New ‘Stylized facts’ about Hedge Funds and Database Selection Bias*

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; November 2012

Abstract

This paper presents new stylized facts about hedge fund performance and database selection biases based on a novel database aggregation. By highlighting economically important effects of database selection bias on previously documented results we aim to improve the ability of researchers in this literature to compare results across different studies. We carefully motivate and test a set of eight hypotheses regarding the impact of database selection biases on stylized facts. We document significant 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 databases. Performance persistence results are 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 multivariate analysis shows that funds imposing lockups do not deliver significantly higher risk-adjusted returns. Several stylized facts are sensitive to the choice of the database which highlights the importance of using a consolidated database that is more representative of the aggregate industry.

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

*We would like to thank, Turan Bali, Bill Fung, Bruno Gerard 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 thank Mikko Kauppila and Mikko Perttunen for helping us with the data processing. 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.
The usual disclaimer applies.

1. Introduction

This paper presents new stylized facts about hedge fund performance and database selection biases based on a novel database aggregation and a comprehensive analysis of differences between the main commercial hedge fund databases. We carefully motivate and test a set of hypotheses regarding the impact of database selection biases on stylized facts. By highlighting the effect of database differences on previously documented results we aim to improve the ability of researchers in this literature to compare results across different studies. Our eight hypotheses link database features such as differences in survivorship bias, attrition rates, percentage of missing return and assets under management information, coverage of cross-sectional variables such as share restrictions and incentive fees, for example, to key stylized facts such as average performance, performance persistence and the cross-sectional relationship between performance and various fund characteristics.

We show that these facts differ in an economically and statistically significant way between databases. Importantly, the stylized facts obtained using one database are often in contrast to results inferred from the consolidated database, which means that certain findings in the literature are sensitive to the choice of database. However, 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 the particular date of our data download and revisions in databases 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 several hedge fund studies build and use a large consolidated database containing multiple databases, there is, to the best of our knowledge, no standard merging methodology in the literature that could be used as a benchmark to help gauge the sensitivity of findings to the databases employed. 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. Our results can be useful in 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.[1] We argue that, for three reasons, the database selection is even more important in the hedge fund than in the mutual fund literature. 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 highly non-overlapping - we find that almost 70% of funds in our consolidated database report only to one of the 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.

Our aggregate database compares favourably with a recent study by Edelman, Fung and Hsieh (2012) that combines non-reporting fund information with three commercial databases. Our data set aggregates information from the BarclayHedge, EurekaHedge, Hedge Fund Research, Morningstar and TASS databases and consists of 30,040 unique hedge funds that report at least 12 monthly return observations. For these hedge funds, 12,283 are active, while 17,757 stopped providing any data to vendors and we classify them as defunct. Edelman, Fung and Hsieh (2012) gather data for non-reporting funds from a variety of private sources so that

[1] Edelman, Fung and Hsieh (2012) find that the bias between the commercially available databases and non-reporting funds is low.

they are able to identify 1,903 hedge funds with return observations that are not included to the commercial databases. Their study is very important for academic research using commercial hedge fund databases since they find that including non-reporting funds does not qualitatively change most insights about hedge fund performance. In 2012, PerTrac, one of the leading providers of analytical platforms for hedge fund analysis, reports that the hedge fund industry contains about 10,800 active funds. Based on these comparisons, we believe that our aggregate database is closest to the true unobservable population of hedge funds, and therefore we contribute to existing literature by investigating the impact of commercial database selection bias in hedge-fund industry.

Using our consolidated database, we first create new stylized facts about average hedge fund performance, performance persistence, and fund-specific characteristics explaining cross-sectional differences in hedge fund performance.

First, using our consolidated database, we provide evidence suggesting that hedge funds deliver superior average risk-adjusted performance. Specifically, for the aggregate equal-weight portfolio, we estimate an annualized average excess return of 7.8 percent and an annualized Fung and Hsieh (2004) alpha of 5.2 percent, a finding that is consistent with previous studies such as Kosowski, Naik and Teo (2007). The average excess returns (alpha) is lower for the aggregate value-weight portfolio, at 7.0 (4.6) percent per year. We document also significant Fung and Hsieh (2004) alphas across hedge fund strategies except Fund-of-Funds that suffer from a double-layered fee-structure.

Second, using the aggregate database, we conduct a series of tests to investigate whether a real-time investor is able to exploit the short-term hedge fund performance persistence. We document marginally significant performance persistence at annual horizons. After taking portfolio rebalancing possibilities into account so that a strategy is investible, we find performance persistence only at quarterly horizon. In addition, our tests show significant performance persistence for small hedge funds, but larger funds'persistence is much weaker which is consistent with the Berk and Green (2004) model. These results highlight the importance of make realistic rebalancing assumptions when conducting persistence tests.

Third, using the consolidated database, we examine cross-sectional differences in hedge fund performance. Our findings show that both smaller firms and funds outperform their large peers. In addition, we find the onshore hedge funds deliver higher performance than offshore registered funds suggesting that domicile effects also explain differences in average performance. Finally, hedge funds with greater managerial incentives deliver superior performance, while multivariate regressions reveal that funds imposing lockups do not provide significantly higher risk-adjusted returns compared to non-lockup funds. However, hedge funds with longer notice periods outperform suggesting that they are able to earn an illiquidity premium.

We next investigate whether these stylized facts inferred from the consolidated database, being a close proxy of unobservable hedge-fund population, are sensitive to the choice of commercial database. We demonstrate that more pronounce data biases related to hedge fund coverage and AuM reporting explain why we observe non-randomly different performance results across commercial databases.

First, we document that hedge fund coverage differences across commercial databases impacts on stylized facts about hedge fund performance. We start by documenting that the number of hedge funds ranges across data vendors from 7,502 for Morningstar to 10,520 for BarclayHedge. Importantly, the proportion of alive and defunct funds show us that BarclayHedge, HFR and TASS (EurekaHedge and Morningstar) contain relatively more (fewer) defunct funds than alive funds. In other words, the attrition rates are remarkably different across data vendors showing that EurekaHedge and Morningstar have very limited information about defunct funds before 2004. In contrast, BarclayHedge, HFR and TASS do not suffer from the same lack of data suggesting that the survivorship and backfilling biases are much higher in EurekaHedge and Morningstar than in other databases. These facts allow us to formulate hypotheses predicting that EurekaHedge and Morningstar should have higher average returns, but weaker performance persistence than the other databases. The rationale is that the EurekaHedge's and Morningstar's bottom deciles may not contain a large number of liquidated funds that deliver poor performance suggesting the spread portfolio between the top and bottom deciles may be indistinguishable from zero.

Using Q3 of 2012 versions of commercial databases, we demonstrate that the conclusions about average performance and its persistence depend on the choice of the data vendor. In terms of equal-weighted (EW) average performance, we show that EurekaHedge and Morningstar outperform BarclayHedge, HFR and TASS. As our hypothesis predicts, the result is driven by more pronounce backfilling and survivorship biases in EurekaHedge and Morningstar. We find significant evidence of short-term performance persistence using BarclayHedge, HFR and TASS. In contrast, we cannot document any evidence of persistence for EurekaHedge and Morningstar. Consistently with our hypothesis, the finding is driven by a large number of missing defunct funds in EurekaHedge and Morningstar, since we rule out the possibility that BarclayHedge, HFR, and TASS contain a set of 'high quality' funds that only report to their databases.

We next demonstrate how AuM reporting differences across commercial databases impact on stylized facts about value-weighed (VW) returns measuring the overall performance of hedge-fund industry. We find that about 30 percent of AuM observations are missing from our aggregate database. The proportion of missing AuM observations varies across data vendors, being lowest for BarclayHedge (12%) and HFR (19%), while significantly higher for EurekaHedge (36%), TASS (35%), and Morningstar (34%). Our findings suggest that average VW performance differs significantly across databases. TASS shows the highest VW average returns of 5.4 percent, being almost a 25 percent higher than the lowest respective counterpart for BarclayHedge. Interestingly, consistently to Ibbotson, Chen and Zhu (2011), we document that

TASS's VW performance is higher compared to respective EW performance. For other databases, in contrast we find that EW performance is higher than VW performance. We show that commercial data vendors' different kind of tendency to record stale AuM observations in their databases explain this striking pattern.[2] We find that VW performance of commercial databases is very similar across data vendors when we apply various specifications of stale AuM reporting standardization among data vendors. Indeed, TASS does not anymore show extreme VW performance that is higher than its respective EW performance.

Finally, we examine whether the results about the cross-sectional relationship between fund characteristics and hedge fund performance are sensitive to commercial database selection. Using portfolio sorts and the Fama and MacBeth (1973) regressions, we demonstrate robustly across commercial database that smaller and younger funds outperform their peers. In contrast, we document that our proxies related managerial incentives and illiquidity premium do not consistently explain hedge funds' cross-sectional returns. Using the TASS, we find the strongest evidence that strict share restrictions are associated superior performance, whereas using other commercial and even aggregate databases the conclusion is much weaker. In addition, we find that conclusion about importance of managerial incentives is database sensitive. For example, the significance of high-mark provision changes wildly across commercial databases suggesting that the conclusion about the impact of managerial incentives on hedge fund performance varies based on the used data vendor.

The paper is structured as follows. Section 2 relates the paper to the existing literature and develops the hypotheses that we test. Section 3 describes the data and methodology. Section 4 summarizes the stylized facts about average fund performance based on different databases. Section 5 reports stylized facts about performance persistence. Section 6 describes stylized facts about hedge fund performance and cross-sectional characteristics. Section 7 concludes.

[2] AuM observation is defined as a stale if it equals to previous month's observation.

2. Related literature and hypothesis development

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 is 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 stylized facts systematically differ among commercial databases. Indeed, they 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, 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 commerical databases making a conclusion about hedge fund performance.

Third, we contribute to the literature by examining effect of the database selection bias on the stylized facts of the hedge fund performance. 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 Ibbotson (1999), Agarwal and Naik (2000), and Liang

(2000)) find evidence of short-term persistence. We confirm results shown in previous studies by showing that (i) hedge funds add value on average, and (ii) performance of hedge funds persists at short horizon. We also document that persistence vanishes quickly when share restrictions are realistically taken into account.

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 funds and funds with greater 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.

Before discussing the data and methodology, we carefully motivate a set of hypotheses regarding how data selection biases affect stylized facts about hedge fund average performance, performance persistence and the cross-sectional performance fund characteristic relationship. Survivorship bias in the databases can differ depending on when the different databases started and how diligent the database vendors were in including defunct funds. Differences in survivorship bias lead us to our first hypothesis.

> *Hypothesis 1(survivorship bias, attrition and average performance): Significant differences in survivorship bias between different databases affects the average performance of funds since surviving funds tend to have higher returns than dead funds.*

Databases do not just different in the coverage of dead funds but also in the relatively coverage of small and large funds. There is evidence in the literature that small funds perform better than large funds. Therefore, we test the following hypothesis.

Hypothesis 2 (coverage of small funds and average performance): Databases with a more comprehensive coverage of small funds relative to large funds will have higher equal and value-weight performance than databases that have a lower proportion of small funds.

Databases may differ not just in the percentage of missing return, but also in the percentage of missing AuM observations for a given set of return observations. There may be significant differences between value and equal weight performance. This difference may not just be due to the size of funds in a given database (hypothesis 2 above) but also in the completeness of AuM information. One way to test the effect is to compare value-weight performance after filling in AuM observations that are missing for a given return observation by filling them in using the last reported (stale) AuM observations. If the value-weighted returns become more similar across databases this indicates that differences in missing AuM observations are driving average performance differences. Hypothesis 3 captures this reasoning:

Hypothesis 3 (missing AuM observations and value-weight performance): After filling in missing AuM observations for existing return observations by using the last reported AuM observation, the value-weighted average performance differences are reduced across databases.

Apart from average performance differences addressed by the hypotheses above, it is of great interest how style specific performance differs across funds, since every data vendor applies a different style or investment objective classification. Therefore differences in style classifications may affect average results. Hypothesis 4 tests for this:

Hypothesis 4 (hedge fund styles and average performance): There are significant performance differences by hedge fund style across databases.

Domicile effects have been shown to be different across countries. Many hedge fund studies that examine the relationship between performance and cross-sectional fund characteristics have however often not controlled for domicile of the fund or firm. Databases are different in their location classifications for both management firms and funds. Therefore we test hypothesis 5:

Hypothesis 5 (<u>domicile and fund performance</u>): Differences in databases' firm and fund domicile classification affect average performance results.

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 Ibbotson (1999), Agarwal and Naik (2000), and Liang (2000)) find only evidence of short-term persistence.

The coverage of small versus large funds in the databases may not just affect average performance but also performance persistence. Databases with survivorship bias are likely to contain a higher proportion of larger funds which tend to perform worse than small funds and may exhibit less performance persistence. In addition, backfilling bias in databases make difficult to separate skilled funds from unskilled, since some databases do not contain poorly performing funds. This leads to hypothesis 6:

Hypothesis 6a (<u>size distribution and performance persistence)</u> : The fund size distribution differs across funds affects performance persistence (since smaller funds outperform).

Hypothesis 6b (<u>Backfilling and performance persistence)</u>: Backfilling bias difference between databases' affects performance persistence (since unskilled defunct funds underperform).

Data selection biases in databases may not just affect average performance and performance persistence but also the cross-sectional relationship between fund characteristics and fund performance. Recent literature documents that hedge funds limiting investor liquidity are able to earn an illiquidity premium. Indeed, Aragon (2007) shows that hedge funds with strict share restrictions are deliver higher risk-adjusted return compared to funds allowing better liquidity terms to investors. In addition, Agarwal, Daniel and Naik (2009) document that hedge funds with better managerial incentives deliver superior performance.

The proportion of funds with share restrictions and incentive fees may different across databases: This leads us to the following additional hypotheses:

Hypothesis 7a (share restrictions, average performance and performance persistence): Databases with fewer funds that have share restrictions may show lower average performance since there is evidence of a liquidity/performance trade-off but may make portfolio rebalancing and performance persistence results practically more feasible.

Hypothesis 7b (cross-sectional share restrictions and performance relationship): Databases with fewer liquid funds may exhibit a weaker share restriction-performance relationship.

Hypothesis 8a (incentive fees and average performance): Databases with a lower percentage of funds with high incentive fees exhibit a lower average performance since high incentive fee funds may perform better.

Hypothesis 8b (cross-sectional incentive fee/performance relationship: These databases also exhibit a weaker performance/incentive fee relationship.

In the next section we discuss the data and methodology that we will use to test the hypotheses developed above.

3. Data and Methodology

3.1. Merging approach

In this paper, we propose an *industry standard* in constructing a consolidated hedge fund database. To construct a consolidated database, we merge five commercial hedge fund databases

(BarclayHedge, EurekaHedge, Hedge Fund Research (HFR), Morningstar, and TASS) consisting of over 60,000 share classes. This number does not provide a true picture of unique hedge funds because there is a significant duplication of information, as multiple providers often cover the same fund. Our merging approach is based on the transparent procedure that can be easily replicated almost automatically on a regular basis. We update our consolidated database each quarter making the data real time applicable. Easily replicable procedure implies that other researchers can follow it in constructing their own data set or even use the same aggregate database.

It is not a trivial task to merge several commercial hedge fund databases and to separate unique hedge funds from multiple share class structures. 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. Therefore, few of the existing papers provide transparent and detailed explanations of how their database is constructed. Notable exceptions are Patton and Ramadorai (2011) and Aggarwal and Jorion (2010). 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. To highlight issue, Bali, Brown, and Caglayan (2011) show that approximately 16% of share classes in TASS are duplicates that should be removed from the sample in order to conduct reliable research. Thus, it is important to remove duplicate share classes even if a study is based on only one of the databases.

We develop a statistical procedure that is used to separate unique hedge funds from the share classes. The goal of the procedure is find out which of the share classes employ exactly the same underlying investment process. Our merging methodology is based on the assumption that multiple share classes should exhibit highly correlated returns. We run therefore a statistical algorithm consisting of estimating correlation coefficients of each pair of share classes that exist within unique management firms. We classify correlated share classes into groups based on the correlation limit of 0.99. We select a main share class from each group of share classes to represent a unique hedge fund. Our criteria is as follows: we select the share class with (1) the longest return time series, (2) largest average AuM, (3) offshore domicile, or (4) USD currency. Online Appendix provides a detailed description of the methodology.

Figure 1 shows the Venn diagram describing the proportions of multiple share classes from the union of the five databases. We can report that over 67% of all share classes are covered by only one of the databases. Due to the differences in the coverage of share classes across databases, our universe of share classes provides a fertile setting to examine effects of database selection on hedge fund performance. The consolidated database has 11,217 unique management firms and 30,040 unique hedge funds obtained from the union of five databases. Figure 2 documents that the total reported AuM of single-manager hedge funds was approximately $2 trillion at the end of 2011.[5] Similar stylized facts of the total AuM are reported in latest papers and surveys (e.g., PerTrac and HFR).[6] This suggests that we can provide a reliable estimate of the industry size using the consolidated database.

3.2. Properties of databases

Due to the fact that a large proportion of hedge funds are covered by only one of the commercial databases, we carefully compare properties of commercial databases. To understand the differences in commercial databases alleviates us to investigate why conclusions about hedge fund performance can be sensitive to the commercial database selection.

It is essential to compare the coverage of defunct funds across databases because survivorship bias creates an upward bias in performance results (e.g., Ackermann, McEnally and

[5] We use end of December AuMs to estimate aggregate AuM of hedge fund-industry given that they most reliable. See discussion in Edelman, Fung and Hsieh (2012).

[6] PerTrac 2012 survey documents that the total reported AuM was about $1.892 trillion at the end of the year. HFR documents aggregate AuM of $2.01 trillion at 4Q 2011.

Ravenscraft (1999), Liang (2000), and Fung and Hsieh (2000, 2009)).[7] Panel A of Table 1 shows attrition rates defined as the ratio of the number of defunct funds to the number that existed at the start of the year. We find that the average attrition rate is almost zero from 1994 to 2002 in EurekaHedge and Morningstar databases, while it is over 8% in TASS, HFR, and BarclayHedge databases. Thus, EurekaHedge and Morningstar have a low coverage of defunct funds, and therefore, a large bias towards active funds. These findings are associated with our hypotheses 1 suggesting that EurekaHedge and Morningstar should outperform other databases in terms of average performance. Consistently with hypothesis 6a, performance persistence should be weaker in databases having a low number of defunct funds, because of difficulties to separate skilled funds from defunct unskilled funds. Hence, it is extremely interesting to test these hypotheses in the next sections.

Panel B of Table 1 provides summary statistics of time-series of returns and AuMs. BarclayHedge is the largest database in terms of the number of funds (10,520). EurekaHedge is the largest data vendor in terms of number (4,765) of active hedge funds. Table shows that EurekaHedge and Morningstar contain relatively large funds that have survived, but small dead funds are missing from their databases. This finding is associated with our performance persistence hypothesis 6b suggesting that differences in size distribution affects so that EurekaHedge and Morningstar should not exhibit relatively low persistence.

Panel B of Table 1 presents that significant amount of AuM observations for a given return observation is missing.[8] BarclayHedge (12%) and HFR (19%) deliver relatively comprehensive amount of AuM data, while EurekaHedge's (36%), TASS's (35%), and Morningstar's (34%) AuM coverage is significantly lower. However, when we standardize AuM data across databases

[7] For instance, according to Liang (2000), the difference in performance between surviving funds and all funds is 0.39% per year in HFR (1993-1997) and 2.24% per year in TASS (1994-1998). Consequently, HFR database outperforms TASS database due to smaller coverage of defunct funds.

[8] We calculate all AuM statistics conditional on the restriction of 12 non-missing returns for each fund.

by excluding the stale AuM observations,[9] it seems that differences between databases are not anymore so dramatic. Indeed, the amount of missing AuM data ranges from 46.51% (Morningstar) to 31.12% (BarclayHedge). We link the poor coverage of AuM observations to our hypothesis 3 about performance of hedge-fund industry as a whole. We expect that value-weighted average performance should be sensitive to database's AuM coverage.

Table 2 provides statistics of returns including normality, serial correlation and smoothing. Our overall findings show that statistical properties and misevaluation behavior of hedge funds are very similar across commercial databases. Recent papers (e.g., Agarwal and Naik (2004), Gupta and Liang (2005), and Lo (2001)) pose that hedge fund returns are frequently non-gaussian exhibiting unusual levels of skewness and kurtosis, and raise doubts on the validity of the standard deviation as the risk measure. In consolidated database, over half of the funds have non-normal returns (negative skewness and excess kurtosis) based on the Jarque-Bera test of normality (5% level of significance) and over one fifth of funds have serially correlated returns based on the Ljung-Box test (5% level of significance). Results of normality and serial correlation are similar across databases.

To test the fact that hedge funds misreport returns (e.g., Bollen and Pool (2008, 2009) and Agarwal, Daniel and Naik (2011)), we estimate model of conditional smoothing as well as misreporting measure proposed by Jylhä (2011). We report that 5.4% of funds exhibit higher first-order serial correlation when the returns are below the long-term average that is consistent with Bollen and Pool (2006). Estimated measures proposed by Jylhä (2011) reveal that defunct have higher estimates of smoothing than alive funds. Finally, we find evidence of the December Spike as documented by Agarwal, Daniel and Naik (2011). Difference in Fung and Hsieh (2004) alphas between average December and January-November values is statistically significant across databases after correcting for clustering at the fund-level.

[9] Stale AuM observations mean that the missing AuM observations are filled with previous non-missing AuM observations. BarclayHedge has the largest amount of stale AuM observations (19.13%).

4. Average performance

In this section, we provide new stylized facts of average hedge fund performance. Based on our carefully motivated hypotheses, we investigate whether the commercial database choice has an impact on the conclusion about average hedge fund performance.

4.1 Baseline

An important question is whether hedge funds add value on average after fees charged from the investors. A common approach to examine issue is to estimate the alpha or abnormal return – the value added (after fees and trading costs) not explained by exposures to common systematic risk factors. As a benchmark model in our performance analysis, we use the Fung and Hsieh (2004) seven-factor model that is the standard workhorse in hedge fund performance evaluation studies. We regress the net-of-fee monthly returns (in excess of risk-free rate) of a hedge fund portfolio *i* $(R^{e}_{i,t})$ against buy-and-hold equity- and bond-orientated as well as primitive trend-following factors

$$R^{e}_{i,t} = \alpha_i + \sum_{k=1}^{7} \beta_{i,k} f_{k,t} + e_{i,t}\,, \qquad (1)$$

where these *k* 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*). The intercept (α_i) is defined as the Fung and Hsieh (2004) alpha providing an estimate for hedge fund portfolio *i*'s average abnormal performance.

