones. In addition, the performance of a VW portfolio can also give some hints how AuM reporting affects performance. It is important to note that we control for the potential impact of a duplicate bias, since all commercial databases are merged and cleaned from multiple share classes using a statistical algorithm proposed in the Appendix of the paper.

Figure 5 shows the cumulative excess returns of EW and VW portfolios. EW portfolios show that the cumulative excess return of our aggregate database is very close to the respective return for TASS, HFR, and Barclay. However, EurekaHedge and Morningstar clearly outperform other databases suggesting that there are differences across databases. VW portfolios show the highest cumulative return pattern for TASS. Figure 5 suggests that the aggregate returns of the databases are highly correlated with each other, but the average levels of returns vary between databases. This suggests that the same common factors may drive the hedge fund return across databases.

Table 3 shows summary statistics for excess returns of EW and VW portfolios from January 1994 to December 2010. Our aggregate data indicates that hedge funds add value on average. The annualized average EW (VW) excess return for the aggregate data is 8.45% per year (7.36% per year). EurekaHedge and Morningstar databases show the highest annualized average excess returns (10.28% and 9.75%). TASS has the lowest average excess return being 7.92% per year. To highlight differences between databases, we find that EurekaHedge has approximately 23 percent higher average excess return than TASS. Average excess returns show much more variation between databases than annualized standard deviations. Hence, risk levels do not differ significantly across databases. The results are consistent with our hypothesis suggesting that EurekaHedge and Morningstar have the highest returns due to the fact that these two databases have the lowest attrition rates and proportions of defunct funds.

In contrast to the results of EW portfolios, Table 3 shows the highest VW return for TASS (8.56%). Except for TASS, all databases show lower value-weight average returns if compared to equal-weight average returns. This is consistent with Teo (2010), who documents that smaller funds have higher average returns. We find that TASS has a large amount of missing AuM observations that equals to 34.4%. The results may potentially explain why TASS's VW returns are so high. Perhaps, relatively large and high-performing funds report AuM data to TASS, but poor performing funds skip reporting. We address this issue rigorously in section 3.2. It is interesting to note that BarclayHedge has the lowest VW return of 6.91% per year. BarclayHedge provides the best AuM coverage with the longest time-series and the smallest amount of missing observations. The relatively low average return of BarclayHedge may be explained by the fact that it is able to gather AuM information from poor performing hedge funds. The value-weighted return for BarclayHedge is close to the VW return that is obtained for our consolidated database, being 7.36% per year. Therefore, we believe that BarclayHedge's VW portfolio is the closest proxy among the commercial databases for the overall performance of hedge-fund industry.

Next, we investigate whether an average hedge fund is capable to deliver positive abnormal performance. Kosowski, Naik, and Teo (2007) conclude that the average abnormal return across hedge funds from 1994 to 2002 is 0.42% per month (5.04% per year), while Fung, Hsieh, Naik, and Ramadorai (2008) document that hedge fund performance and risk-taking differ across time-periods. As a benchmark model, we use the Fung and Hsieh (2004) seven-factor model that is the standard workhorse in hedge fund performance evaluation studies. [8] We regress the net-of-fee monthly excess returns (in excess of risk-free rate) of a hedge fund against traditional buy-and-hold and primitive trend-following factors

$$R_{i,t} = a + b_1(SP - R_f) + b_2(RL - SP) + b_3(TY - R_f) + b_4(BAA - TY) + b_5(PTFSBD - R_f) + b_6(PTFSFX - R_f) + b_7(PTFSCOM - R_f) + e, \quad (1)$$

where the risk factors are defined as the excess return of the S&P 500 index (*SP-RF*), the return of the Russell 2000 index minus the return of the S&P 500 index (*RL-SP*), the excess return of ten-year Treasuries (*TY-RF*), the return of Moody's BAA corporate bonds minus ten-year Treasuries (*BAA-TY*), the excess returns of look-back straddles on bonds (*PTFSBD-RF*), currencies (*PTFSFX-RF*), and commodities (*PTFSCOM-RF*). Fung and Hsieh (2004) show that their seven-factor model considerably explains time series variation in hedge fund returns.

Table 3 presents the estimates of time series regression (1) for EW (panel A) and VW (panel B) portfolios of hedge funds in our sample from January 1994 to December 2010. The intercepts of the Fung and Hsieh (2004) model summarize the average abnormal performance of hedge funds (EW portfolios) and the abnormal performance of aggregate wealth invested in hedge funds (VW portfolios). The alpha terms for net-of-fees returns show us whether hedge funds have sufficient private information to cover costs that they impose on investors. Based on the annualized intercept estimates of the Fung and Hsieh (2004) model, all hedge fund databases show positive abnormal performance. The result is consistent with previous literature like Kosowski, Naik, and Teo (2007).

The results in Table 3 show that the aggregate hedge fund EW and VW alphas equal to 5.87% and 4.88% per year, respectively. The results for abnormal performance suggest that hedge funds add value on average after adjusting for common systematic risk factors. Table 3 shows

[8] In an unreported robustness test, we augment the model with liquidity, currency and carry-trade risk factors. The main inference remains unchanged.

substantial variation in abnormal returns between databases. The highest EW abnormal return estimate is found for EurekaHedge and Morningstar databases (7.61% and 7.28%). TASS has the lowest EW alpha estimate equaling to 5.31% per annum. Again, the EW returns are the highest for the databases with the lowest attrition rates. As for the excess returns, BarclayHedge has the lowest VW alpha, while the respective alpha is relatively high for TASS. The differences between the average EW and VW aggregate alphas can be seen from Figure 6 that shows EW and VW cumulative abnormal returns of hedge funds for January 1994 – December 2010. Figure 6 shows clearly that TASS database underperforms the other databases in terms of EW returns but outperforms in terms of VW returns.

Since the factor loadings and *R*-squares of the Fung and Hsieh (2004) model in Table 3 do not show significant variation between aggregate and across databases, it implies that differences in risk taking do not explain significant variation in alphas. We also include in unreported robustness checks additional factors such as liquidity, carry, and currency risk factors. We find that the levels of alphas are insignificantly lower, but the *t*-statistics of alphas are slightly higher since the risk factors explain better the residual variance. Therefore, we argue that differences in the survivorship bias and AuM coverage across commercial databases are driving the alpha differences between databases.

3.2. Backfilling- and smoothing-adjustment, size categories, and missing AuM observations

To address rigorously our arguments about the determinants that are driving performance difference across databases, we examine the impact of backfilling bias, serial correlation, fund size and missing AuM observations. We find that small hedge funds, which size is below 10 million dollars, deliver extremely high performance. It is the most important reason why we document so high average hedge fund performance earlier using EW and VW portfolios. Differences in survivorship bias remain as an explanation why equally-weighted average performance varies significantly across commercial databases. We demonstrate two main reasons why value-weighted performance differs considerably across data vendors. TASS's superior VW performance is explained by the fact that TASS contains large hedge funds having relatively high performance. More importantly, TASS has a significant number of missing AuM observations exactly at the same time when respective hedge funds deliver poor returns. Other commercial databases do not share this property. Therefore, we argue that it is important determinant in explaining VW performance differences among data vendors.

Panel C in Table 3 presents the average performance results that address the impact of backfilling bias. We demonstrate that after excluding the first 12 return observations the average performance is significantly lower.[9] To compare performance with the baseline results presented in Panel A of Table 3, we conclude that the backfilled average performance is significantly higher across databases. Using an aggregate database, we find backfilled alpha (*t*-statistic of alpha) is over 20% (25%) higher than the respective non-backfilled measures. Among commercial databases, the impact of backfilling bias is more pronounce for mature databases like TASS, HFR, and BarclayHedge. The difference between backfilled and non-backfilled returns is not so remarkable for EurekaHedge and Morningstar. This may be due to the fact that these databases suffer from significant survivorship bias, which is difficult to disentangle from backfilling bias.

Panel D in Table 3 presents the EW average performance results that address the impact of the performance smoothing that is a consequence of serial correlation in hedge fund returns. Recent papers in the related literature argue that serial correlation in hedge fund returns is either due to (i) holdings of illiquid assets (Getmansky, Lo and Makarov (2004) or (ii) misreporting (Bollen and Pool 2008, 2009). First, to measure the impact of serial correlation on average performance, we add a MA(2) process to the Fund and Hsieh (2004) model's error term. Second, we estimate the econometric model proposed by Getmansky, Lo and Makarov (2004) and adjust the standard deviations of excess returns and Sharpe ratios for serial correlation. Our results in Panel D suggest that the adjustment leads to lower alpha *t*-values in Fung and Hsieh (2004) model. However, the t-values remain statistically significant at the 5% level. EurekaHedge and Morningstar databases show the highest average alphas (7.68% per year and 7.35% per year). Thus, the database ranking based on the average performance remains the same as shown in Panels A and B. The results suggest also that the adjustment for serial correlation leads to lower Sharpe ratios and higher standard deviation of excess returns. These conclusions are consistent with Getmansky, Lo and Makarov (2004).

[9] Another possibility to address backfilling bias would be to use information from the 'Date added to database' –field. However, all of the databases do not provide such data field. Hence, for consistence, we opt not to adjust backfilling bias using it.

To examine whether hedge funds are capable to deliver superior performance consistently through time, we divide our aggregate database into subperiods following Hsieh, Naik and Ramadorai (2008). Panel E in Table 3 presents the results showing that hedge fund risk taking is changing through time, but hedge funds are able to deliver alpha during all the subperiods expect during the period from January 1997 to September 1998. Finally, Table A1 in Appendix (Panel E) shows that the return differences between databases are smaller after 2004 when the survivorship and backfilling biases are not so pronounced in EurekaHedge and Morningstar. It implies that the properties of databases seem to converge over the time.

Panel F in Table 3 presenting the average performance across size categories show that the average performance is strongly related to fund size. Small hedge funds deliver an extremely high performance, while the large ones are not able to deliver statistically significant alpha. Indeed, the hedge funds having AuM below 10 million dollars deliver outstanding performance with alpha (*t*-statistic of alpha) of 7.25% per year (6.80). The magnitude of the average alpha and its statistical significance are implausible high. Therefore, we argue that small hedge funds are associated with pronounce data biases. In addition, large hedge funds, which size is above 250 million dollars, are not able to deliver significant alpha. There also seems to be monotonic relationship between hedge fund size and performance. It is interesting to note that hedge funds, which do not report even a single AuM observation, provide superior performance. We see it as a one of the reasons to explain why EW returns are higher than VW returns, since also the hedge funds having solely missing AuM observations are included in EW portfolios. To sum up, these results shed further light why we document so high performance for equally- and value-weighted portfolios.

Table 4 presents convincing evidence why the average performance of EW and VW portfolios differs among data vendors. Panel A in Table 4 shows that large hedge funds' average returns are the highest for TASS when compared to two other mature databases, HFR and BarclayHedge. Specifically, average returns are consistently the highest for TASS when fund size is above 500 million dollars. BarclayHedge delivers the smallest average returns among large funds, respectively. This is one of the reasons why TASS has the highest VW performance, whereas BarclayHedge delivers the lowest VW returns. Panel B in Table 4 shows another important reason in explaining performance differences. We demonstrate that missing AuM observations show different behavior for TASS if compared to other mature databases. Since the AuM coverage of TASS is poor if especially compared to BarclayHedge, we estimate its impact on average returns. Only for TASS, but not for other databases, we find that when AuM observations are missing from the middle or the end of AuM time-series, then the respective returns tend to be extremely poor. Specifically, the associated mean return is 3.48% per annum in TASS, while the mean return is 6.28% in BarclayHedge being almost as twice high. The magnitude is also important, since TASS contains 2,972 hedge funds that have missing AuM observations at the middle or at the end of AuM time-series. Clearly, missing AuM observations also shed some light why the TASS's VW performance is superior over its EW performance, since EW portfolio contains those returns that do not have the respective AuM observations. We conclude that differences in VW returns among data vendors can be explained by non-randomly missing AuM observations.

3.3. Investment objectives

Next, we turn on the average performance of hedge fund strategies by examining whether hedge funds grouped by investment objective add value on a net-of-fees basis. Table 5 reports the results for the consolidated database, while Table A2 in Internet Appendix displays the results for each of the commercial databases. We classify hedge funds into 12 categories: CTA, Emerging Markets, Event Driven, Global Macro, Long/Short, Long Only, Market Neutral, Multi-Strategy, Relative Value, Short Bias, Sector and Others. The Internet Appendix provides the classification of hedge fund strategies. Figure 3 presents the proportions of hedge funds by investment objective. The proportions are similar across data vendors. This implies that our merging approach delivers an accurate strategy matching and multiple share class deletion approach.

Table 5 presents the average performance of hedge fund strategies for the aggregate database. Overall, we find that hedge fund strategies are capable to deliver significantly positive risk-adjusted returns. The only exception is Long Only strategy, which has a positive but insignificant alpha (t-statistic=1.76). The results in Table A2 show that the strategy-level average performance varies significantly between commercial databases. It seems that for the most of the hedge fund strategies, EurekaHedge and Morningstar (BarclayHedge, HFR and TASS) are among top (bottom) performing databases. For example, the annual average alpha of Emerging Markets strategy equals to 12.63 (5.23) percent for Morningstar (TASS). Therefore, the strategy-level results give us further evidence that differences in survivorship bias are driving the performance differences between commercial databases. Importantly, we also demonstrate significant differences in the average performance between TASS, HFR and BarclayHedge. For example, Emerging Markets strategy shows the annual average alpha of 9.03 % per year for HFR, while TASS's and BarclayHedge's respective alphas are remarkable lower being 5.23% per year and 6.92 % per year. Hence, not only the average hedge fund performance differs across databases, as we documented earlier, but the strategy-level average alphas differ even across mature commercial databases.

Panel B in Table 5 address the effect of the fund size on the strategy-level average performance using the aggregate database. The results indicate that for ten of the twelve indices (all except sector and short bias) small funds outperform large funds. For example, in Emerging Markets category, small (large) funds exhibit the average alpha of 11.88% per year (0.86% per year). Hedge funds are sorted into terciles each December based on fund-level monthly non-missing AuM observations. Portfolio returns are calculated for equal-weight portfolios monthly using 12-month holding period. Results in Panel B support the stylized facts shown in Table 3 and Table 4. Panel B in Table A2 describes the effect of the fund size on the strategy-level performance using the five individual databases. The results are consistent with results shown in Table 5 (Panel B). For the most of the strategies and databases, small funds outperform large funds in terms of the annual average performance.

3.4. Domiciles

We investigate whether hedge funds add value, on average, across domiciles. The prior literature such as Aragon, Liang, and Park (2011) documents that onshore hedge funds registered in USA deliver higher performance than the offshore hedge funds. We extend their work by examining the hedge fund average performance around the whole world. The domicile regions of hedge funds and management firms are divided to two groups: (1) onshore; and (2) offshore. United States and Canada are classified as onshore regions. Other domicile regions are classified into four groups: (1) Asia and Pacific; (2) Caribbean; (3) Europe; (4) Rest of world. Figure 4 shows the pie charts of the proportions of funds grouped by fund-level domicile region. In BarclayHedge database, 46% of funds are onshore funds. In other databases, most of the funds are domiciled in Caribbean (37% in the aggregate database). Overall, the proportion of funds between onshore and offshore seems to be similar across all databases.

Table 6 shows results of the number of unique funds and firms in domicile region groups as well as the average performance results of funds grouped by domiciles. Panel A in Table 6 shows the number of unique hedge funds in domicile regions. Results are obtained using Hedge Fund Research, BarclayHedge, and EurekaHedge, since these data vendors provide information on domicile of the management firms. Most of the firms are established in 'onshore regions' (64%) and most of the funds are domiciled as onshore vehicles (38%). Hedge funds that are established in the Caribbean (Europe) account for 35% (16%) of all hedge funds (15,805). The number of unique funds suggests that the domicile region of both firms and funds provide the same type of classification of offshore and onshore domicile.

Panel B in Table 6 presents the results of the average performance grouped by fund domicile. All portfolios are equal-weighted monthly. The results show that there are significant differences in average performance across domicile groups. On average, we find that onshore based funds outperform offshore based funds. Onshore (offshore) category has the annualized Fung and Hsieh (2004) alpha of 7.45% per year (4.74% per year). Indeed, Europe based hedge funds deliver the poorest performance (alpha equals to 3.21% per year). The aggregate database shows the highest annual average alpha for Asia-Pacific group, 8.16 % with the *t*-statistic of 3.99. All domicile groups show statistically significant average alpha with *t*-statistics above the 5% level. Finally, the results in Table A3 in Appendix show that EurekaHedge and Morningstar seem to outperform mature databases that do not suffer pronounce data biases. Specifically, in the following groups: (1) all funds; (2) onshore; (3) all offshore funds ; (4) Caribbean; (5) rest of world, EurekaHedge and Morningstar outperform BarclayHedge, HFR and TASS. This is consistent with the fact that EurekaHedge and Morningstar have smaller attrition rates and higher average performance if compared to BarclayHedge, HFR and TASS databases. Panel C in Table 6 provides the results of the average performance grouped by the firm-level domicile region. On average, onshore funds outperform offshore funds in terms of the average alphas (6.88% per year and 5.14% per year). We document also the average performance grouped by cities where management firms are legally established. New York and London based management companies generate similar levels of the average alphas (6.33% per year and 6.16% per year).

4. Hedge fund performance persistence

In this section, we examine hedge fund performance persistence. When the abnormal performance of hedge funds is due to manager skills, then the same top performing hedge funds should have a high return year after year. If some hedge fund managers have access to superior information, sorting funds on past performance should indicate whether past winners are on average future winners and past losers are future losers. For investors, the performance persistence is crucial since hedge funds typically restrict capital withdrawals by imposing lockup, advance notification, and redemption periods.[10] All these restrictions indicate that new investors are not able to withdraw from hedge funds in a timely fashion. Therefore, hedge funds that are able to add value after fees consistently through time is a rewarding feature for investors. The pioneering literature (e.g., Agarwal and Naik (2000), Brown, Goetzmann, and Ibbotson (1999), and Liang (1999)) documents that hedge fund performance only persists for short periods and it disappears at the annual horizons. However, recently, using a sophisticated econometric approaches, Jagannathan, Malakhov, and Novikov (2010), and Kosowski, Naik, and Teo (2007) show that top abnormal performance of hedge funds persists even at annual horizons. The stylized fact is that hedge funds are able to deliver economically significant performance persistence.

The overall results documented in Table 7 show that hedge fund performance persists, but the persistence is driven by small funds. Using an aggregate database, we document that post-ranking Fung and Hsieh (2004) alpha estimates are positive and statistically significant for the top decile across different holding periods and size terciles. Hence, hedge fund investors that invest in top decile hedge funds are able to earn continuously statistically and economically significant alpha even at annual horizons. However, the post-ranking alpha estimates for spread portfolio, calculated as top decile minus bottom decile, are not always significantly positive. For the aggregate database, equal-weight spread portfolios show positive persistence even at annual horizon, while value-weighted spread portfolios only indicate significant persistence on a monthly basis. Hence, hedge fund performance persistence seems to be driven by small funds. By dividing hedge funds into three size terciles, we confirm the issue by showing that performance persistence is the strongest for small funds and weakest for the large funds. The conclusion about the performance persistence changes when one draws the inference using only one of the commercial databases. Using mature databases (BarclayHedge, HFR and TASS), we find significant evidence about performance persistence. In contrast, using EurekaHedge and Morningstar, we are not able to document significant performance persistence. Differences in

[10] Hedge funds can impose a lockup provision specifying a time period during which new investors are not able to withdraw their shares. Investors can withdraw their shares at the end of the lockup period by giving an advance notice. When the notice is given, investors have to wait until the pre-specified redemption interval is at hand. About 25% of hedge funds apply one year lockup, while a typical hedge fund imposes a 30-day's notice and allows quarterly redemptions.

performance persistence can be explained by the fact that BarclayHedge, HFR and TASS are not associated with pronounce data biases, while EurekaHedge and Morningstar suffer significantly from survivorship and backfilling biases.

We start by cleaning commercial and aggregate databases from multiple share classes. To investigate hedge fund performance persistence, we follow the standard way in performance evaluation literature (e.g. Carhart (1997)). We sort hedge funds based on the *t*-statistic of Fung and Hsieh (2004) alpha. Kosowski Timmermann, Wermers and White (2006) describe several attractive features of *t*-statistic. Although alpha measures the size of the abnormal performance, those estimates can be sensitive to outliers. Due to short time series, the alphas are also estimated imprecisely. The *t*-statistic of alpha provides a correction, since it is measured as a ratio of alpha estimate and the estimated precision of alpha estimate. Specifically, we sort hedge funds into decile portfolios based on their *t*-statistics of alpha obtained from the Fung and Hsieh (2004) model that is estimated over the previous two years. We report persistence results using four different holding periods: (i) monthly, (ii) quarterly, (iii) semiannual, and (iv) annual over the period from January 1994 to December 2010. For example, if a quarterly holding period is used, we sort hedge funds on March, June, September and December of each year into decile portfolios. Following Kosowski, Naik, and Teo (2007), the post-ranking portfolio excess returns are created monthly, so the weights are readjusted whenever a fund disappears. Hedge funds with the highest past two-year alpha's *t*-statistic comprise decile 10 (top decile), while funds with the lowest past two-year alpha's *t*-statistic form decile 1 (bottom decile). We estimate also the abnormal return of the spread portfolio (decile 10 minus decile 1). A significant difference in abnormal performance between top and bottom decile provides evidence of performance persistence during the selected holding period. As a baseline, we form equally-weighted (EW) post-ranking portfolios. To study whether the performance persistence is driven by small funds, we also construct value-weighted (VW) portfolios.

