

**The Dow Jones
Hedge Fund Manager Universe
Creation and Maintenance**

Scott Lang

August 2005

The Dow Jones Hedge Fund Manager Universe

Dow Jones Hedge Fund Indexes (“DJHFI”) has built and maintains a database containing the universe of hedge fund managers. This universe serves two main functions:

1. It is the basis for conducting quantitative analysis on potential candidates for inclusion in the existing and new Dow Jones Hedge Fund Strategy Benchmarks;
2. It is the source of hedge fund manager data for the purposes of conducting empirical research on hedge funds and the industry.

This document describes the process used in the creation and maintenance of the hedge fund manager universe.

I. Raw Universe

DJHFI currently subscribes to six publicly available databases to represent the hedge fund manager universe. These databases are merged into a single master database to produce the Raw Hedge Fund Manager Universe.

All information and performance contained in the databases is self-reported by the fund managers. By reconciling the information across the six databases ensures that the impact of reporting errors is minimized. Also, the use of all six databases guarantees the broadest coverage possible using publicly available sources.

Many hedge fund managers report their returns to several of the six databases. As a result, duplicate funds must be removed from the merged universe to create a Unique Universe. In order to maintain a consistent selection process, the databases are ranked in a hierarchy to determine the order of preference in the case the same fund appeared in multiple databases. The rankings were based on two factors:

1. Database Size

Larger databases include a larger data set and are more representative of the hedge fund universe than smaller databases.

2. Number of Strategies Represented

Most of the research conducted by DJHFI depends on fund strategy. Therefore, data providers tracking a larger diversity of hedge fund strategies are placed at the top of the hierarchy.

Over time, the hierarchy may be updated based on additional factors not known at the time of the initial ranking. To prepare for a potential shift in rankings, a reference table of duplicate funds and their source is maintained. If a database rises in the hierarchy, the fund originally flagged as a duplicate from that database can be easily switched and saved to the Unique Universe. Conversely, databases that move down in the hierarchy will have its funds removed from the Unique Universe and replaced with a duplicate fund, if one exists, from the next highest ranked database.

II. Cleaning the Data

Four of the six databases are downloaded using the PerTrac Data Manager Application. The others are downloaded from their respective web sites.

Although each database has the same structure, data is organized differently across providers. As a result, each data provider includes a file listing the database field mappings.

After merging each database into one central location, the data cleaning process begins. Because of the purpose served by our Unique Universe, we excluded funds that were indexes, dead or no longer reporting to the provider, and cases in which the fund had not reported returns within 6 months prior to the database download. Note: funds excluded on the basis of this last criteria would be included in the unique database if, for instance, in the next update they were to update their reporting.

After applying the three criteria listed above, the Raw Universe had been reduced from 21,065 funds to 17,650. This represents 83.79% of the Raw Universe. The number of remaining funds from each database is presented in Table 1.

Table 1
Funds Remaining after Indexes, Dead Funds, and Late-reporting Funds are Removed Ending 3Q2004

Provider	Number of Funds	Number of Candidate Funds	%
A	4,909	2,880	58.67%
B	3,893	3,486	89.55%
C	2,941	2,618	89.02%
D	3,607	3,360	93.15%
E	2,348	2,120	90.29%
F	3,367	3,186	94.62%
Totals	21,065	17,650	83.79%

Source: DJHFI

III. Removing Duplicate Funds

Locating duplicates is based mostly on a fund's name, name of the company running the fund, and the company's address. Unfortunately, this information is not standardized across databases, and as a consequence, finding duplicates becomes a complicated and tedious exercise. Though we have tried to keep the process as objective and consistent as possible, identifying duplicates is also a measure of subjective judgment as highlighted by these examples:

1. *Abbreviations*

CompanyName Equity Statistical Arbitrage LP
CompanyName ESA

In one database, the entire fund name is spelled out completely, while another database abbreviates part of the name. We tag these funds as duplicates after verifying the history of returns.

2. *Same fund name, different companies*

The XYZ Fund managed by company *ABC*
 XYZ Fund managed by company *DEF*

At first glance these funds appear to be duplicates. However, upon further investigation, two different companies operate the funds, and their return series are poorly correlated, indicating the two funds are unique.

