Data Appendix for Hedge Fund Performance: Are Stylized Facts Sensitive to Which Database One Uses?

Juha Joenväärä, Mikko Kauppila, Robert Kosowski, and Pekka Tolonen

This version: 31 March 2019

Contents

1. Overview ... 2
2. Database-Level Steps ... 2
   2.1 Harmonizing Fund Characteristics ... 2
   2.2 Fund Filters ... 4
3. Constructing Harmonized Identifiers ... 4
   3.1 Firm Identifier (Firm ID) ... 4
   3.2 Fund Identifier (Fund ID) ... 5
   3.3 Share Class Identifier (*Share Class ID*) ... 6
      3.3.1 Automated Share Class Clustering ... 7
      3.3.2 Manual Corrections ... 7
4. Fund-Level Aggregation ... 8
   4.1 Static Characteristics ... 8
   4.2 Assets Under Management (AUM) ... 9
   4.3 Returns ... 9
5. Bias Correction ... 9
6. Linking Fund of Funds Holdings ... 11
   6.1 Holdings Data ... 11
   6.2 Linking Procedure ... 12
   6.3 Quarterly Returns ... 12
   6.4 Discretionary Liquidity Restrictions ... 13

Bibliography... 14

1. Overview

Our hedge fund database consolidates seven commercial databases over the period 1994 to 2016: BarclayHedge, EurekaHedge, eVestment, Hedge Fund Research (HFR), Morningstar, Lipper TASS, and Preqin. The preliminary database-level cleaning steps are described in Section 2.

We assign each database-level fund a harmonized firm identifier, fund identifier, and share class identifier (Section 3). Each firm has one or more funds, and each fund has one or more share classes (e.g., onshore and offshore classes). Using these assignments, we can eliminate duplicate database-level funds and aggregate their data at firm, fund, or share class levels—all within and across databases.

Specifically, we show how we combine fund-level information across multiple databases to get the most complete per-fund coverage (Section 4). For example, a fund may have reported its early returns to Lipper TASS, but later switched to HFR; our method allows the combining of such disjointed information. Finally, to correct for survivorship and backfill bias, we utilize all available information on fund listing dates (Section 5).

2. Database-Level Steps

In this section we describe the steps applied to each of the seven databases before merging them.

2.1 Harmonizing Fund Characteristics

We first harmonize the hedge fund characteristics in each database. We harmonize database-specific investment styles into nine standard styles that follow the broad styles in SEC Form PF: Fund of funds, Credit, Equity, Event driven, Macro, Managed futures / CTA, Multi-strategy, Relative value, and Other (the funds of funds will be omitted at analysis stage). The mapping rule is given in Table A1. For databases other than BarclayHedge and Preqin, we can simply map the main strategy into

Form PF style using the primary rule given in Panel A. Preqin sometimes lists multiple main strategies, in which case we apply the primary rule to each listed main strategy, and use the resulting mapping only if it is unique. For BarclayHedge, the primary rule is often undefined, especially for emerging market funds. Therefore, Panel B shows a secondary mapping rule for BarclayHedge, which utilizes the secondary strategy. Finally, Panel C gives database-specific mapping exceptions, which override the primary and secondary rules. The database-specific style mapping may be unavailable (for example, for emerging markets funds in Lipper TASS), in which case we set the database-level Form PF style as missing. However, we resolve this issue later by combining style information across databases (Section 4.1).

We harmonize fund domicile into eight broad domiciles: Asia, Caribbean, Europe, Pacific, North America, Central America, South America, and Others. The mapping rule is given in Table A2.

We convert all management and incentive fees into fractions of assets under management (AUM), and convert liquidity periods (subscription, notice, redemption, and lockup periods) into years. For funds with a soft lockup, we set the lockup period to zero. Although funds may employ different hurdle rates, we construct a hurdle rate dummy which takes a value of one if the fund employs any non-zero hurdle rate. We measure the use of leverage both with a dummy variable, and average leverage defined as percentage margin-to-equity ratio, which takes a value of zero for non-levered funds and a value of 100 for 2:1 levered funds.