Importantly, to understand the impact of data biases between commercial databases, we calculate dropout rates for each of the decile portfolios used in persistence tests. The dropout rate is the percentage of hedge funds dropping out from the underlying decile portfolio during the holding period. We expect that dropout rates are relatively low (high) for the top (bottom) decile. This is due to the fact that poor performing hedge funds close their operations, while top ones continue to operate. In addition, since the attrition rates in Table 2 are considerable lower for EurekaHedge and Morningstar, we hypothesize that across decile portfolios dropout rates vary significantly more to BarclayHedge, HFR and TASS than to EurekaHedge and Morningstar. We also expect that the spread between the top and the bottom deciles is wider for BarclayHedge, HFR and TASS than for EurekaHedge and Morningstar. Hence, BarclayHedge, HFR and TASS should show more significant performance persistence than EurekaHedge and Morningstar.

Table 7 shows the results of EW, VW, and size tercile hedge fund performance persistence tests. Table reports the annualized Fung and Hsieh (2004) alphas, *t*-statistics of alphas, and dropout rates for bottom, top, and spread portfolios. Panel A in Table 7 presents the persistence tests for the equally-weighted portfolios. Using the aggregate database, we find a significantly positive alpha for spread portfolios across different holding periods. In particular, using the monthly rebalancing, the spread portfolio delivers an annual post-rank alpha of 5.80% with the *t*-statistic of 4.10. For a yearly holding period, the respective post-rank alpha is 2.84% per year, being statistically significant with the *t*-statistic of 1.97. Hence, the results suggest that hedge fund performance persists even at annual horizons.

To examine whether small hedge funds are driving the performance persistence, we construct post-ranking portfolios on value-weighted basis. According to Panel B in Table 7, the *t*-statistics of VW post-rank spread portfolio alpha reveal performance persistence only at monthly horizon for the aggregate database. Specifically, the spread portfolio between the top and the bottom deciles of the aggregate database shows the annual alpha of 4.62% with *t*-statistic equaling to 2.89. The aggregate database do not show statistically significant spread portfolio alpha for quarterly, semiannual, and yearly holding periods. The results suggest that performance persistence is driven by relatively small hedge funds that have smaller weight in VW portfolios relative to large funds. Overall, results of persistence suggest that the results depend on weighting schema and performance persistence of VW portfolios disappear faster compared to EW portfolios.

To confirm the performance persistence is driven by small funds, we investigate hedge fund performance persistence among different size categories including small, median, and large funds.[11] It would be important to find performance persistence among large hedge funds, since investors could exploit it in practice. In addition, the data biases are more pronounce among small

[11] Table A5 in Appendix presents the number of hedge funds and AuM cut points in these three terciles. The first size group contains hedge funds that are, on average, smaller than 10 million dollars, while the second size group includes funds, which average size is between 10 and 50 million dollars. The third group’s average hedge fund size is above 50 million dollars.

funds than large funds. Panel C in Table 7 presents the persistence test results for aggregate database across three size groups. The results show that small funds are driving the performance persistence. Specifically, the spread portfolio alphas are almost consistently the highest for small funds and the lowest for large funds. Indeed, using only the group of small funds, we document significant performance persistence even at annual horizons. For the median and large hedge funds, we find significant performance persistence at semiannual horizons, but not at annual horizons. However, the top decile alphas are still economically and statistically significant even for large funds at annual horizon. This implies that hedge fund investors may be able to earn significant alpha by investing in funds based on the *t*-statistic of alpha.

According to Table 7, the conclusion about hedge fund performance persistence varies significantly across commercial and aggregate databases. We cannot document significant performance persistence at annual horizons for any of the commercial databases. This suggests that the conclusion changes if one relies only on one of the commercial databases. On an equally-weighted basis, Morningstar and EurekaHedge do not show any performance persistence. Indeed, we find that their spread portfolio's post-rank alphas are indistinguishable from zero even with monthly rebalancing. In contrast, for TASS, HFR, and BarclayHedge, the results show performance persistence at monthly, quarterly and semiannual horizons. Even these mature databases do not show performance persistence at annual horizons, as our aggregate database does. On a value-weighted basis, only the HFR database shows also persistence at monthly horizon. Hence, using commercial databases, performance persistence seems to vanish very quickly.

To investigate why performance persistence results differ across databases, we calculate dropout rates. First, we find that dropout rates are remarkably wider for mature databases than to younger databases. At annual holding period, Table 7 presents that the difference in dropout rates between top and bottom portfolios is 12.61% (BarclayHedge), 13.80% (HFR), and 15.27% (TASS) for mature databases, whereas they are significantly lower 4.19% (Morningstar), and 6.97% (EurekaHedge) for younger databases. This is consistent with the findings that EurekaHedge and Morningstar databases have small attrition rates for 1994-2004 periods if compared to other databases. Second, the alphas of bottom portfolios vary significantly across commercial databases being the highest for EurekaHedge and Morningstar. This is shown clearly in Figure 7 that plots the EW annualized Fung and Hsieh (2004) alphas for persistence portfolios using four different holding periods. Based on the graph, EurekaHedge and Morningstar have considerably higher levels of average alphas for bottom portfolios if compared to other databases. The levels of the average alphas of the other databases are close to each other. In contrast, the top portfolios' alphas are remarkable similar across databases. Based on these two reasons, we argue that relatively low attrition rates in Morningstar and EurekaHedge databases are driving the differences in persistence test among databases.

In our robustness checks, we rule out the possibility that differences in performance persistence are driven by heterogeneity in risk exposures. Table A4 in Appendix presents persistence tests at annual level and reports annualized average excess returns and standard deviations as well as the alpha, risk-loadings, *R*-square with respect to Fung and Hsieh (2004) model, and dropout rate. Table A4 shows that Sharpe ratios, risk loadings and R^2:s of the seven-factor Fund and Hsieh (2004) model are very similar across individual databases and the aggregate databases. Therefore, we argue that performance persistence differences cannot be explained by risk exposures among databases.

Finally, we perform a performance persistence test for the aggregate database to measure whether a real-time investor is able to exploit the short-term performance persistence of hedge funds. We build feasible rebalancing strategies by taking into account fund-specific share restrictions within each rebalancing horizon. We exclude funds having lockup, notice, or redemption periods longer than the rebalancing period in question. For the aggregate database, results suggest EW persistence only at quarterly horizon when funds are restricted to have share restrictions smaller than six months. Without these constraints, our baseline result (Table 7, Panel A) shows EW persistence at semiannual horizon.

We sort hedge funds into portfolios based on their t-statistics of alphas using the recent 24-months of returns preceding the evaluation period. Portfolios are based on four rebalancing horizons: (1) monthly; (2) quarterly; (3) semiannually; and (4) yearly. Within each rebalancing horizon, decile portfolios are formed using only the feasible information taking into account fund-specific share restrictions. For instance, for the feasible quarterly rebalancing strategy, we exclude funds that have lockup, redemption, or notice periods longer than three months. This implies that, for this strategy, we use 3-month lagged information to estimate persistence in order to mitigate effects on look-ahead bias. Table 8 shows the annualized EW post-rank alphas for persistence portfolios. Portfolio returns are calculated for EW portfolios monthly, so the weights are rebalanced whenever a fund disappears.

Results in Table 8 suggest that the length of share restrictions affects to the estimates of performance persistence. Panel A in Table 7 shows performance persistence (Equal-weight) for the aggregate database at semiannual level (t-statistic=3.48). According to Table 8, the feasible strategy for the semiannual holding horizon does not suggest significant performance persistence. Thus, the results indicate that funds with notice period longer than six month show performance persistence, and the real-time investor cannot exploit this predictability in semiannual rebalancing strategies.

5. Hedge fund performance and fund characteristics

In this section, we examine the cross-sectional relationship between hedge fund characteristics and performance. Specifically, we explain the cross-sectional variation in hedge fund returns and alphas by fund-specific characteristics related to managerial incentives, liquidity, and capacity constraints. Several academic papers have documented that hedge fund-specific characteristics explain cross-sectional differences in fund performance. First, Ackermann, McEnally, and Ravenscraft, (1999) and Liang (1999) find a positive relation between incentive fees and Sharpe ratio. Using a comprehensive database (a union of CISDM, HFR, MSCI, and TASS), Agarwal, Daniel, and Naik (2009) document that hedge funds with greater manager's option delta deliver superior performance. The findings suggest that managerial incentives are associated with superior performance. Second, Aragon (2007) argues that share restrictions allow hedge funds to manage illiquid assets and earn an illiquidity premium. Aragon uses the TASS database for January 1994 through December 2001 and documents that hedge funds with a lockup period deliver approximately 4% higher annual returns than their peers. Finally, using a union of TASS and HFR databases, Teo (2010) shows that small hedge funds outperform large ones by 2.75% per year after adjusting for risk. Using the TASS database, Aggarwal and Jorion (2010) show that due to particularly strong financial incentives, emerging hedge funds are able to outperform their peers.

To understand how fund-specific characteristics differ across databases, Panel A in Table 9 presents the descriptive statistics of hedge fund characteristics. The cross-sectional averages and standard deviations of fund characteristics are quite similar across commercial databases. The average incentives fee between databases ranges from 17 to 19%, while management fee is consistently about 2% across data vendors. The high-water mark indicator varies significantly among databases. Using HFR, we find that 88% of hedge funds impose a high-water provision, while BarclayHedge suggests that only 61% of hedge funds use high-water mark. The number of hedge funds imposing a lockup varies significantly across databases. According to EurekaHedge, only 19% of hedge funds use a lockup, but based on Morningstar, we find that 45% of hedge funds use lockups. Table also reports the proportion of hedge funds having a missing value for a specific variable. The results indicate that the coverage of hedge fund characteristics differ between databases. Morningstar has the highest amount of missing observations for both compensation and share restrictions variables. The other four databases have almost the similar amount of missing observations. BarclayHedge has relatively large number of missing redemption and lockup period observations, while all of the four databases have almost the perfect coverage of compensation variables. Overall, these results imply that the properties of compensation and share restriction variables differ remarkably among commercial databases. Therefore, it is interesting to investigate cross-sectional performance differences across commercial databases and compare results to ones obtained using the aggregate database.

We first address this issue by forming decile portfolios based on fund age, size, and flow, inventive fee as well as lockup, notice and redemption periods across databases and size categories. We rebalance portfolios on a yearly basis, and estimate the spread between top and bottom deciles in order to examine whether the average performance differs between the extreme realizations of a specific fund characteristic. Panel B in Table 9 presents the Fung and Hsieh (2004) alphas and the associated *t*-statistics across fund characteristics and databases. The portfolio sorts across databases show that younger and smaller hedge funds outperform, while hedge funds with significant short term flows do not deliver superior future performance. Managerial incentives seem to be important, since hedge funds with great incentive fees outperform the funds with low fees across databases. Most interestingly, we find that share restrictions seem not to explain consistently hedge fund performance across database. As Aragon (2007), we document that hedge funds imposing strict lockups are associated with significant outperformance. However, notice and redemption periods seem not to explain consistently cross-sectional differences in hedge fund risk-adjusted returns. Using TASS and HFR, we document a significant and positive relation between notice period and risk-adjusted average return, while other commercial databases and even the aggregate database show insignificant relation between notice period and risk-adjusted returns. The relationship is extremely strong in TASS, since the average spread between the top and bottom deciles is 4.30% per year with the t-statistics of 4.66. Hence, the conclusion about the impact of notice period on hedge fund performance varies significantly across databases. The unreported tests suggest that results are robust even after we exclude CTAs from our sample.

Using the Fama-McBeth (1973) approach, we examine which of the fund-specific characteristics are the most important variables in explaining the cross-section of hedge fund performance. Formally, the Fama-McBeth (1973) procedure can be expressed as

$$R_{i,t} = \lambda_0 + \lambda_1' Y_{i,t} + \lambda_2' Z_i + u \quad (2)$$

where $R_{i,t}$ refers to excess return (alpha) of a hedge fund i at the time t, λ_1 is a vector representing the slope coefficients for time-variant characteristics, and λ_2 is a vector representing the slope coefficients for time-invariant characteristics. The vector of time-variant characteristics ($Y_{i,t}$) includes hedge fund size, flow, and age, while the vector of time-invariant characteristics (Z_i) contains management and incentive fees, high-water mark provision and share restrictions in the form of lockup, notice, and redemption periods. We control for the strategy and domicile fixed effects, and adjust standard errors for autocorrelation and heteroskedasticity following Newey and West (1987).[12]

Panel A in Table 10 reports Fama-McBeth regression results using hedge fund excess returns from January 1994 to December 2010, while Panel B of Table 10 presents the results for the Fung and Hsieh (2004) alphas. For the aggregate database, we find evidence that small, young and the funds with greater managerial incentives outperform their peers. We cannot document that strict share restrictions are associated with superior performance. The only share restriction variable that shows weak significance is lockup, but after controlling for the role of common risk factors, the relation between the lockup and future alphas is also insignificant.

Table 10 demonstrates that the results differ significantly across databases suggesting that the conclusion about which fund characteristics explain cross-sectional differences between fund returns may be database specific. Specifically, Table 10 shows that the results for fund size, age

---

[12] It is important to control for domicile fixed effects, since Aragon, Liang, and Park (2011) document that the impact of share restriction on hedge fund performance varies across domiciles. We find very similar results when we adjust standard errors for within-cluster correlation, heteroskedasticity, and autocorrelation.

and flow are very robust across databases as the previous literature (e.g., Teo, (2010, 2011) and Aggarwal and Jorion (2010)) suggests. However, the variables that are related to managerial incentives do not explain hedge fund cross-sectional returns consistently across databases. The sign and significance for the high-water mark is changing wildly. However, all of the variables related to incentives are significant for aggregate database. Therefore, we conclude that hedge funds with greater managerial incentives outperform. Finally, we find very little evidence that share restrictions in the form of lockup, notice and redemption periods are related to higher risk-adjusted returns when we control for the role of other characteristics in multivariate regression. As Aragon (2007), we find the strongest evidence using the TASS database, whereas using other databases and even the aggregate database the conclusion is much weaker. For example, we find a significantly positive coefficient for notice period only using TASS, but none of the other databases.

Panel C of Table 10 presents results for aggregate database across three size groups. The size groups are the same as in persistence tests. The results show that the relationship between size and age is not significant consistently within size groups. None of the coefficients for fund size is significant implying that cross-sectional performance difference may not be monotonic. The fund's age seems to be important only for small funds. This may be explained by more pronounced backfilling bias.[13] The relationship between the fund flow and performance is consistent across size groups. A positive (negative) past month flow is associated with a higher (lower) next's month performance. All of the coefficients related to compensation structure are positive among size groups. However, managerial incentives seem to be more important for small funds, since the coefficient is the highest for them. Finally, share restrictions are weakly associated with hedge fund performance also across size groups. The coefficients for lockup are significantly positive for median and large funds. However, none of the other share restriction variables are significantly related to performance. This suggests that share restrictions are not able to explain cross-sectional differences in hedge fund performance when the role of other characteristics is controlled for. Overall, the results for size categories are in the line with the findings that especially managerial incentives are important for emerging hedge funds.

[13] We define fund age as a length of return time-series. Aggarwal and Jorion (2010) examine the relationship between fund age and performance using a novel event study methodology. We do not follow this methodology, since only some of the databases provide 'date added to database' –variable that is needed in order to follow their methodology.

6. Conclusion

This paper provides new stylized facts about the hedge fund industry using a large consolidated hedge fund database. We document the sensitivity of hedge fund performance and data biases to different data bases.

We propose a novel, easily repeatable methodology that can be used to identify unique hedge funds from multiple share class structures containing more than two units linked to the same risky investment portfolio. The major advantage of the approach is that we obtain as long as possible time-series for unique hedge funds. Therefore, we can measure hedge fund performance more accurately and mitigate the impact on several biases more efficiently. We highlight the importance of using multiple databases in order to achieve more reliable proxy of the unobservable population of hedge funds. After merging commercial databases, one of the striking observations is that 67.7% of all share classes are covered exclusively by only one database with BarclayHedge having the largest share of unique share classes (15.6%).

We demonstrate importance of using a large consolidated database by investigating hedge fund average performance, performance persistence, and the cross-sectional relation between fund characteristics and hedge fund performance. First, our aggregate database shows that hedge funds add value after fees: average excess returns and alphas are economically significant on both equal- and value-weight basis across databases. We document differences in average performance measures between databases suggesting that hedge fund performance results are subject to database selection. Second, using our aggregate database, TASS, HFR, and BarclayHedge we find that hedge fund performance persists at annual horizon, if portfolios are created on an equal-weight basis. Based on value-weighted portfolios, the aggregate data and HFR show persistence at monthly horizon. Hence, the results suggest that performance persistence vanishes faster on a value-weight basis, a result that is likely to be driven by small funds. For EurekaHedge and Morningstar, we do not find persistence at all. This is due to the fact that EurekaHedge and Morningstar have lower attrition rates and higher a survivorship bias creating an upward bias in results.

The stylized facts proposed in the paper suggest that one should be careful in interpreting the results of hedge fund performance if the analysis is conducted using only one of the commercial databases. Our study of databases shows differences in coverage of returns, assets under management, and number of funds in the graveyard module.

The comparison of commercial databases to our aggregate database allows us to evaluate whether all hedge fund databases contain the same level of information and whether differences between databases induce biased inference. The major finding is that the conclusions based on the consolidated database are qualitatively different from those based on the individual databases. To avoid major biases, we argue that it is important to use a consolidated database because stylized facts inferred from individual databases may differ from the true population.

References

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

Agarwal, V., Daniel N. D., and Naik N. Y., 2009. Role of managerial incentives and discretion in hedge fund performance. The Journal of Finance 64(5), 2221-2256.

Agarwal, V., and Naik, N.Y., 2000. Multi-period performance persistence analysis of hedge funds. Journal of Financial and Quantitative Analysis 53, 327-342.

Agarwal, V., and Ray S., 2011. Determinants and implications of fee changes in the hedge fund industry. Georgia State University.

Aggarwal, R., K. and Jorion, P., 2010. The performance of emerging hedge funds and managers. Journal of financial economics. 96, 238-256.

Aragon, G. O., 2007. Share restrictions and asset pricing: Evidence from the hedge fund industry. Journal of Financial Economics 83(1), 33.58.

Aragon, G. O. Liang, B., and Park H., 2011. Onshore and offshore hedge funds: Are they twins? Arizona State University.

Aragon, G. O. and Martin, J. S., 2010. A unique view of hedge fund derivatives usage: Safeguard or speculation? forthcoming at the Journal of Financial Economics.

Aragon, G. O. and Nanda V., 2011. Strategic delays and clustering in hedge fund reported returns. Arizona State University and Georgia Institute of Technology.

Bollen, N. P.B., and Pool, V. K., 2009. Do hedge fund managers misreport returns? Evidence from the pooled distribution , Journal of Finance 64, 2257-2288.

Bollen, N. P.B., and Pool, V. K., 2008. Conditional return smoothing in the hedge fund industry , Journal of Financial and Quantitative Analysis 43, 267-298.

Brown, S., Goetzmann, W., Ibbotson, R., 1999. Offshore hedge funds: Survival and performance, 1989-95. Journal of Business 72, 91-117.

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

Elton, E. J., Gruber M. J., and Blake C. R., 2001. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases. Journal of Finance 56(6), 2415-2430.

Fama, E. F., and MacBeth J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607-636.

Fung, W., Hsieh, D., 2000. Performance characteristics of hedge funds and CTA funds: Natural versus spurious biases. Journal of Financial and Quantitative Analysis 35, 291-307.

Fung, W., Hsieh, D., 2004. Hedge fund benchmarks: a risk based approach. Financial Analyst Journal 60, 65-80.

Fung, W. and Hsieh D., 2009, Measurement biases in hedge fund performance data: An update, Financial Analysts Journal 65, 36-38.

Fung, W., Hsieh, D., Naik, N., Ramadorai, T., 2008. Hedge funds: performance, risk, and capital formation. Journal of Finance, 63 (4), 1777-1803.

Getmansky, M., Lo, A., Makarov, I., 2004. An econometric model of serial correlation and illiquidity of hedge fund returns. Journal of Financial Economics 74, 529-610.