Panel A of Table 3 provides the stylized facts about average hedge fund performance inferred from the consolidated database. Consistent with previous studies (e.g., Liang (1999), Fung and Hsieh (2004), and Kosowski, Naik, and Teo (2007)) we confirm that hedge funds add positive value even after fees. We document economically and statistically significant equal-weighted (EW) Fung and Hsieh (2004) alpha in terms of net-of-fees and gross-of-fees returns, 5.23% and 10.77% per year. The average fee (5.54%) that investors pay to beat the market is really extraordinary. As a reference, using the HFR database from 1996 to 2007, French (2008) concludes that the average fee investors pay is 4.26%.

We document that superior average EW performance even after adjusting for backfilling bias and return smoothing.[10] Following Malkiel and Saha (2005), we define the fund-level backfilling period as the difference between the date when the fund was added to database and its inception date.[11] We find that EW alpha is considerably lower, but remain significant after returns are adjusted for backfilling bias. After smoothing returns using the Getmansky, Lo and Makarov (2004) algorithm, we find that standard deviation is higher, but we still document a significant Fung and Hsieh (2004) alpha, while Share ratio is considerably lower given the adjusted volatility estimate.

Building on the work of Fung, Hsieh, Naik and Ramadorai (2008), we estimate the average performance for subperiods. We find across subperiods that hedge fund average performance is time-varying. Consistent with Aiken, Clifford and Ellis (2012), we document small and only marginally significant average alpha for the last sub sample from January 2005 to December 2011 equaling to 3.68% with *t*-value of 1.69.

---

[11] The average backfilling period across all databases is 32 months. We exclude therefore 32 months of returns from fund-level time series to control for backfill bias. Table A1 provides details of backfill adjustment.

Panel A of Table 3 shows hedge-fund industry as a whole has deliver superior average performance. We document that the VW alpha is 4.64% per year being economically and statistically significant. The difference between EW and VW alphas suggests that small funds outperform large funds that mirror results of Teo (2010). The finding is robust, since we report almost as high VW alpha estimate after adjusting for AuM stale observations. Specifically, we fill in each missing AuM observations using the previous non-missing AuM observation (if available).[12]

We next examine how commercial database selection impacts on hedge fund average performance. As first evidence, Figure 3 plots the cumulative excess returns of EW and VW portfolios across databases showing the first supporting evidence for our data bias hypotheses explaining differences in average performance between databases. Consistent with findings of attrition rates in Panel A of Table 1, in terms of cumulative EW returns, EurekaHedge and Morningstar outperform TASS, HFR, and BarclayHedge. Figure 3 also indicates that TASS's relative ranking changes dramatically, since its cumulative VW returns are highest among the databases, but respective EW returns are lowest. Other databases behave consistently in terms of their EW and VW rankings. Overall, Figure 3 suggests that the database choice and reporting behavior of AuM data affects to the hedge fund performance.

Panel B1 of Table 3 compares EW average performance estimates and Fung and Hsieh (2004) risk exposures across commercial databases. We find clear evidence that EurekaHedge and Morningstar outperform other databases in terms of average EW performance. Since Sharpe ratios, risk loadings and R^2's of the Fung and Hsieh (2004) seven-factor model are very similar across commercial databases, the average performance differences across databases cannot be

[12] We implement various specifications to fill in AuM observations and find very similar results. We opt to use simple method without any interpolation due to potential look-ahead bias.

explained by risk exposures.[13] Hence, these differences across databases are driven by the different levels of survivorship bias in the commercial databases. Consistent with our hypothesis regarding to the equal-weight average performance are the highest (lowest) for EurekaHedge and Morningstar (BarclayHedge, HFR and TASS), which has the lowest (highest) amount of defunct funds.[14]

We finally examine our third hypothesis how AuM reporting differences are associated with value-weighted average performance among commercial databases. Panel B2 in Table 3 shows that TASS delivers the highest value-weighted average returns of 5.4 percent per year. Other mature databases, BarclayHedge and HFR, provide significantly lower VW performance than TASS. Consistently to previous EW results, Morningstar and EurekaHedge deliver relatively high VW performance due to more pronounce survivorship bias. In each of the databases except TASS, average EW performance is always higher than respective average VW performance. TASS results are consistent with Ibbotson, Chen and Zhu (2011) showing that TASS's VW performance is higher that its EW performance. In contrary, using consolidated database, we find that EW performance is higher than VW performance. To understand why TASS's VW results are not consistent, we generate average VW results using the stale AuM standardization.[15] After VW returns are calculated using adjusted for stale observations, we find that TASS behave in a similar way as other mature databases (BarclayHedge and HFR). Hence, consistently with our hypothesis 3, the low coverage of AuM observations and other databases' higher tendency to record state AuM observations in their database's explain VW performance

[13] 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.

[14] Table A2 reports that after returns across commercial databases are adjusted for the backfilling bias and return smoothing, results of the average performance rank databases similarly.

[15] In unreported robustness tests, we execute various interpolation and extrapolation techniques to build AuM time-series. We find that results very similar for each of the techniques. We opt to simple way to fill in AuM observations in order to avoid look-ahead bias and unnecessary complexity.

differences even between mature commercial databases.

4.2. Size

We next investigate how fund and firm size explain average hedge fund performance. In recent literature of active portfolio management, the result of declining performance with fund size is connected to capacity constraints, holdings of illiquid securities, and organizational diseconomies related to hierarchy costs (Teo (2010), Chen, Hong, Huang and Kubik (2004)). From theoretical motivation, Berk and Green (2004) set an equilibrium model showing that funds with positive alphas face costs that are an increasing convex function of fund size. A fund with positive alpha received inflows until its size reaches the point where expected alpha, net of costs, is zero. In their equilibrium, all active funds have positive expected alpha before costs and zero expected alpha net of costs.

Table 4 shows that small funds and firms outperform the large ones. Using the aggregate database, we conduct our analysis by sorting hedge funds (firms) into portfolios based on the nominal AuM limits. We then estimate performance measures using the monthly rebalanced size portfolios. Panel A (Panel B) shows results for sorts based on fund-level AuM (Firm-level AuM). The Fung and Hsieh (2004) alpha of the small funds (AuM is less than $10 million) is 6.47% per year (*t*-statistic = 5.93) while the large funds (AuM larger than $1 billion) have the alpha equaling to 1.67% (*t*-statistic = 1.05). For the firm-size portfolios, the alphas of small and large funds are 7.67% and 1.51% per year, respectively. The difference in the abnormal performance between small and large funds is economically and statistically significant. Our findings are consistent with Berk and Green (2004) since the average alpha of hedge funds decreases with size.

We document quantitatively similar results across the commercial databases suggesting that there is a strong relationship between size and performance. We do not find any significant size distribution differences between databases. Consistently with our hypothesis 2, we show that size and performance relationship is very similar across commercial database.

Figure 4 presents the results of an experiment that we conduct in the spirit of Dichev and Yu (2011). To estimate returns what investors really earn, Dichev and Yu (2011) focus on so-called dollar-weighted instead of time-weighted returns. The rationale is the fact that dollar-weighed returns can be interpreted as reflecting the ability of investors to time their investments into hedge funds better than simple time-weighted returns would do. Specifically, at December 2011, we sort funds into nominal dollar groups as described in Table A3 and use the full sample of excess returns to form a monthly rebalanced EW portfolio. We form also percentiles of the number of funds that belong in each of the nominal size groups. We apply these percentile limits, sort hedge funds into portfolios every December using the respective AuM observation. We finally estimate size portfolios' Fung and Hsieh (2004) alphas that are referred as "Backward-looking" and "Forward-looking" alphas. Figure 4 show that Forward-looking alphas support the view that small funds clearly outperform large ones with a significant spread in alphas. Backward-looking alphas suggest that large funds outperform in the end-of-the sample: only the most successful large funds continue reporting to the database. This shows us that some outperforming hedge funds grow very fast over the time, but they cannot anymore deliver superior performance as Berk and Green (2004) model predicts. Our findings are also consistent with Dichev and Yu (2011) showing that investors could improve their timing ability in allocating to funds.

4.3 Strategies

We turn next to the average performance of hedge fund strategies by examining whether hedge funds grouped by investment objective add value on a net-of-fees basis. 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. Table

A10 shows the proportions of funds grouped by the strategies. We find that commercial databases are similar in terms of proportion of fund following a specific investment strategy.

Table 5 presents the average performance of hedge fund strategy based on our aggregate database. Hedge funds generate economically and statistically significant risk-adjusted returns across investment strategies except Fund-of-Funds. The annualised Fung and Hsieh (2004) alphas range from 8.08% (Sector) to 0.70% (Fund-of-funds[17]). All strategies have statistically significant exposure to the total stock market factor. For instance, Long/Short strategy's the market beta is 0.45. Relative value funds have positive loadings to bond factors: 0.13 (TY-RF) and 0.38 (BAA-TY). The bond and FX PTFS factors provide some explanatory power for CTA funds. Table A4 presents results of strategy-level Fung and Hsieh (2004) alphas across size categories. Within each strategy, funds are sorted every December based on the monthly AuM. Time series of excess returns of portfolios are calculated using 12-month holding period and monthly rebalancing. According to results, in every style group except fund of funds, small funds outperform large funds. For example, in Emerging Markets category, small (large) funds exhibit the average alpha of 7.9% per year (0.53% per year). We document also that performance of style indexes differs across databases. For example, Table A5 shows that performance for emerging markets index ranges from 4.13% per year to TASS and 11.01% to Morningstar. Typically, EurekaHedge and Morningstar outperform that is consistent with hypothesis 1.

4.4. Domicile

Aragon, Liang, and Park (2011) documents that onshore hedge funds registered in USA deliver higher average performance than the registered in offshore locations. The domicile regions of hedge funds and management firms are divided to two groups: (1) onshore; and (2)

[17] We merge five databases containing fund of funds in a similar way as we merge single hedge funds.

offshore. United States and Canada are classified as onshore regions[18]. Other domicile regions are classified into four groups: (1) Asia and Pacific; (2) Caribbean; (3) Europe; (4) Rest of world. Table 10 shows the proportions of funds grouped by the fund-level domicile region. In BarclayHedge database, 46% of funds are onshore funds. In other databases, most of the funds are domiciled in Caribbean (38% in the aggregate database). Overall, the proportion between onshore and offshore funds is similar across commercial databases.

Table 6 presents the number of unique hedge funds and management firms in domicile regions (Panel A) and the average performance by hedge fund domicile (Panel B).[19] Most of the firms are established in North America (62%) and most of the funds are domiciled as offshore Caribbean vehicles (36%). Hedge funds that are established in the onshore (Europe) account for 35% (18%) of all hedge funds (19,490). Hence, there are a significant number of onshore management firms that establish both onshore and offshore hedge funds.

Panel B in Table 6 shows significant differences in average hedge fund 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 6.91% per year (4.17% per year). Europe based hedge funds deliver the poorest performance (alpha equals to 2.51% per year). Asia-Pacific domiciled funds deliver the highest average alpha, being 7.20 % with the *t*-statistic of 3.6. Table A6 shows results of average performance grouped by firm domicile region. We find that onshore funds outperform offshore based funds. Table A7 shows results of average performance grouped by fund domicile across databases. We find that onshore funds outperform offshore funds consistently across databases.

---

[18] HFR database reports a dummy variable Offshore_Vehicle (1 for offshore and 0 otherwise). We find most of the funds having described as offshores vehicles are legally established in North America. Therefore, we classify funds that legally established in North America as onshore funds.

[19] We also examine domicile performance using commercial databases, but not find significant performance differences across domiciles.

5. Performance persistence

The previous section shows that an average hedge fund delivers superior risk-adjusted performance. If some hedge fund managers have access to superior information, sorting funds on past performance should indicate whether, on average, past winners are 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 notice, and redemption periods.[20] All these restrictions indicate that new investors are not able to withdraw capital from hedge funds in a timely fashion. Therefore, hedge funds that are able to add value after fees consistently is a rewarding feature for investors.

We investigate performance persistence using standard methodology. In the spirit of Carhart (1997), we sort hedge funds into decile portfolios based on their past Fung and Hsieh (2004) seven-factor alpha *t*-statistics that are estimated over the prior two years data. Given superior statistical properties of the alpha *t*-statistic, the performance persistence is expected to be stronger than in case we sort on fund alpha.[21] We use three different portfolio rebalancing periods: (i) quarterly, (ii) semiannual, and (iii) annual. Across rebalancing horizons, we calculate buy-and-hold returns for each of the decile portfolios.[22] Thereafter, we estimate the Fung and Hsieh (2004) spread between the top and the bottom decile portfolios. 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 funds dropping out from

[20] 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.

[21] Funds with a short history of monthly net returns will tend to generate alphas that are outliers. The alpha *t*-statistic provides a correction for outliers by normalizing the fund alpha by the estimated precision of the fund alpha (e.g. Kosowski Timmermann, Wermers and White (2006), Kosowski, Naik and Teo (2007)).

[22] If a fund stops reporting during the holding period, the fund is assumed to be in liquidation and the proceeds are reinvested in Treasury bills. Results provided in Table A8 shows that our conclusions are not sensitive to our assumptions of performance persistence tests.

the underlying decile portfolio during the holding period. We expect high dropout rates for the bottom decile given that the poorly performing funds either dead or stop reporting to commercial databases. Since EurekaHedge and Morningstar cover a larger portion of active funds than other databases, our hypothesis 6 suggests that their bottom deciles should contain relatively few defunct funds, and therefore exhibit low dropout rates. Therefore, we our hypothesis 6 suggests relatively low performance persistence should be inferred from these two databases.

5.1. Stylized facts about performance persistence

Using consolidated database, our overall findings reveal that EW and VW portfolios' performance persists at short horizons Small hedge funds show persistence even at annual horizon, but short-term persistence is difficult to exploit due to share restrictions. Our main findings confirm the pioneering literature (e.g., Agarwal and Naik (2000), Brown, Goetzmann, and Ibbotson (1999), Liang (1999), Baquero, Horst, and Verbeek (2005)) that document short-term persistence. Using 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. We highlight the importance of accounting share restrictions of hedge funds in persistence tests.

For the aggregate database, Panel A of Table 7 shows that the top decile portfolio generates an annualized EW alpha of 6.02% (10.27%) for an annual (quarterly) holding period. Using quarterly holding period, the aggregate database has the alpha for the spread between top and bottom equaling to 7.22% ($t = 2.59$) which is statistically different from zero at 5% level of significance suggesting that the sorting based on the alpha *t*-statistic separates good from bad funds. For longer holding periods, the alpha for the spread portfolio is only marginally significantly different from zero.

We find that the top decile portfolio generates a VW alpha of 4.22% (8.06%) for an annual (quarterly) holding period. The alphas of VW are lower than the EW alphas suggesting that small funds outperform large funds (Table 4). Using quarterly holding period, the aggregate database has the alpha for the spread between top and bottom equaling to 7.88% ($t = 2.75$) which is statistically different from zero at 5% level of significance.

The difference in performance between EW and VW top extremes suggests that the fund size affects persistence results. In Panel B of Table 7 we examine this issue further and show persistence results for the aggregate database within size terciles. The top decile portfolio of small (large) funds generates an EW alpha of 8.91% (4.83%) at an annual holding period. Small and large funds show statistically significant spread alphas for semiannual and annual holding period. Using an annual holding period, small (large) funds have the spread alpha equaling to 6.36% with *t*-statistic of 3.61 (2.31% with t=1.66). Consistent with Boyson (2008), we document that small funds exhibit the strongest performance persistence. However, it is interesting to note that the largest hedge funds are able to deliver statistically significant alphas that even show some persistence.

Using the aggregate database, we conduct a series of performance persistence tests to investigate whether a real-time investor is able to exploit the short-term performance persistence. Within each rebalancing horizon, we conduct persistence tests only using the feasible information so that portfolio can be rebalanced in practice. Test exploits only the information set available at the moment when the investor makes her fund selection decision. 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 all funds that do not provide quarterly liquidity for investors are dropped out from the analysis. In addition, we estimate post-rank alphas taking realistically notice periods into account so that we mitigate look-ahead bias. For example, if the fund has 90 days advance notice, then its ranking is based on alpha that is estimated using the 3-month lagged return information.

For our feasible strategies, Panel C of Table 7 shows that the top decile alphas are positive and statistically significant at 5% level across investors' different liquidity terms. However, we cannot document significant performance persistence, since the Fung and Hsieh (2004) alpha spreads are almost consistently insignificant across investor liquidity levels. Specifically, the feasible strategies that provide quarterly liquidity deliver the highest top decile alpha and the most significant performance persistence. Based on the 1-month notice period, the quarterly rebalanced feasible strategy's top decile alpha is 8.28% (t=3.84), being slightly lower compared to the baseline strategy's respective alpha of 10.27 (t=5.26). Furthermore, we find that the alphas of the spread portfolios are 6.36% (t=1.91) for the feasible strategy and 7.76% (t=2.91) for the baseline strategy. Hence, we can document quarterly performance persistence even after controlling properly the role of share restrictions. However, our persistence test cannot separate good from bad for longer horizons suggesting that performance persistence tests are very sensitive to the realistic assumptions.

5.1. Database selection and persistence

We next investigate whether the conclusion about performance persistence is sensitive to commercial database selection. Figure 5 highlights that the conclusion about the performance persistence varies significantly across commercial databases. Figure shows that BarclayHedge, HFR and TASS reveal significant performance persistence, while EurekaHedge's and Morningstar's performance seem not to persist. Indeed, EurekaHedge's and Morningstar's bottom portfolio alphas are significantly higher compared to respective alphas in other commercial databases. On the other hand, top decile alphas are very similar across commercial databases. These findings provide first support for our hypothesis 6 suggesting that databases containing larger funds and higher survivor and backfilling biases should reveal relatively lower performance persistence.

According to Table 8, the conclusion about hedge fund performance persistence varies significantly across commercial and aggregate databases. We cannot document any evidence about 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 and weighted basis, Morningstar and EurekaHedge do not reveal any performance persistence. Indeed, we find that their spread portfolio's post-rank alphas are indistinguishable from zero even with quarterly portfolio rebalancing. In contrast, for TASS, HFR, and BarclayHedge, the results show performance persistence at quarterly horizons.

To investigate why performance persistence results differ across databases, we calculate dropout rates. Our findings provide support for our hypothesis 6, since we find that dropout rates are remarkably wider for mature databases than to younger databases. At annual holding period, Table 8 presents that the difference in dropout rates between top and bottom portfolios is 10.84% (BarclayHedge), 11.69% (HFR), and 13.03% (TASS) for mature databases, whereas they are significantly lower 5.02% (Morningstar), and 5.67% (EurekaHedge) for younger databases. This is consistent with the findings that EurekaHedge and Morningstar databases have almost zero attrition rates for 1994-2004 periods if compared to other databases.

6. Hedge fund performance and fund characteristics

In this section, we examine how hedge fund-specific characteristics explain cross-sectional differences in fund performance. First, we test whether hedge funds imposing strict share restrictions are able to earn an illiquidity premium. Second, we test whether hedge funds with greater managerial incentives are associated with superior performance.

To understand how fund-specific characteristics explain fund performance, we conduct univariate sorts across databases. We calculate buy-and-hold portfolio returns using 12-month holding period, and then estimate the spread between top and bottom categories in order to examine whether the average risk-adjusted returns differs between the extreme realizations of a specific fund characteristic.

We first create sorts based on the length of lockup and notice periods. The previous studies document that hedge funds with strict redemption restrictions are associated with superior performance. Indeed, Aragon (2007) argues that share restrictions allow hedge funds to manage illiquid assets and earn an illiquidity premium. Using TASS for January 1994 to December 2001, Aragon (2007) shows that hedge funds with a lockup period deliver approximately 4% higher annual risk-adjusted returns than their peers. In addition, using multivariate regressions, Aragon (2007) and Agarwal, Daniel, and Naik (2009) find that share restrictions are important in explaining cross-sectional difference in hedge fund performance.

To test how hedge fund managerial incentives are associated with superior performance, we next form portfolios based on incentive fee and high-water mark categories. The prior literature on managerial incentives suggests that there is a positive relation between compensation structure variables and hedge fund performance. Specifically, 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 a high-water provision and greater manager's option delta deliver superior performance.

Using the multivariate Fama-McBeth (1973) approach, we thereafter 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

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

where $R_{i,t}$ refers to the Fung and Hsieh (2004) alpha of a hedge fund i at the time t, λ_1 is a vector representing the slope coefficients for time-invariant characteristics containing management and incentive fees, high-water mark provision and share restrictions in the form of lockup, notice, and redemption periods, and λ_2 is a vector representing the slope coefficients for time-variant characteristics, which control for the role of fund size, flow and age that are found to be important by Teo (2010), Teo (2011) and Aggarwal and Jorion (2010).We control for the strategy and domicile fixed effects, and adjust standard errors for autocorrelation and heteroskedasticity following Newey and West (1987).[23]

We start by investigating whether hedge funds imposing strict share restrictions are associated with superior performance. Using the consolidated database, Table 9 shows that there is a monotonically increasing relationship between share restrictions and Fung and Hsieh (2004) alphas. However, we can observe that the conclusion about relationship between share restrictions and performance differ significantly across commercial databases. Panel B of Table 9 shows that differences in lockup periods seem not to be important for EurekaHedge, but we can infer to four other databases that funds with longer lockups outperform. In addition, Panel A shows that notice period seems not to explain consistently cross-sectional differences in hedge fund alphas. Using BarclayHedge and TASS, we document a marginally significant alpha spread between the top and bottom notice period portfolios, while three other commercial databases show insignificant respective alpha spread. Consistent with our hypothesis 7b, finding is driven by the fact that these databases contain fewer 'liquid' funds providing a daily liquidity. We find that funds having over 60 days notice period deliver similar level of average alphas across commercial databases. In contrast, 'liquid' funds in HFR, Eureka, and Morningstar databases have higher alphas compared to TASS and BarclayHedge. This may be associated with the fact that average notice periods are higher in HFR, Eureka, and Morningstar compared to ones

[23] 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.

obtained using TASS and BarclayHedge. To summarize, the conclusion about the impact of and lockup and notice periods on hedge fund performance varies significantly across databases, but using consolidated databases we find support for well economically motivated hypothesis by Aragon (2007) suggesting that funds with strict share restrictions can earn illiquidity premium cannot be rejected based on univariate sorts.