We convert monthly fund returns into USD and monthly AUM into millions of USD using spot currency rates downloaded from Thomson Reuters Datastream. If the USD-converted return R doesn't satisfy $-1 < R < 3$, then we set the return as missing. We set the AUM as missing if AUM ≤ 0.

2.2 Fund Filters

We remove non-investable indices, which are contained in many databases. We remove funds that report gross-of-fees returns: this indicator is available in the Lipper TASS, Hedge Fund Research, and BarclayHedge databases. We also remove funds that don't report returns at a monthly frequency; such an indicator is only available in Lipper TASS, so for robustness we also remove funds that do not report returns on any non-quarter-end month.

At database level, we only retain funds with returns reported over 1994 to 2016. Although our database snapshots contain early 2017 returns (the snapshots are dated March 2017 at earliest), ending the sample on December 2016 alleviates the concern of strategic reporting delays (Aragon and Nanda 2017). At this stage we still include funds of funds.

3. Constructing Harmonized Identifiers

In this section we describe how we construct our harmonized firm, fund, and share class identifiers.

3.1 Firm Identifier (Firm ID)

For each database-level fund, all databases include the associated firm name, i.e., the name of the hedge fund's advisor. However, different databases have slightly different naming conventions (e.g., "Catharsis Capital Management LLC" and "Catharsis Capital, L.L.C.").

To create the harmonized firm identifier that is common across databases, we manually cluster all database firm names into name-based firm clusters. As an additional context clue, we use the firm name's most common fund name appearing in the databases.[1] We refer to the resulting firm identifier as *Firm ID*.

[1] Technically, we define the most common fund name as the name with minimum median string distance to rest of the names. As string distance metric, we use the restricted Damerau-Levenshtein distance, and pre-process all names by disregarding their case and removing periods, commas, and initial word "the."

3.2 Fund Identifier (Fund ID)

To create a harmonized fund identifier that is common among share classes of the same fund, and also common across databases, we first manually cluster all database fund names into tentative name-based fund clusters. As additional context clues, we use the firm name, domicile, currency, and investment style.

However, unlike in firm name clustering, we don't use the fund name clusters as final fund identifiers, because this clustering is prone to two types of errors. First, funds with similar names may sometimes implement different strategies (e.g., "ABC Fund I, LLC" and "ABC Fund II, LLC"), which we cannot detect during manual clustering. Second, databases may contain very different naming conventions for the same fund (e.g., "ABC Fund I, LLC" and "Amber, Brown, and Collins Fund I, LLC"), and our manual matching may miss these kinds of hard-to-detect differences in naming conventions. To correct for errors introduced during manual name clustering, we employ two successive graph-based algorithms that first break and then join the tentative name-based fund clusters using return correlations.

To resolve the first type of error, we break down each cluster into sub-clusters. To do this, we form an undirected graph of the cluster's database-level funds, where two funds are considered linked if they meet any of the following three conditions:

1. Their return correlation is at least 0.8. We calculate both Spearman and Pearson correlation for both base currency and USD-converted returns, and finally use the maximum of the four correlations. We require at least four common return observations.
2. Their *Firm ID*s and fund names are equal. When comparing fund names, we pre-process the names by disregarding their case and removing periods, commas, and initial word "the".
3. Their Preqin fund identifiers are equal. The Preqin database is different from the others because uniquely it has information on share classes: in Preqin two database-level funds

can have the same fund identifier but different share class identifiers. In all other databases, two database-level funds always have distinct fund identifiers.

After forming this graph and its links, we use its connected subcomponents as the new clusters. However, this algorithm treats funds with non-overlapping return series (less than four common return observations) as distinct unless conditions 2 or 3 are met, or unless the funds are linked transitively, which can produce excessive breaking of clusters. For this reason, we only break the cluster if there is at least one correlation below 0.8 where conditions 2 and 3 are also not met.