Goetzmann, W. N., Brown, S., Liang, B., and Schwarz C., 2009. Estimating Operational Risk for Hedge Funds: The Omega-Score, Financial Analysts Journal, 5(1), 43-53,

Goetzmann, W.N., Brown, S., Liang, B., and Schwarz C., 2008. Mandatory Disclosure and Operational Risk: Evidence from Hedge Fund Registration, Journal of Finance, 63(6), 2785-2815.

Güner, B., Rachev, S. T., Edelman, B., and Fabozzi, F. J., 2010. Bayesian inference for hedge funds with stable distribution of returns. Ch.4 in Rethinking Risk Measurement and Reporting: Volume II, Klaus Böcker, eds.

Jagannathan, R., Malakhov A., and Novikov D., 2010. Do hot hands exist among hedge fund managers? An empirical evaluation. The Journal of Finance 65(1), 217.255.

Jylhä P., (2011). Hedge fund return misreporting: Incentives and effects. Aalto University.

Liang, B., 1999. On the performance of hedge funds. Financial Analysts Journal 55(4), 72.85.

Liang, B., 2000. Hedge funds: The living and the dead. Journal of Financial and Quantitative Analysis 35(3), 309.326.

Kosowski, R., Naik N. Y., and Teo M., 2007. Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. Journal of Financial Economics 84(1), 229.264.

Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. Journal of Finance, 61(6).

Newey, W., West, K., 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

Patton, A. J. and Ramadorai, T., 2011. On the Dynamics of Hedge Fund Risk Exposures. Forthcoming in Journal of Finance.

Patton, A. J., Ramadorai, T, and Streatfield M., 2011. The reliability of voluntary disclosures: Evidence from hedge funds. Duke University and University of Oxford.

Teo, M., 2009. Geography of hedge funds. Review of Financial Studies, 22, 3531-3561

Teo, M., 2010. Does size matter in hedge fund industry? Singapore Management University - School of Business.

Teo, M., 2011. The Liquidity Risk of Liquid Hedge Funds, Journal of Financial Economics, 100, 24–44.

Table 1

Summary statistics of hedge fund returns and asset under management [January 1994 - June 2011]

Table provides summary statistics of hedge funds with at least 12 monthly return observations. Results are reported separately for each database. Aggregate database is a merged database of single databases including TASS, Hedge Fund Research, BarclayHedge, EurekaHedge, and Morningstar. Databases are merged using a novel statistical algorithm. First, all databases are grouped based on the management company name that is cleaned from errors and abbreviations. Second, correlation coefficients are estimated for each share class pair within each management company. Each share class pair in correlation analysis is required to have at least 12 non-missing return observations. Third, multiple share classes within management companies are grouped based on 0.99 correlation coefficient limit. Within groups, the share class having the longest return time series represents a unique hedge fund. If a single group has more than one share class left with the same length of returns, the largest share class is selected based on the asset under management to represent a unique hedge fund. If more than one share class has the same amount of returns and assets, the US dollar currency share class is selected. The row "aggregate" contains results of the merged database that applies all five databases as a source data. The aggregate database is removed from multiple share classes than can exist either between databases or within management companies. Single hedge fund databases are removed from multiple share classes using the same statistical procedure.

Panel A shows summary statistics of returns and asset under management. Panel shows the number of unique share classes (column "# share classes"). The name matching algorithm (presented in Appendix) is used to merge databases at the share class level in order to obtain the number of unique share classes in the aggregate data. Next column shows the number of unique management firms in each database. For simple returns, Panel reports the number of funds (column 4), annualized average return (column 5), and annualized standard deviation (column 6). Percentage amount of missing simple returns are reported in column 7. The next two columns consider the mean and standard deviation of the length of simple return time series. Summary statistics are also reported for asset under management (in millions of US dollars). Panel B shows the statistical properties of hedge fund returns. The first column describes the amount of funds in each database. The next three columns show the statistics of normality including skewness, kurtosis and the number of funds for which the Jarque-Bera test of normality is rejected at the 5% level. The next three columns show the first two order serial correlations and the number of funds for which the Ljung-Box test of serial correlation is rejected at the 5% level. The results of conditional smoothing are based on the regression proposed by Bollen and Pool (2006):

$$R_t = a + b_1^+ R_{t-1} + b_1^- (1 - I_{t-1}) R_{t-1} + \eta_t ,$$

where $I_{t-1} = 1$ if the return in month *t-1* is greater than its mean and zero otherwise. The next four columns show the number of funds having statistically significant positive and negative estimates of smoothing profiles based on the 5% level of significance. The next two columns show the results of misreporting based on a measure proposed by Jylhä (2011). The measure formalizes the fact that the number of small gains exceeds the number of small losses documented by Bollen and Pool (2009). Panel presents the measure of discontinuity (DC(%)) and z statistic (z(DC)) for testing the existence of discontinuity across databases. The final columns show the difference between the average December values and the average of January-November values and the p-value for the test that this difference equals zero after correcting standard errors for clustering at the fund-level. These measures test the result that on average, the December values are higher than January-November values as documented by Agarwal, Daniel and Naik (2011).

A. Summary statistics

All funds			Returns						Asset Under Management					
							Lenght of time series						Lenght of time series	
Database	# share classes	# firms	# funds	Mean % pa	Std % pa	Missing (%)	Mean	Std	# funds	Mean	Std	Missing (%)	Mean	Std
Lipper TASS	10073	3338	8220	8.98	4.19	0.68	64.45	44.49	6719	101.92	312.29	34.40	42.57	43.76
Hedge Fund Research	11423	3710	9648	10.05	4.08	0.98	65.15	45.36	7929	107.22	635.96	19.43	53.01	48.38
BarclayHedge	11358	4224	9719	10.47	4.52	0.94	64.04	45.15	9308	112.24	502.23	11.49	57.22	45.17
EurekaHedge	8632	3040	7164	10.37	3.91	0.29	66.81	45.41	6399	141.45	397.82	37.12	42.14	38.91
Morningstar	8274	2817	6772	9.31	3.69	2.58	70.33	48.38	5570	122.41	418.65	32.23	48.92	48.79
Aggregate	37592	8353	24749	9.66	4.46	1.36	62.03	44.37	21308	110.08	519.81	27.56	45.56	43.62

Alive			Returns						Asset Under Management					
							Lenght of time series						Lenght of time series	
Database	# share classes	# firms	# funds	Mean % pa	Std % pa	Missing (%)	Mean	Std	# funds	Mean	Std	Missing (%)	Mean	Std
Lipper TASS	4144	1386	3110	11.95	3.20	0.35	74.69	50.33	2413	123.17	356.34	37.75	46.66	50.81
Hedge Fund Research	4674	1716	3822	11.81	3.27	0.39	78.31	52.65	2943	170.94	1007.65	19.36	63.39	58.20
BarclayHedge	4256	1755	3536	12.44	3.44	0.34	78.15	52.80	3394	167.62	679.50	10.73	70.00	53.79
EurekaHedge	5441	2133	4452	12.39	3.62	0.27	73.15	48.78	4003	174.27	482.06	35.60	47.24	42.49
Morningstar	4369	1698	3558	10.99	3.33	2.14	77.58	52.25	2838	141.46	505.08	32.31	53.66	53.34
Aggregate	12929	4191	8512	12.15	3.71	0.35	72.45	50.61	7328	160.20	751.83	26.45	53.48	51.12

Defunct			Returns						Asset Under Management					
							Lenght of time series						Lenght of time series	
Database	# share classes	# firms	# funds	Mean % pa	Std % pa	Missing (%)	Mean	Std	# funds	Mean	Std	Missing (%)	Mean	Std
Lipper TASS	5929	1952	5110	7.18	4.61	0.94	58.23	39.25	4306	90.01	283.99	31.80	40.09	38.64
Hedge Fund Research	6749	1994	5826	8.90	4.50	1.52	56.52	37.43	4986	69.61	200.45	19.49	46.20	39.22
BarclayHedge	7102	2469	6183	9.34	5.01	1.42	55.98	37.84	5914	80.45	359.58	12.09	49.91	37.49
EurekaHedge	3191	907	2712	7.05	4.17	0.34	56.40	36.98	2396	86.62	172.25	40.34	33.76	30.39
Morningstar	3905	1119	3214	7.44	3.98	3.19	62.29	42.27	2732	102.62	302.68	32.13	43.67	42.60
Aggregate	24663	4162	16237	8.35	4.77	2.03	56.57	39.64	13980	83.82	336.99	28.29	41.40	38.48

B. Statistical properties of hedge fund returns

All funds

Database	# funds	Skew	Normality: Kurt	Normality: JB (% Rejection)	Serial Correlation: ρ_1	Serial Correlation: ρ_2	Serial Correlation: LB (% rejection)	Conditional Smoothing b+: Pos (%)	Conditional Smoothing b+: Neg (%)	Conditional Smoothing b-: Pos (%)	Conditional Smoothing b-: Neg (%)	Discontinuity: DC (%)	Discontinuity: z(DC)	December Spike: Avg (Dec minus Jan-Nov) %/mth	December Spike: Diff (p-value)
TASS	8072	-0.22	3.56	54.98	0.12	0.04	23.27	12.57	87.43	5.93	94.07	9.21	14.70	0.28	< 0.01
Hedge Fund Research	9508	-0.16	3.26	54.29	0.12	0.03	24.61	14.20	85.80	5.04	94.96	7.92	14.39	0.20	<0.01
Barclay	9588	-0.07	3.09	53.90	0.10	0.02	22.53	12.99	87.01	5.09	94.91	7.92	14.57	0.27	<0.01
Eureka	7009	-0.15	2.92	53.09	0.12	0.04	24.17	12.53	87.47	5.68	94.32	6.86	10.53	0.32	<0.01
Morningstar	6681	-0.22	3.28	54.42	0.13	0.03	25.92	13.73	86.27	7.07	92.93	7.87	12.04	0.36	<0.01
Aggregate	24768	-0.16	3.02	51.36	0.10	0.03	21.44	11.34	88.66	5.77	94.23	7.91	20.05	0.28	<0.01

Alive

Database	# funds	Skew	Normality: Kurt	Normality: JB (% Rejection)	Serial Correlation: ρ_1	Serial Correlation: ρ_2	Serial Correlation: LB (% rejection)	Conditional Smoothing b+: Pos (%)	Conditional Smoothing b+: Neg (%)	Conditional Smoothing b-: Pos (%)	Conditional Smoothing b-: Neg (%)	Discontinuity: DC (%)	Discontinuity: z(DC)	December Spike: Avg (Dec minus Jan-Nov) %/mth	December Spike: Diff (p-value)
TASS	2963	-0.19	3.28	57.75	0.13	0.06	27.07	12.69	87.31	6.41	93.59	7.22	7.89	0.29	<0.01
Hedge Fund Research	3682	-0.15	3.33	58.17	0.13	0.04	29.88	15.92	84.08	5.57	94.43	8.14	10.61	0.22	<0.01
Barclay	3400	-0.07	3.15	57.47	0.11	0.03	27.15	14.26	85.74	5.88	94.12	7.38	9.40	0.27	<0.01
Eureka	4296	-0.12	3.02	55.49	0.12	0.04	26.05	13.08	86.92	5.61	94.39	6.86	8.87	0.30	<0.01
Morningstar	3463	-0.16	3.11	54.87	0.12	0.02	27.40	13.84	86.16	7.48	92.52	7.52	9.17	0.38	<0.01
Aggregate	10644	-0.14	2.83	52.02	0.10	0.03	23.97	11.34	88.66	6.28	93.72	7.21	13.17	0.30	<0.01

Defunct

Database	# funds	Skew	Normality: Kurt	Normality: JB (% Rejection)	Serial Correlation: ρ_1	Serial Correlation: ρ_2	Serial Correlation: LB (% rejection)	Conditional Smoothing b+: Pos (%)	Conditional Smoothing b+: Neg (%)	Conditional Smoothing b-: Pos (%)	Conditional Smoothing b-: Neg (%)	Discontinuity: DC (%)	Discontinuity: z(DC)	December Spike: Avg (Dec minus Jan-Nov) %/mth	December Spike: Diff (p-value)
TASS	5109	-0.24	3.72	53.38	0.12	0.03	21.06	12.50	87.50	5.64	94.36	10.48	13.67	0.28	<0.01
Hedge Fund Research	5826	-0.18	3.22	51.84	0.12	0.03	21.28	13.11	86.89	4.70	95.30	9.23	13.31	0.19	<0.01
Barclay	6188	-0.07	3.06	51.94	0.09	0.02	19.99	12.28	87.72	4.65	95.35	8.02	12.01	0.27	<0.01
Eureka	2713	-0.18	2.77	49.28	0.12	0.04	21.19	11.65	88.35	5.79	94.21	7.54	7.47	0.35	<0.01
Morningstar	3218	-0.29	3.47	53.95	0.15	0.03	24.33	13.62	86.38	6.62	93.38	9.18	10.01	0.33	<0.01
Aggregate	14124	-0.17	3.16	50.86	0.11	0.02	19.53	11.34	88.66	5.39	94.61	8.72	17.42	0.26	<0.01

Table 2
Attrition rates

Table presents attrition rates for each hedge fund database. Aggregate database is a merged database of single databases including Lipper TASS, Hedge Fund Research, BarclayHedge, EurekaHedge, and Morningstar. Databases are merged with a novel statistical procedure that is described in Appendex of the paper. Single databases are removed from multiple share classes using the same statistical procedure. Attrition rate is calculated as the ratio of the number of dissolved funds to the number that existed at the start of the year. Table includes year, number of funds that existed at the start of the year (Start), number of new defunct funds, and attrition rate in percentage (AR %).

	TASS			Hedge Fund Research			BarclayHedge			EurekaHedge			Morningstar			Aggregate		
Year	Start	Exit	AR %	Start	Exit	AR %	Start	Exit	AR %	Start	Exit	AR %	Start	Exit	AR %	Start	Exit	AR %
1994	.	9	.	.	4	.	.	9	.	.	.	.	.	.	.	.	22	.
1995	772	48	6.22	1100	38	3.45	1085	138	12.72	234	.	.	283	.	.	2420	207	8.55
1996	941	97	10.31	1398	130	9.30	1331	132	9.92	308	.	.	384	.	.	2979	302	10.14
1997	1112	72	6.47	1616	152	9.41	1501	108	7.20	418	.	.	517	.	.	3407	289	8.48
1998	1336	108	8.08	1786	261	14.61	1721	112	6.51	546	.	.	668	.	.	3838	405	10.55
1999	1488	145	9.74	1882	180	9.56	1965	95	4.83	715	1	0.14	849	.	.	4240	365	8.61
2000	1770	167	9.44	2112	245	11.60	2319	236	10.18	1037	.	.	1115	.	.	4895	552	11.28
2001	1985	189	9.52	2314	213	9.20	2552	191	7.48	1321	.	.	1391	.	.	5356	424	7.92
2002	2244	191	8.51	2594	212	8.17	2944	537	18.24	1715	9	0.52	1739	.	.	6148	609	9.91
2003	2548	190	7.46	2974	250	8.41	2997	250	8.34	2243	35	1.56	2217	1	0.05	7042	499	7.09
2004	2965	242	8.16	3406	264	7.75	3480	326	9.37	2885	178	6.17	2775	27	0.97	8347	662	7.93
2005	3471	334	9.62	3998	389	9.73	3939	413	10.48	3454	398	11.52	3471	113	3.26	9833	982	9.99
2006	3892	406	10.43	4499	477	10.60	4380	534	12.19	3737	374	10.01	4130	359	8.69	11072	1266	11.43
2007	4251	631	14.84	4859	671	13.81	4669	660	14.14	4143	401	9.68	4537	576	12.70	12132	1633	13.46
2008	4283	898	20.97	4963	986	19.87	4745	1058	22.30	4434	670	15.11	4619	927	20.07	12648	2520	19.92
2009	3935	616	15.65	4526	677	14.96	4278	594	13.88	4354	498	11.44	4240	672	15.85	11907	1807	15.18
2010	3724	580	15.57	4321	610	14.12	4160	631	15.17	4394	615	14.00	3951	805	20.37	11624	1729	14.87

Table 3

Hedge fund average performance [January 1994 - December 2010].

Table presents results of the average performance of hedge funds for all databases. Aggregate database is a merged database of single databases including TASS, Hedge Fund Research, BarclayHedge, EurekaHedge, and Morningstar. Databases are merged using a novel statistical procedure that is described in Appendix of the paper. Single databases are removed from multiple share classes using the same statistical procedure. Panel A presents results of the average performance for equal-weight portfolios and Panel B presents results for value-weight portfolios. Panel C shows the backfilling-adjusted results of the average performance for equal-weight portfolios. Panel D shows performance smoothing-adjusted results of the average performance for equal-weight portfolios. The annualized Fung and Hsieh (2004) alpha is adjusted for serial correlation by adding the MA(2) process to the model's error term. The annualized standard deviations of the excess returns and the Sharpe ratios are adjusted for serial correlation using the methodology proposed by Getmansky, Lo and Makarov (2004). Panel E shows results of the equal-weight average performance of the aggregate database in sub periods that are selected based on Fung, Hsieh, Naik and Ramadorai (2008). The first sub period is from January 1994 to December 1996. The second sub period is January 1997 - September 1998. The third sub period is October 1998 - March 2000. The fourth sub period is April 2000 - December 2004. The last sub period is January 2005 - December 2010. Panel F shows results of the equal-weight performance of the aggregate database in size groups. Funds are sorted into size groups every December based on monthly fund-level non-missing AUM observations (reported in millions of US dollars). "Missing AUM" portfolio consists of sorting of funds into portfolio every December based on the fund-level monthly missing AUM observations. This group includes funds that report at least one missing AUM observation. Equal-weight returns are calculated for each portfolio monthly, so the portfolio weights are adjusted monthly.

In each Panel, the first column describes the name of the database. The second column presents the number of funds in each portfolio. The third column presents the percentage amount of funds that are defunct. Annualized mean and standard deviation are reported in the next two columns. Values are reported in percentages. Sharpe ratio is annualized and defined as the average excess return divided by the standard deviation of return. Alpha is the measure of abnormal return estimated from the seven factor model proposed by Fung and Hsieh (2004). Alpha is annualized and reported in percentage. Risk factors are: the excess return of the S&P 500 index (SP-RF), the return of the Russell 2000 index minus the return of the S&P 500 index (RL-SP), the excess return of ten-year Treasuries (TY-RF), the return of Moody's BAA corporate bonds minus ten-year Treasuries (BAA-TY), the excess returns of look-back straddles on bonds (PTFSBD-RF), currencies (PTFSFX-RF), and commodities (PTFSCOM-RF). RSQ is the R-square of the model. Values of t-statistics are reported in parentheses.