However, we cannot conclude that strict share restrictions are consistently associated with superior performance when the role of other fund characteristics and fixed domicile effects are properly controlled for. Across databases, we find that parameter estimates for lockup and notice periods are almost consistently lower when we take domicile fixed effects into the account. The inclusion of domicile fixed effects is motivated by Aragon, Liang, and Park (2011) documenting that US domiciled funds impose lockups based on regulatory reasons, not necessarily economically reasons associated with illiquidity premium. However, as they argue US regulation does not suggest that US domiciled funds should impose longer notice periods. This view is supported by results obtained using the consolidated database. We find that the coefficient of lockup period is insignificant when the role of domicile effects are controlled for, but the coefficient for notice period remains statistically significant suggesting that funds having a long notice period are able to earn illiquidity premium even after the role of fixed domicile effects is taken into the account. We find the strongest evidence using the TASS database, whereas using other databases and even the aggregate database the conclusion is much weaker.

Using the consolidated database, our stylized facts show that hedge funds with greater managerial incentives tend to outperform. According the Panel C of Table 9, there is a monotonically increasing relationship between incentive fee and average Fung and Hsieh (2004) alpha. In addition, hedge funds that impose a high-water mark provision deliver significantly higher average Fung and Hsieh (2004) alphas compared to funds that do not use high-water mark provisions. Multivariate results presented in Table 10 show that even after we control for the role of other variables hedge funds with higher managerial incentives are associated with superior performance. Hence, our results suggest that hedge funds with greater managerial incentives as proxied by incentive fee and high-water mark provision may help to align the incentives of investors and managers.

We next show that this conclusion is sensitive to the commercial database selection due to the fact that the coverage of different types of hedge funds varies across databases. We can observe from Table 9 that the average Fung and Hsieh (2004) alpha spread between incentive fee top and bottom portfolios is only marginally significant for TASS, EurekaHedge, and Morningstar. Consistent to hypothesis 8, these three databases contain a lower portion of high incentive fee funds than BarclayHedge and HFR. Sorts also show that hedge funds imposing a high-water mark provision deliver only marginally higher risk-adjusted performance if we base our inference on a sample drawn from EurekaHedge. This finding is associated with the fact that EurekaHedge reports the highest proportion of funds imposing a high-water mark provision. Indeed, Table 9 shows that the Fung and Hsieh (2004) alpha spread is the highest (lowest) for commercial databases having the lowest (highest) proportion of funds with high-water mark.

The multivariate results provide more evidence that results are sensitive to the commercial database selection. Table 10 shows that slope coefficient for incentive fee is insignificant for TASS, but significant for other four commercial databases. In addition, the sign and significance for the high-water mark is changing wildly across commercial databases but remains significantly positive for the consolidated database. Specifically, the coefficients for high-water mark are only consistently significant for TASS, but not for other four databases. Indeed, we can observe that coefficients are even negative for EurekaHedge and Morningstar, which is contrast to the conclusion that we inferred using the consolidated database. We run cross-sectional regressions using fund-level excess returns and report results in Table A9. Findings are consistent with Table 10 reporting that effects of incentive fee and share restriction on performance are database specific.

7. Concluding remarks

Using a novel database aggregation and a comprehensive analysis of differences between the main commercial hedge fund databases (BarclayHedge, EurekaHedge, HFR, Morningstar, and TASS), we highlight the effects of database differences and biases on previously documented ‘stylized facts’ such as hedge funds’ (i) average performance, (ii) persistence, and (iii) the cross-sectional relation between fund-level characteristics and risk-adjusted returns. We show that documented ‘stylized facts’ of hedge fund performance are sensitive to the choice of the database and, thus, any single commercial database may lead to biased conclusions since it only a small portion of the hedge fund population.

Our study of databases shows differences in coverage of returns, assets under management, and number of funds in the graveyard module (attrition). The comparison of individual 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 stylized facts based on the consolidated database are qualitatively different from those based on the individual databases. We show how different conclusions regarding value-weight average performance as well as performance persistence can be traced back to differences in AuM coverage as well as survivorship bias and the coverage of small funds and large funds. There is evidence of performance persistence in the aggregate database and some of the individual data bases, but not all of them. Similarly, the incentive-performance and liquidity-performance tradeoffs are sensitive to the choice of data base since some databases do not contain the same information on share restrictions as others.

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., Naik, N.Y., 2011. Do hedge funds misreport their reported returns? Review of Financial Studies. 24 (10).

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

Agarwal, V., Fos, V., Jiang, W., 2010. Inferring Reporting-Related Biases in Hedge Fund databases from Hedge Fund Equity Holdings. Forthcoming in Management Science.

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

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

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

Aiken, A. L., Clifford, C. P., Ellis, J. A., 2012. Out of the dark: Hedge fund reporting biases and commercial databases. Review of Financial Studies. Forthcoming.

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

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

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

Baquero, G., Horst, J., Verbeek, M., 2005. Survival, look-ahead bias, and the persistence in hedge fund performance. Journal of financial and quantitative analysis, 40 (3), 493-517.

Bali, T.G., Brown, S.J., Caglayan, M. O., 2011. Do hedge funds' exposures to risk factors predict their future returns? Journal of financial economics. 101, 36-68.

Berk, J.B., Green, R.C., 2004. Mutual fund flows and performance in rational markets. Journal of political economy. 112 (6), 1269-1295.

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

Boyson, N., M., 2008. Hedge fund performance persistence: a new approach. Financial Analyst Journal. 64 (6).

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.

Chen, J., Hong, H.G., Huang, M., Kubik, J.D., 2004. Does fund size erode mutual fund performance? The role of liquidity and organization. American Economic Review, 94 (5).

Dichev, I.D., Yu, G., 2011. Higher risk, lower return: What investors really earn. Journal of Financial Economics. 100 (2), 248-263.

Edelman, D., Fung, W., Hsieh, D.A., 2012. Exploring uncharted territories of the hedge fund industry: empirical characteristics of mega hedge fund firms. Alternative investment solutions, London Business School, and Fuqua School of Business, Duke University.

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

French, K., 2008. Presidential address: the cost of active investing. Journal of Finance. 63 (4), 1537-1573.

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., 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.

Gupta, A., Liang, B., 2005. Do hedge funds have enough capital? A value-at-risk approach. Journal of Financial Economics. 77 (1), 219-253.

Güner, B., Rachev, S.T., Edelman, B., 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.

Harris, R.S., Jenkinson, T., Kaplan, S.N., 2012. Private equity performance: what do we know? University of Virginia, University of Oxford and University of Chicago.

Ibbotson, R. G., Chen, P., Zhu, K. X., 2011. The ABCs of hedge funds: alphas, betas and costs. Financial Analyst Journal. 67 (1).

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.

Lo, A.W., 2001. Risk management for hedge funds: Introduction and overview. Financial Analyst Journal, 57 (6), 16-33.

Malkiel, B. G., Saha, A., 2005. Hedge funds: risk and return. Financial Analyst Journal. 61 (6), 80-88.

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., 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. Attrition rates and summary statistics

Table shows the yearly attrition rates for hedge fund databases. Aggregate database is a merged database using a correlation algorithm that separates unique investment portfolios from multiple share classes. Details of the methodology are described in Appendix. Panel A shows yearly attrition rates that are measured as the ratio of the number of dissolved funds to the number that existed at the start of the year. Panel B shows summary statistics of excess returns and asset under management. Table reports the number of unique management firms (# Firms), hedge funds (# Funds) and share classes (# Share classes). Columns Missing (%) and Length show the number of missing excess returns and the average length of fund-level time series. For fund-level AuM time series, Panel B shows the cross-sectional averages of the fund-level average AuM (in millions of US dollars) and the length of the fund-level time series. Conditional on the fund-level non-missing excess returns, Panel reports the total number of missing AuM observations in column Original. This value is further divided into three parts: (1) before the first non-missing AuM observations (Begin (%)), (2) in the middle of the first and last non-missing AuM observations (middle (%)), and (3) after the last non-missing AuM observation (end (%)). Column Stale (%) refers to the amount of stale AuM data (hedge funds have reported the same AuM for two consecutive months or more). Finally, column Total shows the number of missing AuMs after stale observations are excluded.

A. Attrition rates

	Aggregate			TASS			Hedge Fund Research			BarclayHedge			EurekaHedge			Morninstar		
Year	Start	Exit	AR (%)	Start	Exit	AR (%)	Start	Exit	AR (%)	Start	Exit	AR (%)	Start	Exit	AR (%)	Start	Exit	AR (%)
1995	2 608	208	7.98	799	47	5.88	1 108	38	3.43	1 110	139	12.52	250	0	0	284	0	0
1996	3 244	323	9.96	978	99	10.12	1 413	130	9.20	1 360	133	9.78	334	0	0	385	0	0
1997	3 738	300	8.03	1 162	73	6.28	1 636	152	9.29	1 541	108	7.01	449	0	0	519	0	0
1998	4 262	416	9.76	1 401	108	7.71	1 818	263	14.47	1 771	110	6.21	586	0	0	673	0	0
1999	4 777	379	7.93	1 574	150	9.53	1 928	181	9.39	2 035	96	4.72	764	1	0.13	851	0	0
2000	5 452	575	10.55	1 797	171	9.52	2 128	245	11.51	2 343	239	10.20	1 054	0	0	1 114	0	0
2001	6 046	458	7.58	2 013	191	9.49	2 336	213	9.12	2 582	192	7.44	1 338	0	0	1 395	0	0
2002	6 991	653	9.34	2 271	192	8.45	2 622	211	8.05	2 978	543	18.23	1 732	9	0.52	1 744	0	0
2003	8 083	533	6.59	2 586	192	7.42	3 011	249	8.27	3 041	252	8.29	2 269	35	1.54	2 223	1	0.04
2004	9 698	734	7.57	3 021	244	8.08	3 457	262	7.58	3 530	326	9.24	2 928	180	6.15	2 771	27	0.97
2005	11 455	1 109	9.68	3 529	334	9.46	4 058	389	9.59	4 007	413	10.31	3 505	396	11.30	3 470	115	3.31
2006	12 943	1 410	10.89	3 963	406	10.24	4 570	475	10.39	4 462	533	11.95	3 818	371	9.72	4 136	360	8.70
2007	14 230	1 844	12.96	4 330	621	14.34	4 960	670	13.51	4 793	665	13.87	4 254	403	9.47	4 533	569	12.55
2008	14 851	2 936	19.77	4 399	899	20.44	5 109	983	19.24	4 908	1 045	21.29	4 586	668	14.57	4 653	915	19.66
2009	14 029	2 059	14.68	4 086	618	15.12	4 744	676	14.25	4 501	593	13.17	4 575	491	10.73	4 323	666	15.41
2010	13 765	1 961	14.25	3 930	563	14.33	4 607	606	13.15	4 462	628	14.07	4 750	587	12.36	4 117	724	17.59
2011	13 696	2 197	16.04	3 770	662	17.56	4 477	739	16.51	4 342	733	16.88	4 825	727	15.07	3 926	617	15.72

B. Summary statistics of number of products, returns and asset under management

Database	Number of products			Returns		Asset Under Management							
								Missing AuM observations (in %s)					
	# Firms	# share classes	# Funds	Missing (%)	Length	Mean	Length	Original	Begin	Middle	End	Stale	Total
All funds													
Aggregate	5 661	42 386	30 195	0.54	62.2	120.0	51.5	29.71	15.02	1.63	13.06	11.37	41.09
TASS	2 860	12 092	8 788	0.32	65.1	102.6	52.5	34.99	12.86	3.54	18.59	3.80	38.79
Hedge Fund Research	3 399	13 247	10 332	0.35	65.5	100.4	64.7	19.48	3.96	0.03	15.49	14.24	33.72
BarclayHedge	3 437	13 930	10 520	0.36	64.0	116.2	58.7	11.99	8.17	0.24	3.58	19.13	31.12
EurekaHedge	2 655	12 311	8 149	0.23	65.7	148.3	46.7	36.21	27.34	0.63	8.24	10.30	46.51
Morningstar	2 394	16 682	7 504	1.24	67.8	121.6	57.1	34.25	13.74	3.68	16.82	6.10	40.35
Alive													
Aggregate	2 828	18 063	12 300	0.48	70.3	172.5	58.8	32.25	16.35	1.46	14.44	7.83	40.08
TASS	1 086	4 834	3 026	0.31	75.4	128.4	61.1	39.69	14.85	2.28	22.56	2.56	42.25
Hedge Fund Research	1 547	5 099	3 959	0.24	78.8	154.5	81.6	19.85	2.15	0.01	17.70	10.80	30.66
BarclayHedge	1 437	4 658	3 528	0.23	76.9	175.8	70.3	11.56	8.77	0.19	2.59	10.75	22.31
EurekaHedge	1 759	7 707	4 774	0.21	71.4	190.9	52.0	34.60	25.74	0.71	8.15	9.56	44.16
Morningstar	1 266	8 059	3 725	1.26	72.4	132.7	65.9	35.82	12.59	3.98	19.25	4.88	40.70
Defunct													
Aggregate	4 037	24 323	17 895	0.60	56.6	86.6	46.8	27.55	13.87	1.78	11.89	14.40	41.95
TASS	2 086	7 258	5 762	0.33	59.7	90.6	48.4	31.87	11.53	4.38	15.96	4.62	36.49
Hedge Fund Research	2 344	8 148	6 373	0.45	57.3	69.6	55.1	19.16	5.51	0.05	13.61	17.17	36.33
BarclayHedge	2 461	9 272	6 992	0.44	57.4	85.7	52.8	12.28	7.76	0.27	4.25	24.78	37.06
EurekaHedge	1 256	4 604	3 375	0.27	57.6	88.2	39.1	39.04	30.14	0.50	8.40	11.59	50.63
Morningstar	1 432	8 623	3 779	1.22	63.3	112.7	49.9	32.48	15.05	3.35	14.07	7.48	39.95

Table 2. Return statistics

Table presents summary statistics of the hedge fund excess returns including the number of funds, the annualized mean return and standard deviation, skewness, and excess kurtosis. Next two columns show the proportion of funds that have non-normal return distibution and serial correlation in returns measured using the Jarque-Beta test of normality and the Ljung-Box test of serial correlation (5% level of significance). Next four columns show the results of conditional smoothing (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. Columns Pos (%) and Neg (%) show the number of funds that have statistically significant beta coefficients based on the 5% level of significance. The next two columns show the Bollen and Pool's (2008) discontinuity results based on the measure proposed by Jylhä (2011) (discontinuity (DC(%)) and z statistic (z(DC)) that test the existence of discontinuity across databases. The final columns show the difference between the average December gross returns and the average of the January-November gross returns 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 returns are higher than January-November returns as documented by Agarwal, Daniel and Naik (2011).

	Statistics of normality and serial correlation							Conditional Smoothing				Discontunuity		December Spike	
Database	# Funds	Mean ER %	Std ER %	Skew	Kurt	JB-test (% reject.)	LB-test (% reject.)	b+ Pos (%)	b+ Neg (%)	b- Pos (%)	b- Neg (%)	DC (%)	z(DC)	Avg (Dec - Jan-Nov)%/mth	Diff (p-value)
All Funds															
Aggregate	30 195	5.41	16.80	-0.18	3.01	51.5	21.4	11.5	2.4	5.4	4.7	19.9	65.0	0.31	<0.01
TASS	8 788	5.06	16.53	-0.23	3.52	55.1	22.7	12.6	2.4	5.8	4.5	18.8	36.9	0.35	<0.01
Hedge Fund Research	10 332	6.22	15.99	-0.17	3.23	53.9	24.3	14.0	1.8	4.5	5.0	14.8	31.2	0.25	<0.01
BarclayHedge	10 520	6.66	17.20	-0.07	3.03	53.1	22.0	12.9	2.4	4.7	6.0	21.3	46.3	0.33	<0.01
EurekaHedge	8 149	6.28	16.33	-0.15	2.81	51.4	22.5	12.1	2.0	5.1	4.6	10.9	20.9	0.31	<0.01
Morningstar	7 504	5.03	15.95	-0.22	3.10	51.8	24.7	13.6	1.9	5.9	5.1	12.6	23.7	0.37	<0.01
Alive															
Aggregate	12 300	6.32	17.62	-0.17	2.74	51.5	23.7	12.0	1.8	4.8	4.5	13.7	32.6	0.23	<0.01
TASS	3 026	7.04	17.46	-0.21	3.14	56.8	25.1	12.9	1.9	5.0	4.0	15.3	20.6	0.29	<0.01
Hedge Fund Research	3 959	7.20	16.47	-0.15	3.17	56.3	29.1	15.5	1.2	4.3	5.0	12.2	19.0	0.21	<0.01
BarclayHedge	3 528	7.83	17.38	-0.08	2.92	54.4	25.3	13.8	2.0	4.5	5.8	13.2	20.0	0.28	<0.01
EurekaHedge	4 774	7.42	17.59	-0.13	2.80	52.2	23.7	12.4	1.8	4.6	4.3	10.0	16.3	0.24	<0.01
Morningstar	3 725	5.24	17.14	-0.15	2.72	49.6	25.3	14.1	1.6	4.9	5.5	9.0	13.4	0.28	<0.01
Defunct															
Aggregate	17 895	4.79	16.23	-0.18	3.20	51.5	19.9	11.2	2.7	5.7	4.8	22.4	56.6	0.38	<0.01
TASS	5 762	4.01	16.05	-0.24	3.72	54.3	21.4	12.4	2.6	6.2	4.7	19.1	30.3	0.39	<0.01
Hedge Fund Research	6 373	5.61	15.68	-0.18	3.26	52.3	21.4	13.2	2.2	4.6	5.0	15.5	25.1	0.30	<0.01
BarclayHedge	6 992	6.07	17.11	-0.07	3.08	52.4	20.4	12.5	2.7	4.7	6.1	24.6	43.6	0.37	<0.01
EurekaHedge	3 375	4.68	14.54	-0.17	2.82	50.3	20.9	11.7	2.4	5.8	5.0	10.6	13.2	0.45	<0.01
Morningstar	3 779	4.83	14.78	-0.28	3.47	53.9	24.0	13.0	2.2	6.7	4.7	13.8	18.5	0.46	<0.01

Table 3. Results of average hedge fund performance

Table shows the results of the average hedge fund performance across commercial databases. Each panel shows the number unique hedge funds (# Funds), the percentage of dead funds (% of Dead), the annualized mean (Mean ER %), the standard deviation (Std ER %) and the Sharpe ratios (Sharpe). Next Columns include the results of the Fung and Hsieh (2004) model including the annualized alpha (Alpha %) and the risk loadings of the seven risk factors: 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 Fung and Hsieh (2004) model. All portfolios are rebalanced monthly using either equal- or value-weighting schema. Panel A reports the results for the aggregate data including the value-weight portfolios after stale AuM observations are included (missing AuM observations are filled with the previous non-missing AuM observations) and backfill-adjusted portfolios after the first 32 months of fund-level returns are excluded to adjust for backfill bias. Performance smoothing adjusted results are based on the methodology proposed by Getmansky, Lo and Makarov (2004). The gross returns are calculated as proposed by Feng (2011). Sub periods are selected based on Fung, Hsieh, Naik and Ramadorai (2008). Panel B compares equal-weight and value-weight results across commercial databases.

A. Aggregate database

Database	# funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Equal-weight:														
Baseline	30 195	59.26	7.77	7.34	1.05	5.23	0.30	0.16	0.08	0.28	0.00	0.01	0.01	0.67
						(5.05)	(14.15)	(6.43)	(1.90)	(5.71)	-(0.13)	(2.95)	(1.34)	
Gross returns	14 652	59.15	13.56	8.64	1.56	10.77	0.35	0.20	0.09	0.29	0.00	0.02	0.01	0.65
						(8.52)	(13.79)	(6.40)	(1.85)	(5.00)	(0.61)	(3.03)	(1.55)	
Backfill-adjusted	20 923	57.90	5.83	7.89	0.74	2.98	0.31	0.17	0.09	0.29	0.00	0.02	0.01	0.71
						(2.59)	(14.15)	(6.27)	(2.07)	(5.81)	-(0.22)	(2.61)	(0.88)	
Smoothing-adjusted	30 195	59.26	7.77	9.13	0.85	5.23	0.31	0.15	0.08	0.25	0.00	0.01	0.01	0.684
						(4.13)	(14.31)	(6.78)	(1.83)	(6.69)	(0.04)	(2.70)	(1.93)	
Jan 1994 - Dec 1996	4 130	83.15	9.68	4.33	2.19	8.56	0.26	0.15	0.03	0.04	0.00	0.01	0.04	0.64
						(3.74)	(3.75)	(2.30)	(0.33)	(0.13)	(0.11)	(1.72)	(3.20)	
Jan 1997 - Sep 1998	5 158	80.03	4.90	7.52	0.65	3.11	0.38	0.19	-0.30	0.23	0.02	0.01	0.04	0.88
						(1.05)	(6.79)	(2.74)	-(1.33)	(0.63)	(0.85)	(0.38)	(1.31)	
Oct 1998 - Mar 2000	6 244	75.96	22.15	6.91	3.18	15.72	0.30	0.27	0.41	0.55	0.05	0.00	-0.02	0.87
						(4.18)	(4.20)	(5.81)	(1.69)	(2.03)	(1.86)	(0.11)	-(1.03)	
Apr 2000 - Dec 2004	14 137	71.95	7.91	5.61	1.45	5.69	0.27	0.17	0.24	0.16	0.00	0.03	0.01	0.88
						(5.20)	(12.01)	(6.50)	(6.27)	(2.20)	(0.60)	(5.24)	(1.10)	
Jan 2005 - Dec 2011	25 594	51.97	4.49	9.05	0.49	3.68	0.38	-0.07	-0.05	0.25	0.00	0.01	0.01	0.70
						(1.69)	(8.17)	-(0.89)	-(0.54)	(3.48)	(0.02)	(0.46)	(0.95)	
Value-weight:														
Baseline	25 779	61.12	7.05	6.70	1.04	4.64	0.25	0.12	0.13	0.23	-0.01	0.01	0.02	0.55
						(4.19)	(10.95)	(4.62)	(3.01)	(4.53)	-(1.26)	(2.63)	(2.19)	
Stale-adjusted VW	25 780	61.12	6.78	6.78	0.99	4.38	0.25	0.13	0.12	0.23	-0.01	0.01	0.02	0.56
						(3.94)	(11.24)	(4.72)	(2.75)	(4.40)	-(1.34)	(2.57)	(2.16)	

B. Commercial databases

B1. Equal-weight

Database	# funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
TASS	8 788	65.57	7.30	7.31	0.99	4.72	0.29	0.17	0.08	0.27	0.00	0.01	0.01	0.66
						(4.48)	(13.49)	(6.55)	(2.05)	(5.59)	-(0.68)	(2.79)	(1.24)	
Hedge Fund Research	10 332	61.68	8.05	7.19	1.11	5.58	0.31	0.18	0.06	0.22	0.00	0.01	0.01	0.71
						(5.84)	(15.73)	(7.88)	(1.57)	(5.02)	-(0.52)	(2.70)	(1.18)	
BarclayHedge	10 520	66.46	8.29	6.71	1.23	6.16	0.27	0.14	0.06	0.25	0.00	0.02	0.01	0.65
						(6.29)	(13.66)	(5.99)	(1.63)	(5.39)	(0.71)	(3.80)	(1.92)	
EurekaHedge	8 149	41.42	9.48	7.69	1.22	6.89	0.31	0.16	0.08	0.29	0.00	0.02	0.01	0.65
						(6.18)	(13.76)	(6.04)	(1.80)	(5.51)	(0.06)	(2.99)	(1.48)	
Morningstar	7 504	50.36	8.84	7.12	1.23	6.46	0.29	0.17	0.07	0.24	0.00	0.01	0.01	0.66
						(6.30)	(13.97)	(6.72)	(1.80)	(5.08)	-(0.03)	(2.96)	(1.67)	
Aggregate	30 195	59.26	7.77	7.34	1.05	5.23	0.30	0.16	0.08	0.28	0.00	0.01	0.01	0.67
						(5.05)	(14.15)	(6.43)	(1.90)	(5.71)	-(0.13)	(2.95)	(1.34)	

	B2. Value-weight						B3. Stale-adjusted value-weight					
Database	# funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	# funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %
TASS	7 110	68.23	7.96	7.27	1.09	5.39	7 110	68.23	6.90	7.43	0.92	4.35
						(4.14)						(3.40)
Hedge Fund Research	8 452	63.69	7.25	6.23	1.15	5.14	8 452	63.69	7.31	6.25	1.16	5.19
						(5.20)						(5.21)
BarclayHedge	10 129	66.22	6.63	5.96	1.11	4.81	10 129	66.22	6.60	5.94	1.10	4.77
						(4.46)						(4.44)
EurekaHedge	7 332	41.45	7.52	7.06	1.06	5.18	7 332	41.45	7.49	7.08	1.05	5.14
						(4.37)						(4.33)
Morningstar	5 936	55.24	7.71	6.21	1.23	5.61	5 938	55.22	7.36	6.33	1.15	5.25
						(5.43)						(5.00)
Aggregate	25 779	61.12	7.05	6.70	1.04	4.64	25 780	61.12	6.78	6.78	0.99	4.38
						(4.19)						(3.94)

Table 4. Average performance of hedge fund size portfolios

Table provides the results of the average performance for the size portfolios based on the hedge fund-level AuM (Panel A) and the management firm-level AuM (Panel B). Nominal AuM (in millions of US dollars) groups are shown in the first Column (AuM group). Funds are sorted into nominal groups every December and excess returns of equal-weight portfolios are constructed using 12-month holding period and monthly rebalancing. Descriptions of other Columns are the same as in Table 3. Time period of analysis is December 1994 - December 2011.

