



Can the variable "crsp_company_key" in your mutual fund holdings data be reused over time? Since quite a few portfolio compositions do not have "PERMNO" but have "crsp_company_key," can I use "crsp_company_key" in addition to "permno" to identify a security across portfolios and across time?

PERMNO is a security identifier. Certain mutual funds trade instruments that are not securities, so those funds would not have PERMNOs in their portfolio composition. Crsp_company_keys are assigned to all holdings in the Mutual Fund Database. Crsp_company_keys should match up 1:1 with portfolio holdings and not be reused. There are some securities in the Mutual Fund database that have erroneously been assigned multiple crsp_company_keys. CRSP is working to correct these errors.

It appears that for most FUNDNOs, there is only one PORTNO through time; but for some FUNDNOs the PORTNO changes at some point in time. Why is that?

PORTNO is a unique identifier for holdings in a fund's portfolio. The PORTNO could change if the fund changed its parent ID (primary fund), this can happen if the primary fund is deactivated (like in cases of liquidation or merger), or if the management company decided to change the primary fund.

What would cause a change within a PORTNO?

Securities can and will drop out of PORTNOs if the fund sells all of the shares for that security.

I've noticed a case where several FUNDNOs share the same PORTNO. Why?

A FUNDNO is given to each fund or fund class so that class-specific information can be captured in the database. In most cases the different classes of a fund note differences in either the target market (institutional or consumer) or in the structure of the fees being levied. For example, one class may be classified as a no-load fund, usually with higher ongoing expense fees, and another class may be identified as a front-loaded product, which usually implies more money up front but smaller ongoing expenses. Because each class of the fund will all hold the same portfolio of securities, a single CRSP_PORTNO is created that describes the holdings that apply to each of the different classes.

The CRSP_PORTNO_MAP file provides a map between FUNDNOs and PORTNOs, but so do the FUND_HDR and the FUND_HDR_HIST files. What is the difference between the two mappings?

CRSP recommends primary usage with the portno_map. The portno_map holds the same information that can be found in the header and header hist files, and was created specifically for ease of use.

We are looking for a way to group the holdings in groups of similar types (e.g., US large caps together, International government fixed income with low ratings together, etc); How can we do this in the Mutual Fund Database?

Use the CRSP Style Code, described in our mutual fund guide, to identify the funds that fit your criteria.

US equity large cap, for example, would be E(equity), D(Domestic), L(Large Cap) and I(Fixed Income), G(government), Foreign(F), etc.

Once funds are identified, related PORTNOs can be found in the portno_map file. Using PORTNO, the associate holdings can be extracted from the Holdings_info file.

How can I access CRSP Mutual Fund data?

The CRSP Survivor-Bias Free US Mutual Fund Database is provided in either SAS Data Sets or ASCII files. All SAS data sets may be accessed with SAS software version 8 or above. Almost all of the ASCII files are too large to be opened and parsed in Excel. Rather than using Excel or Microsoft Access, CRSP recommends that our subscribers load the ASCII data into a commercial-grade relational database or statistical package. CRSP includes generic create and load procedures with the ASCII data. These procedures, found in the file, mfdb_create_load_procedure.txt, may be used as a starting point for loading the data into the relational database of your choice. When accessing the Mutual Funds data files, note that the Fund Header table is the central table for the database. It contains the most recent information for active and delisted funds. Researchers may find it useful to link information in this table to the other tables where information is categorized by data type, such as holdings, returns, and NAVs.

What are CRSP_FUNDNO and CRSP_PORTNO and what is their relationship?

CRSP_FUNDNO is a unique identifier created by CRSP for each mutual fund followed in the CRSP US Survivor-Bias-Free Mutual Fund Database. A CRSP_FUNDNO is given to each fund or fund class combination so that class specific information can be captured in the database. In most cases the different classes of a fund note differences in either the target market (institutional or consumer) or in the structure of the fees being levied. For example, one class may be classified as a no-load fund, usually with higher ongoing expense fees, and another class may be identified as a front-loaded product, which usually implies more money up front but smaller ongoing expenses. CRSP_PORTNO was created by CRSP to minimize the size of the holdings files. In the situation where a fund may have multiple classes, a separate CRSP_FUNDNO is assigned to each class. Because each class of the fund will all hold the same portfolio of securities, a single CRSP_PORTNO is created that describes the holdings that apply to each of the different classes. The current CRSP_PORTNO may be found in the Fund Header Table. The value for this variable would be the same for each CRSP_FUNDNO representing the different classes of a fund. This

eliminates the need to report the holdings for each class individually and keeps the size of the Holdings Table more manageable than it otherwise would be.