To resolve the second type of error, we join clusters into super clusters. For robustness, we only do this within each *Firm ID*. For each cluster we select its cluster-representative database-level fund, defined as the fund with longest return series, and resolve ties by preferring USD-denominated funds. Next, we form an undirected graph of the cluster-representative funds within the *Firm ID*, where two funds are considered linked if they have at least four common return observations with a Pearson correlation (maximum between base currency and USD-converted returns) of at least 0.99. (Using Spearman correlations sometimes leads to false positives, so we don't use it at the joining stage.) Again, we use the connected subcomponents of this graph as the new clusters.

Finally, the fund clusters formed by correcting the initial name-based fund clustering with the two successive graph-based algorithms define the harmonized fund identifier, which we refer to as *Fund ID*. This identifier will still be manually refined in Section 3.3.2.

3.3 Share Class Identifier (*Share Class ID*)

Of all seven commercial hedge fund databases, only Preqin identifies share classes within a fund (e.g., onshore and offshore classes). We want to create a similar *Share Class ID* that maps each database-level fund within a *Fund ID* to a common share class.

3.3.1 Automated Share Class Clustering

Our share class identification is based on clustering of AUM series, with the logic that different share classes should have different investor bases and thus different AUMs. We start by collecting the 1994 to 2016 USD-converted year-end AUMs of all database-level funds. For simplicity we currently exclude funds of funds (FOF) from share class clustering. We identify a *Fund ID* as a FOF if any of its database-level funds is identified as a FOF based on its investment style.

We measure the error between two AUM observations $X > 0$ and $Y > 0$ by their relative difference $|X - Y| / [(X + Y) / 2]$. Within each *Fund ID*, we create a graph of its database-level funds, where two funds are linked if their median error is below 0.1. The connected subcomponents of this graph define the share class identifier, referred to as *Share Class ID*.

3.3.2 Manual Corrections

We manually verify the correctness of the *Share Class ID*s produced by the automatic clustering. During this step, we make manual corrections mostly to *Share Class ID*s but also to some *Fund ID*s. Once the manual corrections are complete, we re-run automated share class clustering (Section 3.3.1) so that manual *Fund ID* corrections are properly imposed, and the manual *Share Class ID* corrections are used as overriding constraints in the clustering algorithm.

During this manual correction step, we also detect *Share Class ID*s that represent the master share class of their respective *Fund ID*s, i.e., a calculated share class whose AUM series is an aggregate of the whole fund. To aid this process, we automatically detect potential master share classes using the following algorithm. First, we calculate the monthly USD-converted AUM of each *Share Class ID* as the median AUM of its database-level funds, disregarding stale AUMs. We then denote the sum of these *Share Class ID*-level AUMs within a *Fund ID* as F. If a *Share Class ID* is a master share class, its AUM S satisfies $F = 2S$. To allow for some imprecision, we mark a *Share Class ID* as a potential master class if the median of $|F - 2S| / S$ is below 0.2.

We also manually detect database-level funds with erroneously reported currencies. The most common case is where the base currency AUM, reportedly denominated in a non-USD currency, is actually denominated in USD. To aid this process, we again use an automatic heuristic to detect potential errors which are then manually verified. For each non-USD-denominated database-level fund, we calculate a distance metric against each USD-denominated database-level fund within the same *Fund ID*. This distance metric is the median of |NonUSD – USD| / [(NonUSD + USD) / 2], where NonUSD and USD are the monthly AUM observations of the non-USD-denominated and USD-denominated funds, disregarding stale AUMs. We mark the non-USD-denominated fund as potentially erroneous if this distance is below 0.05.

We pay extra attention to *Fund ID*s with more than one *Share Class ID*. For completeness, at this stage we also calculate a different version of *Share Class ID* calculated not within *Fund ID*, but within *Firm ID*, and pay extra attention to cases where the same *Share Class ID* appears in multiple *Fund ID*s of the same *Firm ID*; that is, cases where two *Fund ID*s contain similar-looking AUM series. Such cases often require corrections in both *Fund ID*s and *Share Class ID*s.