A. Fund-level AuM portfolios

AuM group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
0 <= AUM <= 10	11 652	68.88	9.06	7.27	1.24	6.47	0.29	0.17	0.08	0.24	0.01	0.02	0.01	0.66
						(5.93)	(13.71)	(6.79)	(1.86)	(4.83)	(1.23)	(4.08)	(1.58)	
10 < AUM <= 50	12 661	62.48	6.95	7.71	0.90	3.94	0.32	0.19	0.07	0.25	0.00	0.01	0.01	0.71
						(3.71)	(15.26)	(7.54)	(1.56)	(5.29)	-(0.77)	(2.96)	(1.20)	
50 < AUM <= 250	9 746	55.89	5.08	7.54	0.67	1.93	0.30	0.15	0.09	0.29	-0.01	0.01	0.01	0.68
						(1.76)	(13.81)	(5.98)	(2.13)	(5.96)	-(1.40)	(2.09)	(1.55)	
250 < AUM <= 500	3 271	47.11	4.73	7.55	0.62	1.52	0.27	0.13	0.12	0.35	-0.01	0.01	0.01	0.63
						(1.28)	(11.46)	(4.69)	(2.53)	(6.51)	-(1.34)	(1.69)	(1.23)	
500 < AUM <= 1000	1 802	43.06	5.31	7.05	0.75	2.26	0.25	0.09	0.10	0.29	-0.02	0.01	0.01	0.58
						(1.92)	(10.67)	(3.25)	(2.20)	(5.52)	-(2.52)	(1.94)	(1.13)	
AUM > 1000	876	40.87	5.40	8.54	0.63	1.67	0.28	0.07	0.19	0.34	-0.02	0.02	0.01	0.48
						(1.05)	(9.11)	(1.89)	(3.06)	(4.70)	-(1.85)	(2.05)	(1.25)	

B. Firm-level AuM portfolios

Firm AuM Groups	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
0 <= AUM <= 10	6 733	70.53	10.04	6.93	1.44	7.67	0.29	0.18	0.07	0.20	0.01	0.02	0.01	0.67
						(7.51)	(14.31)	(7.38)	(1.66)	(4.33)	(1.53)	(4.33)	(1.78)	
10 < AUM <= 50	8 536	64.00	8.05	7.54	1.06	5.28	0.32	0.19	0.05	0.22	0.00	0.02	0.01	0.71
						(5.04)	(15.35)	(7.81)	(1.21)	(4.70)	-(0.11)	(3.32)	(1.41)	
50 < AUM <= 250	9 827	59.96	6.58	7.70	0.85	3.47	0.31	0.17	0.08	0.30	0.00	0.01	0.01	0.71
						(3.24)	(14.58)	(6.91)	(1.91)	(6.29)	-(0.70)	(2.66)	(1.25)	
250 < AUM <= 500	5 490	57.12	5.53	7.90	0.70	2.20	0.31	0.15	0.09	0.32	-0.01	0.02	0.01	0.68
						(1.91)	(13.71)	(5.41)	(2.07)	(6.18)	-(1.56)	(3.82)	(1.38)	
500 < AUM <= 1000	4 751	54.45	4.39	7.45	0.59	1.27	0.27	0.15	0.12	0.28	-0.01	0.01	0.02	0.59
						(1.03)	(10.97)	(5.16)	(2.43)	(5.00)	-(1.71)	(1.72)	(1.98)	
AUM > 1000	5 838	54.01	4.87	7.87	0.62	1.51	0.29	0.11	0.15	0.32	-0.01	0.02	0.01	0.58
						(1.15)	(11.17)	(3.67)	(2.79)	(5.51)	-(1.15)	(3.14)	(1.54)	

Table 5. Results of average performance by investment styles

Panel A presents the results of the average performance of hedge fund style portfolios using the aggregate database. Hedge funds are grouped into portfolios based on the reported investment styles. Equal-weight excess returns of the portfolios are calculated using monthly rebalancing. Time period is January 1994 - December 2011. Descriptions of columns are the same as in Table 3.

Style	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
CTA	2 964	59.28	7.55	7.38	1.02	6.72	0.09	0.04	0.19	0.14	0.01	0.04	0.03	0.30
						(4.41)	(3.07)	(1.02)	(3.19)	(1.98)	(1.70)	(5.39)	(3.01)	
Emerging markets	3 756	37.51	10.33	15.64	0.66	5.88	0.53	0.24	-0.02	0.45	-0.02	0.01	0.00	0.50
						(2.16)	(9.65)	(3.56)	-(0.16)	(3.57)	-(1.26)	(0.52)	-(0.10)	
Event driven	1 539	66.41	7.40	6.41	1.15	5.00	0.22	0.15	0.02	0.30	-0.02	0.01	0.00	0.75
						(6.29)	(13.77)	(7.50)	(0.49)	(8.18)	-(3.64)	(1.50)	-(0.80)	
Fund of funds	12 968	61.23	3.23	7.18	0.45	0.70	0.22	0.13	0.12	0.32	-0.01	0.01	0.01	0.51
						(0.56)	(8.88)	(4.37)	(2.55)	(5.51)	-(1.46)	(2.28)	(1.45)	
Global macro	1 887	64.44	6.44	5.41	1.18	4.69	0.19	0.07	0.14	0.15	0.00	0.02	0.01	0.45
						(4.72)	(9.56)	(2.78)	(3.56)	(3.26)	(0.12)	(3.24)	(1.99)	
Long only	814	32.31	7.45	11.96	0.62	3.17	0.55	0.26	0.05	0.38	0.00	0.01	0.00	0.81
						(2.46)	(21.22)	(8.27)	(0.98)	(6.40)	(0.04)	(0.97)	(0.04)	
Long/Short	8 430	63.31	8.78	9.90	0.88	5.55	0.45	0.30	0.04	0.18	0.00	0.01	0.01	0.75
						(4.51)	(17.92)	(9.97)	(0.93)	(3.13)	-(0.57)	(1.63)	(0.65)	
Market neutral	1 626	65.99	5.53	4.40	1.23	4.01	0.13	0.04	0.10	0.15	-0.01	0.01	0.00	0.38
						(4.71)	(7.68)	(1.78)	(2.90)	(3.85)	-(1.46)	(2.26)	(0.32)	
Multi-strategy	3 912	59.97	7.33	7.23	1.01	5.83	0.15	0.07	0.16	0.30	0.02	0.02	0.03	0.32
						(3.98)	(5.07)	(1.87)	(2.72)	(4.45)	(1.89)	(3.47)	(2.94)	
Others	1 265	81.03	7.49	6.54	1.14	5.18	0.26	0.15	0.11	0.20	0.00	0.01	0.01	0.60
						(5.06)	(12.32)	(6.18)	(2.65)	(4.12)	-(0.50)	(1.93)	(0.95)	
Relative value	3 117	60.15	5.87	5.20	1.13	3.57	0.13	0.06	0.13	0.38	-0.01	0.00	0.00	0.64
						(4.65)	(8.57)	(3.17)	(4.23)	(10.70)	-(2.23)	(0.65)	-(0.38)	
Sector	753	62.82	11.90	13.05	0.91	8.08	0.56	0.43	0.04	0.14	0.00	0.01	0.00	0.69
						(4.48)	(15.31)	(9.84)	(0.56)	(1.68)	-(0.14)	(0.94)	(0.09)	
Short bias	132	75.00	2.30	11.57	0.20	5.05	-0.50	-0.38	-0.05	0.22	0.00	0.01	0.01	0.57
						(2.69)	-(13.07)	-(8.41)	-(0.74)	(2.51)	-(0.02)	(0.87)	(0.42)	

Table 6. Domicile and average performance

Panel A describes the number of unique management firms grouped by the manager domicile region. # Firms is the number of unique firms in each manager domicile group. Next five columns show the number of hedge funds grouped by the fund domicile region. For example, hedge fund managers that are domiciled in Asia/Pacific region have 31 unique hedge funds that are legally established in North America (USA or Canada). Table describes performance results of the equal-weight domicile portfolios that are created based on the fund domicile region (Panel B). All results are obtained using the aggregate database.

A. Number of hedge funds grouped by manager country and fund domicile region

		Number of Funds by fund region				
Manager region	# Firms	USA	Asia/Pacific	Caribbean	Europe	Rest of World
Asia/Pacific	759	31	304	975	58	63
Caribbean	266	46	0	429	7	11
Europe	1 750	73	5	3 039	3 220	323
Rest of world	403	48	0	346	49	460
USA	5 147	6 720	6	2 205	146	926

B. Results of the average performance by fund domicile

Domicile	# Funds	% Of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
All funds	30 195	59.26	7.77	7.34	1.05	5.23	0.30	0.16	0.08	0.28	0.00	0.01	0.01	0.67
						(5.05)	(14.15)	(6.43)	(1.90)	(5.71)	-(0.13)	(2.95)	(1.34)	
USA (Onshore)	9 499	66.66	8.91	6.65	1.33	6.91	0.29	0.19	0.02	0.17	0.00	0.01	0.01	0.73
						(8.11)	(16.95)	(9.01)	(0.62)	(4.38)	(0.66)	(3.69)	(1.86)	
All offshore funds	20 696	55.87	6.95	7.77	0.89	4.17	0.29	0.15	0.10	0.33	0.00	0.01	0.01	0.62
						(3.53)	(12.23)	(5.15)	(2.22)	(5.92)	-(0.66)	(2.50)	(1.07)	
Caribbean	11 484	62.94	6.86	7.38	0.92	4.27	0.27	0.15	0.08	0.29	-0.01	0.01	0.01	0.60
						(3.72)	(11.55)	(5.46)	(1.82)	(5.43)	-(1.27)	(2.17)	(0.93)	
Europe	5 334	43.74	5.57	8.84	0.63	2.51	0.31	0.13	0.17	0.40	0.01	0.02	0.01	0.55
						(1.71)	(10.57)	(3.60)	(2.91)	(5.81)	(0.60)	(2.92)	(1.13)	
Asia/Pacific	1 105	26.88	10.13	10.47	0.97	7.20	0.28	0.08	0.12	0.52	0.01	0.01	0.01	0.40
						(3.60)	(7.02)	(1.58)	(1.49)	(5.57)	(0.98)	(1.13)	(0.76)	
Rest of world	2 773	61.49	8.45	8.46	1.00	5.52	0.32	0.16	0.08	0.35	0.00	0.01	0.00	0.63
						(4.34)	(12.37)	(5.12)	(1.51)	(5.85)	-(0.65)	(2.25)	(0.53)	

Table 7. Results of hedge fund perfomance persistence

Using t-statistic of the Fung and Hsieh (2004) alpha, hedge funds are sorted into decile portfolios that are rebalanced quarterly, semiannually, or annually. The alpha t-statistics are estimated using 24 the most recent return observations. We calculate buy-and-hold returns for each of the portfolios. If a fund stops reporting return during the holding period, the fund is assumed to be in liquidation and the proceeds are invested in risk-free asset. Table shows the Fung and Hsieh (2004) alphas (per year) in percentages and the alpha t-statistics in parentheses for each of the portfolios. Column DropOut (%) describes the average number of funds that drop from each of the equal-weight portfolio during the holding period. For spread portfolio (top minus bottom portfolio), drop out rate contains the difference in drop out rates between equal-weight top and bottom portfolios. Panel A provides the baseline results of persistence including equal-weight and value-weight results. Panel B provides results of persistence by size terciles. In Panel C, portfolios are formed using only feasible information by taking into account the fund-specific share restrictions. For instance, using the quarterly rebalanced portfolios, we exclude hedge funds having lockup, redemption and notice periods longer than three months. Therefore, depending on the length of the notice period, we sort funds using either 1-month or 3-month lag in alpha t-statistic due to look-ahead bias. Column Feasible (Baseline) includes the Fung and Hsieh (2004) alpha and alpha t-statistic for feasible EW portfolios (Baseline EW portfolios).

A. Baseline

	Quarterly			Semiannual			Annual		
Database	FH alpha		DropOut	FH alpha		DropOut	FH alpha		DropOut
	EW	VW	(%)	EW	VW	(%)	EW	VW	(%)
Top	10.26	8.06	2.24	7.87	6.02	4.89	6.05	4.21	9.61
	(5.38)	(4.53)		(5.93)	(5.27)		(6.47)	(4.62)	
Bottom	2.50	0.17	5.37	4.25	3.44	11.93	3.60	2.40	21.67
	(1.04)	(0.07)		(2.22)	(1.88)		(2.78)	(1.74)	
Spread	7.76	7.88	3.13	3.62	2.58	7.03	2.45	1.82	12.07
	(2.91)	(2.75)		(1.85)	(1.29)		(1.86)	(1.27)	

B. Size terciles
(Equal-weight)

	Quarter		Semiannual		Annual	
Portfolio	FH Alpha (%)	DropOut (%)	FH Alpha (%)	DropOut (%)	FH Alpha (%)	DropOut (%)
Small-Top	13.51	3.11	10.32	7.23	8.92	14.10
	(4.28)		(4.75)		(6.55)	
Medium-Top	10.19	1.90	6.90	4.24	5.28	9.88
	(4.47)		(4.59)		(5.01)	
Large-Top	10.23	2.28	7.23	4.64	4.83	8.99
	(5.78)		(6.42)		(5.40)	
Small-Bottom	3.43	8.04	4.04	17.56	2.55	31.33
	(1.02)		(1.73)		(1.73)	
Medium-Bottom	2.98	4.72	4.46	11.01	4.31	21.22
	(1.18)		(2.12)		(3.02)	
Large-Bottom	1.06	3.35	3.06	7.39	2.52	14.43
	(0.44)		(1.66)		(1.91)	
Small-Spread	10.08	4.94	6.28	10.33	6.37	17.23
	(2.55)		(2.24)		(3.62)	
Medium-Spread	7.21	2.82	2.44	6.77	0.97	11.34
	(2.25)		(1.04)		(0.60)	
Large-Spread	9.17	1.07	4.17	2.75	2.31	5.44
	(2.84)		(2.04)		(1.66)	

C. Feasible strategies
(Equal-weight)

Top

	Quarter		Semiannual		Annual	
	Fung and Hsieh Alpha %		Fung and Hsieh Alpha %		Fung and Hsieh Alpha %	
Notice	Feasible	Baseline	Feasible	Baseline	Feasible	Baseline
Notice <= 1 month	8.28	10.26	6.27	7.87	5.01	6.05
	(3.84)	(5.38)	(4.06)	(5.93)	(4.65)	(6.47)
Notice <= 3 months	8.56		6.58		5.33	
	(5.01)		(5.39)		(6.39)	
Notice <= 6 months			5.26		3.70	
			(4.09)		(4.19)	
Notice <= 12 months					3.74	
					(4.42)	

Bottom

	Quarter		Semiannual		Annual	
	Fung and Hsieh Alpha %		Fung and Hsieh Alpha %		Fung and Hsieh Alpha %	
Notice	Feasible	Baseline	Feasible	Baseline	Feasible	Baseline
Notice <= 1 month	1.92	2.50	4.88	4.25	3.67	3.60
	(0.73)	(1.04)	(2.11)	(2.22)	(2.40)	(2.78)
Notice <= 3 months	5.18		6.06		4.38	
	(1.51)		(2.14)		(2.37)	
Notice <= 6 months			5.92		4.65	
			(2.40)		(2.74)	
Notice <= 12 months					5.42	
					(3.37)	

Spread

	Quarter		Semiannual		Annual	
	Fung and Hsieh Alpha %		Fung and Hsieh Alpha %		Fung and Hsieh Alpha %	
Notice	Feasible	Baseline	Feasible	Baseline	Feasible	Baseline
Notice <= 1 month	6.36	7.76	1.38	3.62	1.34	2.45
	(1.91)	(2.91)	(0.57)	(1.85)	(0.86)	(1.86)
Notice <= 3 months	3.38		0.53		0.95	
	(0.99)		(0.19)		(0.54)	
Notice <= 6 months			-0.66		-0.95	
			-(0.25)		-(0.56)	
Notice <= 12 months					-1.67	
					-(1.07)	

Table 8. Results of hedge fund perfomance persistence

Table provides persistence results for commercial databases including equal-and value-weight results. Using t-statistic of the Fung and Hsieh (2004) alpha, hedge funds are sorted into decile portfolios that are rebalanced quarterly, semiannually, or annually. The alpha t-statistics are estimated using 24 the most recent return observations. We calculate buy-and-hold returns for each of the portfolios. If a fund stops reporting return during the holding period, the fund is assumed to be in liquidation and the proceeds are invested in risk-free asset. Table shows the Fung and Hsieh (2004) alphas (per year) in percentages and the alpha t-statistics in parentheses for each of the portfolios. Column DropOut (%) describes the average number of funds that drop from each of the equal-weight portfolio during the holding period. For spread portfolio (top minus bottom portfolio), drop out rate contains the difference in drop out rates between equal-weight top and bottom portfolios.

	Quarterly			Semiannual			Annual		
Database	FH Alpha		DropOut	FH alpha		DropOut	FH alpha		DropOut
	EW	VW	(%)	EW	VW	(%)	EW	VW	(%)
Top									
TASS	9.33	8.75	2.25	7.22	6.60	4.71	5.73	4.81	9.28
	(4.94)	(5.25)		(5.31)	(5.86)		(6.05)	(5.32)	
Hedge Fund Research	10.38	7.44	2.33	7.56	5.62	4.92	5.51	3.65	9.61
	(6.14)	(3.98)		(6.82)	(4.69)		(7.07)	(4.34)	
BarclayHedge	10.58	8.94	1.83	8.07	6.54	4.69	6.11	4.77	10.01
	(4.93)	(4.38)		(5.87)	(5.09)		(6.18)	(4.20)	
EurekaHedge	9.91	8.57	1.30	7.52	5.80	2.94	6.09	2.99	6.01
	(4.14)	(2.69)		(4.25)	(2.97)		(5.11)	(2.30)	
Morningstar	10.14	8.20	2.39	7.57	6.61	4.62	5.53	4.17	9.19
	(6.32)	(4.50)		(7.10)	(6.16)		(6.65)	(4.73)	
Aggregate	10.26	8.06	2.24	7.87	6.02	4.89	6.05	4.21	9.61
	(5.38)	(4.53)		(5.93)	(5.27)		(6.47)	(4.62)	
Bottom									
TASS	0.92	1.00	5.09	4.36	3.46	11.85	3.10	1.12	22.31
	(0.33)	(0.27)		(2.12)	(1.39)		(2.22)	(0.57)	
Hedge Fund Research	2.72	2.90	5.57	5.01	3.76	12.12	4.29	3.74	21.56
	(1.03)	(1.32)		(2.50)	(2.03)		(3.31)	(2.85)	
BarclayHedge	2.94	-0.90	5.19	4.02	3.28	11.64	3.91	2.97	20.85
	(1.26)	-(0.34)		(2.29)	(1.71)		(3.22)	(2.08)	
EurekaHedge	9.51	8.87	2.83	10.35	8.48	6.27	8.68	5.77	11.68
	(2.85)	(2.50)		(4.09)	(3.51)		(5.04)	(3.33)	
Morningstar	7.70	6.53	4.21	8.70	7.76	8.72	7.02	5.03	14.21
	(2.50)	(2.22)		(3.62)	(3.44)		(4.40)	(3.12)	
Aggregate	2.50	0.17	5.37	4.25	3.44	11.93	3.60	2.40	21.67
	(1.04)	(0.07)		(2.22)	(1.88)		(2.78)	(1.74)	
Spread									
TASS	8.41	7.75	2.84	2.86	3.14	7.14	2.62	3.69	13.03
	(3.04)	(1.95)		(1.37)	(1.23)		(1.84)	(1.90)	
Hedge Fund Research	7.66	4.54	3.24	2.55	1.86	7.20	1.22	-0.09	11.96
	(2.58)	(1.67)		(1.21)	(0.90)		(0.89)	-(0.06)	
BarclayHedge	7.64	9.84	3.37	4.05	3.26	6.96	2.19	1.80	10.84
	(2.67)	(3.08)		(2.01)	(1.52)		(1.59)	(1.10)	
EurekaHedge	0.40	-0.31	1.52	-2.83	-2.68	3.33	-2.59	-2.78	5.67
	(0.13)	-(0.06)		-(1.13)	-(0.90)		-(1.50)	-(1.35)	
Morningstar	2.44	1.68	1.81	-1.13	-1.15	4.10	-1.49	-0.86	5.02
	(0.78)	(0.49)		-(0.47)	-(0.47)		-(0.92)	-(0.53)	
Aggregate	7.76	7.88	3.13	3.62	2.58	7.03	2.45	1.82	12.07
	(2.91)	(2.75)		(1.85)	(1.29)		(1.86)	(1.27)	

Table 9. Statistics of characteristics

Table shows statistics of the fund-level characteristics including the mean, median, the number of funds and the proportion of hedge funds having the missing observation for each of the variables (Missing (%)). Results of univariate sorts (the number of funds and the annualized Fung and Hsieh (2004) alphas) are reported for each of the characteristics including incentive fee, high-water mark and share restrictions. Buy-and-hold returns of univariate portfolios are created using a 12-month holding period.