crsp_fundno	crsp_portno	Fund Name
004514	1000022	Allianz Funds: OCC Growth Fund; Class A Shares
004515	1000022	Allianz Funds: OCC Growth Fund; Class B Shares
004516	1000022	Allianz Funds: OCC Growth Fund; Class C Shares
004517	1000022	Allianz Funds: OCC Growth Fund; Class D Shares
004518	1000022	Allianz Funds: OCC Growth Fund; Class R Shares
004519	1000289	Allianz Funds: OCC Equity Premium Strategy Fund; Institutional Class Shares
004520	1000289	Allianz Funds: OCC Equity Premium Strategy Fund; Administrative Class Shares

Can one PERMNO have multiple CRSP_COMPANY_KEYS?

Yes. For equities that are included in CRSP Stock Database, a new CRSP_COMPANY_KEY is assigned to the security whenever the name, ticker, or cusip changes. The PERMNO and PERMCO in the records will remain the same. For example, Red Hat, Inc. changed the ticker from RHAT to RHT. Therefore, a new CRSP_COMPANY_KEY (11555) was assigned. In this case, PERMNO 87184 are associated with two CRSP_COMPANY_KEYS, 11554 and 11555. Below is the name event history for PERMNO 87184 in CRSP Stock Database.

Namedt	Enddt	NCUSIP	Ticker	Company Name	SH	EX	SIC
19990811	20000731	75657710	RHAT	RED HAT INC	11	3	7370
20000801	20010823	75657710	RHAT	RED HAT INC	11	3	7372
20010824	20060713	75657710	RHAT	RED HAT INC	11	3	7372
20060714	20061211	75657710	RHAT	RED HAT INC	11	3	7372
20061212	20081128	75657710	RHT	RED HAT INC	11	1	7372

Below are the CRSP_COMPANY_KEYS associated with PERMNO 87184. RHAT

CRSP_COMPANY_KEY	SECURITY_NAME	CUSIP	PERMNO	
11554	RED HAT INC	75657710	87184	16753
11555	RED HAT INC	75657710	87184	16753 RHT

It appears that exp_ratio can be less than the mgmt_fee, even though exp_ratio includes the mgmt_fee. Why?

The expense ratio includes the management expense. What you do not see is any waivers that have been applied to bring the expense ratio down to a certain point. Currently CRSP does not track waivers.

The fiscal_yearend date always precedes begdt. If the fees are valid for the fiscal year ending on fiscal_yearend, what is the purpose of begdt and enddt?

The begin and end dates signify the dates for which the information in that row are valid. It is similar to the header table in the stock database. Any time one of the fields changes, a new record is created with the new information and a new set of begin and end dates. For instance, if a fund reports the same fund fee values with each data item the same for three months, then changes one of them, there would be one record with a three-month range between the begin and end date, and another record with a one-month range.

What is the difference between the Report Date and the Effective Date?

The Report Date (report_dt) is the date of the holdings as reported by CRSP's sources. The Effective Date (eff_dt) is the date CRSP received the information from our vendor. As a result, all effective dates are equal to or more current than report dates in the database.

Why are there negative values in the asset allocation fields (such as per_com, per_pref, per_conv, per_corp, per_muni, per_govt, per_oth, per_cash, and per_bond) in the fund_summary dataset?

The negative percentage values are due to derivatives and securities held in short position.

Why are there negative values in the mgmt_fee in the fund_fees dataset?

Reimbursements can lead to negative Management Fees. The fee is calculated using ratios based on the line items reported in the Statement of Operations. The management fee can be offset by fee waivers and/or reimbursements which will make this value differ from the contractual fees found in the prospectus.

What is the unit of measurement for mgmt_fee?

The management fees are represented in percent format so that 0.5 represents .5%. Management fees are always a function of assets under management and not NAVs. For instance, if you have 100K invested in a mutual fund, over the course of a year you would pay $100,000 * .29\% = \$290$ in management fees. To look at the total fees collected by a fund, multiply the Total Net Assets by the management fee. This is what the fund expects to receive in fees annually, not every month.