A. Equal-weight

Dataset	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
TASS	8072	63.29	7.92	7.07	1.12	5.31	0.28	0.16	0.10	0.28	0.00	0.01	0.01	0.66
						(5.13)	(13.32)	(6.65)	(2.59)	(5.77)	-(0.55)	(2.61)	(1.74)	
Hedge Fund Research	9508	61.27	8.72	7.01	1.24	6.20	0.30	0.18	0.08	0.22	0.00	0.01	0.01	0.71
						(6.53)	(15.60)	(8.02)	(2.07)	(5.04)	-(0.29)	(2.46)	(1.56)	
BarclayHedge	9588	64.54	9.02	6.51	1.38	6.87	0.26	0.14	0.08	0.25	0.01	0.02	0.01	0.65
						(7.08)	(13.37)	(6.04)	(2.20)	(5.52)	(1.00)	(3.60)	(2.36)	
EurekaHedge	7009	38.71	10.28	7.57	1.35	7.61	0.31	0.16	0.10	0.29	0.00	0.01	0.01	0.66
						(6.85)	(13.66)	(6.18)	(2.25)	(5.65)	(0.15)	(2.80)	(1.95)	
Morningstar	6681	48.17	9.75	7.01	1.38	7.28	0.28	0.16	0.09	0.26	0.00	0.01	0.01	0.67
						(7.14)	(13.86)	(6.62)	(2.26)	(5.44)	(0.05)	(2.54)	(2.00)	
Aggregate	24768	57.03	8.45	7.11	1.19	5.87	0.29	0.16	0.10	0.28	0.00	0.01	0.01	0.69
						(5.83)	(14.32)	(6.56)	(2.48)	(6.01)	(0.04)	(2.76)	(1.88)	

B. Value-weight

Dataset	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
TASS	6644	65.32	8.56	7.24	1.18	5.97	0.25	0.16	0.15	0.18	-0.02	0.01	0.02	0.50
						(4.60)	(9.55)	(5.23)	(3.01)	(3.06)	-(2.13)	(1.70)	(1.79)	
Hedge Fund Research	7844	63.64	7.73	6.19	1.24	5.59	0.23	0.15	0.09	0.18	-0.01	0.01	0.01	0.58
						(5.48)	(11.44)	(5.96)	(2.20)	(3.73)	-(1.17)	(1.97)	(1.76)	
BarclayHedge	9193	64.68	6.91	5.68	1.21	5.01	0.20	0.11	0.12	0.20	0.00	0.01	0.02	0.53
						(5.07)	(10.14)	(4.71)	(3.14)	(4.34)	-(0.16)	(2.51)	(2.59)	
EurekaHedge	6310	38.92	7.72	6.86	1.12	5.48	0.23	0.11	0.11	0.30	0.00	0.01	0.02	0.53
						(4.62)	(9.82)	(3.79)	(2.30)	(5.37)	-(0.11)	(1.56)	(2.86)	
Morningstar	5695	49.53	8.07	6.14	1.31	6.04	0.21	0.15	0.12	0.20	0.00	0.01	0.02	0.54
						(5.73)	(10.09)	(5.76)	(2.93)	(4.11)	-(0.50)	(1.45)	(3.00)	
Aggregate	21587	58.39	7.36	6.61	1.11	4.88	0.25	0.13	0.12	0.24	-0.01	0.01	0.01	0.60
						(4.61)	(11.78)	(5.21)	(2.96)	(4.79)	-(1.40)	(2.07)	(2.03)	

C. Backfill-adjusted results of the average performance (Equal-weight)

Dataset	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
TASS	7963	63.23	7.40	7.30	1.01	4.32	0.28	0.17	0.10	0.28	0.00	0.01	0.01	0.68
						(3.96)	(13.35)	(6.66)	(2.38)	(5.82)	-(0.61)	(2.46)	(1.54)	
Hedge Fund Research	9384	61.11	8.08	7.37	1.09	5.06	0.31	0.19	0.08	0.23	0.00	0.01	0.01	0.73
						(4.95)	(15.55)	(7.99)	(1.98)	(5.15)	-(0.41)	(2.39)	(1.36)	
BarclayHedge	9464	64.41	7.91	6.80	1.16	5.13	0.27	0.15	0.10	0.26	0.00	0.02	0.01	0.69
						(5.10)	(13.99)	(6.38)	(2.66)	(5.85)	(0.50)	(3.75)	(2.22)	
EurekaHedge	6924	38.65	9.93	7.91	1.25	6.73	0.31	0.17	0.10	0.30	0.00	0.01	0.01	0.68
						(5.65)	(13.52)	(6.19)	(2.20)	(5.72)	(0.11)	(2.59)	(1.58)	
Morningstar	6635	48.21	9.23	7.24	1.27	6.28	0.29	0.17	0.09	0.27	0.00	0.01	0.01	0.68
						(5.79)	(13.61)	(6.62)	(2.18)	(5.58)	-(0.14)	(2.54)	(1.65)	
Aggregate	24420	56.91	7.74	7.45	1.04	4.57	0.30	0.16	0.11	0.29	0.00	0.01	0.01	0.71
						(4.29)	(14.52)	(6.67)	(2.58)	(6.19)	-(0.29)	(2.74)	(1.69)	

D. Smoothing-adjusted results of the average performance (Equal-weight)

Database	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
TASS	8072	63.29	7.92	9.31	0.85	5.42	0.29	0.15	0.11	0.24	0.00	0.01	0.01	0.68
						(4.01)	(12.92)	(7.21)	(2.22)	(6.46)	-(0.36)	(2.46)	(2.39)	
Hedge Fund Research	9508	61.27	8.72	9.10	0.95	6.27	0.31	0.17	0.09	0.19	0.00	0.01	0.01	0.73
						(5.19)	(16.88)	(9.12)	(1.96)	(5.30)	-(0.24)	(2.12)	(2.37)	
BarclayHedge	9588	64.54	9.02	8.04	1.12	6.89	0.27	0.13	0.09	0.23	0.01	0.01	0.02	0.66
						(6.08)	(12.80)	(6.43)	(1.95)	(6.18)	(0.87)	(3.59)	(2.83)	
EurekaHedge	7009	38.71	10.28	9.87	1.04	7.68	0.31	0.15	0.10	0.26	0.00	0.01	0.02	0.68
						(5.41)	(13.43)	(6.25)	(1.95)	(6.13)	(0.10)	(2.67)	(2.87)	
Morningstar	6681	48.17	9.75	8.95	1.08	7.35	0.29	0.15	0.09	0.23	0.00	0.01	0.02	0.69
						(5.89)	(13.51)	(6.74)	(2.04)	(6.19)	(0.11)	(2.43)	(2.69)	
Aggregate	24768	57.03	8.45	9.13	0.92	5.93	0.30	0.15	0.10	0.25	0.00	0.01	0.02	0.70
						(4.71)	(13.87)	(6.87)	(2.18)	(6.81)	(0.13)	(2.67)	(2.53)	

E. Performance of the aggregate database in sub periods (Equal-weight)

Sub period	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Jan 1994 - Dec 1996	3771	82.84	9.70	4.38	2.16	8.59	0.25	0.16	0.04	0.04	0.00	0.01	0.04	0.62
						(3.61)	(3.56)	(2.26)	(0.34)	(0.13)	(0.12)	(1.58)	(3.11)	
Jan 1997 - Sep 1998	4635	79.03	4.89	7.57	0.64	2.97	0.39	0.18	-0.30	0.25	0.02	0.01	0.04	0.88
						(1.00)	(6.77)	(2.67)	-(1.32)	(0.66)	(0.85)	(0.47)	(1.28)	
Oct 1998 - Mar 2000	5627	73.75	21.87	6.73	3.23	15.82	0.29	0.26	0.41	0.56	0.05	0.00	-0.02	0.88
						(4.35)	(4.23)	(5.85)	(1.77)	(2.12)	(1.78)	(0.15)	-(1.05)	
Apr 2000 - Dec 2004	12319	68.71	7.81	5.64	1.42	5.71	0.28	0.17	0.23	0.16	0.00	0.03	0.01	0.88
						(5.23)	(12.22)	(6.65)	(6.16)	(2.14)	(0.47)	(5.15)	(1.14)	
Jan 2005 - Dec 2010	20432	47.94	6.03	8.85	0.68	5.66	0.37	-0.09	-0.02	0.25	0.01	0.00	0.02	0.73
						(2.50)	(8.05)	-(1.24)	-(0.23)	(3.51)	(0.43)	-(0.14)	(1.48)	

F. Performance of the aggregate database in size groups (Equal-weight)

Size group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Missing AUM observations	9110	54.31	9.37	8.22	1.14	5.91	0.32	0.17	0.12	0.32	-0.01	0.01	0.01	0.67
						(4.72)	(13.02)	(5.93)	(2.35)	(5.80)	-(0.70)	(1.68)	(1.27)	
$0M <= AUM <= $10M	10152	65.57	9.89	7.03	1.41	7.25	0.28	0.17	0.10	0.24	0.01	0.02	0.02	0.67
						(6.80)	(13.65)	(6.80)	(2.37)	(5.02)	(1.38)	(3.87)	(2.32)	
$10M < AUM <= $50M	10532	59.60	7.81	7.37	1.06	4.71	0.31	0.18	0.09	0.25	0.00	0.01	0.01	0.73
						(4.63)	(15.54)	(7.78)	(2.25)	(5.54)	-(0.55)	(2.88)	(1.83)	
$50M < AUM <= $250M	7750	52.58	5.73	7.35	0.78	2.48	0.29	0.15	0.11	0.31	-0.01	0.01	0.01	0.70
						(2.32)	(13.86)	(6.05)	(2.68)	(6.45)	-(1.13)	(1.98)	(1.90)	
$250M < AUM <= $500M	2521	44.03	4.84	7.32	0.66	1.59	0.26	0.13	0.13	0.34	-0.01	0.01	0.01	0.63
						(1.34)	(11.21)	(4.63)	(2.90)	(6.49)	-(1.33)	(1.50)	(1.37)	
$500M < AUM <= $1000M	1313	41.20	5.84	6.83	0.85	2.80	0.24	0.09	0.12	0.29	-0.01	0.01	0.01	0.58
						(2.39)	(10.66)	(3.25)	(2.61)	(5.48)	-(1.91)	(1.92)	(1.18)	
AUM > $1000M	626	42.81	5.45	8.75	0.62	1.58	0.30	0.10	0.18	0.30	-0.02	0.01	0.01	0.50
						(0.97)	(9.47)	(2.56)	(2.75)	(4.16)	-(2.35)	(1.20)	(1.27)	

Table 4

Average performance in size groups [January 1994 - December 2010]

Table shows results of the average performance in size groups. Results are reported for each database. Aggregate database is a merged database including TASS, Hedge Fund Research, BarclayHedge, EurekaHedge, and Morningstar. Databases are merged using a novel statistical procedure that is described in Appendix of the paper. Single databases are removed from multiple share classes using the statistical algorithm. All funds are required to have at least 12 monthly return observations. In panel A, funds are sorted into size groups every December based on monthly fund-level non-missing AUM observations (reported in millions of US dollars). "Missing AUM" portfolio consists of sorting of funds into portfolio every December based on the fund-level monthly missing AUM observations. This group includes funds that report at least one missing AUM observation. In panel B, three subsamples are created for each database. The first subsample consists of funds that do not report AUM observations. The second subsample consists of all fund-level AUM observations that are missing in the beginning of the time series (conditional on non-missing returns) until the first reported non-missing AUM observation appears. The third subperiod consists of all fund-level missing AUM observations that are reported after the first reported AUM observation. All portfolios are equal-weighted monthly.

In Panel A and B, column "# funds" is the number of funds that are included in each portfolio. "Dead %" column tells the amount of dead funds in percentage. Columns "Mean ER" and "Std ER" are the annualized mean and standard deviation of excess returns (in percentages). "Alpha" is the measure of abnormal return estimated from the seven factor model proposed by Fung and Hsieh (2004). Alpha is annualized and reported in percentage. Risk factors are: the excess return of the S&P 500 index (SP-RF), the return of the Russell 2000 index minus the return of the S&P 500 index (RL-SP), the excess return of ten-year Treasuries (TY-RF), the return of Moody's BAA corporate bonds minus ten-year Treasuries (BAA-TY), the excess returns of look-back straddles on bonds (PTFSBD-RF), currencies (PTFSFX-RF), and commodities (PTFSCOM-RF). RSQ is the R-square of the model. Values of t-statistics are reported in parentheses. In Panel B, "Obs" is the number of excess returns in each portfolio and Sharpe ratio (Column "Sharpe") is annualized and defined as the average excess return divided by the standard deviation of return.

A. Average performance in size groups

Size group	TASS # funds	TASS Dead %	TASS Mean ER	TASS Std ER	TASS Alpha	Hedge Fund Research # funds	Hedge Fund Research Dead %	Hedge Fund Research Mean ER	Hedge Fund Research Std ER	Hedge Fund Research Alpha	BarclayHedge # funds	BarclayHedge Dead %	BarclayHedge Mean ER	BarclayHedge Std ER	BarclayHedge Alpha
Missing AUM observations	3804	60.17	9.08	7.87	5.87	2485	63.06	9.80	7.83	6.49	1558	62.20	9.77	8.48	6.26
					(4.49)					(5.30)					(5.30)
$0M <= AUM <= $10M	3138	70.94	9.35	7.48	6.53	4016	67.06	11.25	7.68	8.41	4947	69.90	10.53	6.39	8.46
					(5.70)					(7.79)					(7.83)
$10M < AUM <= $50M	3507	65.10	7.49	7.24	4.55	4723	62.95	8.07	7.16	5.25	4710	62.97	8.60	6.87	5.82
					(4.35)					(5.42)					(5.96)
$50M < AUM <= $250M	2683	60.23	5.76	7.05	2.82	3419	55.60	5.95	7.24	2.86	3602	56.91	5.93	6.92	2.85
					(2.64)					(2.73)					(2.75)
$250M < AUM <= $500M	851	53.47	4.59	7.65	1.52	1116	43.55	5.02	7.03	2.04	1246	49.28	5.45	6.88	2.34
					(1.19)					(1.80)					(1.92)
$500M < AUM <= $1000M	458	53.93	6.64	8.32	3.20	579	37.82	5.14	7.00	2.25	680	45.74	5.10	6.44	2.59
					(2.03)					(1.81)					(2.16)
AUM > $1000M	205	60.49	6.27	11.43	2.14	269	31.97	5.99	7.84	2.81	290	43.10	4.40	6.07	1.88
					(0.84)					(1.91)					(1.66)

Size group	EurekaHedge # funds	EurekaHedge Dead %	EurekaHedge Mean ER	EurekaHedge Std ER	EurekaHedge Alpha	Morningstar # funds	Morningstar Dead %	Morningstar Mean ER	Morningstar Std ER	Morningstar Alpha	Aggregate # funds	Aggregate Dead %	Aggregate Mean ER	Aggregate Std ER	Aggregate Alpha
Missing AUM observations	3780	38.78	10.87	8.03	7.55	3260	49.33	8.84	7.34	5.69	9110	54.31	9.37	8.22	5.91
					(6.29)					(4.66)					(4.72)
$0M <= AUM <= $10M	2149	42.86	12.98	7.95	10.28	2581	51.38	13.28	7.86	10.52	10152	65.57	9.89	7.03	7.25
					(7.59)					(9.00)					(6.80)
$10M < AUM <= $50M	2931	40.09	9.50	9.08	6.50	2999	49.68	10.01	7.36	7.08	10532	59.60	7.81	7.37	4.71
					(4.11)					(6.73)					(4.63)
$50M < AUM <= $250M	2488	36.70	8.21	7.76	4.92	2261	45.78	7.36	7.47	4.18	7750	52.58	5.73	7.35	2.48
					(3.71)					(3.68)					(2.32)
$250M < AUM <= $500M	926	29.81	7.52	8.73	4.52	761	41.52	6.56	7.48	4.01	2521	44.03	4.84	7.32	1.59
					(2.56)					(2.74)					(1.34)
$500M < AUM <= $1000M	542	29.34	5.85	7.54	2.91	404	37.87	4.80	7.57	2.08	1313	41.20	5.84	6.83	2.80
					(1.92)					(1.34)					(2.39)
AUM > $1000M	249	23.69	3.11	8.28	1.36	197	42.64	5.66	7.31	2.98	626	42.81	5.45	8.75	1.58
					(0.83)					(2.08)					(0.97)

B. Average performance conditional on reporting of missing AUM observations

	TASS							Hedge Fund Research						
Category	Obs	# funds	Dead %	Mean ER	Std ER	Sharpe	Alpha	Obs	# funds	Dead %	Mean ER	Std ER	Sharpe	Alpha
No AUM reported	204	1435	53.59	9.44	8.76	1.07	6.63	204	1666	50.06	9.93	8.04	1.23	6.96
							(4.00)							(5.66)
Missing AUM in the beginning	203	2181	59.70	10.93	8.04	1.35	8.07	201	927	85.87	10.79	5.70	1.88	8.93
							(6.06)							(9.93)
Missing AUM in the middle-end	203	2972	72.07	3.48	7.58	0.46	0.93	199	220	95.00	9.49	12.74	0.74	8.29
							(0.82)							(2.61)

	BarclayHedge							EurekaHedge						
Category	Obs	# funds	Dead %	Mean ER	Std ER	Sharpe	Alpha	Obs	# funds	Dead %	Mean ER	Std ER	Sharpe	Alpha
No AUM reported	204	427	57.14	9.44	8.41	1.12	6.64	204	784	33.16	9.94	7.06	1.40	7.33
							(5.34)							(6.75)
Missing AUM in the beginning	203	1939	72.98	11.71	8.14	1.44	8.93	203	3474	41.02	11.62	7.72	1.50	9.09
							(7.77)							(8.07)
Missing AUM in the middle-end	199	410	79.51	6.28	9.50	0.66	2.47	187	576	43.58	5.27	16.93	0.31	2.99
							(1.54)							(0.80)

	Morningstar							Aggregate						
Category	Obs	# funds	Dead %	Mean ER	Std ER	Sharpe	Alpha	Obs	# funds	Dead %	Mean ER	Std ER	Sharpe	Alpha
No AUM reported	204	1036	38.42	8.43	7.58	1.11	5.75	204	3399	44.81	8.95	7.79	1.14	6.11
							(4.31)							(5.03)
Missing AUM in the beginning	203	1705	52.73	9.65	6.88	1.40	7.31	203	6472	58.27	10.95	8.04	1.36	8.07
							(6.27)							(6.79)
Missing AUM in the middle-end	201	2697	51.72	9.00	8.05	1.11	6.70	203	4189	63.98	4.66	7.76	0.60	2.02
							(4.69)							(1.65)

Table 5

Average performance of the aggregate database by investment objective [January 1994 - December 2010]

Table shows results of the average performance grouped by the investment objective. Results are reported for the aggregate database including TASS, Hedge Fund Research, BarclayHedge, EurekaHedge, and Morningstar. Databases are merged using a statistical procedure that is described in Appendix of the paper. All funds are required to have at least 12 monthly return observations. Panel A shows results of the average performance where hedge funds are categorized based on the investment objectives. Classification of strategies is provided in Appendix of the paper. In Panel B, hedge funds are sorted into terciles each December based on monthly AuM observations and results of the average performance are reported for each size tercile and investment objective. In Table, the second column describes the number of funds in each portfolio and Column "% of Dead" tells the percentage amount of defunct funds if compared to all funds. The next two columns include the annualized average excess return and the standard deviation of excess returns. Sharpe ratio is annualized and defined as the average excess return divided by the standard deviation of return. Alpha is the measure of abnormal return estimated from the seven factor model proposed by Fung and Hsieh (2004). Alpha is annualized and reported in percentage. Risk factors are: the excess return of the S&P 500 index (SP-RF), the return of the Russell 2000 index minus the return of the S&P 500 index (RL-SP), the excess return of ten-year Treasuries (TY-RF), the return of Moody's BAA corporate bonds minus ten-year Treasuries (BAA-TY), the excess returns of look-back straddles on bonds (PTFSBD-RF), currencies (PTFSFX-RF), and commodities (PTFSCOM-RF). RSQ is the R-square of the model. Values of t-statistics are reported in parentheses.

A. Performance of the aggregate database by the investment objective (Equal-weight)

Investment objective	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
All	24768	57.03	8.45	7.11	1.19	5.87	0.29	0.16	0.10	0.28	0.00	0.01	0.01	0.69
						(5.83)	(14.32)	(6.56)	(2.48)	(6.01)	(0.04)	(2.76)	(1.88)	
CTA	2413	59.22	7.59	6.87	1.11	6.74	0.09	0.03	0.21	0.17	0.02	0.04	0.03	0.34
						(4.78)	(3.13)	(0.74)	(3.76)	(2.65)	(1.87)	(5.70)	(3.13)	
Emerging Markets	2724	33.11	11.79	15.72	0.75	7.06	0.53	0.23	-0.01	0.46	-0.02	0.00	0.00	0.49
						(2.50)	(9.30)	(3.42)	-(0.08)	(3.52)	-(1.32)	(0.30)	-(0.07)	
Event Driven	1363	63.98	8.15	6.37	1.28	5.60	0.23	0.14	0.02	0.31	-0.02	0.00	0.00	0.75
						(6.98)	(14.03)	(7.54)	(0.77)	(8.25)	-(3.42)	(1.27)	-(0.41)	
Global Macro	1601	61.02	6.82	5.30	1.28	5.10	0.19	0.07	0.16	0.15	0.00	0.01	0.02	0.46
						(5.18)	(9.39)	(3.17)	(4.20)	(3.25)	(0.23)	(2.84)	(2.38)	
Long Only	473	31.08	7.28	12.96	0.56	2.64	0.60	0.30	0.05	0.39	0.00	0.01	0.00	0.79
						(1.76)	(19.78)	(8.44)	(0.90)	(5.53)	(0.15)	(0.88)	(0.34)	
Long/Short	6991	59.30	9.61	9.90	0.97	6.20	0.45	0.30	0.06	0.18	0.00	0.01	0.01	0.76
						(5.02)	(18.24)	(10.22)	(1.29)	(3.10)	-(0.47)	(1.24)	(0.94)	
Market Neutral	1354	63.88	5.80	4.20	1.36	4.28	0.12	0.04	0.11	0.17	-0.01	0.01	0.00	0.41
						(5.25)	(7.57)	(1.98)	(3.43)	(4.42)	-(1.47)	(2.15)	(0.84)	
Multi-Strategy	3480	55.57	8.04	7.27	1.10	6.77	0.13	0.06	0.18	0.32	0.02	0.02	0.04	0.34
						(4.54)	(4.32)	(1.55)	(3.18)	(4.56)	(2.43)	(3.44)	(3.69)	
Others	1084	82.38	7.96	6.62	1.20	5.51	0.26	0.16	0.11	0.20	-0.01	0.01	0.01	0.63
						(5.37)	(12.67)	(6.64)	(2.81)	(4.21)	-(0.98)	(1.62)	(1.08)	
Relative Value	2522	58.29	6.27	4.92	1.28	4.07	0.12	0.05	0.14	0.38	-0.01	0.00	0.00	0.65
						(5.51)	(8.26)	(2.97)	(4.95)	(11.04)	-(1.83)	(0.57)	(0.30)	
Sector	648	62.19	12.65	12.96	0.97	8.74	0.56	0.44	0.05	0.10	0.00	0.01	0.00	0.67
						(4.65)	(14.66)	(9.75)	(0.70)	(1.17)	-(0.28)	(0.61)	(0.34)	
Short Bias	115	74.78	2.14	11.86	0.18	4.93	-0.52	-0.40	-0.03	0.26	0.00	0.01	0.01	0.58
						(2.53)	-(13.17)	-(8.54)	-(0.44)	(2.83)	-(0.08)	(0.68)	(0.82)	

