

CRSP SURVIVOR-BIAS-FREE US MUTUAL FUND DATABASE

APRIL 2012 QUARTERLY UPDATE

These Release Notes accompany the April 2012 quarterly release of the CRSP Survivor-Bias-Free US Mutual Fund Database, and contain data through March 2012.

UPCOMING CHANGES!

The July shipment of the Mutual Funds database is targeted for several noteworthy changes.

- New Class Codes – Long requested, a new fund class identifier, CRSP_CL_GRP is being added to Fund Header and Fund Header History tables.
- 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.
- Access to Mutual Fund data through CRSPSift – The mutual fund database is reformatted for use with 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 or ASCII Data File products.

MORE ABOUT THE CRSP STYLE CODE

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

CRSP has decided to build 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.

CRSP invites you to review the codes at www.crsp.chicagobooth.edu/documentation/MF_stylecode.html and share your thoughts with us. Comments, recommendations, and questions can be sent from the web page. Our timeline for implementation of a new style code schema is short, with a targeted initial release in next quarterly shipment in July 2012.

DATA IMPROVEMENTS

RETURNS CLEANUP

Several dozen very large monthly return values were reviewed and corrections were made to seven funds.

NAV CORRECTIONS

About 100 funds that had abnormal NAVs early in their history (daily NAVs began a month or more before the monthly NAVs) were corrected.

DISTRIBUTIONS

Roughly 6,000 very small dividends had amounts that had been rounded down to zero. These were filled in with the appropriate amounts.

FILE VERSION SPECIFICS

The database has 49,975 total funds. Of these 28,701 are active and 21,274 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	174,077
crsp_portno_map	59,897
daily_nav	73,519,542
daily_returns	73,990,172
dividends	2,242,716
front_load_det	92,961
front_load_grp	31,360
fund_fees	318,419
fund_hdr	49,975
fund_hdr_hist	353,878
fund_style	142,941
fund_summary	1,231,932
holdings	47,987,363
holdings_co_info	530,089
monthly_nav	4,695,844
monthly_returns	4,697,743
monthly_tna	4,417,492
rear_load_det	142,377
rear_load_grp	45,909

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	6.59 GB
ASCII Data	7.04 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 73,990,172 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.