	A. Notice period (Days)											
	Aggregate		TASS		HFR		Barclay		Eureka		Morningstar	
Mean (Median)	30 (30)		31 (30)		34 (30)		26 (30)		33 (30)		35 (30)	
# Funds (Missing %)	24,977 (7.88)		8,788 (0)		9,946 (3.74)		10,520 (0)		7,774 (4.60)		6,010 (19.91)	
	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %
0	5 984	3.84	1 811	3.54	799	5.58	3 588	4.64	1 130	5.53	199	7.35
		(2.84)		(2.20)		(3.91)		(3.66)		(3.42)		(4.14)
(0, 30)	5 519	5.20	1 762	4.67	2 342	5.49	1 462	5.53	1 759	6.90	968	6.59
		(4.01)		(3.18)		(4.53)		(3.89)		(4.16)		(4.72)
30	7 571	6.47	2 605	7.38	3 372	5.64	2 679	6.58	2 184	7.51	1 468	7.32
		(5.51)		(5.30)		(5.31)		(5.90)		(6.08)		(6.16)
(30, 60)	2 537	6.74	824	6.53	1 018	6.59	878	8.43	896	8.53	583	6.80
		(6.08)		(6.01)		(6.05)		(7.66)		(6.98)		(7.36)
60	1 968	6.81	683	7.19	987	6.44	781	7.82	476	8.80	497	7.64
		(6.58)		(6.68)		(6.30)		(7.26)		(6.38)		(7.49)
(60, …]	2 772	7.07	896	6.37	1 056	7.10	869	7.30	1 003	7.72	554	7.68
		(7.07)		(6.45)		(7.39)		(6.91)		(7.32)		(7.14)
Spread		3.24		2.84		1.52		2.66		2.19		0.33
		(3.11)		(2.27)		(1.58)		(2.11)		(1.80)		(0.20)

B. Lockup period (Months)

	Aggregate		TASS		HFR		Barclay		Eureka		Morningstar	
Mean (% uses lockup)	2.76 (23.07)		2.76 (22.04)		3.24 (26.05)		3.12 (25.56)		2.64 (20.36)		3.72 (31.75)	
# Funds (Missing %)	23,455 (13.49)		8,788 (0)		9,807 (5.08)		8,753 (16.80)		7,887 (3.22)		5,877 (21.68)	
	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %
0	18 907	5.43	6 679	4.92	6 818	5.28	6 319	6.37	6 361	6.87	2 079	6.69
		(4.67)		(3.98)		(5.00)		(5.66)		(5.41)		(6.40)
(0, 12)	1 285	6.98	357	6.12	439	6.57	466	7.63	261	7.68	494	8.30
		(6.58)		(5.86)		(7.40)		(6.58)		(5.32)		(6.36)
12	3 942	7.36	1 337	7.99	1 882	7.50	1 513	8.22	886	8.87	1 125	8.54
		(6.69)		(7.19)		(6.22)		(7.87)		(6.87)		(7.14)
(12, …]	586	8.69	208	10.52	268	8.56	230	9.27	160	7.11	159	9.24
		(6.83)		(7.25)		(6.54)		(5.15)		(4.61)		(5.47)
Spread		3.26		5.60		3.28		2.90		0.24		2.55
		(4.34)		(5.35)		(3.45)		(2.20)		(0.21)		(2.13)

C. Incentive Fee

	Aggregate		TASS		HFR		BarclayHedge		EurekaHedge		Morningstar	
Mean (Median) (%)	17.79 (20.00)		17.06 (20.00)		18.67 (20.00)		18.53 (20.00)		18.50 (20.00)		18.81 (20.00)	
# Funds (Missing %)	26,132 (3.62)		8,517 (3.08)		10,227 (1.02)		10,520 (0.00)		8,052 (1.19)		7,035 (6.25)	
	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %
(0,20%)	3 376	3.59	925	3.47	957	4.31	1 026	3.62	921	5.85	765	4.94
		(2.67)		(2.71)		(3.27)		(2.78)		(3.42)		(3.80)
20 %	20 779	5.97	6 101	6.15	8 035	6.16	7 887	6.10	6 067	7.62	4 911	7.44
		(5.72)		(5.44)		(6.09)		(6.25)		(6.61)		(6.99)
(20%,…]	1 541	6.91	352	5.46	466	7.14	739	8.03	370	9.25	308	8.14
		(4.69)		(3.30)		(5.19)		(5.11)		(4.68)		(4.24)
Spread		3.31		2.00		2.82		4.41		3.39		3.20
		(3.08)		(1.66)		(2.61)		(3.54)		(1.70)		(1.95)

	D. High-water mark (1=yes; 0=no)											
	Aggregate		TASS		HFR		BarclayHedge		EurekaHedge		Morningstar	
Mean	0.71		0.62		0.85		0.62		0.87		0.83	
# Funds (Missing %)	25,959 (4.26)		8,725 (0.72)		10,332 (0.00)		10,432 (0.84)		8,090 (0.72)		6,727 (10.35)	
Portfolio	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %	# Funds	Alpha %
0	7 795	3.87	3 177	3.89	1 151	4.14	3 841	4.62	1 091	5.78	696	5.32
		(3.16)		(2.67)		(3.95)		(4.29)		(3.64)		(4.81)
1	19 801	6.58	5 343	7.60	8 951	5.97	6 334	7.41	6 734	7.48	4 637	7.61
		(5.92)		(6.63)		(5.76)		(6.56)		(6.24)		(7.10)
Spread		2.71		3.71		1.83		2.78		1.70		2.29
		(4.27)		(4.03)		(4.06)		(3.58)		(1.92)		(3.54)

Table 10. Cross-sectional regressions

Table shows the results of the cross-sectional regressions based on the Fama-McBeth (1973) procedure. Monthly Fung and Hsieh (2004) alphas 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 the high-water mark and zero otherwise. Fund-level size, age, and flow are lagged one month. Age is measured since inception date of the fund. Alphas are calculated as a sum of the intercept and the time series of the residual estimated from the Fund and Hsieh (2004) model. Time period is January 1994 - December 2011. All parameter estimates are multiplied by 100 and t-statistics are reported in parentheses. Strategy and domicile dummy variables are applied to control for fixed effects. Standard errors are adjusted using Newey and West (1987) procedure.

Variable	Aggregate		TASS		Hedge Fund Research		BarclayHedge		EurekaHedge		Morningstar	
Incentive fee	1.189	1.029	1.031	0.780	1.385	1.296	1.257	0.971	1.154	1.770	1.640	1.321
	(4.24)	(4.15)	(1.60)	(1.28)	(6.03)	(6.96)	(3.28)	(3.96)	(1.79)	(2.83)	(3.90)	(3.98)
High-water mark	0.130	0.066	0.144	0.097	0.099	0.043	0.093	0.107	-0.191	-0.124	-0.014	-0.044
	(4.31)	(2.67)	(3.15)	(2.44)	(3.21)	(1.79)	(1.41)	(2.73)	-(0.79)	-(0.68)	-(0.15)	-(0.68)
Lockup	0.044	0.016	0.071	0.049	0.088	0.050	0.085	0.048	0.095	0.103	0.028	-0.007
	(1.74)	(0.79)	(1.98)	(1.38)	(2.96)	(2.25)	(1.94)	(1.76)	(1.05)	(1.34)	(0.69)	-(0.21)
Notice	0.758	0.475	1.208	0.996	0.454	0.392	0.097	0.084	1.115	0.491	0.738	0.586
	(2.80)	(2.35)	(4.21)	(3.60)	(1.25)	(1.45)	(0.28)	(0.40)	(2.02)	(1.16)	(1.65)	(1.89)
Redemption	0.025	0.005	-0.025	0.045	-0.046	-0.004	0.046	-0.008	0.044	0.080	-0.066	-0.011
	(0.46)	(0.11)	-(0.45)	(0.80)	-(0.72)	-(0.08)	(0.49)	-(0.12)	(0.29)	(0.65)	-(0.88)	-(0.15)
Control variables:												
Management fee	8.098	6.358	11.301	11.218	11.050	7.330	11.705	11.517	20.743	16.135	0.891	6.194
	(3.24)	(2.98)	(3.19)	(3.64)	(3.44)	(2.71)	(3.03)	(3.81)	(4.67)	(3.25)	(0.17)	(1.16)
Lag (Size)		-0.044		-0.061		-0.054		-0.068		-0.052		-0.083
		-(4.51)		-(3.98)		-(5.36)		-(5.89)		-(2.43)		-(5.30)
Lag (Age)		-0.110		-0.092		-0.118		-0.092		-0.086		-0.109
		-(6.21)		-(4.50)		-(6.54)		-(5.15)		-(3.17)		-(4.80)
Lag (Flow)		1.008		1.035		1.128		1.001		1.046		1.231
		(6.42)		(4.99)		(7.52)		(6.46)		(2.79)		(5.02)
Intercept	0.009	0.394	-0.072	0.503	0.205	0.807	0.036	0.405	-0.130	0.318	0.023	0.665
	(0.08)	(2.13)	-(0.48)	(2.00)	(1.16)	(2.07)	(0.35)	(2.08)	-(0.45)	(0.76)	(0.16)	(2.09)
Strategy Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Domicile Fixed Effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Figure 1. Venn diagram

Figure presents Venn diagram showing overlapping between databases and the percentage amount of unique share classes. Databases are merged using a statistical algorithm that is presented in Appendix. First, five databases including all share classes are classified based on the 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. All presented numbers are in percentages.

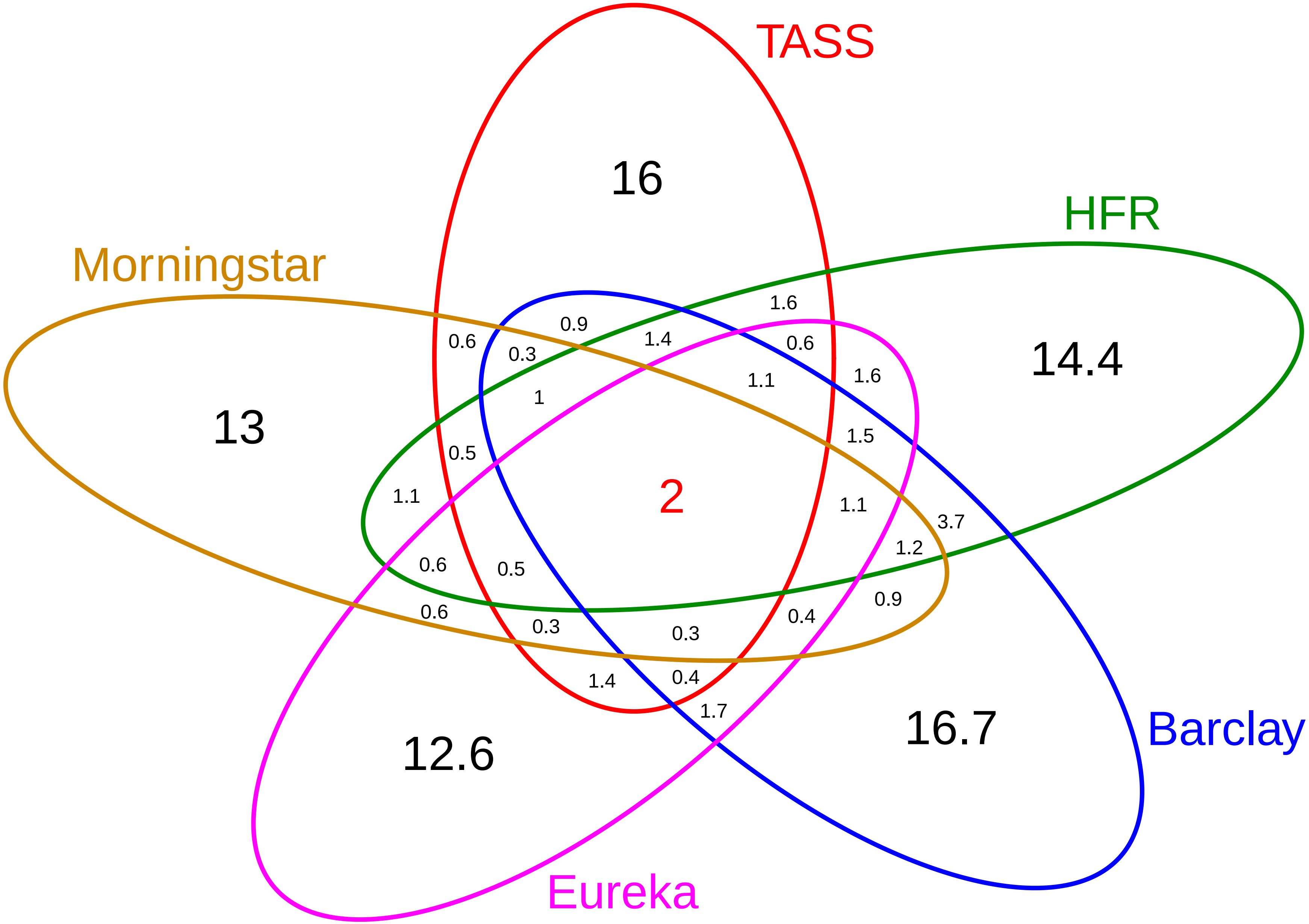


Figure 2
The number of funds and the aggregate asset under management of hedge funds.

The number of funds and the aggregate asset under management are calculated using the consolidated database including five databases (TASS, Hedge Fund Research, BarclayHedge, EurekaHedge, and Morningstar).

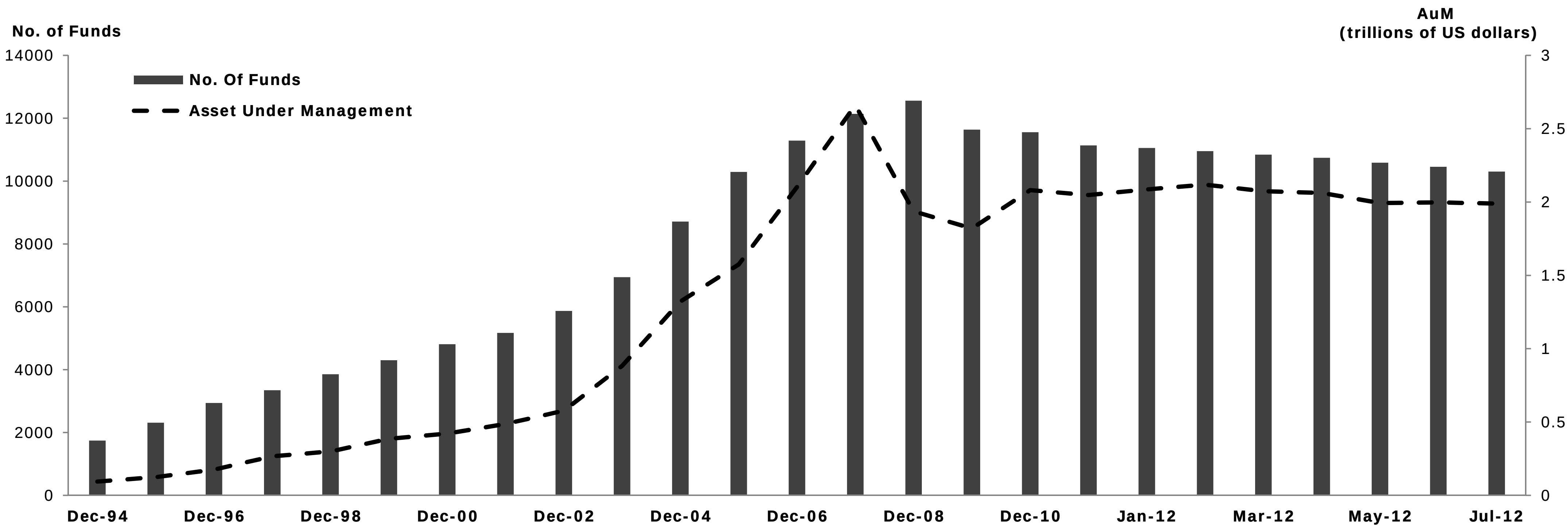


Figure 3. Cumulative excess returns

Figure shows cumulative excess returns of equal-weight (left) and value-weight (right) portfolios. All portfolios are rebalanced monthly and all funds are required to have at last 12 monthly returns. Time period is January 1994 - December 2011.

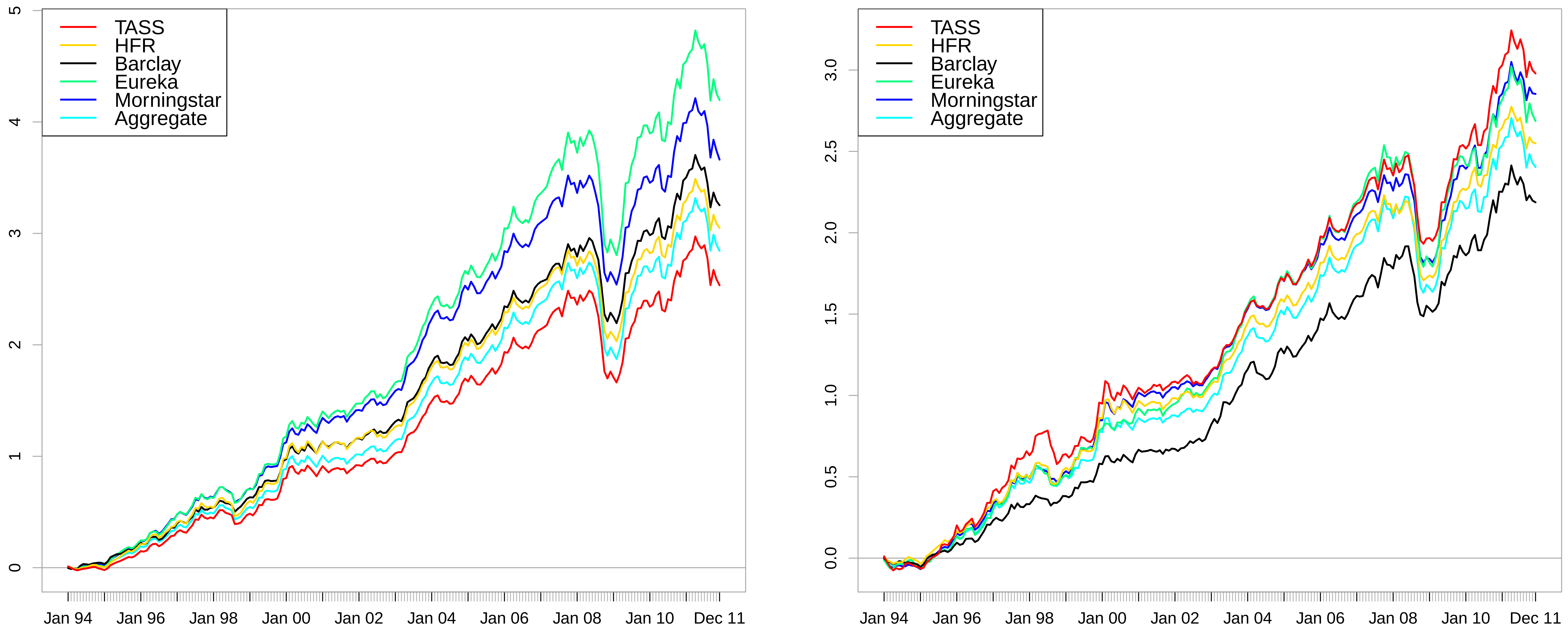


Figure 4. Annualized Fung and Hsieh (2004) alphas for size portfolios

At December 2011, hedge funds are grouped based on the monthly asset under management. Funds are grouped into nominal size (in millions of US dollars) groups as presented in the x-axis (where character "(" means ">" and character "]" refers to "<="). Full sample of the fund-level excess returns are used to calculate the excess return of the equal-weight (monthly rebalanced) portfolio. These portfolios are referred as "backward-looking" because grouping is based on the last non-missing fund-level AuM. At December 2011, we form percentiles of the number of funds that belong in each nominal size groups. We apply these percentile limits and sort hedge funds into portfolios each December using monthly AuMs. Equal-weight excess returns are calculated (monthly rebalanced) for each portfolio using 12-month holding period. These are referred as "Forward-looking". Performance of these portfolios provide information on how small funds outperform if compared to large funds over time.