B. Average performance of the investment objectives in size groups (aggregate data)

CTA

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	1464	63.39	8.73	7.00	1.25	7.56	0.08	0.02	0.20	0.16	0.02	0.04	0.02	0.31
						(4.91)	(2.82)	(0.62)	(3.30)	(2.34)	(2.29)	(5.31)	(2.11)	
Med	911	59.60	5.28	6.87	0.77	3.95	0.08	0.04	0.21	0.13	0.01	0.04	0.03	0.36
						(2.72)	(2.67)	(1.30)	(3.70)	(2.05)	(1.44)	(6.36)	(2.74)	
Large	528	47.73	3.50	7.13	0.49	2.12	0.07	0.01	0.21	0.14	0.01	0.04	0.02	0.27
						(1.31)	(2.24)	(0.32)	(3.33)	(1.91)	(0.78)	(4.98)	(2.43)	

Emerging markets

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	998	39.08	16.40	18.65	0.88	11.88	0.58	0.25	-0.07	0.59	0.02	0.01	0.01	0.44
						(3.22)	(8.09)	(2.94)	-(0.51)	(3.58)	(0.80)	(0.38)	(0.55)	
Med	1340	37.01	11.86	15.96	0.74	6.55	0.53	0.23	0.01	0.46	-0.03	0.00	0.00	0.50
						(2.18)	(8.97)	(3.34)	(0.09)	(3.42)	-(1.57)	(0.28)	(0.03)	
Large	849	30.86	6.50	16.14	0.40	0.86	0.47	0.22	0.03	0.56	-0.04	0.00	0.01	0.48
						(0.28)	(7.84)	(3.07)	(0.23)	(4.10)	-(2.37)	(0.04)	(0.29)	

Event driven

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	470	69.79	11.54	8.57	1.35	8.55	0.30	0.22	0.04	0.20	-0.01	0.00	-0.01	0.58
						(5.79)	(10.37)	(6.58)	(0.75)	(3.06)	-(1.08)	(0.42)	-(0.88)	
Med	687	62.88	8.32	7.25	1.15	5.23	0.25	0.16	0.01	0.34	-0.02	0.01	-0.01	0.77
						(5.65)	(14.04)	(7.64)	(0.37)	(8.17)	-(3.42)	(1.53)	-(1.09)	
Large	634	60.88	6.58	6.33	1.04	3.91	0.18	0.12	0.01	0.34	-0.02	0.00	0.00	0.70
						(4.30)	(10.41)	(5.75)	(0.15)	(8.32)	-(4.18)	(0.55)	(0.28)	

Global macro

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	858	63.64	6.65	5.58	1.19	5.14	0.16	0.07	0.08	0.17	0.01	0.00	0.03	0.38
						(4.41)	(7.17)	(2.52)	(1.75)	(3.27)	(1.09)	(0.44)	(3.48)	
Med	567	60.85	4.97	5.87	0.85	2.71	0.16	0.09	0.20	0.20	0.00	0.02	0.01	0.42
						(2.29)	(7.00)	(3.33)	(4.29)	(3.78)	(0.26)	(4.45)	(1.99)	
Large	408	54.41	5.40	5.88	0.91	2.67	0.18	0.09	0.23	0.14	-0.01	0.02	0.01	0.41
						(2.24)	(7.65)	(3.43)	(4.86)	(2.66)	-(1.98)	(3.14)	(1.55)	

Long only

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	144	36.81	11.35	13.77	0.82	8.14	0.58	0.19	-0.03	0.38	0.03	0.01	-0.01	0.67
						(3.81)	(14.00)	(3.94)	-(0.36)	(3.98)	(2.04)	(0.98)	-(0.85)	
Med	197	32.49	8.48	14.58	0.58	2.77	0.66	0.32	0.09	0.39	0.00	0.01	0.00	0.77
						(1.51)	(18.56)	(7.56)	(1.19)	(4.74)	-(0.22)	(0.67)	(0.40)	
Large	186	25.81	7.04	13.30	0.53	1.48	0.59	0.27	0.11	0.41	-0.01	0.00	0.01	0.77
						(0.88)	(17.94)	(7.04)	(1.66)	(5.53)	-(1.20)	(0.22)	(0.51)	

Long/Short

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	3008	62.77	11.44	11.29	1.01	7.46	0.52	0.35	0.01	0.14	0.00	0.01	0.00	0.79
						(5.41)	(19.52)	(11.09)	(0.16)	(2.34)	(0.07)	(1.46)	(0.33)	
Med	3245	59.11	9.27	10.10	0.92	5.57	0.46	0.32	0.03	0.16	0.00	0.01	0.00	0.80
						(4.60)	(19.69)	(11.35)	(0.72)	(2.98)	(0.05)	(1.17)	(0.32)	
Large	2183	54.10	6.78	10.34	0.65	2.81	0.46	0.29	0.06	0.17	-0.01	0.01	0.01	0.73
						(1.98)	(16.61)	(9.00)	(1.08)	(2.63)	-(1.07)	(0.90)	(1.10)	

Market neutral

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	581	68.85	8.39	5.46	1.51	7.07	0.08	0.04	0.09	0.11	-0.01	0.01	0.01	0.14
						(5.27)	(3.23)	(1.18)	(1.69)	(1.83)	-(0.95)	(1.76)	(1.76)	
Med	643	69.05	4.71	4.11	1.12	3.15	0.11	0.03	0.08	0.17	0.00	0.01	0.00	0.38
						(3.68)	(6.78)	(1.47)	(2.50)	(4.48)	-(0.81)	(2.40)	(0.58)	
Large	450	60.22	3.60	4.02	0.87	2.12	0.14	-0.02	0.08	0.10	-0.01	0.00	0.01	0.38
						(2.54)	(8.29)	-(1.24)	(2.43)	(2.82)	-(1.24)	(1.27)	(1.19)	

Multi-strategy

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	1673	61.98	8.69	8.78	0.99	7.30	0.10	0.07	0.23	0.22	0.03	0.04	0.05	0.34
						(3.87)	(2.77)	(1.70)	(3.08)	(2.65)	(2.26)	(4.31)	(4.34)	
Med	1413	55.56	7.30	7.85	0.93	5.49	0.15	0.08	0.24	0.22	0.02	0.03	0.04	0.37
						(3.32)	(4.73)	(2.10)	(3.65)	(3.04)	(2.28)	(3.94)	(3.98)	
Large	965	48.70	6.27	7.21	0.86	4.58	0.13	0.04	0.22	0.29	0.02	0.02	0.03	0.33
						(2.92)	(4.18)	(1.05)	(3.51)	(4.17)	(2.55)	(3.07)	(2.85)	

Others

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	387	84.24	9.63	8.27	1.17	6.32	0.30	0.20	0.11	0.22	-0.01	0.02	0.00	0.56
						(4.38)	(10.70)	(5.89)	(1.99)	(3.46)	-(1.22)	(2.35)	-(0.11)	
Med	414	78.99	6.77	8.24	0.82	3.62	0.34	0.20	0.07	0.21	0.00	0.01	0.01	0.65
						(2.81)	(13.38)	(6.67)	(1.45)	(3.72)	-(0.52)	(1.29)	(0.87)	
Large	270	82.96	5.66	5.90	0.95	3.32	0.18	0.16	0.10	0.19	-0.01	0.00	0.01	0.54
						(3.15)	(8.96)	(6.47)	(2.41)	(4.14)	-(1.50)	-(0.33)	(1.68)	

Relative value

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	942	59.45	8.38	5.04	1.67	6.08	0.13	0.07	0.16	0.35	0.00	0.00	0.00	0.60
						(7.17)	(7.71)	(3.74)	(4.96)	(9.21)	-(0.09)	(0.85)	(0.45)	
Med	1153	59.58	5.57	5.02	1.11	3.23	0.12	0.06	0.11	0.38	-0.01	0.00	0.00	0.66
						(4.15)	(8.05)	(3.10)	(3.54)	(10.83)	-(2.12)	(0.46)	(0.10)	
Large	986	57.30	4.57	5.68	0.80	2.05	0.11	0.03	0.12	0.46	-0.01	0.00	0.00	0.62
						(2.22)	(6.04)	(1.22)	(3.30)	(11.09)	-(2.37)	-(0.94)	(0.26)	

Sector

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	321	67.60	11.91	16.66	0.71	6.77	0.62	0.51	0.04	0.15	-0.01	0.00	0.00	0.57
						(2.35)	(11.12)	(7.62)	(0.37)	(1.18)	-(0.58)	-(0.27)	(0.04)	
Med	361	60.94	12.68	14.01	0.90	8.51	0.63	0.45	0.01	0.01	0.01	0.00	0.00	0.69
						(4.11)	(15.60)	(9.43)	(0.08)	(0.12)	(0.80)	(0.39)	-(0.10)	
Large	218	48.62	7.90	13.01	0.61	3.64	0.51	0.45	0.07	0.16	0.00	0.00	0.01	0.65
						(1.78)	(12.81)	(9.53)	(0.85)	(1.73)	-(0.06)	(0.34)	(0.62)	

Short bias

Group	No. Of Funds	% of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	60	70.00	0.43	16.01	0.03	3.58	-0.72	-0.39	0.18	0.35	-0.01	-0.01	0.02	0.56
						(1.27)	-(13.20)	-(5.96)	(1.62)	(2.79)	-(0.67)	-(0.66)	(0.90)	
Med	74	70.27	6.58	17.43	0.38	11.30	-0.64	-0.59	-0.10	0.04	-0.01	0.01	0.02	0.53
						(3.59)	-(10.45)	-(8.09)	-(0.80)	(0.27)	-(0.32)	(0.61)	(0.77)	
Large	33	72.73	-3.53	14.70	-0.24	-0.73	-0.70	-0.45	0.18	0.49	-0.01	0.00	0.00	0.64
						-(0.31)	-(15.19)	-(8.28)	(1.96)	(4.66)	-(0.37)	(0.36)	-(0.24)	

Table 6

Average performance of the aggregate database by domicile region [January 1994 - December 2010]

Table shows results of the average performance grouped by the domicile region. Panel A shows the number of unique hedge funds divided by the manager firm domicile. Panel B provides results of the average performance where funds are grouped based on the fund-level domicile region. Panel C shows the results of the average performance where funds are grouped based on the firm domicile region. The domicile regions are divided to two groups: (1) onshore; (2) offshore. United States and Canada are classified as onshore regions. Other domicile regions are classified as offshore and they are further divided to four sub categories: (1) Asia and Pacific; (2) Caribbean; (3) Europe; (4) Rest of World. In Panel A, "# of firms" is the number of unique management firms in each firm domicile group. Numbers in Panel A are calculated using Hedge Fund Research, BarclayHedge, and EurekaHedge databases, since only these databases report information on the domicile of the management firm. In Panel A, within each firm domicile, the number of unique hedge funds are reported in each fund domicile group. In Panel B, the second column describes the number of funds in each portfolio and Column "% of Dead" tells the percentage amount of defunct funds if compared to all funds. The next two columns include the annualized average excess return and the standard deviation of excess returns. Sharpe ratio is annualized and defined as the average excess return divided by the standard deviation of return. Alpha is the measure of abnormal return estimated from the seven factor model proposed by Fung and Hsieh (2004). Alpha is annualized and reported in percentage. Risk factors are: the excess return of the S&P 500 index (SP-RF), the return of the Russell 2000 index minus the return of the S&P 500 index (RL-SP), the excess return of ten-year Treasuries (TY-RF), the return of Moody's BAA corporate bonds minus ten-year Treasuries (BAA-TY), the excess returns of look-back straddles on bonds (PTFSBD-RF), currencies (PTFSFX-RF), and commodities (PTFSCOM-RF). RSQ is the R-square of the model. Values of t-statistics are reported in parentheses. In Panel B, results are reported for the aggregate database including TASS, Hedge Fund Research, BarclayHedge, EurekaHedge, and Morningstar. Panel C shows results of the average performance grouped by (i) the management company domicile region; (ii) the firm city. BarclayHedge is the only database that provides information on the city of the management company.

A. Number of hedge funds in management firm and fund domicile groups

Firm domicile		Fund domicile (total number of funds = 15,805)				
Region	# of firms	Onshore	Asia and Pacific	Caribbean	Europe	Rest of world
Onshore	4229	5846	4	1923	101	854
Asia and Pacific	565	24	240	760	43	44
Caribbean	212	39	.	370	9	11
Europe	1317	54	5	2237	2274	251
Rest of world	325	38	.	268	36	374
Sum	6648	6001	249	5558	2463	1534

B. Average performance of the aggregate database by fund domicile (Equal-weight)

Domicile	No. Of Funds	% Of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
All	24768	57.03	8.45	7.11	1.19	5.87	0.29	0.16	0.10	0.28	0.00	0.01	0.01	0.69
						(5.83)	(14.32)	(6.56)	(2.48)	(6.01)	(0.04)	(2.76)	(1.88)	
Onshore	8202	64.61	9.52	6.56	1.44	7.45	0.29	0.19	0.03	0.17	0.00	0.01	0.01	0.73
						(8.68)	(16.87)	(9.03)	(0.89)	(4.26)	(0.86)	(3.47)	(2.31)	
Offshore (all)	16566	53.27	7.58	7.53	1.01	4.74	0.28	0.14	0.13	0.34	0.00	0.01	0.01	0.63
						(4.12)	(12.20)	(5.16)	(2.80)	(6.29)	-(0.62)	(2.31)	(1.55)	
Caribbean	9304	59.18	7.29	7.25	1.01	4.61	0.26	0.15	0.10	0.30	-0.01	0.01	0.01	0.61
						(4.05)	(11.48)	(5.64)	(2.27)	(5.68)	-(1.21)	(1.94)	(1.27)	
Asia and Pacific	730	29.59	11.19	10.47	1.07	8.16	0.28	0.06	0.15	0.55	0.01	0.01	0.01	0.40
						(3.99)	(6.69)	(1.31)	(1.89)	(5.79)	(0.81)	(1.11)	(0.97)	
Europe	4150	41.35	6.18	8.56	0.72	3.21	0.30	0.12	0.19	0.41	0.01	0.02	0.02	0.55
						(2.22)	(10.26)	(3.43)	(3.33)	(6.10)	(0.80)	(2.71)	(1.81)	
Rest of world	2382	58.23	9.49	7.99	1.19	6.57	0.30	0.14	0.09	0.35	-0.01	0.01	0.01	0.62
						(5.29)	(11.88)	(4.58)	(1.95)	(6.15)	-(0.77)	(2.09)	(1.11)	

C. Average performance of the aggregate database by firm domicile (Equal-weight)

Firm domicile region (Aggregate database):

Domicile	No. Of Funds	% Of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
All	15935	58.17	8.80	6.97	1.26	6.33	0.29	0.16	0.09	0.26	0.00	0.01	0.01	0.69
						(6.50)	(14.80)	(6.77)	(2.25)	(5.82)	(0.37)	(2.98)	(1.92)	
Onshore	8794	67.05	8.84	6.08	1.45	6.88	0.26	0.17	0.04	0.16	0.00	0.01	0.01	0.71
						(8.29)	(15.85)	(8.42)	(1.27)	(4.24)	(0.58)	(3.45)	(2.07)	
Offshore (all)	7141	47.25	8.12	8.39	0.97	5.14	0.31	0.14	0.11	0.38	0.00	0.01	0.01	0.60
						(3.84)	(11.46)	(4.35)	(2.16)	(6.06)	-(0.21)	(2.20)	(1.38)	
Caribbean	435	62.53	11.05	8.59	1.28	8.57	0.28	0.13	0.06	0.25	-0.01	0.00	0.01	0.42
						(5.21)	(8.34)	(3.37)	(0.96)	(3.32)	-(0.64)	(0.64)	(1.19)	
Asia and Pacific	1124	46.44	8.43	10.03	0.84	5.67	0.36	0.16	0.07	0.32	0.01	0.01	0.01	0.49
						(3.14)	(9.97)	(3.67)	(1.04)	(3.77)	(1.38)	(1.20)	(0.77)	
Europe	4861	47.07	6.88	7.98	0.86	4.07	0.28	0.13	0.15	0.38	0.00	0.02	0.01	0.57
						(3.07)	(10.40)	(4.15)	(2.87)	(6.16)	(0.17)	(3.06)	(1.57)	
Rest of world	721	40.50	11.75	12.82	0.92	7.44	0.42	0.16	0.02	0.51	-0.03	0.00	0.00	0.55
						(3.43)	(9.64)	(3.11)	(0.24)	(5.03)	-(2.07)	-(0.16)	(0.31)	

Firm City (BarclayHedge):

Domicile	No. Of Funds	% Of Dead	Mean ER % pa	Std ER % pa	Sharpe (pa)	Alpha % pa	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
New York	1650	66.85	8.71	6.65	1.30	6.33	0.30	0.19	0.02	0.16	-0.01	0.01	0.00	0.78
						(8.09)	(19.04)	(10.28)	(0.61)	(4.51)	-(1.29)	(2.58)	(0.69)	
London	1435	59.30	8.78	8.30	1.06	6.16	0.27	0.12	0.08	0.35	0.00	0.02	0.01	0.50
						(4.14)	(9.16)	(3.45)	(1.43)	(5.09)	-(0.11)	(2.22)	(1.07)	
San Franciso	197	70.56	10.24	8.87	1.15	8.06	0.35	0.30	-0.06	0.10	0.00	0.01	0.01	0.64
						(5.96)	(13.01)	(9.18)	-(1.16)	(1.53)	(0.52)	(0.83)	(0.73)	

Table 7

Results of persistence for holding periods [January 1994 - December 2010]

Hedge funds are sorted into portfolios on (1) monthly; (2) quarterly; (3) Semiannually; and (4) december each year based on their t-statistics of alphas estimated using the seven factor model of Fung and Hsieh (2004). Databases are merged and cleaned from multiple share classes using a novel statistical procedure (see Appendix). The t-statistics of alphas are estimated using the most recent 24 months of returns preceding the evaluation period. Portfolio returns are calculated for equal- and value-weight portfolios monthly, so the weights are readjusted whenever a fund disappears. Results of persistence are reported for bottom, top and spread portfolios. Table shows annualized alphas in percentages and t-values in parentheses. The column "DO (%)" shows the drop out rate (in percentage) for each portfolio describing the average number of funds that drop from each portfolio during the specified holding period. For spread portfolios, the column "Do (%) (diff) contains the difference in drop out rates between the bottom and the top portfolio. The positive number suggests the higher drop out rate for the bottom portfolio if compared to the top portfolio. Risk factors are: the excess return of the S&P 500 index (SP-RF), the return of the Russell 2000 index minus the return of the S&P 500 index (RL-SP), the excess return of ten-year Treasuries (TY-RF), the return of Moody's BAA corporate bonds minus ten-year Treasuries (BAA-TY), the excess returns of look-back straddles on bonds (PTFSBD-RF), currencies (PTFSFX-RF), and commodities (PTFSCOM-RF). Panel A shows the results of the equal-weight persistence portfolios and Panel B shows the results of the value-weight portfolios. Panel C shows the results of the equal-weight persistence portfolios for the aggregate data base in the size terciles. First, funds are sorted into size terciles based on monthly AUM observations. Second, within size terciles, funds are sorted into deciles based on t-statistics of alphas.