4. Fund-Level Aggregation

We can use our harmonized identifiers (Section 3) to combine database information across databases at fund level, resulting in the most complete per-fund coverage. For example, a fund may have reported its early returns to Lipper TASS, but later switched to HFR. Our method allows for the combining of such disjointed information.

4.1 Static Characteristics

To aggregate investment style, fund domicile, and high-water mark dummy, we use the most frequent non-missing value within the *Fund ID*. As a special case, a fund is marked as a fund of fund if *any* database identifies it as such; these funds of funds are then removed in our analyses.

To aggregate management and incentive fees; lockup, notice, and redemption periods; and fund inception date; we use the median within the *Fund ID*.

4.2 Assets Under Management (AUM)

For each fund month, we first aggregate USD-denominated AUM at *Share Class ID* level by using the median AUM across databases. If any of the share classes is marked as a master share class (Section 3.3.2), we use its AUM as the fund-level AUM (if non-missing). Otherwise, we sum over the (non-master) share class AUMs.

As a preliminary step, we remove all eVestment AUMs from November 2011 through June 2012, because this period has numerous hard-to-correct outliers in AUM reporting—even for long-liquidated funds—most likely attributable to eVestment's merger history.

4.3 Returns

To calculate fund-month level equal-weighted (EW) returns, we simply take the median return across databases.

To calculate fund-month level value-weighted (VW) returns, we first calculate the respective *Share Class ID*-level returns as their median over databases. If any of the share classes is marked as a master share class (Section 3.3.2), we use its return as the fund-level VW return (if non-missing). Otherwise, we take an average of the (non-master) share class returns, weighted by the one-month-lagged share class AUM (Section 5. In the special case that one class has lagged AUMs but not returns, and vice-versa for another class, these steps may fail to produce a non-missing VW return; in this case, as a final fallback we set the VW return equal to the EW return.

5. Bias Correction

To optimally correct for survivorship and backfill bias, we want to remove all information prior to the *listing date* of a fund, i.e., the day the fund first started reporting to a database (or databases). We first calculate these listing dates at database level. Four of the databases (TASS, HFR, EurekaHedge, and eVestment) contain this listing date as-is.

For BarclayHedge and Preqin databases, we use the algorithm of Jorion and Schwarz (2017) to impute the listing dates. This algorithm relies on the fact that the database IDs are assigned approximately in the order of listing. The steps of the algorithm are:

1. Sort database funds by database ID and divide the funds into non-overlapping 20-fund chunks.
2. For each chunk, let $N(t) \leq 20$ denote the number of funds reporting returns on month t, and let $M = \max N(t)$ denote its maximum.
3. For each chunk, let $t_1 < t_2 < \ldots < t_K$ denote the $K \geq 1$ months with $N(t_i) = M$, $i = 1,\ldots,K$.
4. For each chunk, the imputed listing date is $t_{\min(3, K)}$.

Jorion and Schwarz (2017) demonstrate that this algorithm works well for the BarclayHedge database. For the Preqin database, which uniquely contains both fund and share class identifiers, we calculate the listing dates at fund identifier level, such that in step 2 a fund is assumed to have reported a return if any of its share classes have reported a return.

We cannot implement the Jorion-Schwarz algorithm for the Morningstar database, because Morningstar Direct provides no running fund identifier; in fact, the only fund identifier is the fund name.

To calculate the listing date at *Fund ID* level (Section 3.2), we take the minimum of its database-level listing dates, giving preference to reported listing dates (TASS, HFR, EurekaHedge, eVestment) over imputed listing dates (BarclayHedge, Preqin).

This *Fund ID*-level listing date denotes the first date that the fund has publicly started advertising its returns. In other words, this is the date that an investor with access to all seven databases (or their precursors) first gained access to the non-backfilled returns. To correct for biases at *Fund ID* level, we remove all time series observations before this listing date. Notice that funds reporting only to Morningstar do not have a *Fund ID*-level listing date and are completely removed in bias correction.

