

CRSP SURVIVOR-BIAS-FREE US MUTUAL FUND DATABASE

JULY 2012 QUARTERLY UPDATE

This July shipment of the Mutual Funds database includes several noteworthy changes. For more information, please see the *CRSP Survivor-Bias-Free US Mutual Fund Guide for SAS and Ascii* at www.crsp.chicagobooth.edu/documentation/pdfs/MFDB_Guide.pdf, or the *CRSP Survivor-Bias-Free US Mutual Fund Guide for CRSPSift* at www.crsp.chicagobooth.edu/documentation/pdfs/MF_Sift_Guide.pdf.

MUTUAL FUND INSTALLATION CHANGES

With the addition of the CRSPAccess version of the Mutual Fund database, the installation has changed to include an additional layer of folders that identify the format of the data.

For example, if all three versions of the mutual funds product were to be installed into the default of `c:\crspdata`, an MFZ201206 folder will be created with a sub folder for each format installed.

If all three formats were to be installed, the following directory structure would be created:

```
C:\crspdata
  MFZ201206
    ASCII
    SAS
    Database
```

If you have any automated processes, be aware of this change and edit accordingly.

NEW CLASS CODES

Long requested, a new fund class identifier, CRSP_CL_GRP, is included in the Fund Header and Fund Header History tables in the SAS and ASCII versions of the data, and in the same-named categories in CRSPSift.

The Class Group ID is used to associate different classes of a fund, allowing them to be treated as a single entity. The ID is represented by a 7-digit number beginning with 2000000. Coverage currently begins on August 31, 1998 and future work will be done to extend the group ID deeper back in time.

The ID is assigned through a name parsing and clean-up exercise and follows these rules:

The name is parsed and separated into the class identifier (i.e. "Class A") and the base name.

- If a group of funds have the same resulting base after parsing and cleanup, for any overlapping time period, then they are considered to be sub-classes of the same fund during the period of overlap. If there is no overlap of date ranges, the funds are not considered part of the same group.
- If two groups share one or more funds, then they are assigned the same CRSP Class Group ID. That is, the Class Group ID will generally remain the same even if the fund name changes. However:
- If a group overlaps in time with another group with the same Class Group ID but a different base name, then one of the groups is assigned a different ID. This accounts for cases where different funds with multiple subclasses are combined. Each fund will share the Class Group when they share a base name, but a different group when they do not.

Because the ID is derived principally from the fund name, anomalies in the name information can potentially cause funds to not be grouped together properly. Almost 900 fund names were corrected during the Class Group ID generation process. Additional corrections will be made if necessary, so do not hesitate to contact CRSP with your feedback.

NEW STYLE CODES

CRSP has created a continuous history of style codes through creating a new code that links codes from different sources through time.

The CRSP US Survivor-Bias-Free Mutual Funds database includes style and objective codes from three different sources over the life of the database. No single source exists for its full time range.

- Wiesenberger Objective codes are populated 1962-1993.
- Strategic Insight Objective codes are populated 1993-1998.
- Lipper Objective codes begin 1998.

CRSP has provided continuity within the database by building our own style code that is based on and is consistent with those codes provided by our different sources.

The CRSP Style Code consists of up to four characters, with each position defined. Reading left to right, the four codes represent an increasing level of granularity. For example, a code for a particular mutual fund is EDYG, where:

E = Equity, D = Domestic, Y = Style, G = Growth.

Codes with less than four characters exist, and it simply means that they are defined to a less granular level.

For a detailed description of the style codes, please see the *CRSP Survivor-Bias-Free US Mutual Fund Guide, Appendix A*, or online at http://www.crsp.chicagobooth.edu/documentation/pdfs/MFDB_Guide.pdf.

CRSPSIFT ACCESS

The mutual fund data files are reformatted for use with CRSPSift, CRSP's Windows interface, simplifying access to the data and allowing data files to be generated in a variety of formats. This new format is provided in addition to, not in place of, the current Mutual Funds SAS and ASCII formats.

FILE VERSION SPECIFICS

The database has 50,611 total funds. Of these 28,989 are active and 21,622 are delisted.

The mutual fund database is available in both SAS and ASCII formats. Due to file sizes, the SAS and ASCII formats of the database are shipped on DVD only.

NUMBER OF ROWS SUMMARY

The following table identifies the number of rows for each table in the database.

TABLE NAME	ROWS
contact_info	176,013
crsp_portno_map	60,522
daily_nav	75,321,096
daily_returns	75,803,508
dividends	2,271,346
front_load_det	93,657
front_load_grp	31,477
fund_fees	323,472
fund_hdr	50,611
fund_hdr_hist	359,499
fund_style	200,928
fund_summary	1,260,875
holdings	52,623,297
holdings_co_info	553,383
monthly_nav	4,782,231
monthly_returns	4,784,162
monthly_tna	4,503,790
rear_load_det	143,250
rear_load_grp	46,224

SYSTEM REQUIREMENTS

- SAS version 8, SAS version 9.1, or a relational database of your choice.
- Sufficient disk space to load and manipulate the database:

DATABASE FORMAT	DISK SPACE REQUIRED
SAS V 9.1	6.92 GB
ASCII Data	7.52 GB
CRSPAccess	4.53 GB

INSTALLATION

The data are installed with a new InstallAnywhere wizard.

To install the data insert the CD and click on the appropriate setup icon.

GENERAL NOTES:

- The table with the largest number of rows is the Daily Returns table, containing 75,803,508 rows.
- A blank field means either that data was not applicable to a fund or that data is not available.
- A field with a “?” in it is a flag to indicate further research is required.
- A field has a zero in it when 1) The fund did not exist; 2) No value could be found; or 3) A calculated value could not be calculated because information was missing.