3. *Company name missing in the fund name*

CompanyName TRADING (DFMSP)
 Diversified Futures Market Sector Portfolio

Most of the funds listed in the Raw Universe contain the company’s name in the title of the fund. However, this is not a safe assumption since the company’s name is excluded in the second one. As a result, duplicates like the example above can be missed. Notice also the first fund abbreviates the entire name of the fund, further complicating the issue.

By conducting multiple searches for fund name, company name and address greatly improves the chances of resolving these issues. Once a set of potential duplicates were found, the funds were then compared by testing for correlation and calculating average monthly return. If the correlation between funds was greater than 0.99 and the difference between the average returns during the same time periods was less than 0.01, the funds were considered duplicates. As noted earlier, these cut-offs are subjective.

Table 2 displays the number of funds in the Unique Universe after removing all duplicate funds.

Table 2
Total Number of Unique Funds by Provider
Ending 3Q2004

Provider	Number of Funds	% of Unique Universe
A	2,337	33.78%
B	1,785	25.80%
C	740	10.70%
D	964	13.93%
E	434	6.27%
F	658	9.51%
Total in Unique Universe	6,918	100.00%

IV. Funds Exclusive to Each Database

The creation of the Unique Universe is best understood by viewing the number of useful funds in each database by those that are exclusive to one database and those that exist in more than one database. Table 3 presents the candidate funds in each database as presented in Table 1 and breaks them down by funds that are exclusive and those that exist in more than one database. Of the 17,650 useful funds, 2,484 (18.09%) are exclusive to one database.

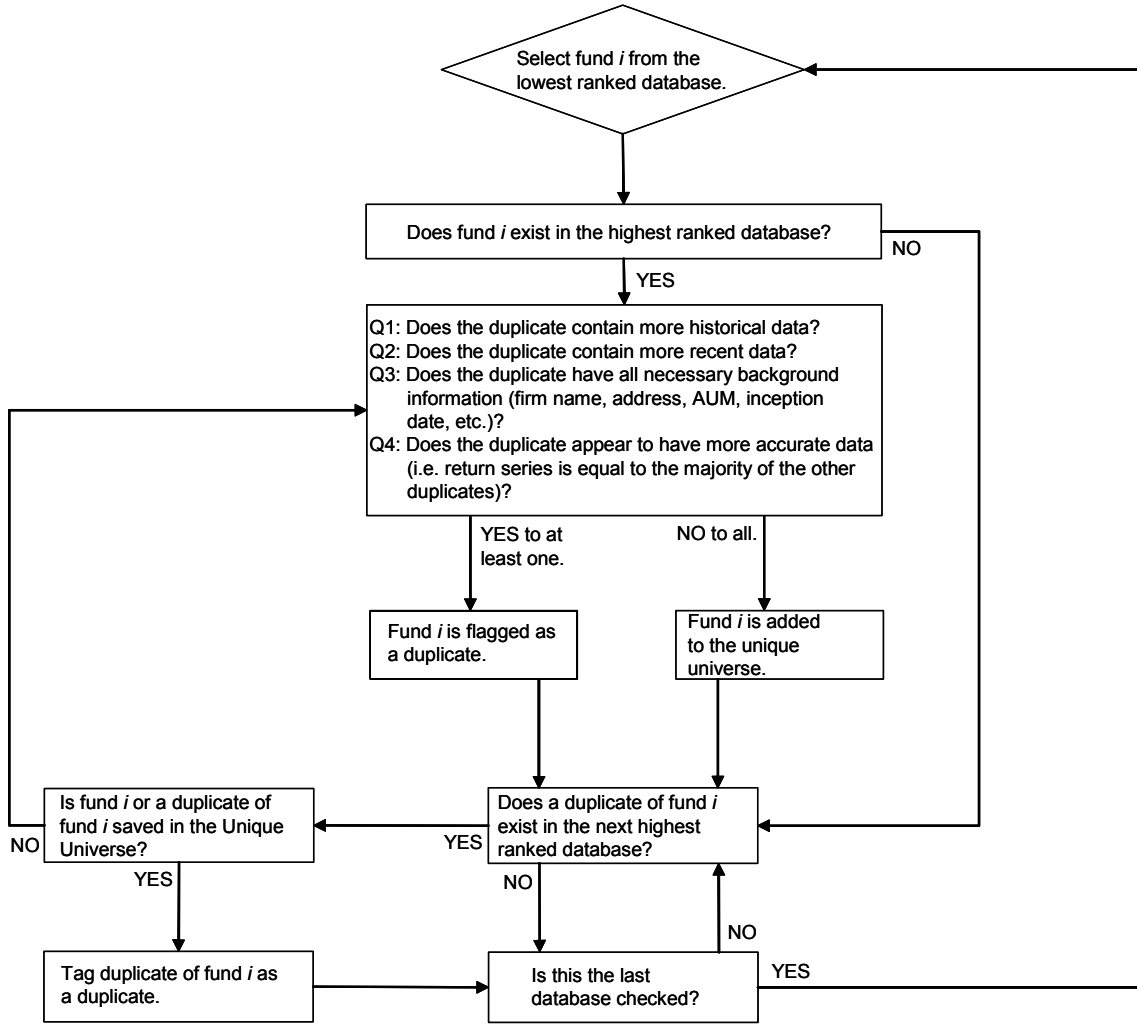
Table 3
Funds Exclusive to a Database
Ending 3Q2004