6. Linking Fund of Funds Holdings

6.1 Holdings Data

Following Aiken, Clifford, and Ellis (2013), we construct a dataset of quarterly holdings of funds of (hedge) funds (FOF). First, we use EDGAR to find all companies (identified by their SEC-assigned Central Index Key, or CIK) that have filed each of three types of reports: a registration statement of a closed-end investment management company (N-2 or its amendments); a semi-annual report for registered investment companies (NSAR-A or NSAR-B, or their amendments); and shareholder reports (N-30D, N-CSR, N-CSRS, or N-Q, or their amendments).[2] This yields a list of 1,216 CIKs. We then manually identify the CIKs whose registration statement identifies the company as a fund of hedge funds, and whose semi-annual reports identify the fund as closed-end. This leaves us with 120 closed-end FOFs.

The FOFs' holdings are listed in their shareholder reports, which have appeared online semi-annually since 2000, and quarterly since 2004. We download all these filings and parse the holdings semi-automatically. Each holding contains the fund's name, cost basis, and market value. In addition, most holdings contain the fund's style, and many contain the fund's liquidity period plus additional footnotes. We manually remove non-hedge fund holdings such as occasional stocks, options, ETFs, mutual funds, and cash positions.

[2] As a special case, we also include the CIK 1218126 corresponding to J.P. Morgan Atlas Global Long/Short Equity Fund L.L.C., even though the company lacks a registration statement.

6.2 Linking Procedure

We map each fund name in the FOF holdings into a *Fund ID* as follows. First, for each fund name A in the holdings we automatically find its closest database fund name B.[3] We then manually verify and correct the match. In the special case A doesn't report to databases, as detected by our inability to detect a proper database counterpart either automatically or manually, we assign A (and all its variants appearing in the FOF holdings) a new, distinct *Fund ID*. Otherwise, given the (possibly corrected) closest database counterpart name B', we assign the *Fund ID* of B' to also be the *Fund ID* of A.

6.3 Quarterly Returns

The FOF holdings can be used to calculate quarterly fund returns that are not necessarily voluntarily reported to commercial databases. Specifically, following Aiken, Clifford, and Ellis (2013), we set

$$R_{i,j,t} = \frac{V_{i,j,t} - (C_{i,j,t} - C_{i,j,t,-1})}{V_{i,j,t-1}} - 1,$$

where $R_{i,j,t}$ is the net-of-fees return on fund with *Fund ID* = j realized by FOF with CIK = i from quarter $t-1$ to t; and $V_{i,j,t}$ and $C_{i,j,t}$ are the market value and cost basis of the holding, respectively. We then calculate the *Fund ID*-level return as the median across returns realized by each FOF:

$$R_{j,t} = \underset{i}{\operatorname{median}}\, R_{i,j,t}.$$

We follow the constraint of Aiken, Clifford, and Ellis (2013) of requiring a non-changing cost basis, i.e., $C_{i,j,t} = C_{i,j,t-1}$, because a quickly changing cost basis often results in return outliers. Finally, we remove the top and bottom 1% of *Fund ID*-level returns to ensure a lack of outliers.

[3] As string distance metric, we use the restricted Damerau-Levenshtein distance, and pre-process all names by disregarding their case and removing periods, commas, and initial word "the". Unless an exact match is found, we make the matching computationally feasible and more robust by only considering the two alphabetic neighbors (previous and following).

6.4 Discretionary Liquidity Restrictions

Following Aiken, Clifford, and Ellis (2015), we use the footnotes in the holdings data to construct a fund-quarter-level indicator variable for whether the fund has imposed discretionary liquidity restrictions (DLR). This variable takes a value of one if any FOF reports a position for the underlying hedge fund that is (1) in a side pocket (either completely or partially), (2) subject to investor-level gates, (3) liquidating, (4) organized as a special purpose vehicle or special liquidating vehicle, or (5) explicitly said to be illiquid or having its liquidity restricted.

Bibliography

Aiken AL, Clifford CP & Ellis JA (2013) "Out of the dark: Hedge fund reporting biases and commercial databases." *Review of Financial Studies* 26(1): 208–243.

Aiken AL, Clifford CP & Ellis JA (2015) "Hedge funds and discretionary liquidity restrictions." *Journal of Financial Economics* 116(1): 197–218.