A. Equal-weight

Bottom

Database	1-month		3-month		6-month		12-month	
	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)
TASS	1.91	3.99	2.17	10.97	2.07	19.12	2.79	31.87
	(1.26)		(1.38)		(1.36)		(1.78)	
Hedge Fund Research	1.43	3.81	2.83	10.70	2.84	18.45	4.17	30.45
	(0.98)		(1.84)		(1.98)		(3.09)	
BarclayHedge	2.78	3.68	3.99	10.15	3.52	17.78	4.49	29.08
	(2.14)		(2.78)		(2.51)		(3.37)	
EurekaHedge	6.87	3.15	8.41	8.74	7.72	15.45	9.54	27.03
	(3.66)		(4.41)		(4.16)		(5.21)	
Morningstar	5.54	2.71	7.17	7.96	6.40	15.54	7.58	27.02
	(3.38)		(3.97)		(3.80)		(4.71)	
Aggregate	1.23	3.75	2.14	10.44	1.85	18.49	3.05	31.13
	(0.90)		(1.44)		(1.32)		(2.24)	

Top

Database	1-month		3-month		6-month		12-month	
	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)
TASS	6.19	1.38	6.00	3.96	5.67	8.18	5.22	16.60
	(6.22)		(6.21)		(5.58)		(5.20)	
Hedge Fund Research	7.27	1.54	6.88	4.38	6.40	8.76	5.55	16.66
	(7.89)		(7.83)		(7.31)		(6.47)	
BarclayHedge	7.12	1.49	7.13	4.31	6.94	8.59	5.82	16.47
	(7.27)		(7.36)		(6.83)		(5.46)	
EurekaHedge	7.56	1.94	7.55	5.43	7.20	10.55	5.74	20.07
	(5.99)		(6.22)		(5.30)		(4.17)	
Morningstar	7.18	2.07	7.02	6.09	6.42	11.96	5.57	22.83
	(7.91)		(8.08)		(7.14)		(6.13)	
Aggregate	7.03	1.58	7.08	4.48	6.84	9.06	5.89	17.51
	(7.15)		(7.31)		(6.74)		(5.84)	

Spread

Database	1-month		3-month		6-month		12-month	
	Alpha % pa	DO (%) (diff)	Alpha % pa	DO (%) (diff)	Alpha % pa	DO (%) (diff)	Alpha % pa	DO (%) (diff)
TASS	4.28	2.61	3.83	7.01	3.60	10.94	2.43	15.27
	(2.78)		(2.44)		(2.29)		(1.51)	
Hedge Fund Research	5.84	2.27	4.04	6.32	3.56	9.69	1.38	13.80
	(3.62)		(2.47)		(2.32)		(0.93)	
BarclayHedge	4.34	2.19	3.14	5.84	3.42	9.19	1.33	12.61
	(2.99)		(2.01)		(2.27)		(0.84)	
EurekaHedge	0.69	1.21	-0.86	3.31	-0.52	4.90	-3.81	6.97
	(0.35)		-(0.45)		-(0.27)		-(1.97)	
Morningstar	1.64	0.64	-0.15	1.87	0.02	3.58	-2.01	4.19
	(0.94)		-(0.08)		(0.01)		-(1.18)	
Aggregate	5.80	2.16	4.93	5.96	4.99	9.44	2.84	13.63
	(4.10)		(3.27)		(3.48)		(1.97)	

B. Value-weight

Bottom

	1-month		3-month		6-month		12-month	
Database	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)
TASS	4.52	6.11	5.61	13.68	6.55	21.95	4.48	34.91
	(2.01)		(2.65)		(2.75)		(2.00)	
Hedge Fund Research	2.41	3.97	3.82	11.24	4.20	19.56	5.88	31.48
	(1.62)		(2.72)		(2.77)		(4.24)	
BarclayHedge	4.02	3.95	5.41	10.81	5.76	18.80	4.85	30.52
	(2.50)		(3.20)		(3.47)		(3.15)	
EurekaHedge	6.02	4.43	8.11	12.47	6.88	21.65	7.42	35.10
	(3.43)		(4.55)		(4.01)		(4.30)	
Morningstar	8.10	4.19	8.15	10.17	7.02	18.33	7.52	29.64
	(3.90)		(4.00)		(3.68)		(4.13)	
Aggregate	1.60	4.73	2.99	12.25	4.66	21.04	5.02	33.94
	(1.01)		(1.88)		(2.68)		(3.38)	

Top

	1-month		3-month		6-month		12-month	
Database	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)
TASS	5.84	4.28	4.96	6.80	4.73	12.37	4.40	19.54
	(5.40)		(4.74)		(4.69)		(4.54)	
Hedge Fund Research	6.01	1.74	5.68	4.97	5.49	9.83	4.54	18.51
	(6.59)		(6.62)		(6.97)		(5.37)	
BarclayHedge	5.94	1.73	5.89	4.63	6.37	9.24	4.94	17.37
	(5.41)		(5.74)		(6.39)		(4.51)	
EurekaHedge	6.12	2.87	5.64	7.86	4.78	15.20	4.16	27.90
	(4.82)		(4.57)		(3.74)		(3.34)	
Morningstar	5.87	3.28	5.81	7.74	5.41	14.35	4.92	25.42
	(6.31)		(6.23)		(5.56)		(4.47)	
Aggregate	6.21	2.60	5.68	5.93	5.49	11.33	4.48	19.82
	(6.33)		(6.05)		(6.01)		(4.90)	

Spread

	1-month		3-month		6-month		12-month	
Database	Alpha % pa	DO (%) (diff)	Alpha % pa	DO (%) (diff)	Alpha % pa	DO (%) (diff)	Alpha % pa	DO (%) (diff)
TASS	1.32	1.83	-0.65	6.88	-1.81	9.58	-0.08	15.37
	(0.59)		-(0.30)		-(0.75)		-(0.03)	
Hedge Fund Research	3.61	2.23	1.87	6.27	1.29	9.73	-1.34	12.96
	(2.29)		(1.28)		(0.85)		-(0.91)	
BarclayHedge	1.92	2.22	0.48	6.17	0.61	9.56	0.09	13.16
	(1.07)		(0.26)		(0.34)		(0.05)	
EurekaHedge	0.10	1.55	-2.47	4.60	-2.10	6.45	-3.27	7.20
	(0.05)		-(1.24)		-(1.07)		-(1.66)	
Morningstar	-2.23	0.92	-2.34	2.43	-1.61	3.99	-2.59	4.22
	-(1.09)		-(1.11)		-(0.84)		-(1.35)	
Aggregate	4.62	2.13	2.69	6.32	0.83	9.71	-0.54	14.12
	(2.89)		(1.65)		(0.48)		-(0.34)	

C. Equal-weight persistence of the aggregate data in size groups

Bottom

	1-month		3-month		6-month		12-month	
Size group	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)
Small	1.36	5.09	2.04	14.27	1.72	24.97	2.37	40.41
	(0.90)		(1.23)		(1.04)		(1.45)	
Med	1.12	3.24	2.41	9.44	2.30	16.99	4.27	29.98
	(0.77)		(1.50)		(1.43)		(2.87)	
Large	-0.09	2.52	1.55	7.32	1.70	12.83	2.13	22.88
	-(0.06)		(1.01)		(1.16)		(1.51)	
All	1.24	3.70	2.31	10.34	1.98	18.31	3.15	30.79
	(0.91)		(1.55)		(1.40)		(2.28)	

Top

	1-month		3-month		6-month		12-month	
Size group	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)	Alpha % pa	DO (%)
Small	7.14	2.15	7.98	6.04	9.13	11.76	9.37	22.31
	(5.58)		(5.76)		(5.82)		(5.98)	
Med	7.25	1.49	7.16	4.40	6.63	8.93	5.38	18.72
	(6.84)		(6.83)		(6.32)		(5.06)	
Large	5.97	1.52	5.77	4.41	5.04	8.45	3.80	16.61
	(6.54)		(6.15)		(5.34)		(3.82)	
All	6.97	1.52	7.03	4.31	6.73	8.66	5.82	16.75
	(7.07)		(7.20)		(6.61)		(5.76)	

Spread

	1-month		3-month		6-month		12-month	
Size group	Alpha % pa	DO (%) (diff)	Alpha % pa	DO (%) (diff)	Alpha % pa	DO (%) (diff)	Alpha % pa	DO (%) (diff)
Small	5.79	2.94	5.94	8.23	7.41	13.21	7.00	18.11
	(3.21)		(3.01)		(3.49)		(3.37)	
Med	6.12	1.75	4.75	5.04	4.33	8.06	1.11	11.26
	(3.81)		(2.83)		(2.60)		(0.67)	
Large	6.06	1.00	4.22	2.91	3.34	4.38	1.67	6.27
	(4.03)		(2.68)		(2.21)		(1.08)	
All	5.73	2.18	4.72	6.03	4.75	9.65	2.67	14.04
	(4.00)		(3.09)		(3.29)		(1.84)	

Table 8

Feasible portfolio rebalancing and performance persistence [January 1994 - December 2010]

Table shows the annualized Fung and Hsieh (2004) alphas for portfolios based on (1) monthly; (2); quarterly; (3) semiannually; and (4) yearly rebalancing horizons. Hedge funds are sorted into portfolios based on their t-statistics of alphas estimated using the seven factor model proposed by Fung and Hsieh (2004). The t-statistics of alphas are estimated using the most recent 24 months of returns preceding the evaluation period. Portfolio returns are calculated for equal-weight portfolios monthly, so the weights are readjusted whenever a fund disappears. Within each rebalancing horizon decile portfolios are formed using only the feasible information taking into account fund-specific share restrictions. For instance, for the feasible semiannual (annual) rebalancing strategy, we exclude the funds that have redemption and lockup periods longer than 6 (12) months. Also, for semiannual (annual) rebalancing strategies, we exclude funds having notice periods longer than 6 (12) months. For the 1-month rebalancing strategy, we exclude funds having notice period longer than 1-month. This implies that we use at least 1-month lagged information to estimate persistence in order to mitigate look-ahead bias. Notice period determines the lag length. The annualized alphas of feasible rebalancing strategies are reported in Column "Feasible". Columm "Baseline" shows the results of persistence without restrictions for lockup and redemption periods. Results of persistence are reported for bottom, top and spread portfolios. Databases are merged and cleaned from multiple share classes using a novel statistical procedure that is described in Appendix of the paper.

	Redemption ≤ 1 m Lockup ≤ 1 m		Redemption ≤ 3 m Lockup ≤ 3 m		Redemption ≤ 6 m Lockup ≤ 6 m		Redemption ≤ 12 m Lockup ≤ 12 m	
Bottom	Feasible	Baseline	Feasible	Baseline	Feasible	Baseline	Feasible	Baseline
	1m		3m		6m		12m	
	Alpha % pa		Alpha % pa		Alpha % pa		Alpha % pa	
Notice ≤ 1 m	0.89	1.23	0.70	2.14	1.53	1.85	2.85	3.05
	(0.56)	(0.90)	(0.44)	(1.44)	(0.98)	(1.32)	(1.86)	(2.24)
Notice ≤ 3 m	.	.	2.40	.	1.98	.	3.05	.
	.	.	(1.51)	.	(1.15)	.	(1.60)	.
Notice ≤ 6 m	.	.	.	.	3.44	.	3.31	.
	.	.	.	.	(2.19)	.	(2.06)	.
Notice ≤ 12 m	.	.	.	.	.	.	4.64	.
	.	.	.	.	.	.	(2.97)	.
Top	Feasible	Baseline	Feasible	Baseline	Feasible	Baseline	Feasible	Baseline
	1m		3m		6m		12m	
	Alpha % pa		Alpha % pa		Alpha % pa		Alpha % pa	
Notice ≤ 1 m	5.44	7.03	5.43	7.08	4.83	6.84	4.80	5.89
	(4.21)	(7.15)	(4.27)	(7.31)	(3.70)	(6.74)	(4.10)	(5.84)
Notice ≤ 3 m	.	.	5.95	.	5.12	.	5.01	.
	.	.	(5.68)	.	(5.09)	.	(5.25)	.
Notice ≤ 6 m	.	.	.	.	4.36	.	3.41	.
	.	.	.	.	(4.23)	.	(3.35)	.
Notice ≤ 12 m	.	.	.	.	.	.	3.68	.
	.	.	.	.	.	.	(3.53)	.
Spread	Feasible	Baseline	Feasible	Baseline	Feasible	Baseline	Feasible	Baseline
	1m		3m		6m		12m	
	Alpha % pa		Alpha % pa		Alpha % pa		Alpha % pa	
Notice ≤ 1 m	4.55	5.80	4.73	4.93	3.31	4.99	1.95	2.84
	(2.61)	(4.10)	(2.75)	(3.27)	(1.93)	(3.48)	(1.18)	(1.97)
Notice ≤ 3 m	.	.	3.55	.	3.14	.	1.96	.
	.	.	(2.28)	.	(1.85)	.	(1.03)	.
Notice ≤ 6 m	.	.	.	.	0.92	.	0.10	.
	.	.	.	.	(0.54)	.	(0.06)	.
Notice ≤ 12 m	.	.	.	.	.	.	-0.96	.
	.	.	.	.	.	.	(-0.62)	.

Table 9

Summary statistics of characteristics and performance results of univariate sorts

Table describes statistics of fund-level characteristics and the results of performance based on univariate sorts. Results are reported for each database. Aggregate database is a merged database of single databases including TASS, Hedge Fund Research, BarclayHedge, EurekaHedge, Morningstar. Databases are merged using a novel statistical algorithm that is described in Appendix of the paper. All hedge funds are required to have at least 12 monthly non-missing returns. Single databases are removed from multiple share classes using the same statistical procedure. In Panel A, all values of characteristics are based on hedge fund database snapshot variables that include one observation for each fund. Table contains variables of fees including management and incentive fee (in percentages) and high-water mark binary variable (one for funds having high-water mark and zero otherwise). Second, table includes variables of share restrictions (reported in years) including lockup period, advance notification period and redemption frequency. Columns include a total number of hedge funds in question (# of Funds) and a number of hedge funds that have missing values for each variable (Missing %). Statistics include cross-sectional mean and standard deviation. In Panel B, hedge funds are sorted into five portfolios on December each year based on one of the time-varying variables (age, size (AuM), or Flow). Equal-weight and value-weight portfolio returns are calculated for each month, so the weights are readjusted whenever a fund disapprears. Also, hedge funds are sorted into portfolios based on time-invariant characteristics including incentive fee, lockup period, notification period, and redemption period. Within each group, hedge funds in the aggregate data are sorted into terciles using monthly AuM observations. In Panel B, annualized Fung and Hsieh (2004) alphas and t-values of alphas are reported for each portfolio.

A. Summary statistics

Incentive Fee

Database	# of Funds	Missing (%)	Mean	Std
TASS	8072	3.47	0.17	0.07
Hedge Fund Research	9508	1.89	0.19	0.05
BarclayHedge	9588	0.01	0.19	0.06
EurekaHedge	7009	1.95	0.19	0.05
Morningstar	6681	12.72	0.19	0.05
Aggregate	24768	5.20	0.18	0.06

Lockup period (in years)

Database	# of Funds	Missing (%)	Mean	Std
TASS	8072	0.00	0.24	0.53
Hedge Fund Research	9508	7.43	0.29	0.55
BarclayHedge	9588	17.48	0.27	0.54
EurekaHedge	7009	2.18	0.19	0.48
Morningstar	6681	46.79	0.45	0.62
Aggregate	24768	17.10	0.25	0.53

Management Fee

Database	# of Funds	Missing (%)	Mean	Std
TASS	8072	0.94	0.02	0.01
Hedge Fund Research	9508	1.31	0.02	0.01
BarclayHedge	9588	0.01	0.02	0.01
EurekaHedge	7009	1.73	0.02	0.01
Morningstar	6681	13.29	0.02	0.01
Aggregate	24768	3.66	0.02	0.01

Notification period (in years)

Database	# of Funds	Missing (%)	Mean	Std
TASS	8072	0.00	0.08	0.09
Hedge Fund Research	9508	5.55	0.10	0.08
BarclayHedge	9588	0.01	0.07	0.10
EurekaHedge	7009	6.15	0.10	0.09
Morningstar	6681	39.60	0.10	0.08
Aggregate	24768	10.57	0.08	0.09

High-Water mark (1=yes, 0=no)

Database	# of Funds	Missing (%)	Mean	Std
TASS	8072	0.94	0.63	0.48
Hedge Fund Research	9508	0.00	0.88	0.32
BarclayHedge	9588	0.67	0.61	0.49
EurekaHedge	7009	0.49	0.87	0.34
Morningstar	6681	25.25	0.86	0.35
Aggregate	24768	5.98	0.70	0.46

Redemption frequency (in years)

Database	# of Funds	Missing (%)	Mean	Std
TASS	8072	9.32	0.17	0.19
Hedge Fund Research	9508	2.77	0.17	0.22
BarclayHedge	9588	24.52	0.19	0.22
EurekaHedge	7009	2.01	0.13	0.17
Morningstar	6681	33.59	0.18	0.21
Aggregate	24768	19.02	0.16	0.21

B. Univariate sorts

Fund age

Portfolio	TASS		HFR		Barclay		Eureka		Morningstar				Aggregate		
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	Small	Med	Large
Young	7.63	9.82	9.39	9.92	9.58	6.16	9.63	11.72	9.45	9.78	8.23	7.45	9.77	6.50	5.85
	(6.17)	(6.75)	(8.24)	(7.14)	(8.75)	(3.71)	(5.38)	(5.02)	(7.82)	(5.85)	(7.31)	(5.34)	(8.11)	(5.56)	(4.94)
	5.28	6.92	5.47	4.96	6.79	6.10	6.44	4.16	8.61	5.99	5.49	5.19	8.76	5.95	6.09
	(4.73)	(5.53)	(4.74)	(4.21)	(6.15)	(4.64)	(3.74)	(2.70)	(6.57)	(4.46)	(4.85)	(4.42)	(6.97)	(5.38)	(6.14)
	4.79	5.90	4.84	6.12	5.53	4.94	8.98	7.52	7.51	7.70	4.99	6.17	6.48	6.27	5.59
	(4.46)	(5.55)	(4.64)	(4.89)	(5.45)	(4.98)	(6.09)	(4.89)	(6.62)	(7.17)	(4.77)	(5.53)	(5.22)	(5.36)	(5.08)
	4.47	7.40	5.03	5.03	4.96	5.38	7.25	6.64	7.54	6.95	4.39	5.30	4.96	6.39	4.90
	(4.15)	(4.12)	(4.98)	(4.12)	(4.78)	(4.82)	(5.37)	(4.85)	(6.98)	(5.93)	(4.19)	(3.97)	(3.70)	(5.88)	(4.36)
Mature	4.16	5.97	4.70	6.21	4.13	5.78	6.55	6.16	6.21	7.48	3.96	5.52	4.41	5.52	5.22
	(3.65)	(3.55)	(4.63)	(5.74)	(3.46)	(4.97)	(4.49)	(4.67)	(5.13)	(5.68)	(3.63)	(4.61)	(2.86)	(4.67)	(5.02)
Spread	3.46	3.85	4.69	3.71	5.45	0.38	3.08	5.56	3.24	2.31	4.27	1.92	5.36	0.98	0.63
	(4.12)	(3.09)	(5.98)	(3.71)	(6.09)	(0.29)	(2.20)	(3.02)	(3.52)	(1.60)	(5.65)	(1.82)	(3.72)	(0.86)	(0.81)

Fund size

Portfolio	TASS		HFR		Barclay		Eureka		Morningstar				Aggregate		
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	Small	Med	Large
Small	7.32	12.60	9.19	13.33	9.35	10.67	10.64	12.52	11.64	13.43	8.28	11.22	11.84	12.77	9.99
	(5.80)	(9.16)	(8.15)	(7.91)	(7.99)	(8.00)	(7.22)	(7.89)	(9.56)	(9.47)	(7.21)	(8.20)	(9.32)	(10.38)	(8.67)
	7.22	12.46	7.33	11.53	7.94	12.51	11.21	15.01	9.56	13.66	7.14	11.38	8.38	8.13	7.33
	(6.31)	(8.85)	(6.88)	(8.83)	(6.73)	(8.61)	(8.03)	(9.87)	(7.74)	(9.37)	(6.42)	(8.49)	(6.41)	(7.75)	(6.46)
Medium	5.22	9.29	6.08	9.10	6.74	9.62	5.95	10.68	7.80	10.18	5.57	9.02	6.54	6.12	5.33
	(4.78)	(7.54)	(5.98)	(8.07)	(6.68)	(8.82)	(3.48)	(6.27)	(7.24)	(8.89)	(5.35)	(7.68)	(4.83)	(5.60)	(4.98)
	3.99	7.66	4.19	7.27	5.01	7.37	6.44	8.66	5.90	9.06	4.10	7.30	4.70	3.67	3.70
	(3.58)	(6.41)	(4.01)	(6.54)	(4.80)	(6.68)	(4.24)	(6.33)	(4.99)	(7.71)	(3.75)	(6.26)	(3.92)	(3.27)	(3.56)
Large	3.30	6.28	3.13	5.41	3.04	4.87	5.20	5.81	4.53	6.48	2.88	5.12	-0.43	-0.49	3.03
	(2.93)	(4.20)	(2.93)	(5.00)	(2.80)	(4.63)	(3.62)	(4.58)	(3.99)	(5.80)	(2.67)	(4.49)	-(0.34)	-(0.41)	(2.69)
Spread	4.02	6.31	6.07	7.92	6.31	5.80	5.43	6.70	7.11	6.96	5.39	6.11	12.28	13.27	6.96
	(4.17)	(5.09)	(7.83)	(6.09)	(7.20)	(5.59)	(4.10)	(4.57)	(8.94)	(6.28)	(6.82)	(6.27)	(9.89)	(11.57)	(8.88)