Why do some mutual funds have mixed taxable and untaxed dividends?

Tax-exempt funds may have a small portion of a taxable dividend. For tax-exempt funds, generally speaking, most if not all income dividends are exempt from federal income tax. It is possible that a tax-exempt fund could earn taxable interest income, in which case it would be distributed to shareholders as "taxable" income dividends. In these cases, a fund may be coded as having both DT and DU dis_type codes. DT=Taxable Dividends DU=Untaxed Dividends.

What does a turnover value of 0 represent?

The values for turnover represent a percentage for the fund. For example, a value of 0.54 represents 54% turnover, while a value of 1.39 represents 139% turnover. Almost all the funds with a 0 turnover ratio are money market funds (100% cash allocation). A few funds that have a passive management style may also show very low or zero turnover.

How are inst_fund determined in FUND_HDR table?

CRSP does not directly determine the inst_fund flag. We derive it from a more detailed field from Lipper that describes the primary sales channel of the fund. In the case of inst_fund, we assign a "Y" to funds that fall in Lipper's "Institutional" or "Bank Institutional" categories.

Here are the definitions for those categories:

- **Bank Institutional**

Funds primarily offered to client, agencies and fiduciaries of bank trust departments, commercial banks, thrifts, trust companies, or similar institutions. The bank, bank affiliate or subsidiary acts as advisor or, in some cases, sub-advisor for the funds and the funds are typically marketed as a bank product.

- **Institutional**

Funds primarily targeted at organizations and institutions, including pension funds, 401k plans, profit sharing plans, endowments, or accounts held by institutions in a fiduciary, agency or custodial capacity lower overall expense ratio.

How can different share classes of mutual funds or fund families be identified?

CRSP provides two unique identifiers, crsp_fundno and crsp_portno. Neither identifier is unique to "families" of funds, or groups of fund share classes with the same holdings portfolio and management company. The identifier crsp_portno identifies unique holdings portfolios, but can be associated with multiple management companies because of client and feeder funds.

One observed method of joining together share classes by fund family is to parse common syntax used in the fund_name field in fund_hdr. The following example in SAS creates a simple but imperfect index number by fund family. It is based on the convention of using semicolon (;) to distinguish fund class from fund family. The example assumes you have created library reference called mf that points to the location of your mutual fund database.

```
data fund_family (drop=len pos);
format fund_name $140.;
set mf.fund_hdr;
/* Find the length of each fund name and the position of the last ';' in the name. */
len=length(fund_name);
pos=findc(fund_name, ';', -len);
/* Return the text before and after the ';'. */
if pos gt 0 then do;
    ff_name=substr(fund_name, 1, pos-1);
    class=substr(fund_name, pos+1, len-pos);
end;
run;

/* Sort the data set by likely fund family names. */
proc sort data=fund_family;
by ff_name crsp_fundno;
run;

/* Assign an index number based on the fund family names. */
data fund_family;
set fund_family;
format ffno z6.;
by ff_name;
if ff_name='' then ffno+1;
if first.ff_name then ffno + 1;
run;
```

Other Characters and Methods of Identifying Fund Classes

Fund_name observations from older sources appear to use other characters to distinguish fund share classes besides semicolon. Prior to 1998, many funds appear to be delimited by forward slash (/) to distinguish fund family and fund class. Parsing by text is only a beginning step to distinguishing fund classes and families. Care should be taken to ensure that you are correctly matching funds. When identifying fund families you may want to also check if the funds have the same fund management company, are associated with the same holdings portfolio, and have correlated returns.

What happened to PORTNO when CRSP changed its data provider from S&P/Morningstar to Lipper and Refinitiv in 2008?

The holdings table was rebuilt from 2003 to the present using a new methodology for combining our two data sources. Portfolios that are covered completely or in part by our original source have been added with holdings data from that source. Portfolios that exist only in our current source have holdings data from that source. Where possible, the crsp_portno has remained the same as it was in previous iterations of the database, but in some cases the new methodology may have required the crsp_portno to be changed.

When CRSP Mutual Fund changed its data provider in 2010, CRSP_PORTNO changed for the same fund across the board. What do I need to do to link old CRSP_PORTNO to new CRSP_PORTNO?

You can utilize the CRSP_PORTNO_MAP table to link the old CRSP_PORTNOs to the new CRSP_PORTNOs.