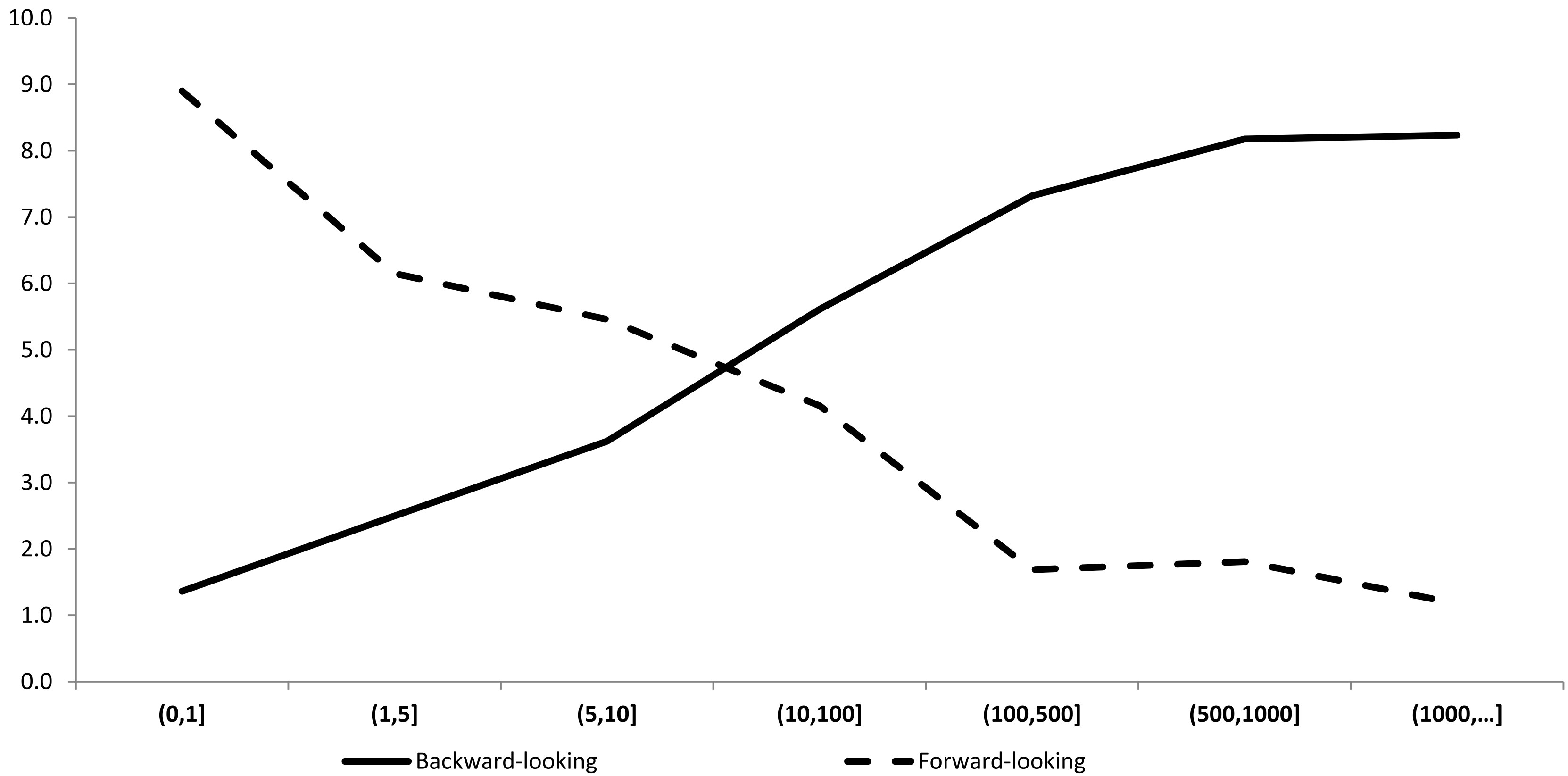
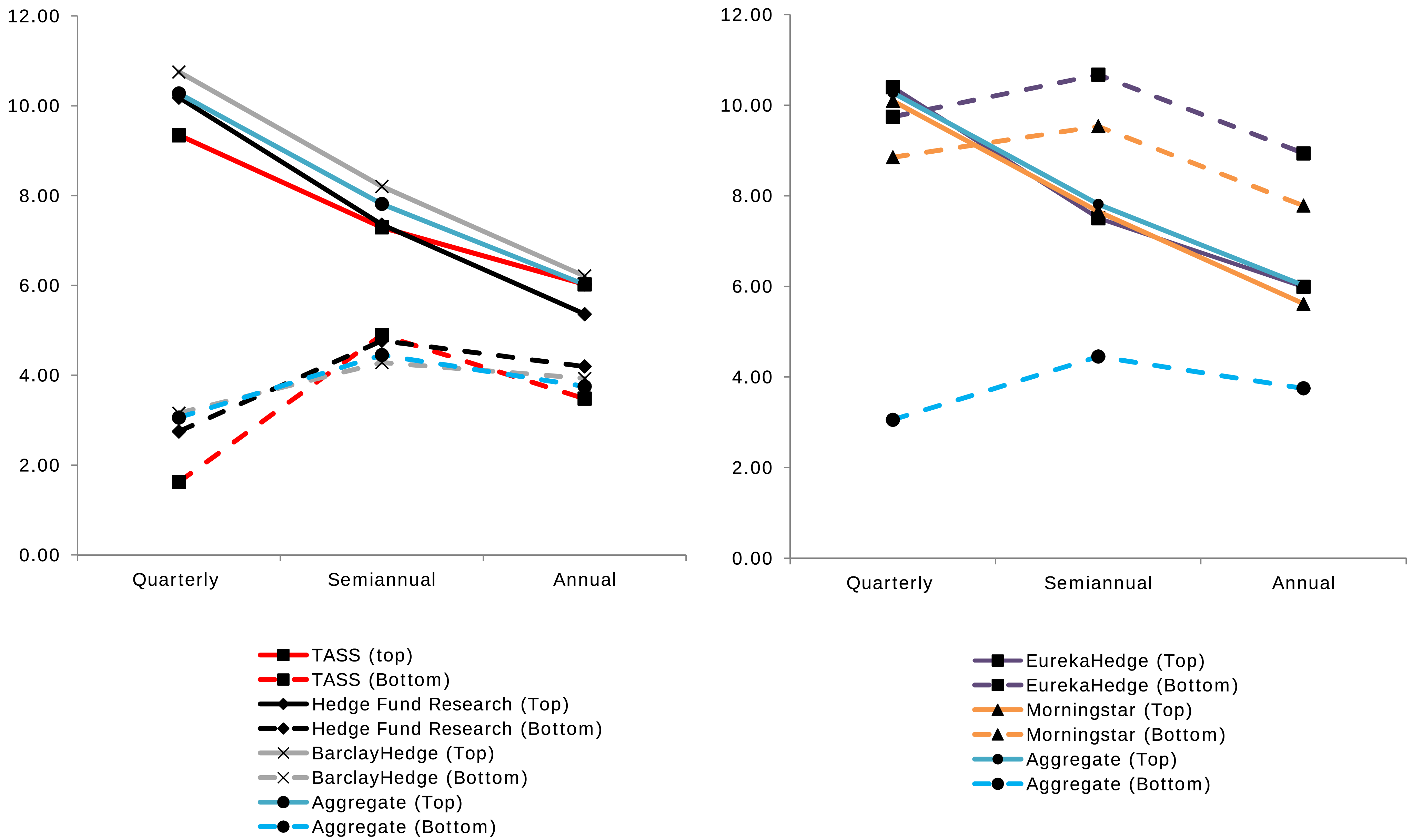


Figure 5. Annualized alphas of equal-weight persistence portfolios

Hedge funds are sorted into decile portfolios based on the t-statistics of alphas obtained using the Fung and Hsieh (2004) model and 24-months of excess returns preceding the evaluation period. Buy-and-hold equal-weight excess returns are calculated for persistence portfolios using three different holding periods (quarterly, semiannual, or annual). Figure shows the annualized Fung and Hsieh (2004) alphas of persistence portfolios including top and bottom deciles.



Online Appendix

This online appendix describes a procedure that we use to create a consolidated hedge fund database and additional analysis to the main tables in 'Stylized facts about Hedge Funds and Database Selection Bias'.

The first section describes the classification of hedge fund strategies. The second section provides methodology used to create the consolidated database of hedge funds that is applied in the paper from the five databases: TASS, Hedge Fund Research, BarclayHedge, EurekaHedge and Morningstar. We apply two methodologies used in merging of databases. First, we describe how to merge databases at the share class level to identify unique share classes. Second, we show details of the methodology to identify unique hedge funds from multiple share class structures. The following tables and figures are supplemental analysis to the main results described in the paper:

Table A1. Estimates of backfill bias
Table A2. Supplementary tables of average performance
Table A3. Sweet spot AuM limits
Table A4. Average performance of hedge fund investment style portfolios grouped by asset under management
Table A5. Results of average performance by investment styles.
Table A6. Average performance grouped by manager domicile
Table A7. Average performance by fund domicile region across databases
Table A8. Results of persistence for monthly rebalanced portfolios
Table A9. Fama-McBeth (1973) regressions using excess returns
Table A10. Proportions of hedge funds by investment styles and fund domicile regions
Figure A1. Histograms of pairwise correlation coefficients of share classes

A. Classification of investment strategies

Original strategies provided by data vendors are in the first column. The name of the database is in the second column. Our style class is in the final column.

Original Strategy	**Database**	**Strategy Key**
ACTIVIST	BarclayHedge	Event Driven
ARBITRAGE	BarclayHedge	Relative Value
ARBITRAGE	EurekaHedge	Relative Value
ASIA/PACIFIC LONG/SHORT EQUITY	Morningstar	Long/Short
BALANCED (STOCKS & BONDS)	BarclayHedge	Multi-Strategy
BEAR MARKET EQUITY	Morningstar	Short Bias
BOTTOM-UP	EurekaHedge	Long Only
CHINA LONG/SHORT EQUITY	Morningstar	Emerging Markets
CLOSED-END FUNDS	BarclayHedge	Others
COMMODITY - AGRICULTURE	Hedge Fund Research	CTA
COMMODITY - ENERGY	Hedge Fund Research	CTA
COMMODITY - METALS	Hedge Fund Research	CTA
COMMODITY - MULTI	Hedge Fund Research	CTA
CONVERTIBLE ARBITRAGE	BarclayHedge	Relative Value
CONVERTIBLE ARBITRAGE	Morningstar	Relative Value
CONVERTIBLE ARBITRAGE	TASS	Relative Value
CONVERTIBLE ARBITRAGE - CREDIT	BarclayHedge	Relative Value
CONVERTIBLE ARBITRAGE - VOLATILITY	BarclayHedge	Relative Value
CREDIT ARBITRAGE	Hedge Fund Research	Relative Value
CTA / MANAGED FUTURES	EurekaHedge	CTA
CURRENCY	Morningstar	CTA
CURRENCY - DISCRETIONARY	Hedge Fund Research	CTA
CURRENCY - SYSTEMATIC	Hedge Fund Research	CTA
DEBT ARBITRAGE	Morningstar	Relative Value
DEDICATED SHORT BIAS	TASS	Short Bias
DISCRETIONARY	BarclayHedge	Global Macro
DISTRESSED DEBT	EurekaHedge	Long Only
DISTRESSED SECURITIES	BarclayHedge	Event Driven
DISTRESSED SECURITIES	Morningstar	Event Driven
DIVERSIFIED ARBITRAGE	Morningstar	Relative Value
DIVERSIFIED DEBT	EurekaHedge	Long Only

DUAL APPROACH	EurekaHedge	Long Only
EMERGING MARKETS	TASS	Emerging Markets
EMERGING MARKETS - ASIA	BarclayHedge	Emerging Markets
EMERGING MARKETS - EASTERN EUROPE/CIS	BarclayHedge	Emerging Markets
EMERGING MARKETS - GLOBAL	BarclayHedge	Emerging Markets
EMERGING MARKETS - LATIN AMERICA	BarclayHedge	Emerging Markets
EMERGING MARKETS - MENA	BarclayHedge	Emerging Markets
EMERGING MARKETS - OTHER	BarclayHedge	Emerging Markets
EMERGING MARKETS LONG/SHORT EQUITY	Morningstar	Emerging Markets
EQUITY 130-30	BarclayHedge	Long/Short
EQUITY DEDICATED SHORT	BarclayHedge	Short Bias
EQUITY HEDGE	Hedge Fund Research	Long/Short
EQUITY LONG ONLY	BarclayHedge	Long Only
EQUITY LONG ONLY - GROWTH ORIENTED	BarclayHedge	Long Only
EQUITY LONG ONLY - OPPORTUNISTIC	BarclayHedge	Long Only
EQUITY LONG ONLY - QUANTITATIVE	BarclayHedge	Long Only
EQUITY LONG ONLY - TRADING ORIENTED	BarclayHedge	Long Only
EQUITY LONG ONLY - VALUE ORIENTED	BarclayHedge	Long Only
EQUITY LONG-BIAS	BarclayHedge	Long/Short
EQUITY LONG-BIAS - GROWTH ORIENTED	BarclayHedge	Long/Short
EQUITY LONG-BIAS - OPPORTUNISTIC	BarclayHedge	Long/Short
EQUITY LONG-BIAS - QUANTITATIVE	BarclayHedge	Long/Short
EQUITY LONG-BIAS - TRADING ORIENTED	BarclayHedge	Long/Short
EQUITY LONG-BIAS - VALUE ORIENTED	BarclayHedge	Long/Short
EQUITY LONG/SHORT	BarclayHedge	Long/Short
EQUITY LONG/SHORT - GROWTH ORIENTED	BarclayHedge	Long/Short
EQUITY LONG/SHORT - OPPORTUNISTIC	BarclayHedge	Long/Short
EQUITY LONG/SHORT - QUANTITATIVE	BarclayHedge	Long/Short
EQUITY LONG/SHORT - TRADING ORIENTED	BarclayHedge	Long/Short
EQUITY LONG/SHORT - VALUE ORIENTED	BarclayHedge	Long/Short
EQUITY MARKET NEUTRAL	BarclayHedge	Market Neutral
EQUITY MARKET NEUTRAL	Hedge Fund Research	Market Neutral

EQUITY MARKET NEUTRAL	Morningstar	Market Neutral
EQUITY MARKET NEUTRAL	TASS	Market Neutral
EQUITY MARKET NEUTRAL - QUANTITATIVE	BarclayHedge	Market Neutral
EQUITY MARKET NEUTRAL - VALUE ORIENTED	BarclayHedge	Market Neutral
EQUITY SHORT-BIAS	BarclayHedge	Short Bias
EUROPE LONG/SHORT EQUITY	Morningstar	Long/Short
EVENT DRIVEN	BarclayHedge	Event Driven
EVENT DRIVEN	EurekaHedge	Event Driven
EVENT DRIVEN	Morningstar	Event Driven
EVENT DRIVEN	TASS	Event Driven
EVENT-DRIVEN	Hedge Fund Research	Event Driven
FIXED INCOME	EurekaHedge	Relative Value
FIXED INCOME - ABS/SEC. LOANS	BarclayHedge	Relative Value
FIXED INCOME - ARBITRAGE	BarclayHedge	Relative Value
FIXED INCOME - ARBITRAGE - CAPITAL STRUCTURE ARB	BarclayHedge	Relative Value
FIXED INCOME - ARBITRAGE - CREDIT DEFAULT SWAPS	BarclayHedge	Relative Value
FIXED INCOME - ARBITRAGE - ONLY CREDIT	BarclayHedge	Relative Value
FIXED INCOME - ASSET BACKED	Hedge Fund Research	Relative Value
FIXED INCOME - ASSET-BACKED LOANS	BarclayHedge	Relative Value
FIXED INCOME - COLLATERALIZED DEBT OBLIGATIONS	BarclayHedge	Relative Value
FIXED INCOME - CONVERTIBLE ARBITRAGE	Hedge Fund Research	Relative Value
FIXED INCOME - CONVERTIBLE BONDS	BarclayHedge	Relative Value
FIXED INCOME - CORPORATE	Hedge Fund Research	Relative Value
FIXED INCOME - DIVERSIFIED	BarclayHedge	Relative Value
FIXED INCOME - HIGH YIELD	BarclayHedge	Relative Value
FIXED INCOME - INSURANCE-LINKED SECURITIES	BarclayHedge	Relative Value
FIXED INCOME - LONG-ONLY CREDIT	BarclayHedge	Relative Value
FIXED INCOME - MORTGAGE BACKED	BarclayHedge	Relative Value
FIXED INCOME - SOVEREIGN	Hedge Fund Research	Relative Value
FIXED INCOME ARBITRAGE	TASS	Relative Value
FUND OF FUNDS	Hedge Fund Research	Fund of Funds
FUND OF FUNDS	TASS	Fund of Funds
FUND OF FUNDS - ARBITRAGE	BarclayHedge	Fund of Funds
FUND OF FUNDS - DEBT	Morningstar	Fund of Funds
FUND OF FUNDS - DIVERSIFIED	BarclayHedge	Fund of Funds
FUND OF FUNDS - EQUITY	Morningstar	Fund of Funds

FUND OF FUNDS - EVENT	Morningstar	Fund of Funds
FUND OF FUNDS - MACRO/SYSTEMATIC	Morningstar	Fund of Funds
FUND OF FUNDS - MULTISTRATEGY	Morningstar	Fund of Funds
FUND OF FUNDS - N/A	BarclayHedge	Fund of Funds
FUND OF FUNDS - OTHER	Morningstar	Fund of Funds
FUND OF FUNDS - RELATIVE VALUE	Morningstar	Fund of Funds
FUND TIMING	BarclayHedge	Event Driven
FUNDAMENTAL - AGRICULTURAL	BarclayHedge	CTA
FUNDAMENTAL - CURRENCY	BarclayHedge	CTA
FUNDAMENTAL - DIVERSIFIED	BarclayHedge	CTA
FUNDAMENTAL - ENERGY	BarclayHedge	CTA
FUNDAMENTAL - FINANCIAL/METALS	BarclayHedge	CTA
FUNDAMENTAL - INTEREST RATES	BarclayHedge	Global Macro
GLOBAL LONG/SHORT EQUITY	Morningstar	Long/Short
GLOBAL MACRO	Morningstar	Global Macro
GLOBAL MACRO	TASS	Global Macro
INVESTABLE INDEX	EurekaHedge	Fund of Funds
LONG / SHORT EQUITIES	EurekaHedge	Long/Short
LONG/SHORT DEBT	Morningstar	Relative Value
LONG/SHORT EQUITY HEDGE	TASS	Long/Short
MACRO	BarclayHedge	Global Macro
MACRO	EurekaHedge	Global Macro
MACRO	Hedge Fund Research	Global Macro
MANAGED FUTURES	TASS	CTA
MERGER ARBITRAGE	BarclayHedge	Event Driven
MERGER ARBITRAGE	Hedge Fund Research	Event Driven
MERGER ARBITRAGE	Morningstar	Event Driven
MULTI-ADVISOR	BarclayHedge	Multi-Strategy
MULTI-STRATEGY	BarclayHedge	Multi-Strategy
MULTI-STRATEGY	EurekaHedge	Multi-Strategy
MULTI-STRATEGY	Hedge Fund Research	Multi-Strategy
MULTI-STRATEGY	TASS	Multi-Strategy
MULTISTRATEGY	Morningstar	Multi-Strategy
NO CATEGORY	BarclayHedge	Others
OPTION STRATEGIES	BarclayHedge	Relative Value
OPTIONS STRATEGY	TASS	Relative Value
OTHER	TASS	Others
OTHERS	EurekaHedge	Others

PIPES	BarclayHedge	Others
RELATIVE VALUE	EurekaHedge	Relative Value
RELATIVE VALUE	Hedge Fund Research	Relative Value
SECTOR - ENERGY	BarclayHedge	Sector
SECTOR - ENERGY/BASIC MATERIALS	Hedge Fund Research	Sector
SECTOR - ENVIRONMENT	BarclayHedge	Sector
SECTOR - FARMING	BarclayHedge	Sector
SECTOR - FINANCIAL	BarclayHedge	Sector
SECTOR - HEALTH CARE/BIOTECH	BarclayHedge	Sector
SECTOR - METALS/MINING	BarclayHedge	Sector
SECTOR - MISCELLANEOUS	BarclayHedge	Sector
SECTOR - NATURAL RESOURCES	BarclayHedge	Sector
SECTOR - REAL ESTATE	BarclayHedge	Sector
SECTOR - TECHNOLOGY	BarclayHedge	Sector
SECTOR - TECHNOLOGY/HEALTHCARE	Hedge Fund Research	Sector
SHORT BIAS	Hedge Fund Research	Short Bias
STATISTICAL ARBITRAGE	BarclayHedge	Market Neutral
STOCK INDEX	BarclayHedge	Global Macro
STOCK INDEX,ARBITRAGE	BarclayHedge	Global Macro
STOCK INDEX,OPTION STRATEGIES	BarclayHedge	Relative Value
SYSTEMATIC	BarclayHedge	Global Macro
SYSTEMATIC DIVERSIFIED	Hedge Fund Research	Multi-Strategy
SYSTEMATIC FUTURES	Morningstar	CTA
TECHNICAL - AGRICULTURAL	BarclayHedge	CTA
TECHNICAL - CURRENCY	BarclayHedge	CTA
TECHNICAL - DIVERSIFIED	BarclayHedge	Multi-Strategy
TECHNICAL - DIVERSIFIED,CURRENCY	BarclayHedge	Multi-Strategy
TECHNICAL - ENERGY	BarclayHedge	CTA
TECHNICAL - FINANCIAL/METALS	BarclayHedge	CTA
TECHNICAL - INTEREST RATES	BarclayHedge	Global Macro
TOP-DOWN	EurekaHedge	Long Only
U.S. LONG/SHORT EQUITY	Morningstar	Long/Short
U.S. SMALL CAP LONG/SHORT EQUITY	Morningstar	Long/Short
UNDEFINED	TASS	Others
VALUE	EurekaHedge	Long Only
VOLATILITY	Hedge Fund Research	Relative Value
VOLATILITY	Morningstar	Relative Value
VOLATILITY TRADING	BarclayHedge	Relative Value

YIELD ALTERNATIVES - ENERGY INFRASTRUCTURE	Hedge Fund Research	Relative Value
YIELD ALTERNATIVES - REAL ESTATE	Hedge Fund Research	Relative Value

B. Identifying unique share classes

As a first step, we select the most reliable variables from admin module of each database for the share class matching process. We do not use compensation structure or share restriction information due to different reporting standards across databases, and the changing nature of these variables.[24] Based on the share class names, we search, identify, and merge commercial databases based on the following information.

1. Legal structure abbreviations (e.g., LP, LLC)
2. Offshore/onshore information: (e.g., Onshore, Cayman, BVI, Bermuda)
3. Leverage (e.g., 2X, 3X, 2X-3X,)
4. Share class letter (e.g., Class A, class B, class A4).
5. Currency code (e.g., USD, GBP, ZAR)

After matching all five databases, we find 43,386 unique share classes in our comprehensive database (Table 1).