Fund Flow

Portfolio	TASS		HFR		Barclay		Eureka		Morningstar		Aggregate				
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	Small	Med	Large
Low	5.78	8.80	6.65	5.71	6.32	5.37	8.20	6.36	8.42	5.45	6.06	6.41	8.98	7.94	7.04
	(4.94)	(3.77)	(5.65)	(4.09)	(5.28)	(4.44)	(4.77)	(3.45)	(6.50)	(3.91)	(5.25)	(4.75)	(5.81)	(6.87)	(5.67)
	4.73	6.27	5.10	5.56	6.34	5.93	7.50	6.52	7.52	5.77	5.24	5.59	7.71	6.39	5.22
	(4.52)	(4.53)	(5.64)	(5.32)	(6.11)	(5.20)	(4.81)	(4.48)	(6.50)	(4.92)	(5.20)	(4.60)	(6.23)	(6.10)	(4.88)
	4.56	5.99	5.61	5.66	6.08	5.61	7.77	7.50	7.32	8.71	4.86	5.62	5.27	6.23	5.30
	(4.29)	(4.94)	(5.65)	(5.34)	(6.82)	(4.12)	(6.57)	(4.74)	(6.88)	(7.24)	(5.06)	(4.72)	(5.13)	(6.25)	(5.33)
	5.44	7.50	5.88	6.67	6.55	5.92	7.75	7.58	7.83	8.04	5.53	5.85	5.93	6.13	5.67
	(4.75)	(5.23)	(5.73)	(5.39)	(5.92)	(5.44)	(5.75)	(5.46)	(6.69)	(6.51)	(5.23)	(4.88)	(4.99)	(5.75)	(5.30)
Great	6.45	5.27	6.51	6.30	6.34	4.88	8.06	5.68	8.41	7.82	6.08	5.17	7.69	4.32	4.70
	(5.13)	(3.44)	(5.42)	(5.10)	(4.68)	(3.83)	(5.10)	(3.79)	(6.86)	(5.65)	(4.73)	(4.37)	(5.33)	(3.19)	(4.03)
Spread	-0.67	3.53	0.14	-0.59	-0.02	0.48	0.14	0.68	0.00	-2.37	-0.02	1.24	1.29	3.62	2.35
	-(0.86)	(1.79)	(0.25)	-(0.63)	-(0.02)	(0.44)	(0.12)	(0.38)	(0.00)	-(1.95)	-(0.03)	(1.28)	(1.14)	(3.93)	(2.80)

Incentive Fee (IF)

Portfolio	TASS		HFR		Barclay		Eureka		Morningstar		Aggregate				
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	Small	Med	Large
IF = 0	2.46	4.81	3.69	3.36	3.55	3.28	5.16	5.63	4.48	3.53	3.26	3.60	0.69	3.50	4.80
	(1.97)	(2.68)	(3.58)	(4.00)	(3.28)	(2.48)	(2.95)	(2.45)	(2.82)	(2.54)	(2.79)	(2.90)	(0.47)	(3.11)	(4.25)
0 < IF< 20	4.21	6.56	5.38	5.49	5.41	4.20	5.96	8.99	5.62	5.98	4.52	4.88	3.46	4.44	7.79
	(3.37)	(4.20)	(4.25)	(4.39)	(4.25)	(3.51)	(3.71)	(4.13)	(4.53)	(3.90)	(3.54)	(3.96)	(2.63)	(3.23)	(5.72)
IF = 20	6.79	7.06	7.25	6.67	7.87	6.08	8.85	6.31	8.70	6.99	7.29	6.18	7.31	7.95	8.13
	(6.81)	(5.33)	(7.67)	(5.89)	(8.51)	(5.93)	(8.35)	(5.49)	(8.93)	(6.64)	(7.59)	(5.65)	(7.04)	(8.19)	(8.33)
IF > 20	8.72	5.96	10.36	7.52	12.30	8.80	13.64	4.94	12.20	9.65	11.47	7.27	13.97	12.40	10.30
	(5.05)	(5.98)	(7.43)	(4.89)	(7.67)	(5.62)	(6.17)	(3.35)	(6.35)	(5.50)	(8.13)	(5.05)	(8.34)	(7.92)	(6.10)
Spread	6.26	1.16	6.66	4.16	8.74	5.52	8.49	-0.70	7.73	6.12	8.22	3.67	13.28	8.90	5.50
	(4.44)	(0.70)	(5.30)	(2.73)	(6.21)	(3.39)	(3.43)	-(0.30)	(3.57)	(3.54)	(6.90)	(2.77)	(7.54)	(6.38)	(3.86)

Lockup Period (LP)

Portfolio	TASS		HFR		Barclay		Eureka		Morningstar		Aggregate				
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	Small	Med	Large
LP= 0	5.47	6.06	6.38	5.62	7.44	6.09	8.26	6.36	8.03	5.81	6.23	5.55	6.34	6.77	7.65
	(4.98)	(4.66)	(6.42)	(5.77)	(7.38)	(5.81)	(6.95)	(4.97)	(7.99)	(5.43)	(5.91)	(5.18)	(5.89)	(6.31)	(7.27)
0 < LP < 1	7.79	8.68	8.54	5.18	9.33	6.67	11.26	6.97	9.73	6.66	9.03	7.18	8.34	10.09	8.07
	(7.48)	(4.57)	(9.77)	(4.30)	(8.38)	(3.01)	(9.11)	(4.38)	(8.41)	(5.01)	(8.84)	(4.15)	(5.51)	(8.84)	(6.62)
LP = 1	8.65	7.07	8.08	7.82	8.55	6.66	8.93	7.45	8.66	7.11	8.07	6.65	6.49	8.67	8.97
	(8.66)	(4.08)	(7.31)	(5.77)	(8.76)	(5.17)	(8.65)	(7.29)	(8.68)	(6.19)	(8.05)	(4.65)	(5.20)	(8.36)	(8.71)
LP > 1	11.88	10.05	12.11	16.36	10.23	5.92	10.92	8.19	11.58	8.89	11.24	10.89	11.56	11.93	11.22
	(8.24)	(5.94)	(8.29)	(8.62)	(6.20)	(2.84)	(7.19)	(4.61)	(7.18)	(5.69)	(8.12)	(5.83)	(3.83)	(6.34)	(6.88)
Spread	6.41	4.00	5.73	10.74	2.80	-0.17	2.65	1.82	3.55	3.08	5.01	5.33	5.11	5.16	3.57
	(6.80)	(2.82)	(5.09)	(6.16)	(2.26)	-(0.09)	(2.36)	(1.10)	(3.04)	(2.20)	(5.16)	(3.09)	(1.86)	(3.45)	(2.92)

Notice Period (NP)

Portfolio	TASS		HFR		Barclay		Eureka		Morningstar		Aggregate				
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	Small	Med	Large
NP < 1	4.13	4.15	5.88	5.26	7.88	5.56	7.36	5.88	9.03	9.16	6.57	5.46	8.88	7.55	6.99
	(3.04)	(2.50)	(4.80)	(2.64)	(6.28)	(4.41)	(4.28)	(1.60)	(4.16)	(3.88)	(5.25)	(4.19)	(6.86)	(5.58)	(5.16)
1 ≤ NP < 30	5.59	6.58	6.82	6.11	6.00	6.24	7.98	6.05	8.65	4.80	6.11	5.99	5.10	6.33	8.16
	(4.06)	(3.11)	(5.92)	(5.80)	(4.81)	(4.38)	(5.23)	(3.61)	(6.29)	(3.83)	(5.10)	(4.25)	(3.84)	(4.94)	(6.63)
NP = 30	7.00	6.34	6.37	5.91	7.20	5.02	8.68	5.98	7.70	6.09	6.73	5.26	6.93	7.51	7.46
	(6.49)	(4.84)	(6.30)	(4.81)	(6.94)	(3.87)	(7.97)	(4.63)	(7.43)	(5.64)	(6.55)	(4.04)	(6.01)	(7.32)	(6.78)
30 < NP < 60	7.59	7.48	7.69	6.11	8.82	6.13	8.78	5.85	7.96	6.93	7.72	6.12	8.08	8.95	7.60
	(7.37)	(6.27)	(6.99)	(4.56)	(8.50)	(5.85)	(7.81)	(4.40)	(8.20)	(7.80)	(7.09)	(5.14)	(4.90)	(7.31)	(6.41)
NP = 60	8.09	8.23	7.54	6.19	8.61	5.69	9.85	8.12	8.43	7.51	7.76	5.70	6.42	7.26	9.09
	(8.08)	(5.61)	(7.34)	(5.02)	(8.72)	(5.01)	(7.48)	(4.74)	(8.36)	(5.81)	(7.77)	(5.48)	(5.43)	(6.80)	(8.09)
NP > 60	8.44	8.18	8.89	10.45	8.57	7.13	9.16	9.90	8.49	6.55	8.42	8.54	8.90	9.17	8.86
	(8.32)	(5.20)	(9.07)	(4.63)	(8.46)	(5.39)	(9.48)	(5.51)	(8.33)	(7.55)	(8.80)	(5.62)	(5.69)	(8.84)	(8.60)
Spread	4.30	4.03	3.01	5.19	0.69	1.57	1.80	4.03	-0.54	-2.61	1.85	3.08	0.02	1.62	1.87
	(4.66)	(2.27)	(2.95)	(2.04)	(0.51)	(1.05)	(1.24)	(1.18)	-(0.27)	-(1.15)	(1.71)	(1.92)	(0.01)	(1.23)	(1.50)

	Redemption Period (RP)														
Portfolio	TASS		HFR		Barclay		Eureka		Morningstar		Aggregate				
	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW	Small	Med	Large
RP ≤ 1	5.23	5.07	6.40	5.32	6.27	4.70	8.14	6.46	8.41	5.75	5.84	4.65	6.00	6.40	6.95
	(4.60)	(3.87)	(6.14)	(5.20)	(5.70)	(4.15)	(6.26)	(4.42)	(7.52)	(5.03)	(5.29)	(4.14)	(5.27)	(5.53)	(6.53)
1 < RP ≤ 3	7.24	8.24	7.58	7.77	8.31	6.52	9.07	6.41	8.18	7.26	7.64	7.63	6.40	8.22	8.87
	(7.33)	(5.24)	(7.76)	(5.79)	(8.58)	(6.09)	(8.58)	(5.71)	(8.41)	(6.54)	(7.76)	(5.73)	(5.92)	(8.37)	(8.00)
RP > 3	8.02	7.63	7.65	7.79	7.87	6.13	9.02	9.45	7.89	8.32	7.62	6.38	7.32	7.37	8.43
	(8.24)	(6.11)	(7.83)	(5.37)	(8.77)	(4.80)	(10.14)	(7.14)	(9.14)	(5.27)	(8.38)	(5.47)	(5.68)	(6.93)	(8.80)
Spread	2.80	2.55	1.26	2.47	1.61	1.44	0.88	3.00	-0.52	2.57	1.78	1.73	1.32	0.97	1.48
	(2.90)	(2.12)	(1.50)	(2.00)	(2.08)	(1.21)	(0.77)	(1.95)	-(0.55)	(1.71)	(2.13)	(1.60)	(1.08)	(1.02)	(1.80)

Table 10

Cross-sectional regressions [January 1994 - December 2010]

Table shows results of the cross-sectional regressions based on Fama-McBeth (1973) procedure. Results are reported for each database. Aggregate database is a merged database of single databases including TASS, Hedge Fund Research (HFR), BarclayHedge, EurekaHedge, and Morningstar. Databases are merged using a novel statistical procedure that is described in Appendix of the paper. Single databases are removed from multiple share classes using the same statistical procedure. In panel A, excess returns are regressed against fund-level characteristics including fund size (AUM in a log scale), age, flow, compensation structure and share restrictions (measured in years). High-water mark is a binary variable that equals one for funds that apply high-water mark and zero otherwise. Fund-level size, age, and flow are lagged one month. Age is measured as a length of the return time series. In panel B, fund-level alphas are regressed against fund-level characteristics. Alpha is calculated as a sum of the intercept and the time series of the residual estimated from the Fund and Hsieh (2004) model. In Panel C, hedge funds are sorted into size terciles based on the AUM observations each December and fund-level excess returns are regressed against fund-level characteristics within size terciles. In panel D, fund-level alphas are regressed against characteristics within size terciles. Time period is from January 1994 - December 2010. All parameter estimates are multiplied by 100.

Variable	A. Excess returns TASS	HFR	BarclayHedge	EurekaHedge	Morningstar	Aggregate	B. Fung and Hsieh (2004) alphas TASS	HFR	BarclayHedge	EurekaHedge	Morningstar	Aggregate
Intercept	1.13	1.83	1.45	1.25	1.64	1.37	0.79	1.39	0.84	0.63	1.19	0.84
	(3.20)	(3.41)	(4.30)	(2.35)	(3.99)	(3.84)	(3.41)	(3.37)	(3.82)	(1.43)	(3.64)	(4.27)
$Size_{t-1}$	-0.06	-0.07	-0.10	-0.03	-0.12	-0.06	-0.05	-0.06	-0.08	-0.03	-0.10	-0.05
	-(3.57)	-(4.70)	-(5.46)	-(1.42)	-(6.58)	-(4.27)	-(4.40)	-(5.43)	-(6.84)	-(1.83)	-(6.47)	-(5.59)
Age_{t-1}	-0.13	-0.17	-0.07	-0.14	-0.16	-0.12	-0.18	-0.24	-0.09	-0.12	-0.19	-0.17
	-(3.19)	-(3.38)	-(1.39)	-(1.42)	-(3.65)	-(3.49)	-(4.51)	-(4.04)	-(1.35)	-(1.48)	-(3.97)	-(5.18)
$Flow_{t-1}$	1.62	1.62	1.28	1.74	1.60	1.52	1.10	1.18	1.00	1.35	1.23	1.10
	(5.19)	(5.80)	(4.36)	(2.69)	(4.19)	(5.50)	(4.96)	(7.22)	(6.17)	(2.61)	(5.04)	(6.91)
Management Fee	10.06	8.10	12.59	15.42	5.25	6.14	10.02	8.73	13.13	21.19	6.84	8.06
	(2.97)	(2.54)	(3.81)	(2.78)	(0.80)	(2.58)	(3.44)	(3.16)	(4.17)	(4.06)	(1.19)	(3.94)
Incentive Fee	0.95	0.68	0.56	0.70	1.38	0.75	1.44	1.37	0.93	1.24	1.26	1.26
	(2.60)	(2.00)	(1.55)	(0.90)	(3.35)	(2.61)	(5.40)	(7.12)	(3.67)	(1.98)	(3.42)	(6.39)
High-Water Mark	0.15	0.04	0.12	-0.18	-0.01	0.10	0.14	0.05	0.11	-0.24	0.00	0.08
	(3.64)	(1.35)	(2.45)	-(0.88)	-(0.08)	(3.81)	(3.66)	(2.15)	(2.66)	-(1.18)	-(0.05)	(3.37)
Lockup period	0.05	0.09	0.09	0.10	0.02	0.08	0.05	0.05	0.06	0.15	0.01	0.04
	(1.11)	(2.67)	(2.30)	(1.25)	(0.43)	(2.38)	(1.23)	(2.25)	(2.10)	(1.82)	(0.16)	(1.67)
Notice period	1.22	0.23	0.00	0.49	0.70	0.27	0.80	0.24	-0.09	0.36	0.60	0.24
	(4.49)	(0.78)	-(0.02)	(1.06)	(2.00)	(1.25)	(2.99)	(0.92)	-(0.41)	(0.83)	(1.84)	(1.19)
Redemption	0.13	0.06	0.06	0.11	0.08	0.10	-0.01	0.01	0.00	0.08	0.05	0.02
	(1.58)	(0.95)	(0.71)	(0.86)	(0.96)	(1.75)	-(0.17)	(0.21)	-(0.02)	(0.63)	(0.69)	(0.35)

	C. Excess return			D. Fung and Hsieh (2004) alphas		
	AUM terciles			AUM terciles		
Variable	[1]	[2]	[3]	[1]	[2]	[3]
Intercept	1.45	0.72	1.07	0.92	0.53	0.82
	(3.02)	(1.58)	(2.40)	(2.76)	(1.95)	(3.31)
Size $_{t-1}$	-0.04	-0.04	-0.02	-0.06	-0.04	-0.02
	-(1.11)	-(1.02)	-(0.59)	-(1.83)	-(1.32)	-(0.90)
Age $_{t-1}$	-0.19	0.01	0.01	-0.33	-0.10	-0.10
	-(2.39)	(0.18)	(0.13)	-(4.03)	-(2.33)	-(1.65)
Flow $_{r-1}$	1.49	1.69	2.22	0.97	1.26	1.64
	(4.29)	(5.03)	(4.75)	(4.09)	(6.44)	(5.36)
Management Fee	11.75	7.97	2.97	22.51	3.21	3.46
	(2.07)	(2.24)	(1.06)	(4.62)	(0.93)	(1.40)
Incentive Fee	1.36	0.67	0.11	2.18	1.12	0.43
	(2.77)	(1.75)	(0.38)	(5.73)	(3.74)	(2.13)
High-Water Mark	0.15	0.13	0.04	0.08	0.08	0.06
	(2.28)	(3.21)	(1.14)	(1.47)	(2.41)	(1.94)
Lockup period	-0.02	0.13	0.07	-0.02	0.09	0.05
	-(0.20)	(3.24)	(2.22)	-(0.34)	(2.79)	(1.85)
Notice period	0.37	0.24	-0.04	0.53	0.20	-0.12
	(1.01)	(0.86)	-(0.17)	(1.49)	(0.84)	-(0.58)
Redemption	0.06	0.03	0.18	-0.02	-0.02	0.08
	(0.50)	(0.38)	(2.13)	-(0.21)	-(0.26)	(1.41)

Figure 1
Histograms of pairwise correlation coefficients of hedge funds

This figure shows histograms of pairwise correlation coefficients of share classes estimated within management companies for each database. Databases are merged using a novel statistical algorithm. First, all databases are grouped based on a management company name that is cleaned from errors and abbreviations. Second, correlation coefficients are estimated for each share class pair within each management company. Each share class pair in correlation analysis is required to have at least 12 non-missing monthly return observations. Share classes that do not have enough return observations for pairwise correlation analysis are automatically excluded from analysis. Share classes have return time series between January 1994 – June 2011.

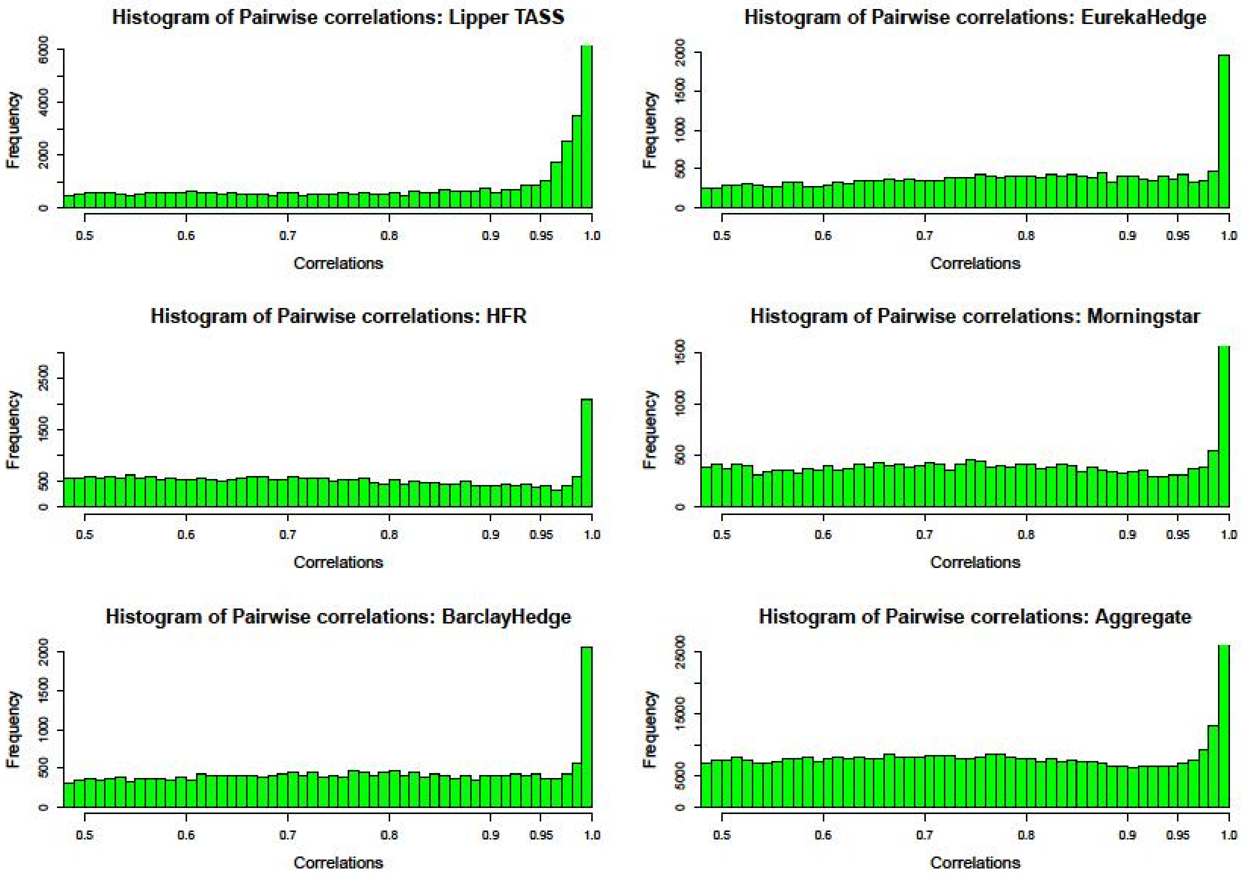

Figure 2
Venn diagram of overlapping and unique share classes

Figure presents Venn diagram showing overlapping between databases and the percentage amount of unique share classes. Databases are merged using a statistical algorithm presented in Appendix of the paper. First, all databases including all share classes are classified based on management company names. Second, pairwise correlations between all possible share class pairs are estimated within management companies. Multiple share classes are identified using a 0.99 correlation limit and correlated share classes are classified in groups. Each group gives information on the amount of databases each share class is reported. In Figure, all presented numbers are in percentages.