Aragon GO & Nanda V (2017) "Strategic delays and clustering in hedge fund reported returns." *Journal of Financial and Quantitative Analysis* 52(1): 1–35.

Jorion P & Schwarz C (2017) "The fix is in: Properly backing out backfill bias." Working paper.

Table A1: Style mapping

This table shows how we map database-specific strategies into eight standardized Form PF broad styles: Credit, Equity, Event driven, Macro, Managed Futures / CTA, Multi-strategy, Relative value, and Other. Panel A shows the primary mapping rule, where "Database Strategy" corresponds to the fund's main strategy. Preqin sometimes lists multiple main strategies, in which case we apply the primary rule to each listed main strategy, and use the resulting mapping only if it is unique. For BarclayHedge, the primary rule is often missing, especially when main strategy is emerging markets. For these cases, we use a BarclayHedge-specific secondary mapping rule, shown in Panel B. Finally, Panel C shows database-specific mapping exceptions, which override the primary and secondary rules of Panels A and B. Wildcard "*" corresponds to any string following the initial string.

Panel A: Primary mapping rule

Form PF Style: Credit

Database	Database Strategy	Database	Database Strategy
BarclayHedge	Fixed Income - ABS/Sec. Loans	Morningstar	EAA HF Global Bond - USD Biased
BarclayHedge	Fixed Income - Asset-Backed Loans	Morningstar	EAA HF Global Bond - USD Hedged
BarclayHedge	Fixed Income - Collateralized Debt Obligations	Morningstar	EAA HF Global Emerging Markets Bond
BarclayHedge	Fixed Income - Diversified	Morningstar	HF Long/Short Debt
BarclayHedge	Fixed Income - High Yield	Morningstar	HF Long-Only Debt
BarclayHedge	Fixed Income - Long/Short Credit	Preqin	Asset-Backed Lending Strategies
BarclayHedge	Fixed Income - Long-Only Credit	Preqin	Fixed Income
BarclayHedge	Fixed Income - Mortgage Backed	Preqin	Long/Short Credit
EurekaHedge	Diversified Debt	Preqin	Mortgage-Backed Strategies
EurekaHedge	Fixed Income	Preqin	Specialist Credit
Morningstar	EAA HF Global Bond - CHF Hedged	eVestment	Credit Long Short
Morningstar	EAA HF Global Bond - EUR Hedged		

Form PF Style: Equity

Database	Database Strategy	Database	Database Strategy
TASS	Dedicated Short Bias	Morningstar	HF Europe Long/Short Equity
TASS	Equity Market Neutral	Morningstar	HF Global Long/Short Equity
TASS	Long/Short Equity Hedge	Morningstar	HF Long-Only Equity
HFR	Equity Hedge	Morningstar	HF U.S. Long/Short Equity
BarclayHedge	Equity*	Morningstar	HF U.S. Small Cap Long/Short Equity
BarclayHedge	Sector*	Preqin	130/30
EurekaHedge	Bottom-Up	Preqin	Equity Market Neutral
EurekaHedge	Long Short Equities	Preqin	Long Bias
EurekaHedge	Value	Preqin	Long/Short Equity
Morningstar	EAA HF Alt - Market Neutral - Equity	Preqin	Sector-Focused
Morningstar	HF Asia/Pacific Long/Short Equity	Preqin	Short Bias
Morningstar	HF Bear Market Equity	Preqin	Statistical Arbitrage
Morningstar	HF China Long/Short Equity	Preqin	Value-Oriented
Morningstar	HF Emerging Markets Long/Short Equity	Preqin	Variable Bias
Morningstar	HF Emerging Markets Long-Only Equity	eVestment	Equity Long Short
Morningstar	HF Equity Market Neutral		