C. Constructing a consolidated database

We construct a consolidated database using the five major hedge fund databases (TASS, Hedge Fund Research, BarclayHedge, EurekaHedge and Morningstar) that consist of monthly returns, asset under management (AUM) and accompanying information on share classes from January 1994 to December 2011. We focus on the post-1994 period to mitigate survivorship bias,

[24] 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.

as most databases started to collect defunct funds only after 1994. We exclude Fund of Funds and restrict each share class to have at least 12 monthly returns. We convert all non-USD[25] currencies to USD using the end-of-month spot rates obtained from Bloomberg. The total number of unique share classes is 42,386 from which 18,063 share classes (42%) are active as of December 2011[26].

Our merging approach is based on the following steps:

1. Assumptions: within each management company, multiple share classes have correlated returns.
2. Requirements: we require each fund to have at least 12 non-missing monthly returns.
3. Grouping: share classes are grouped based on the reported management company names. We create a unique "Management company key" for each firm by parsing the data of management company names for punctuations and spelling errors. We combine management company names into 11,217 management company groups.
4. Estimation: Within each management company, we estimate pairwise correlation using time series of returns of share classes.
5. Identification: Within each management company, we classify correlated share classes into groups based on the estimated correlations (using 0.99 as a limit). Then, we select one share class from each group using the following criteria: (i) the longest time series of returns, (ii) the largest average AuM, (iii) USD currency, and (iv) offshore class[27].

Figure A1 highlights the issue by showing the histograms of pairwise correlation coefficients of share classes estimated within management companies across databases. Figure shows a significant spike at the 0.99 correlation level for each database. This suggests that highly

[25] Returns and asset under management of share classes are reported in up to 29 different currencies and approximately 67% of share classes are reported in USD.
[26] Once a hedge fund decides not to report its performance, is liquidated, restructured, or merged with other hedge funds, the fund is transferred into the Graveyard database. In our sample, 17,757 share classes (58%) are defunct.
[27] We classify share classes that domiciled either in USA or Canada into onshore group.

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. Our methodology is related to Aggarwal and Jorion (2010) who use TASS data (January 1994 - December 2006) and identify multiple share classes that have return time series correlated at the 0.99 level. As a solution, they exclude one of the duplicates. In this paper, we categorize all correlated share classes into groups and select one share class to represent a unique hedge fund. We find that over 60% of the groups we identify have two correlated share classes. The consolidated database has 30,040 unique hedge funds. For single databases, the number of hedge funds ranges from 10,520 (BarclayHedge) to 7,502 (Morningstar). Our estimate of unique hedge funds is consistent with Güner, Rachev, Edelman, and Fabozzi (2010) and PerTrac 2010 study that identify over 20,000 and 23,603 unique hedge funds, respectively.

Formally, the proposed statistical procedure is closely related to clustering analysis. The procedure is based on the three main steps. First, as in clustering, we define a distance function fulfilling four properties:

- Identify: $\boldsymbol{d}(\boldsymbol{i},\boldsymbol{j})=0$ for all i and j
- Non-negativity: $\boldsymbol{d}(\boldsymbol{i},\boldsymbol{j})\geq 0$ for all i and j
- Symmetry: $\boldsymbol{d}(\boldsymbol{i},\boldsymbol{j})=\boldsymbol{d}(\boldsymbol{j},\boldsymbol{i})$ for all i and j
- Triangle inequality: $\boldsymbol{d}(\boldsymbol{i},\boldsymbol{j})\leq\boldsymbol{d}(\boldsymbol{i},\boldsymbol{k})+\boldsymbol{d}(\boldsymbol{k},\boldsymbol{j})$

These properties of the distance function satisfy that the grouping algorithm is not inconsistent suggesting that different share classes are assigned correctly to the respective group. Second, we relate the distance between funds to their correlation coefficient $\rho_{i,j}$. Since the correlation coefficient can have negative values, it is not a distance measure. Thus, we use the following transformation:

$$d(i,j) = \frac{1-\rho_{i,j}}{2}, \tag{A1}$$

which ensures that the distance is never negative. The proposed distance measure is closely related to Euclidian distance that is a very common distance measure. In fact, it is very straightforward to show that the correlation coefficient is inversely related to Euclidean distance between the standardized versions of data.

All share classes that have pairwise correlation above 0.99 are assigned into the same group. The triangle inequality satisfies that groups are formed correctly. Since we impose a correlation limit, we can bypass the most difficult task of any clustering algorithm, namely determining the number of different groups or clusters. As a summary, our statistical procedure identifies multiple share class structures that exist because of (i) multiple share classes within management companies, (ii) duplicate share classes that are reported to multiple databases. Our statistical procedure has major advantages. First, we can automatically provide frequent updates of our comprehensive aggregate database. Second, it is easier to make criteria for database merging using return time series that are much more consistently reported between databases than name information.

Table A1. Backfill bias

Backfill period is defined as the difference between the date the fund was added to the database and the inception date. Panel A shows the cross-sectional averages of the length of the backfill period. The sample average is 32 months. Panel B compares the annualized performance difference between the backfilled and the non-backfilled sample (the first 12 months returns or the first 32 months of returns are excluded).

A. Average length of backfill period across databases

Database	Days	Months	Years
TASS	961.8	31.6	2.63
Hedge Fund Research	896.3	29.4	2.45
BarclayHedge	1059.6	34.8	2.90
EurekaHedge	977.6	32.1	2.68
Average	974	32	3

B. Estimates of backfill bias

	Months excluded		
Database	12-month	32-month	Diff
Aggregate	1.20	1.82	0.63
TASS	1.18	1.76	0.57
Hedge Fund Research	1.21	1.94	0.74
BarclayHedge	1.35	2.15	0.80
EurekaHedge	0.96	1.55	0.58
Morningstar	0.90	1.37	0.47

Table A2. Supplementary tables of average performance

Panel A provides results of the average performance after the first 32 months of fund-level returns are excluded to adjust for backfill bias. Panel B provides smoothing-adjusted results of the average performance using the methodology as proposed by Getmansky, Lo and Makarov (2004). Panel C shows results of the average performance using the gross returns that are calculated as proposed by Feng (2011).

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

Database	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
TASS	6 382	63.73	5.56	7.70	0.72	2.80	0.29	0.17	0.09	0.27	0.00	0.01	0.01	0.69
						(2.39)	(13.22)	(6.34)	(1.91)	(5.33)	-(0.59)	(2.26)	(0.95)	
Hedge Fund Research	7 474	58.88	5.83	7.76	0.75	3.11	0.32	0.19	0.07	0.25	0.00	0.01	0.00	0.75
						(2.97)	(15.97)	(7.78)	(1.67)	(5.42)	-(0.40)	(2.61)	(0.74)	
BarclayHedge	7 432	64.02	6.05	7.14	0.84	3.57	0.28	0.15	0.09	0.26	0.00	0.02	0.01	0.69
						(3.31)	(13.74)	(5.93)	(2.01)	(5.54)	(0.65)	(3.35)	(1.49)	
EurekaHedge	5 876	40.67	7.76	8.15	0.95	4.93	0.31	0.17	0.09	0.31	0.00	0.02	0.01	0.68
						(3.96)	(13.06)	(5.98)	(1.81)	(5.73)	(0.10)	(2.59)	(0.75)	
Morningstar	5 464	51.35	7.14	7.46	0.95	4.57	0.29	0.17	0.08	0.26	0.00	0.02	0.01	0.68
						(4.01)	(13.24)	(6.44)	(1.71)	(5.25)	(0.04)	(2.71)	(0.93)	
Aggregate	20 923	57.90	5.83	7.89	0.74	2.98	0.31	0.17	0.09	0.29	0.00	0.02	0.01	0.71
						(2.59)	(14.15)	(6.27)	(2.07)	(5.81)	-(0.22)	(2.61)	(0.88)	

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

Database	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
TASS	8 788	65.57	7.30	9.19	0.79	4.73	0.30	0.16	0.09	0.25	0.00	0.01	0.01	0.6728
						(3.63)	(13.59)	(7.14)	(1.88)	(6.70)	-(0.35)	(2.47)	(1.72)	
Hedge Fund Research	10 332	61.68	8.05	9.08	0.88	5.57	0.31	0.17	0.07	0.20	0.00	0.01	0.01	0.7242
						(4.67)	(17.26)	(9.04)	(1.70)	(5.57)	-(0.34)	(2.27)	(1.95)	
BarclayHedge	10 520	66.46	8.29	8.11	1.01	6.14	0.28	0.14	0.07	0.23	0.00	0.02	0.02	0.658
						(5.38)	(13.66)	(6.37)	(1.65)	(6.20)	(0.66)	(3.71)	(2.45)	
EurekaHedge	8 149	41.42	9.48	9.65	0.98	6.88	0.32	0.15	0.08	0.26	0.00	0.01	0.01	0.6706
						(4.98)	(13.96)	(6.17)	(1.68)	(6.24)	(0.13)	(2.72)	(2.24)	
Morninstar	7 504	50.36	8.84	8.97	0.98	6.46	0.30	0.16	0.08	0.21	0.00	0.01	0.02	0.6777
						(4.99)	(14.54)	(6.91)	(1.75)	(5.53)	(0.07)	(2.82)	(2.50)	
Aggregate	30 195	59.26	7.77	9.13	0.85	5.23	0.31	0.15	0.08	0.25	0.00	0.01	0.01	0.6837
						(4.13)	(14.31)	(6.78)	(1.83)	(6.69)	(0.04)	(2.70)	(1.93)	

C. Results of the average performance for gross returns (Equal-weight)

Dataset	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
TASS	6 166	69.40	12.66	8.09	1.55	10.01	0.33	0.19	0.08	0.25	0.00	0.02	0.01	0.64
						(8.40)	(13.64)	(6.71)	(1.82)	(4.48)	-(0.14)	(2.95)	(1.55)	
Hedge Fund Research	7 159	63.19	13.03	8.50	1.52	10.34	0.36	0.23	0.06	0.23	0.00	0.01	0.01	0.68
						(8.71)	(14.89)	(7.87)	(1.35)	(4.18)	(0.04)	(2.62)	(1.43)	
BarclayHedge	8 289	67.90	14.58	8.26	1.75	12.03	0.34	0.18	0.09	0.26	0.01	0.02	0.01	0.63
						(9.65)	(13.38)	(5.92)	(1.78)	(4.53)	(0.94)	(3.73)	(1.75)	
EurekaHedge	6 239	40.29	15.06	9.34	1.60	12.09	0.35	0.16	0.09	0.34	0.00	0.02	0.02	0.57
						(7.99)	(11.42)	(4.39)	(1.57)	(4.88)	-(0.35)	(3.04)	(1.66)	
Morninstar	2 742	55.43	14.34	7.99	1.78	11.98	0.33	0.23	0.06	0.20	0.00	0.02	0.01	0.66
						(10.37)	(14.03)	(8.35)	(1.32)	(3.67)	(0.68)	(2.81)	(1.85)	
Aggregate	14 652	59.15	13.56	8.64	1.56	10.77	0.35	0.20	0.09	0.29	0.00	0.02	0.01	0.65
						(8.52)	(13.79)	(6.40)	(1.85)	(5.00)	(0.61)	(3.03)	(1.55)	

Table A3. Nominal and percentile limits of the AuM sweet spots

At December 2011, hedge funds are sorted into nominal size groups based on the fund-level AuM in millions of USD. In Nominal Group column, "(" character refers to the ">" sign and "]" refers to the "<=" sign. The nominal groups are used to construct size portfolios using the full time series of fund-level excess returns. The annualized Fung and Hsieh (2004) alphas of these portfolios are referred as a backward-looking alphas. At December 2011, nominal size groups are used to find corresponding percentiles of the number of funds that are sorted by AuM. For example, the first 5.6% of all funds that belong to the nominal AuM group have AuM between zero and one million USD. To construct forward-looking alphas, hedge funds are sorted into percentiles every December (the same limits as of December 2011) and 12-month holding period returns are estimated for each size portfolio using monthly rebalancing. Estimated annualized Fung and Hsieh (2004) alphas for these portfolios are the forward-looking alphas.

A. Backward-looking			B. Forward-looking		
Nominal Group	# Funds	Alpha %	Percentiles (% of funds)	# Funds	Alpha %
(0,1]	2 574	1.4	(0, 5.6]	2 743	8.9
		(1.19)			(6.73)
(1,5]	4 202	2.5	(5.6, 16.86]	5 839	6.1
		(2.26)			(5.04)
(5,10]	2 907	3.6	(16.86, 26.41]	5 714	5.5
		(3.42)			(4.46)
(10,100]	10 908	5.6	(26.41 ,71]	15 212	4.2
		(5.32)			(3.81)
(100,500]	3 908	7.3	(71, 92.15]	7 415	1.7
		(7.36)			(1.55)
(500,1000]	736	8.2	(92.15,96.62]	2 143	1.8
		(7.52)			(1.58)
(1000,…]	544	8.2	(96.62…]	1 039	1.2
		(7.54)			(0.99)

Table A4. Average performance of hedge fund investment style portfolios grouped by asset under management

Hedge funds are sorted into terciles based on the fund-level AuM each December and equal-weight excess returns of the style portfolios are calculated using each size group, 12-month holding period, and monthly rebalancing. Column A reports the size group. Descriptions of other Columns are the same as in Table 3.

CTA

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	1 268	68.77	8.99	6.71	1.34	7.69	0.10	0.02	0.16	0.14	0.01	0.04	0.02	0.27
						(5.21)	(3.32)	(0.60)	(2.82)	(2.16)	(1.44)	(5.47)	(1.63)	
Median	1 327	61.57	5.77	7.76	0.74	4.77	0.08	0.06	0.18	0.08	0.02	0.04	0.03	0.31
						(2.86)	(2.58)	(1.59)	(2.78)	(1.13)	(2.07)	(5.49)	(2.64)	
Large	927	54.15	3.96	7.40	0.53	2.65	0.08	0.03	0.20	0.09	0.01	0.04	0.02	0.26
						(1.61)	(2.35)	(0.90)	(3.06)	(1.22)	(0.79)	(5.12)	(2.30)	

Emerging markets

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	1 460	49.18	12.59	17.42	0.72	7.90	0.60	0.26	-0.06	0.54	0.00	0.01	0.00	0.52
						(2.54)	(9.75)	(3.56)	-(0.45)	(3.88)	(0.18)	(0.40)	-(0.05)	
Median	1 490	45.37	9.54	16.75	0.57	4.61	0.51	0.24	-0.06	0.50	-0.03	0.01	0.00	0.49
						(1.49)	(8.46)	(3.34)	-(0.48)	(3.63)	-(1.86)	(0.52)	-(0.01)	
Large	1 033	40.46	5.80	16.30	0.36	0.53	0.51	0.22	0.01	0.52	-0.04	0.01	0.00	0.49
						(0.18)	(8.64)	(3.14)	(0.08)	(3.86)	-(2.06)	(0.53)	-(0.09)	

Event driven

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	758	70.71	9.02	7.61	1.18	6.19	0.28	0.22	0.01	0.25	-0.02	0.01	-0.01	0.75
						(6.35)	(14.56)	(9.59)	(0.21)	(5.65)	-(2.83)	(1.93)	-(1.50)	
Median	725	64.00	6.90	6.38	1.07	4.40	0.20	0.13	0.02	0.32	-0.02	0.01	0.00	0.70
						(4.92)	(11.20)	(6.31)	(0.50)	(7.96)	-(3.36)	(1.25)	-(0.81)	
Large	547	63.62	5.65	6.51	0.86	3.11	0.19	0.11	0.00	0.34	-0.02	0.00	0.00	0.70
						(3.38)	(10.58)	(4.91)	-(0.05)	(8.31)	-(4.44)	(0.47)	(0.17)	

Fund of funds

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	5 356	67.79	2.70	7.43	0.36	-0.21	0.25	0.14	0.10	0.31	-0.01	0.02	0.01	0.54
						-(0.16)	(9.76)	(4.47)	(2.05)	(5.27)	-(1.09)	(2.67)	(1.22)	
Median	5 219	58.75	3.12	7.08	0.44	0.30	0.23	0.14	0.10	0.29	-0.01	0.01	0.01	0.54
						(0.24)	(9.61)	(4.68)	(2.07)	(5.20)	-(1.50)	(2.07)	(1.35)	
Large	3 757	51.05	3.81	7.17	0.53	0.92	0.22	0.12	0.12	0.32	-0.01	0.01	0.01	0.51
						(0.71)	(8.58)	(3.91)	(2.30)	(5.54)	-(1.53)	(1.45)	(1.74)	

Global macro

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	730	73.15	6.45	5.71	1.12	5.06	0.18	0.05	0.06	0.11	0.01	0.01	0.02	0.33
						(4.19)	(7.58)	(1.78)	(1.31)	(2.00)	(1.28)	(1.25)	(2.62)	
Median	795	67.80	4.56	5.83	0.78	2.64	0.15	0.06	0.13	0.22	0.00	0.02	0.01	0.35
						(2.18)	(6.36)	(2.13)	(2.80)	(4.06)	(0.54)	(3.60)	(1.44)	
Large	563	57.55	4.45	5.85	0.76	1.95	0.17	0.08	0.20	0.15	-0.01	0.02	0.01	0.38
						(1.65)	(7.52)	(2.80)	(4.25)	(2.89)	-(1.31)	(3.69)	(1.63)	

Long only

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	321	40.50	11.91	12.32	0.96	7.95	0.56	0.26	0.00	0.31	0.01	0.00	0.00	0.77
						(5.18)	(18.61)	(7.28)	(0.02)	(4.46)	(1.63)	-(0.13)	(0.04)	
Median	340	33.24	8.29	13.25	0.62	3.44	0.59	0.21	0.05	0.40	0.00	0.00	0.01	0.73
						(1.92)	(16.68)	(5.08)	(0.77)	(5.04)	-(0.12)	(0.58)	(0.57)	
Large	249	28.11	6.17	12.57	0.49	1.30	0.53	0.29	0.07	0.36	-0.02	0.00	-0.01	0.75
						(0.79)	(16.39)	(7.53)	(1.11)	(4.88)	-(1.59)	(0.40)	-(0.62)	

Long/Short

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	3 732	69.40	10.20	11.29	0.90	6.55	0.52	0.35	-0.01	0.15	0.00	0.01	0.00	0.79
						(4.87)	(19.57)	(11.21)	-(0.18)	(2.45)	(0.16)	(1.81)	(0.06)	
Median	3 711	64.81	7.79	10.06	0.77	4.32	0.45	0.31	0.02	0.18	0.00	0.01	0.00	0.78
						(3.58)	(19.01)	(11.03)	(0.36)	(3.31)	-(0.29)	(1.35)	(0.47)	
Large	2 578	58.30	5.95	10.17	0.58	2.28	0.44	0.28	0.05	0.17	-0.01	0.01	0.01	0.71
						(1.62)	(16.06)	(8.58)	(0.84)	(2.64)	-(1.13)	(1.22)	(1.11)	

Market neutral

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	695	73.96	6.42	4.27	1.46	5.10	0.10	0.06	0.08	0.08	-0.01	0.01	0.00	0.24
						(5.32)	(5.48)	(2.51)	(1.99)	(1.94)	-(1.16)	(2.09)	(0.71)	
Median	730	70.82	3.71	4.40	0.83	2.23	0.12	0.04	0.06	0.13	-0.01	0.01	0.00	0.33
						(2.40)	(6.67)	(1.92)	(1.68)	(3.10)	-(1.28)	(2.05)	(0.49)	
Large	494	61.34	3.72	4.50	0.81	2.11	0.15	-0.03	0.06	0.11	-0.01	0.01	0.01	0.38
						(2.31)	(8.39)	-(1.56)	(1.68)	(2.81)	-(1.93)	(1.32)	(0.88)	

Multi-strategy

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	1 635	68.07	7.92	8.90	0.89	6.21	0.12	0.08	0.22	0.22	0.02	0.04	0.04	0.31
						(3.25)	(3.21)	(1.74)	(2.87)	(2.54)	(1.61)	(4.57)	(3.71)	
Median	1 673	62.82	6.28	7.70	0.81	4.37	0.17	0.08	0.19	0.20	0.02	0.03	0.03	0.33
						(2.69)	(5.19)	(2.14)	(2.99)	(2.79)	(1.68)	(4.09)	(3.24)	
Large	1 087	55.57	5.59	6.84	0.81	3.59	0.14	0.05	0.19	0.29	0.02	0.02	0.02	0.32
						(2.47)	(5.04)	(1.46)	(3.30)	(4.41)	(1.78)	(2.73)	(2.30)	

Others

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	472	81.36	8.95	8.60	1.04	5.80	0.31	0.18	0.09	0.26	-0.01	0.01	0.00	0.55
						(3.92)	(10.59)	(5.09)	(1.58)	(3.97)	-(0.85)	(1.83)	-(0.09)	
Median	469	80.81	6.38	8.38	0.76	3.33	0.33	0.21	0.09	0.21	0.00	0.01	0.01	0.61
						(2.49)	(12.61)	(6.64)	(1.77)	(3.44)	-(0.07)	(1.49)	(0.63)	
Large	345	80.29	4.61	6.05	0.75	2.25	0.20	0.14	0.12	0.20	0.00	0.00	0.01	0.52
						(2.08)	(9.18)	(5.63)	(2.83)	(4.16)	-(0.66)	-(0.10)	(1.36)	

Relative value

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	1 314	67.43	7.83	5.56	1.40	5.42	0.14	0.07	0.11	0.39	0.00	0.00	-0.01	0.64
						(6.32)	(8.45)	(3.65)	(3.34)	(10.20)	-(0.84)	(1.03)	-(1.43)	
Median	1 384	64.67	4.05	5.44	0.74	1.54	0.13	0.06	0.10	0.39	-0.01	0.00	0.00	0.66
						(1.87)	(8.16)	(3.04)	(3.21)	(10.70)	-(2.85)	-(0.14)	(0.06)	
Large	978	60.63	4.40	5.64	0.78	1.95	0.11	0.03	0.10	0.44	-0.01	0.00	0.00	0.62
						(2.17)	(6.15)	(1.35)	(2.88)	(10.95)	-(2.59)	-(1.06)	-(0.04)	

Sector

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	384	70.57	11.44	16.03	0.71	6.58	0.65	0.47	0.04	0.16	0.00	0.00	-0.01	0.64
						(2.64)	(13.23)	(8.09)	(0.43)	(1.44)	-(0.02)	(0.20)	-(0.86)	
Median	366	64.75	10.90	13.93	0.78	6.76	0.61	0.45	0.03	0.04	0.00	0.01	0.00	0.68
						(3.32)	(15.33)	(9.46)	(0.32)	(0.46)	(0.38)	(0.88)	(0.38)	
Large	247	56.28	8.33	13.04	0.64	4.24	0.52	0.42	0.05	0.20	0.00	0.00	0.01	0.66
						(2.17)	(13.60)	(9.26)	(0.60)	(2.24)	(0.08)	(0.42)	(0.84)	

Short bias

Size Group	# Funds	% of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
Small	62	74.19	0.86	15.42	0.06	3.55	-0.68	-0.35	0.21	0.36	0.00	0.00	0.00	0.55
						(1.33)	-(13.08)	-(5.63)	(1.99)	(3.01)	(0.08)	-(0.27)	-(0.18)	
Median	70	74.29	2.70	16.69	0.16	7.07	-0.61	-0.55	-0.11	0.16	0.01	0.02	0.01	0.50
						(2.34)	-(10.26)	-(7.73)	-(0.91)	(1.20)	(0.28)	(1.21)	(0.43)	
Large	51	70.59	-1.43	14.57	-0.10	2.48	-0.68	-0.49	-0.04	0.21	-0.01	0.00	0.01	0.67
						(1.15)	-(15.97)	-(9.65)	-(0.50)	(2.14)	-(1.03)	-(0.08)	(0.50)	

Table A5. Results of average performance by investment styles

Hedge funds are grouped into portfolios based on the reported investment styles. Equal-weight excess returns of the portfolios are calculated using monthly rebalancing. Time period is January 1994 - December 2011. Table reports the number of funds, Sharpe ratio, Fung and Hsieh (2004) alpha (annualized), and R-square of the model.