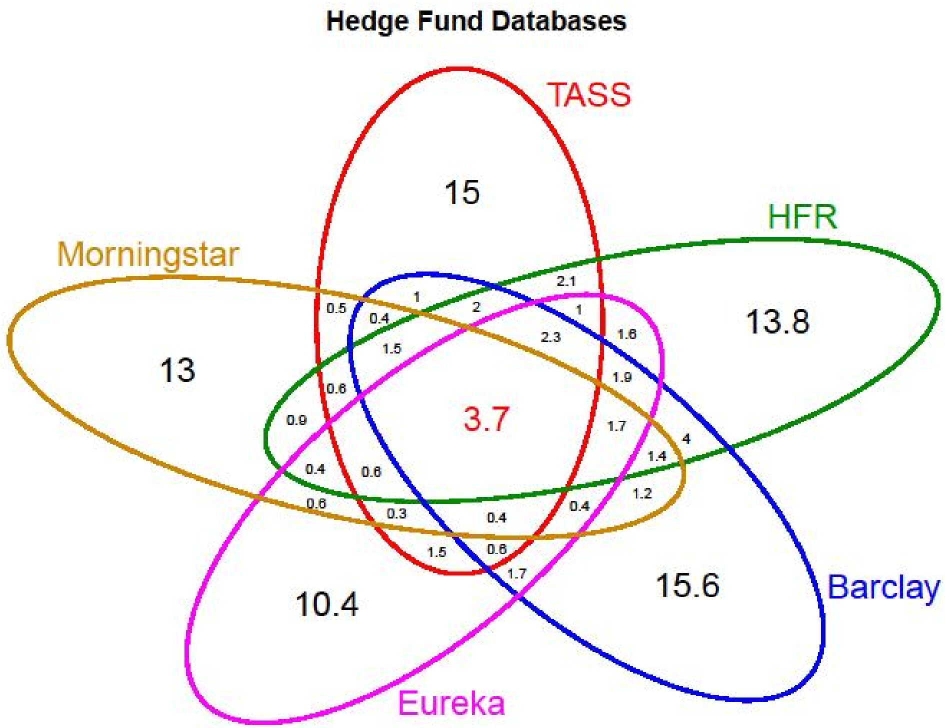

Figure 3
Proportion of hedge funds by investment objective

Figure presents the number of hedge funds by investment objective. Results are reported for each database. All funds are required to have at least 12 monthly return observations. Aggregate database is constructed using a statistical procedure that is described in Appendix. All databases are cleaned from multiple share classes using the same statistical procedure. Numbers in the pie charts are percentage amounts of hedge funds by investment objective. Pie chart of aggregate data includes all main investment objectives. Appendix provides details of the classification of the investment objectives.

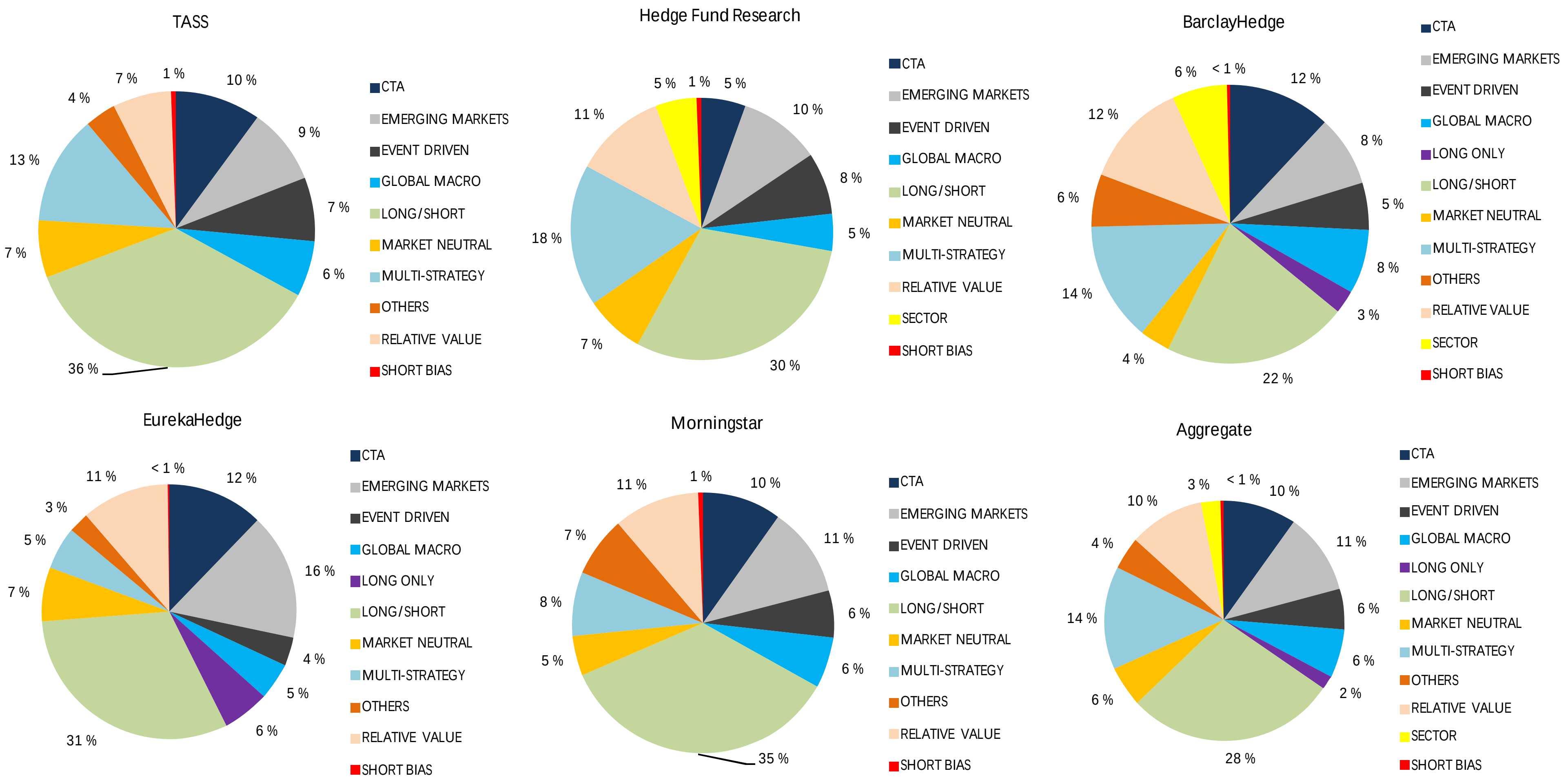

Figure 4
Proportion of hedge funds by domicile region

Figure presents the number of hedge funds by fund-level domicile regions. Results are reported for each database. All funds are required to have at least 12 monthly return observations. Aggregate database is constructed using a statistical procedure (see Appendix for details). All databases are cleaned from multiple share classes using the same statistical procedure. Numbers in the pie charts are percentage amounts of hedge funds by domicile regions that are categorized as (1) Onshore; (2) Asia and Pacific; (3) Caribbean; (4) Europe; and (5) Rest of World. Funds that are legally established in either United States or Canada are classified as onshore funds. Other categories of fund domiciles are classified as offshore funds.

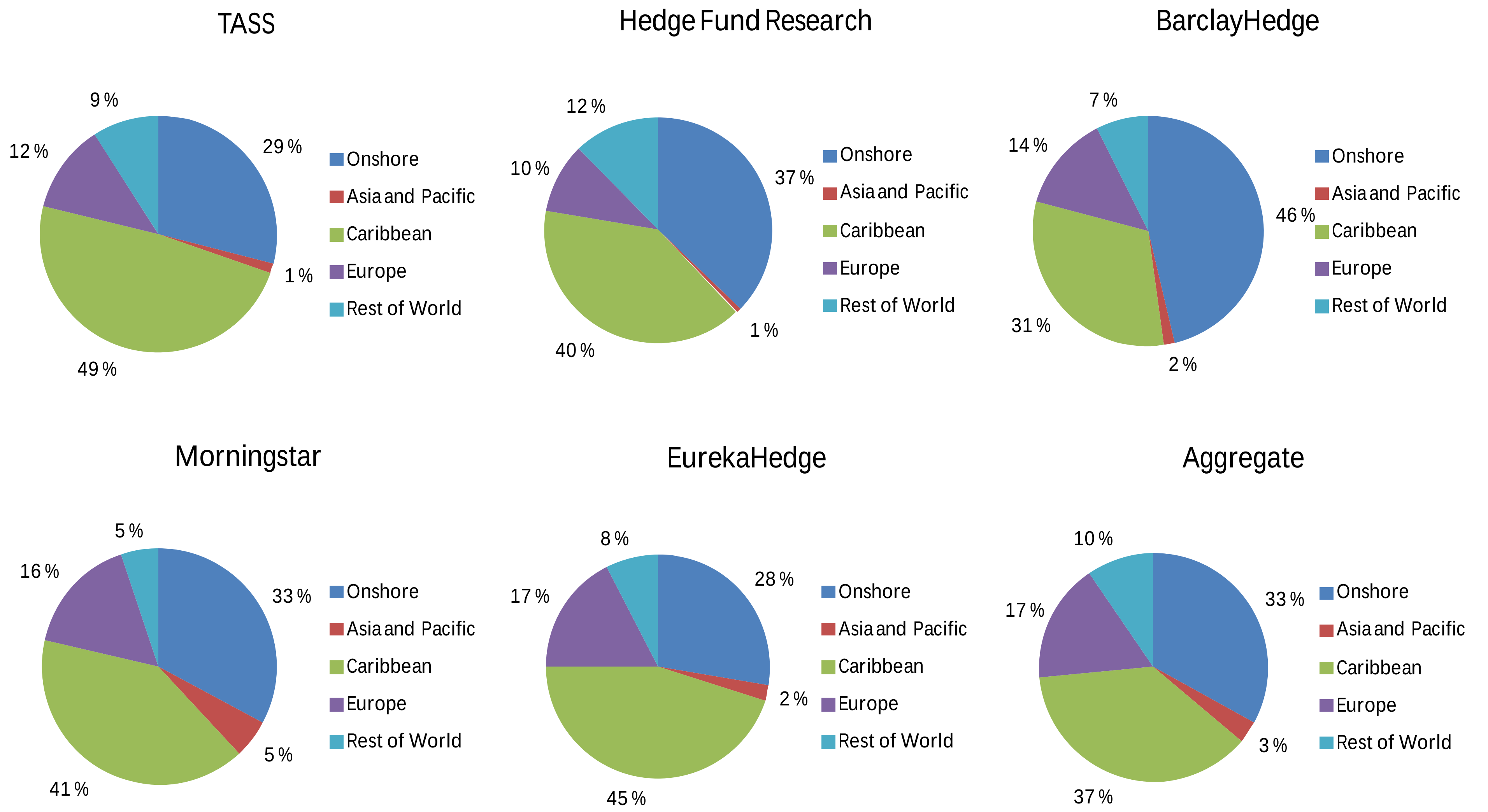

Figure 5
Equal- and value-weight cumulative excess returns of hedge funds [January 1994 – December 2010]

Figure presents equal- and value-weight cumulative excess returns of hedge funds for each database. Databases are removed from multiple share classes that exist either between databases or within management companies. Procedure to exclude multiple share classes is performed using a statistical algorithm (presented in Appendix). Equal- and value-weight portfolio excess returns are calculated using single hedge funds that have at least 12 monthly return observations. Aggregate data is the merged database using all five databases as source datasets.

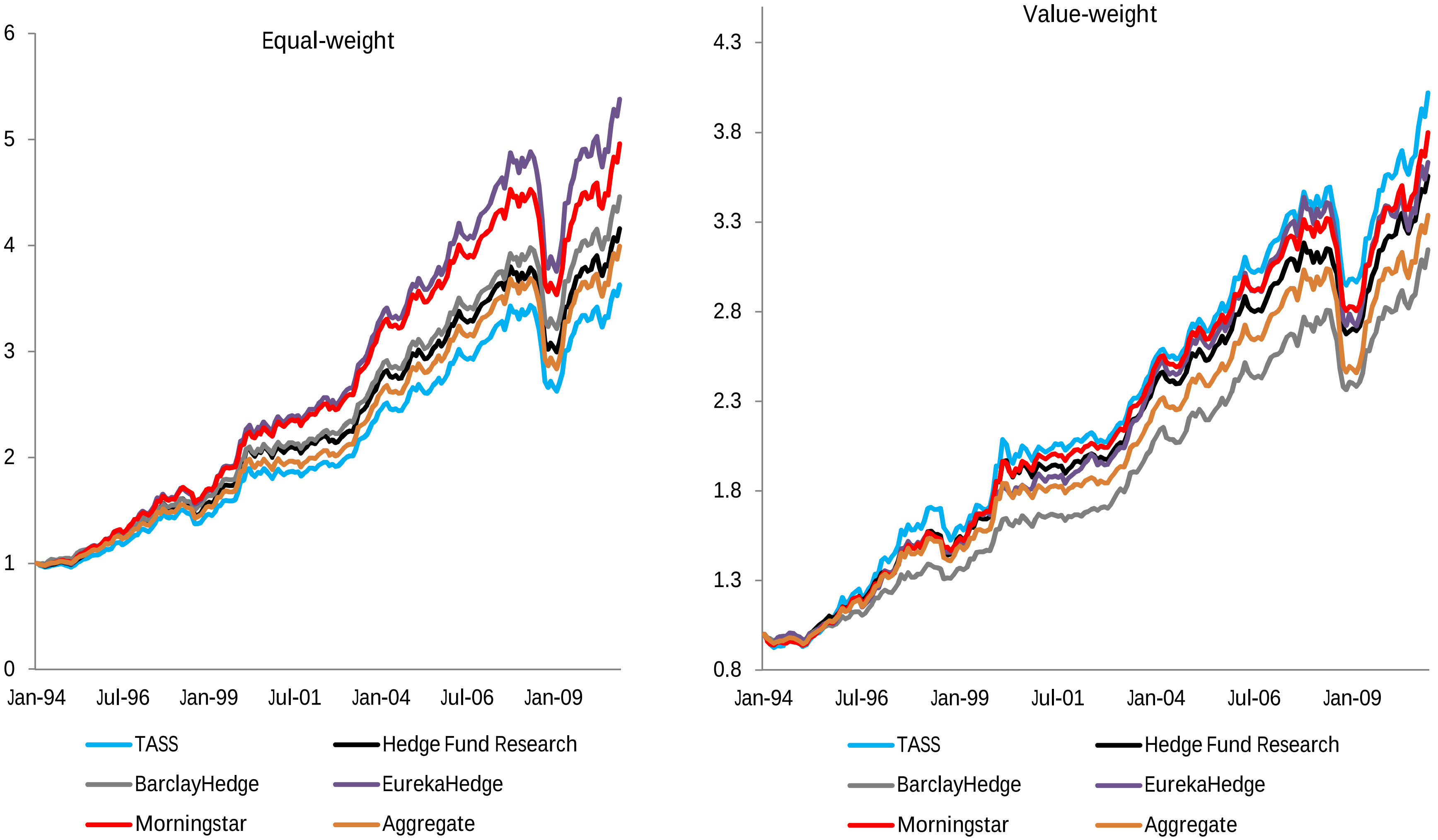

Figure 6

Equal- and value-weight cumulative abnormal returns of hedge funds [January 1994 – December 2010]

Figure presents equal- and value-weight cumulative abnormal returns of hedge funds for each database. Time series of abnormal returns are measured as the model's intercept (alpha) plus the time series of residual. All hedge funds are required to have at least 12 monthly non-missing return observations. Aggregate data is constructed using a statistical procedure (see Appendix for details). Abnormal return is based on the seven factor model proposed by Fung and Hsieh (2004). Risk factors are: the excess return of the S&P 500 index (SP-RF), the return of the Russell 2000 index minus the return of the S&P 500 index (RL-SP), the excess return of ten-year Treasuries (TY-RF), the return of Moody's BAA corporate bonds minus ten-year Treasuries (BAA-TY), the excess returns of look-back straddles on bonds (PTFSBD-RF), currencies (PTFSFX-RF), and commodities (PTFSCOM-RF).

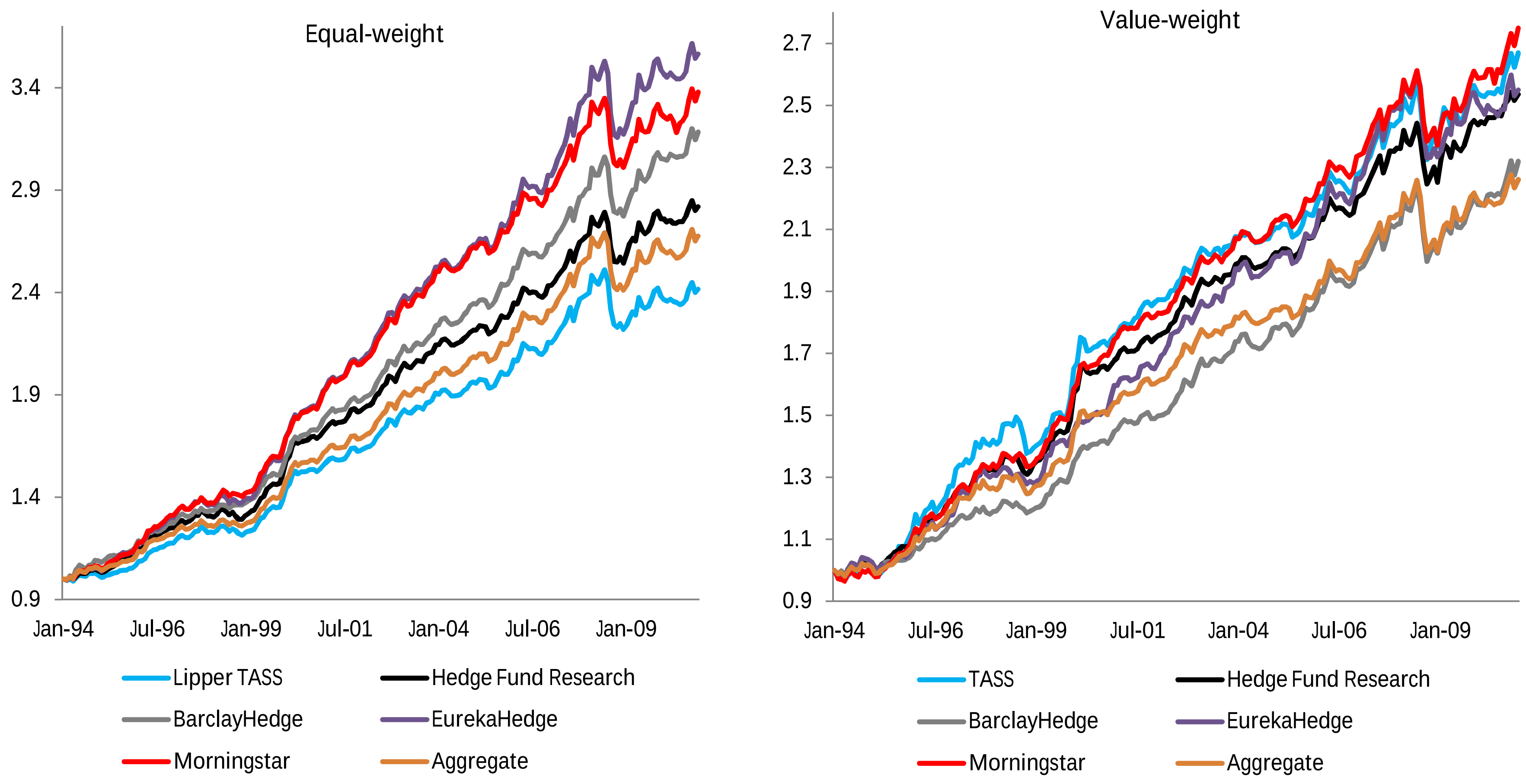

Figure 7
Equal-weight annual alphas of persistence portflios (January 1994 – December 2010)

Figure shows the annualized Fung and Hsieh (2004) alphas for persistence portfolios using four different holding periods. Top (bottom) portfolios are displayed as single (dashed) lines. Results are reported for the five databases that are used in the paper and all hedge funds are required to have at least 12 monthly return observations. The left y-axis shows the annualized Fung and Hsieh (2004) alphas in perentages and x-axis displays the holding period in months. Returns of persistence portfolios are equal-weighted monthly, so the weights are adjusted every month if there is funds that drop from the portfolio during the holding period.