Form PF Style: Event Driven

Database	Database Strategy	Database	Database Strategy
TASS	Event Driven	Morningstar	HF Distressed Securities
HFR	Event Driven	Morningstar	HF Event Driven
BarclayHedge	Distressed Securities	Morningstar	HF Merger Arbitrage
BarclayHedge	Event Driven	Preqin	Distressed
BarclayHedge	Merger Arbitrage	Preqin	Event Driven
BarclayHedge	PIPEs	Preqin	Risk/Merger Arbitrage
EurekaHedge	Distressed Debt	Preqin	Special Situations
EurekaHedge	Event Driven	eVestment	Event Driven

Form PF Style: Macro

Database	Database Strategy	Database	Database Strategy
TASS	Global Macro	Morningstar	EAA HF Commodities - Energy
HFR	Macro	Morningstar	HF Currency
BarclayHedge	Balanced (Stocks & Bonds)	Morningstar	HF Global Macro
BarclayHedge	Macro	Preqin	Commodities
BarclayHedge	Macro - Discretionary	Preqin	Foreign Exchange
BarclayHedge	Stock Index*	Preqin	Macro
EurekaHedge	Dual Approach	eVestment	Commodities
EurekaHedge	Macro	eVestment	Macro
EurekaHedge	Top-Down		

Form PF Style: Managed Futures / CTA

Database	Database Strategy	Database	Database Strategy
TASS	CTA	BarclayHedge	Technical*
TASS	Managed Futures	EurekaHedge	CTA/Managed Futures
BarclayHedge	Fundamental*	Morningstar	HF Systematic Futures
BarclayHedge	Macro - Quantitative	Preqin	Managed Futures/CTA
BarclayHedge	Systematic		

Form PF Style: Multi-Strategy

Database	Database Strategy	Database	Database Strategy
TASS	Multi-Strategy	Morningstar	HF Multi-Strategy
BarclayHedge	Multi-Advisor	Preqin	Diversified
BarclayHedge	Multi-Strategy*	Preqin	Multi-Strategy
EurekaHedge	Multi-Strategy	eVestment	Multi-Strategy

Form PF Style: Relative Value

Database	Database Strategy	Database	Database Strategy
TASS	Convertible Arbitrage	EurekaHedge	Relative Value
TASS	Fixed Income Arbitrage	Morningstar	HF Convertible Arbitrage
HFR	Relative Value	Morningstar	HF Debt Arbitrage
BarclayHedge	Convertible Arbitrage*	Morningstar	HF Diversified Arbitrage
BarclayHedge	Fixed Income - Arbitrage	Morningstar	HF Volatility
BarclayHedge	Fixed Income - Arbitrage - Capital Structure Arb	Preqin	Capital Structure Arbitrage
BarclayHedge	Fixed Income - Convertible Bonds	Preqin	Convertible Arbitrage
BarclayHedge	Volatility Trading	Preqin	Fixed Income Arbitrage
EurekaHedge	Arbitrage	Preqin	Relative Value Arbitrage

Form PF Style: Other

Database	Database Strategy	Database	Database Strategy
TASS	Options Strategy	EurekaHedge	Others
TASS	Other	Morningstar	HF Long-Only Other
BarclayHedge	Algorithmic	Preqin	Insurance-Linked Strategies
BarclayHedge	Closed-end funds	Preqin	Niche
BarclayHedge	Discretionary	Preqin	Real Estate
BarclayHedge	Dividend Capture	eVestment	Absolute Return
BarclayHedge	Fixed Income - Insurance-Linked Securities	eVestment	Insurance
BarclayHedge	Mutual Funds/ETFs	eVestment	Niche
BarclayHedge	Option Strategies	eVestment	Real Estate
BarclayHedge	Tail Risk	eVestment	Volatility

Panel B: BarclayHedge secondary mapping rule

Secondary Strategy	Form PF Style
Fixed Income - ABS/Sec. Loans	Credit
Fixed Income - Asset-Backed Loans	Credit
Fixed Income - Diversified	Credit
Fixed Income - High Yield	Credit
Fixed Income - Long/Short Credit	Credit
Fixed Income - Long-Only Credit	Credit
Equity*	Equity
Sector*	Equity
Distressed Securities	Event Driven
Event Driven	Event Driven
Merger Arbitrage	Event Driven
Macro	Macro
Fundamental*	Managed Futures / CTA
Systematic	Managed Futures / CTA
Technical*	Managed Futures / CTA
Multi-Strategy	Multi-Strategy
Activist	Other
Balanced (Stocks & Bonds)	Other
Closed-end funds	Other
Discretionary	Other
Dividend Capture	Other
Mutual Funds/ETFs	Other
Convertible Arbitrage	Relative Value
Fixed Income - Arbitrage	Relative Value
Volatility Trading	Relative Value