	TASS				Hedge Fund Research				BarclayHedge				EurekaHedge				Morningstar			
MainStrategy	# Funds	Sharpe Ratio	Alpha (%)	R2	# Funds	Sharpe Ratio	Alpha (%)	R2	# Funds	Sharpe Ratio	Alpha (%)	R2	# Funds	Sharpe Ratio	Alpha (%)	R2	# Funds	Sharpe Ratio	Alpha (%)	R2
CTA	866	0.64	5.37	0.27	563	1.11	6.51	0.28	1233	1.48	7.25	0.37	1053	1.03	9.90	0.31	702	0.89	9.69	0.30
			(2.67)				(4.66)				(7.08)				(4.85)				(4.14)	
Emerging markets	786	0.57	4.13	0.50	1091	0.75	8.22	0.45	921	0.67	6.05	0.51	1326	0.71	7.71	0.47	1100	0.80	11.01	0.38
			(1.60)				(2.71)				(2.23)				(2.53)				(3.01)	
Event Driven	618	1.09	4.63	0.73	760	1.18	5.41	0.78	489	1.24	5.63	0.71	284	1.17	6.12	0.73	395	1.32	6.09	0.72
			(5.77)				(7.16)				(6.66)				(6.08)				(7.29)	
Fund of Funds	4508	0.42	0.56	0.49	3566	0.50	1.04	0.53	3523	0.53	1.23	0.53	3163	0.51	1.10	0.51	3802	0.50	1.00	0.50
			(0.42)				(0.91)				(1.05)				(0.91)				(0.81)	
Global Macro	551	0.89	3.61	0.44	461	1.10	4.80	0.42	803	1.50	6.51	0.39	443	1.27	6.00	0.41	443	1.35	6.99	0.30
			(3.01)				(4.07)				(6.84)				(5.03)				(5.69)	
Long Only									370	0.66	4.36	0.86	485	0.54	2.09	0.76	185	0.92	5.80	0.76
											(3.34)				(1.53)				(4.62)	
Long/Short	3092	0.93	6.02	0.73	3081	0.80	5.06	0.78	2192	0.99	6.51	0.77	2429	1.13	7.64	0.70	2440	0.98	6.45	0.77
			(4.78)				(4.25)				(5.67)				(6.14)				(5.62)	
Market Neutral	568	1.22	4.59	0.30	739	1.27	3.73	0.32	367	1.33	3.91	0.24	515	1.26	4.73	0.46	343	1.42	5.12	0.37
			(4.61)				(4.87)				(5.25)				(4.88)				(5.90)	
Multi-Strategy	1177	1.00	4.90	0.56	1840	1.18	6.63	0.32	1448	0.79	7.66	0.29	462	1.37	6.48	0.53	570	1.41	6.52	0.54
			(4.04)				(5.14)				(3.72)				(6.27)				(6.64)	
Others	358	1.34	5.13	0.46					597	1.25	6.02	0.62	232	1.24	9.51	0.34	500	0.83	3.86	0.45
			(5.23)								(6.34)				(4.96)				(3.00)	
Relative Value	581	0.83	2.26	0.60	1190	1.25	4.21	0.68	1365	1.42	4.63	0.65	895	1.44	4.91	0.61	786	1.25	4.35	0.57
			(2.50)				(5.95)				(6.69)				(6.76)				(5.10)	
Sector					536	0.87	8.78	0.65	698	0.94	7.56	0.73								
							(4.17)				(4.98)									
Short Bias	50	0.06	3.99	0.57	70	0.01	4.87	0.74	37	0.16	7.08	0.80	14	0.65	15.29	0.01	38	0.40	5.91	0.33
			(1.86)				(2.27)				(4.18)				(2.51)				(2.55)	

Table A6. Average performance by manager domicile

Table describes performance results of the equal-weight domicile portfolios that are created based on the manager domicile region. All results are obtained using the aggregate database. Panel A reports the firm domicile. Descriptions of other Columns are the same as in Table 3.

Domicile	# Funds	% Of Dead	Mean ER %	Std ER %	Sharpe	Alpha %	SP-RF	RL-SP	TY-RF	BAA-TY	PTFSBD-RF	PTFSFX-RF	PTFSCOM-RF	RSQ
All funds	19 502	58.92	8.14	7.28	1.11	5.66	0.30	0.16	0.07	0.27	0.00	0.01	0.01	0.68
						(5.56)	(14.54)	(6.48)	(1.82)	(5.71)	(0.16)	(3.09)	(1.41)	
USA (Onshore)	10 009	68.26	8.34	6.24	1.32	6.40	0.27	0.17	0.03	0.17	0.00	0.01	0.01	0.71
						(7.77)	(16.22)	(8.49)	(1.03)	(4.47)	(0.42)	(3.61)	(1.77)	
All offshore funds	9 493	49.08	7.50	8.60	0.87	4.56	0.32	0.14	0.10	0.37	0.00	0.02	0.01	0.59
						(3.36)	(11.50)	(4.30)	(1.85)	(5.84)	-(0.17)	(2.39)	(0.92)	
Caribbean	493	64.50	9.61	7.94	1.20	7.30	0.25	0.12	0.06	0.27	0.00	0.01	0.01	0.44
						(5.00)	(8.61)	(3.41)	(1.06)	(3.92)	-(0.58)	(0.77)	(1.20)	
Europe	6 664	48.50	6.53	8.32	0.78	3.68	0.29	0.13	0.13	0.37	0.00	0.02	0.01	0.55
						(2.67)	(10.40)	(3.99)	(2.42)	(5.80)	-(0.04)	(3.11)	(0.98)	
Asia/Pacific	1 432	49.23	7.90	10.10	0.78	5.29	0.37	0.16	0.04	0.31	0.02	0.01	0.01	0.50
						(3.00)	(10.41)	(3.74)	(0.57)	(3.79)	(1.62)	(1.29)	(0.64)	
Rest of world	904	44.69	10.72	12.70	0.85	6.58	0.43	0.18	0.03	0.49	-0.02	0.00	0.00	0.57
						(3.22)	(10.44)	(3.63)	(0.32)	(5.09)	-(1.80)	(0.11)	(0.01)	

Table A7. Average performance by fund domicile region

Hedge funds are grouped into portfolios based on the reported fund domicile regions. Equal-weight excess returns of the portfolios are calculated using monthly rebalancing. Time period is January 1994 - December 2011. Table reports the number of funds, Sharpe ratio, Fung and Hsieh (2004) alpha (annualized), and R-square of the model.

	TASS				Hedge Fund Research				BarclayHedge			
Domicile	# Funds	Sharpe Ratio	Alpha (%)	R-Square	# Funds	Sharpe Ratio	Alpha (%)	R-Square	# Funds	Sharpe Ratio	Alpha (%)	R-Square
ALL	8647	0.99	4.70	0.66	10331	1.11	5.58	0.71	10520	1.23	6.16	0.65
			(4.47)				(5.84)				(6.29)	
USA (Onshore)	2419	1.26	6.74	0.75	3809	1.23	7.00	0.78	4797	1.45	7.26	0.61
			(7.72)				(8.03)				(7.89)	
All Offshore Funds	6228	0.82	3.46	0.59	6522	0.99	4.59	0.62	5723	1.02	5.10	0.64
			(2.93)				(4.24)				(4.59)	
Caribbean Funds	4122	0.88	3.80	0.60	4025	0.94	4.44	0.62	3219	1.07	5.41	0.61
			(3.40)				(3.94)				(4.79)	
Europe	1114	0.44	0.98	0.41	1194	0.80	3.86	0.56	1609	0.76	3.71	0.51
			(0.60)				(2.65)				(2.55)	
Asia/Pacific	127	0.67	4.32	0.41	73	0.52	4.94	0.30	166	1.31	10.30	0.34
			(1.86)				(1.20)				(5.37)	
Rest of World	865	0.80	4.41	0.52	1230	1.17	5.13	0.54	729	0.99	5.26	0.69
			(2.74)				(5.40)				(4.74)	

	EurekaHedge				Morningstar				Aggregate			
Domicile	# Funds	Sharpe Ratio	Alpha (%)	R-square	# Funds	Sharpe Ratio	Alpha (%)	R-square	# Funds	Sharpe Ratio	Alpha (%)	R-square
ALL	8138	1.22	6.89	0.65	7502	1.23	6.46	0.66	30 195	1.05	5.23	0.67
			(6.18)				(6.30)				(5.05)	
USA (Onshore)	2190	1.59	8.45	0.68	2342	1.43	7.99	0.72	9 499	1.33	6.91	0.73
			(9.41)				(8.78)				(8.11)	
All Offshore Funds	5948	1.04	5.95	0.62	5160	1.05	5.29	0.59	20 696	0.89	4.17	0.62
			(4.58)				(4.51)				(3.53)	
Caribbean Funds	3461	1.16	6.45	0.59	2873	1.16	5.84	0.61	11 484	0.92	4.27	0.60
			(5.29)				(5.40)				(3.72)	
Europe	1725	0.67	3.24	0.56	1218	0.72	3.87	0.47	5 334	0.63	2.51	0.55
			(2.04)				(2.22)				(1.71)	
Asia/Pacific	193	0.76	8.97	0.41	711	0.49	5.05	0.23	1 105	0.97	7.20	0.40
			(2.57)				(1.24)				(3.60)	
Rest of World	569	1.00	7.68	0.54	358	1.03	5.40	0.50	2 773	1.00	5.52	0.63
			(4.17)				(4.12)				(4.34)	

Table A8. Results of persistence for monthly rebalanced portfolios

Table shows hedge fund performance persistence returns. Panel A (Panel B) shows results of equal-weight (Value-weight) portfolios. Using t-statistic of Fung and Hsieh (2004) alpha, funds are sorted into decile portfolios that are rebalanced at (i) quarterly (ii) semiannual, or (iii) annual horizons. We estimate alpha t-statistics using 24 the most recent return observations. Monthly rebalanced returns are calculated for each holding period. Portfolio weights are readjusted monthly whenever a fund disappears from the sample. Table shows the annualized alphas in percentages and the t-values in parentheses. The column DropOut (%) 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, Column DropOut (%) contains the difference in drop out rates between the bottom and the top portfolio.

A. Equal-weight

Top	Quarter		Semiannual		Annual	
Database	Monthly	DropOut %	Monthly	DropOut %	Monthly	DropOut %
Aggregate	10.21	2.24	7.66	4.89	5.25	9.61
	(5.30)		(5.77)		(5.56)	
TASS	9.32	2.25	7.12	4.71	5.26	9.28
	(4.73)		(5.17)		(5.45)	
Hedge Fund Research	10.29	2.33	7.39	4.92	4.86	9.61
	(5.96)		(6.62)		(6.04)	
BarclayHedge	10.53	1.83	7.92	4.69	5.20	10.01
	(4.97)		(5.80)		(5.31)	
EurekaHedge	9.87	1.30	7.36	2.94	5.33	6.01
	(3.97)		(4.18)		(4.43)	
Morningstar	10.02	2.39	7.44	4.62	5.09	9.19
	(6.11)		(6.75)		(5.86)	
Bottom						
Aggregate	2.28	5.37	3.93	11.93	2.83	21.67
	(0.93)		(2.09)		(2.32)	
TASS	0.64	5.09	3.97	11.85	2.18	22.31
	(0.22)		(1.94)		(1.63)	
Hedge Fund Research	2.43	5.57	4.74	12.12	3.53	21.56
	(0.91)		(2.42)		(2.91)	
BarclayHedge	2.98	5.19	3.77	11.64	3.30	20.85
	(1.25)		(2.19)		(2.88)	
EurekaHedge	9.38	2.83	10.21	6.27	8.33	11.68
	(2.70)		(4.10)		(4.96)	
Morningstar	7.55	4.21	8.28	8.72	6.45	14.21
	(2.44)		(3.59)		(4.29)	
Spread						
Aggregate	7.93	3.13	3.73	7.03	2.42	12.07
	(2.93)		(1.92)		(1.89)	
TASS	8.68	2.84	3.15	7.14	3.07	13.03
	(3.09)		(1.52)		(2.23)	
Hedge Fund Research	7.86	3.24	2.64	7.20	1.33	11.96
	(2.58)		(1.27)		(1.00)	
BarclayHedge	7.55	3.37	4.15	6.96	1.90	10.84
	(2.60)		(2.07)		(1.40)	
EurekaHedge	0.49	1.52	-2.85	3.33	-3.00	5.67
	(0.15)		-(1.16)		-(1.75)	
Morningstar	2.47	1.81	-0.84	4.10	-1.36	5.02
	(0.77)		-(0.36)		-(0.87)	

B. Value-weight

Top	Quarter		Semiannual		Annual	
Database	Monthly	DropOut %	Monthly	DropOut %	Monthly	DropOut %
Aggregate	9.19	2.28	7.02	5.03	4.84	10.08
	(4.96)		(5.79)		(5.43)	
TASS	9.17	2.25	7.46	4.90	5.78	9.91
	(5.22)		(6.34)		(6.21)	
Hedge Fund Research	8.86	2.36	6.79	5.13	4.61	9.89
	(4.57)		(5.68)		(5.54)	
BarclayHedge	9.79	1.88	7.23	4.82	4.96	10.43
	(4.79)		(5.59)		(4.99)	
EurekaHedge	9.33	1.19	6.68	2.77	4.03	4.58
	(3.19)		(3.51)		(3.19)	
Morningstar	8.96	2.08	7.19	3.85	4.87	8.60
	(4.91)		(6.54)		(5.59)	
Bottom						
Aggregate	1.80	5.80	4.94	12.85	4.01	23.50
	(0.72)		(2.57)		(2.83)	
TASS	2.84	5.64	4.96	12.93	2.24	24.27
	(0.69)		(1.88)		(1.06)	
Hedge Fund Research	4.46	5.67	5.14	12.04	5.34	21.31
	(1.91)		(2.68)		(3.85)	
BarclayHedge	0.58	5.46	4.59	12.44	4.18	22.09
	(0.21)		(2.31)		(2.81)	
EurekaHedge	10.41	2.91	9.54	6.42	7.17	12.01
	(2.93)		(3.89)		(4.06)	
Morningstar	8.18	3.44	8.75	7.56	5.86	12.31
	(2.72)		(3.74)		(3.41)	
Spread						
Aggregate	7.39	3.53	2.08	7.82	0.83	13.42
	(2.50)		(1.01)		(0.56)	
TASS	6.33	3.38	2.50	8.03	3.54	14.36
	(1.47)		(0.94)		(1.69)	
Hedge Fund Research	4.39	3.31	1.65	6.90	-0.73	11.43
	(1.55)		(0.78)		-(0.50)	
BarclayHedge	9.20	3.58	2.64	7.62	0.78	11.66
	(2.91)		(1.21)		(0.48)	
EurekaHedge	-1.08	1.72	-2.85	3.65	-3.14	7.43
	-(0.23)		-(0.97)		-(1.51)	
Morningstar	0.78	1.36	-1.57	3.71	-1.00	3.71
	(0.23)		-(0.65)		-(0.58)	

Table A9. Cross-sectional regressions

Table shows results of the cross-sectional regressions based on the Fama-McBeth (1973) procedure. Monthly 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 since inception date of the fund. Time period is from January 1994 - December 2011. All parameter estimates are multiplied by 100. T-statistics are reported in parenthesis. Strategy and domicile dummy variables are applied to control for fixed effects.

Variable	Aggregate		TASS		Hedge Fund Research		BarclayHedge		EurekaHedge		Morningstar	
Incentive fee	0.766	0.541	0.559	0.325	0.989	0.719	0.834	0.625	0.695	1.316	1.660	1.407
	(2.61)	(1.97)	(0.95)	(0.52)	(2.69)	(2.20)	(2.06)	(1.74)	(0.96)	(1.90)	(4.35)	(3.90)
High-water mark	0.108	0.080	0.161	0.112	0.036	0.024	0.124	0.108	-0.130	-0.128	0.008	-0.023
	(3.39)	(2.67)	(3.42)	(2.61)	(1.12)	(0.80)	(2.72)	(2.35)	-(0.80)	-(0.70)	(0.12)	-(0.36)
Lockup	0.110	0.061	0.084	0.050	0.164	0.093	0.119	0.077	0.094	0.048	0.052	0.015
	(3.21)	(1.94)	(2.11)	(1.27)	(4.19)	(2.64)	(2.91)	(1.99)	(1.25)	(0.59)	(1.21)	(0.33)
Notice	0.414	0.410	1.321	1.205	0.280	0.304	0.052	0.062	0.841	0.399	0.596	0.612
	(1.79)	(1.92)	(4.55)	(4.35)	(0.83)	(0.97)	(0.21)	(0.27)	(1.68)	(0.91)	(1.66)	(1.88)
Redemption	0.088	0.072	0.130	0.165	0.032	0.033	0.046	0.037	0.088	0.085	-0.045	0.009
	(1.30)	(1.32)	(1.55)	(2.07)	(0.52)	(0.61)	(0.47)	(0.47)	(0.60)	(0.64)	-(0.56)	(0.12)
Control variables:												
Management fee	4.212	3.687	10.876	10.500	5.049	5.789	13.037	10.744	13.034	11.887	2.453	4.589
	(1.68)	(1.62)	(2.09)	(2.45)	(1.53)	(1.88)	(4.15)	(3.38)	(2.54)	(2.31)	(0.41)	(0.71)
Lag (Size)		-0.052		-0.066		-0.068		-0.087		-0.060		-0.113
		-(3.81)		-(3.54)		-(4.75)		-(5.30)		-(2.57)		-(6.04)
Lag (Age)		-0.080		-0.063		-0.089		-0.075		-0.067		-0.086
		-(3.59)		-(2.91)		-(4.01)		-(3.44)		-(2.18)		-(3.66)
Lag (Flow)		1.487		1.651		1.562		1.303		1.337		1.626
		(5.16)		(5.13)		(5.83)		(4.53)		(2.93)		(4.33)
Intercept	0.625	1.069	0.397	1.002	0.807	1.411	0.548	1.129	0.770	1.072	0.489	1.110
	(2.47)	(3.16)	(2.33)	(3.41)	(2.07)	(2.62)	(2.09)	(3.21)	(1.84)	(2.08)	(2.24)	(2.71)
Strategy Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Domicile Dummies?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Table A10. Proportions of hedge funds by investment styles and fund domicile regions

Table shows proportions of hedge funds grouped by the reported investment styles and fund domiciles. All hedge funds are required to have at least 12 monthly return observations.

A. Proportions of funds by investment styles

Investment style	Aggregate		TASS		Hedge Fund Research		BarclayHedge		EurekaHedge		Morningstar	
	# Funds	%	# Funds	%	# Funds	%	# Funds	%	# Funds	%	# Funds	%
CTA	2 964	9.8	884	10.1	563	5.4	1 233	11.7	1 059	13.0	702	9.4
Emerging Markets	3 756	12.4	794	9.0	1 091	10.6	921	8.8	1 329	16.3	1 100	14.7
Event Driven	1 539	5.1	630	7.2	760	7.4	489	4.6	284	3.5	395	5.3
Global Macro	1 887	6.2	569	6.5	461	4.5	803	7.6	444	5.4	444	5.9
Long only	814	2.7	-		-		370	3.5	485	6.0	185	2.5
Long/Short	8 430	27.9	3 125	35.6	3 081	29.8	2 192	20.8	2 430	29.8	2 440	32.5
Market Neutral	1 626	5.4	574	6.5	739	7.2	367	3.5	515	6.3	343	4.6
Multi-Strategy	3 912	13.0	1 205	13.7	1 841	17.8	1 448	13.8	462	5.7	570	7.6
Others	1 265	4.2	361	4.1	NA		597	5.7	232	2.8	501	6.7
Relative Value	3 117	10.3	596	6.8	1 190	11.5	1 365	13.0	895	11.0	786	10.5
Sector	753	2.5	-		536	5.2	698	6.6	-		-	
Short Bias	132	0.4	50	0.6	70	0.7	37	0.4	14	0.2	38	0.5
Total	30 195	100	8 788	100.0	10 332	100	10 520	100	8 149	100	7 504	100

B. Proportions of funds by fund domicile

Investment style	Aggregate		TASS		Hedge Fund Research		BarclayHedge		EurekaHedge		Morningstar	
	# Funds	%	# Funds	%	# Funds	%	# Funds	%	# Funds	%	# Funds	%
Onshore	9 499	31.5	2 476	28.2	3 809	36.9	4 797	45.6	2 194	26.9	2 342	31.2
Offshore:												
Asia/Pacific	1 105	3.7	128	1.5	73	0.7	166	1.6	193	2.4	711	9.5
Caribbean	11 484	38.0	4 164	47.4	4 026	39.0	3 219	30.6	3 465	42.5	2 873	38.3
Europe	5 334	17.7	1 136	12.9	1 194	11.6	1 609	15.3	1 728	21.2	1 220	16.3
Rest of world	2 773	9.2	884	10.1	1 230	11.9	729	6.9	569	7.0	358	4.8
Total	20 696	68.5	6 312	71.8	6 523	63.1	5 723	54.4	5 955	73.1	5 162	68.8
Onshore+Offshore	30 195	100	8 788	100	10 332	100	10 520	100	8 149	100	7 504	100

Figure A1. 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 that is presented in Appendix. Each share class pair in correlation analysis is required to have at least 12 non-missing monthly return observations. Share classes have return time series between January 1994 – June 2011.