Provider	Exclusive to the database	Funds in other database(s)	Total
A	321	2,559	2,880
B	475	3,011	3,486
C	285	2,333	2,618
D	515	2,845	3,360
E	375	1,171	2,120
F	513	1,811	3,186
Total	2,484	13,730	17,650

Source: DJHFI

By definition, funds tracked exclusively by one data provider are automatically added to the Unique Universe. This means that of the 6,918 unique funds presented in Table 2, 2,484 (35.91%) are exclusive to only one database. The remaining 4,434 unique funds appear in more than one database. Diagram 1 illustrates the process used to identify these funds. It begins with a non-exclusive fund from the lowest ranked database.

Diagram 1: Flagging Duplicate Funds



Source: DJHFI

The above process repeats for duplicates in the next highest ranked database after all duplicate funds in the lowest ranked database are either added to the Unique Universe or flagged as a duplicate. After exhausting all of the databases, we are left with the Unique Universe as presented in Table 2.

Important Flags

All excluded funds from the Unique Universe contain a flag indicating the reason for its exclusion. Table 4 lists these flags and a description for each:

*Table 4
Fund Flags
Ending 3Q2004*

Flag	Description	# of Records	% of Raw Universe
xIndex	The record was an index.	576	2.73%
xDate	The fund had not reported within 6 months of the database download.	2,206	10.47%
xDuplicate	The fund was a duplicate.	11,544	54.80%
xDead	The fund had liquidated all assets and is no longer operating.	442	2.10%
xOther	The fund contained only quarterly returns.	14	0.07%
	The fund had no data.	23	0.11%
	Incomplete historical data.	7	0.03%
	The fund was a demo fund.	3	0.01%
	Total	14,815	70.52%

NOTE: Maximum number of overlapping funds is 707. The numbers presented are counts after the initial universe was created. As updates are made, these numbers will change. Source: DJHFI

These five flags are useful for a case in which the fund saved in the Unique Universe is no longer preferred and needs to be replaced by one of its duplicates. In these instances, the fund previously saved is removed from the Unique Universe and is replaced by a duplicate from the next highest ranked database. However, to ensure that the newly saved fund, for instance, does not contain missing data, funds are flagged for all criteria that apply. For example, a fund lacking data for the past six months may also be flagged as a duplicate. Funds receiving multiple flags are included in the counts of each flag shown in Table 4. Thus, there is some overlap.

Lastly, a sixth flag indicates whether a fund is a constituent in one of the six Dow Jones Strategy Benchmarks. Currently, 39 of the 40 funds report to at least one of our database subscriptions.