Panel C: Mapping exceptions

Database	Main Strategy	Secondary Strategy	Form PF Style
HFR	Event Driven	Activist	Equity
BarclayHedge	Emerging Markets - Asia	Fixed Income - Convertible Bonds	Credit
BarclayHedge	Statistical Arbitrage	*Missing value*	Macro
BarclayHedge	Activist	*Missing value*	Other
BarclayHedge	Arbitrage	*Missing value*	Other
BarclayHedge	*Missing value*	Fixed Income - Convertible Bonds	Relative Value
EurekaHedge	Arbitrage	Merger Arbitrage*	Event Driven
EurekaHedge	Arbitrage	Mergers & Acquisitions*	Event Driven
EurekaHedge	Fixed Income	Capital Structure Arbitrage*	Relative Value
EurekaHedge	Fixed Income	Convertible Bond Arbitrage*	Relative Value
EurekaHedge	Fixed Income	Fixed Income Arbitrage*	Relative Value
EurekaHedge	Fixed Income	Relative Value Multi-Strategy*	Relative Value
EurekaHedge	Fixed Income	Volatility Arbitrage*	Relative Value
eVestment	Niche	Direct Lending	Credit
eVestment	Credit Long Short	Distressed Credit	Event Driven
eVestment	Macro	CTA/Managed Futures	Managed Futures / CTA
eVestment	Event Driven	Activist	Other
eVestment	Event Driven	Private Placements/Regulation D	Other
eVestment	Credit Long Short	Convertible Arbitrage	Relative Value
eVestment	Macro	Fixed Income Relative Value	Relative Value
eVestment	Volatility	Volatility Arbitrage	Relative Value

Table A2: Domicile mapping

This table shows how we standardize each fund domicile country into one of eight broad domiciles: Asia, Caribbean, Europe, Pacific, North America, Central America, South America, and Others.

Asia	Caribbean	Europe		
Cambodia	Anguilla	Andorra	Iceland	Spain
China	Bahamas	Austria	Ireland	Sweden
Hong Kong	Barbados	Belgium	Isle of Man	Switzerland
India	Bermuda	British Islands	Italy	Turkey
Indonesia	British Virgin Islands	Bulgaria	Jersey	United Kingdom
Japan	British West Indies	Canary Islands	Latvia	Russia
Korea	Cayman Islands	Channel Islands	Liechtenstein	
Labuan	Curacao	Cyprus	Lithuania	
Malaysia	Grenada	Czech Republic	Luxembourg	
Pakistan	Netherlands Antilles	Denmark	Macedonia	
Philippines	Nevis Island	Estonia	Malta	
Singapore	Puerto Rico	Europe	Monaco	
South Korea	St. Kitts and Nevis	Finland	Netherlands	
Taiwan	St. Lucia	France	Nigeria	
	St. Martin	Germany	Norway	
	St. Vincent and the Grenadines	Gibraltar	Poland	
	Turks and Caicos Islands	Guernsey	Portugal	
	United States Virgin Islands	Greece	Romania	
		Hungary	Slovakia	

Pacific
Australia
Christmas Island
Cook Islands
Marshall Islands
New Zealand
Samoa
Vanuatu

North America
Canada
United States

Central America
Belize
Mexico
Panama

South America
Argentina
Brazil
Chile
Colombia
Peru

Others
Benin
Bahrain
Botswana
Bouvet Island
British Indian Ocean Territory
Dubai
Israel
Lebanon
Kuwait
Mauritius
Qatar
Saudi Arabia
Seychelles
South Africa
Swaziland
United Arab Emirates