Search WRDS

Get Data   Analytics   Classroom   Research   Support

Overview of CRSP U.S. Stock Database

The CRSP Stock database provides stock prices, returns and event information.

General Description

The **CRSP U.S. Stock database** contains end-of-day and month-end prices on primary listings for the **NYSE**, **NYSE MKT**, **NASDAQ**, and **Arca** exchanges, along with basic market indices. CRSP databases are characterized by their comprehensive corporate action information and highly accurate total return calculations.

Top of Section

Coverage

All securities listed in CRSP U.S. Stock databases are **equity** securities, not bonds. CRSP U.S. Stock databases **do not include** securities for international companies (i.e. Canadian companies) - unless they are **ADRs**, **cross-listed**, or **traded** on the major stock exchanges mentioned above.

CRSP U.S. Stock database **contains the following information**:

- Price and quote data (e.g. Open, close, bid/low, ask/high, trade-only).
- Holding period returns with and without dividends.
- Excess returns and other derived data items.
- Market capitalization.
- Shares outstanding.
- Trading volume.
- Security delisting information.
- Corporate actions.
- Identifiers, descriptors, and supplemental data items.

CRSP U.S. Stock **data coverage for the various exchanges** includes:

- NYSE: All data series begin on December 31, 1925.
- NYSE MKT: All data series begin on July 2, 1962.
- NASDAQ: All data series begin on December 14, 1972.
- Arca: All data series begin March 8, 2006.

The **market indices** cover:

- Equal and value-weighted returns for CRSP NYSE, NYSE MKT, NASDAQ, and Arca - with and without dividends.
- Composite Indices for S&P 500 and NASDAQ.

For S&P 500 constituent data, refer to to the CRSP Indices product.

At the share type (SHRCD) level, CRSP U.S. Stock includes:

- Common Stocks.
- Certificates.
- ADRs.
- Shares of Beneficial Interest.
- Units (e.g. Depository Units, Units of Beneficial Interest, Units of Limited Partnership Interest, Depository Receipts).
- ETFs.
- Closed-End Mutual Funds.
- Foreign companies traded on NYSE, NYSE MKT, NASDAQ, and NYSE Arca.
- Americus Trust Components (Primes and Scores).
- HOLDRs Trusts.
- REITs (Real Estate Investment Trusts).

For additional information, refer to the "Share Codes Security Type" section of this overview.

Lastly, CRSP U.S. Stock database **excludes**:

- Rights and warrants.
- Preferred shares.
- Units representing common stocks bundled with rights or warrants.
- Over the Counter Bulletin Board Issues.
- When Issued Trading.

Top of Section

Data Collection Process

CRSP U.S. Stock database is collected from multiple sources. For recent years, it is collected from Interactive Data Corporation (IDC). For more information, please refer to pages 1-2 of Stock & Indices Documentation.

Top of Section

Dataset Organization

File Location Description

The table below lists **all CRSP U.S. Stock products** available on WRDS, as well as their respective SAS and Unix locations. For further information, visit the individual product pages on WRDS.

File Location Table

Table 1: Datafile Locations

Product	Primary SAS Libname	Alternate SAS Libname	Unix Location
CSRP U.S. Stock- Annual Update	crsp	crspa	/wrds/crsp/sasdata/a_stock
CRSP U.S. Stock - Quarterly Update	crspq	crspq	/wrds/crsp/sasdata/q_stock
CRSP U.S. Stock - Monthly Update	crspm	crspm	/wrds/crsp/sasdata/m_stock
CSRP U.S. Stock- 1962 Annual Update	crsp	crspa	/wrds/crsp/sasdata/a_stock62
CRSP U.S. Stock 1962 - Quarterly Update	crspq	crspq	/wrds/crsp/sasdata/q_stock62
CRSP U.S. Stock 1962 - Monthly Update	crspm	crspm	/wrds/crsp/sasdata/m_stock62

CRSP 1962 U.S. Stock datasets are **a subset of the regular CRSP U.S. Stock database** that start in 1962, compared to the full database which starts in 1925.

CRSP annual products are **updated each year** in February, and consist of data for the entire previous year.

Datasets Created by WRDS

WRDS maintains the original event tables that come from CRSP (DSE/MSE) and also create DSEALL, MSEALL, and Stocknames to present the information in different ways. See the note on event tables for more information.

<div align="right">Top of Section</div>

Linking to Other Products

Identifiers Used

The primary identifiers in CRSP are `PERMNO` and `PERMCO`. They are historical identifiers, and **allow users to trace companies and securities over time**. Both `PERMNO` and `PERMCO` are assigned by CRSP. It is helpful to know that a company (PERMCO) can have multiple securities (PERMNO) at the same time.

`PERMNO` is a **unique permanent security identification number** assigned by CRSP to each security. PERMNO is currently **a 5-digit integer** for all common securities in the CRSP files. Unlike `CUSIP`, Ticker Symbol, and Company Name, **a `PERMNO` does not change** during an issue's trading history, **nor is it reassigned** after an issue ceases trading. This **allows users track a security through its entire trading history** in CRSP files with a single `PERMNO`, regardless of name or capital structure changes. It is useful to know Stock data are sorted and indexed by this field.

`PERMCO` is a **unique permanent company identification number** assigned by CRSP to all companies with issues on a CRSP File. This number is permanent for all securities issued by this company regardless of name changes.

Aside from these to identifiers, the **Monthly** Stock files (commonly referred to as "msf") and **Daily** Stock files (i.e. "dsf") **additionally contain header CUSIP identifiers** (`CUSIP`) - but **not** Ticker or historical CUSIP (`NCUSIP`) information. The WRDS web query forms for these datasets merges the information on the fly. Refer to tables `mse/dse`, `mseall/dseall`, `dsenames/msenames` or `stocknames` for retrieval; additional information about these specific tables can be found below.

The following list provides information on **how to link CRSP with other databases**.

Database Linking InformationCOMPUSTAT (with CRSP CCM subscription)
[CRSP/COMPUSTAT Merged (CCM) Overview](#)

Audit Analytics
[AuditAnalytics with CRSP and Compustat](#)

Bank Regulatory
[Bank Regulatory Dataset: Linking with CRSP](#)

Blockholders
[Linking Blockholders to CRSP](#)

TAQ
[WRDS Research Macro: TCLINK](#)

OptionMetrics
[WRDS Research Macro: OCLINK](#)

Bankscope
[WRDS Support Articles: BVD](#)

For other databases not listed here, **always use `CUSIP` when it is available**. If the databases of interest have **historic `CUSIP`**, match it with CRSP's `NCUSIP`. **When the `CUSIP` lengths are not the same**, truncate the longer one to make them compatible.

For example, Compustat uses 9-digit `CUSIP`s while CRSP uses 8-digit `CUSIP`s. The **9th digit** in Compustat is a **check digit** that is not meaningful and may be removed. **6-digit `CUSIP`s** may also be used. As **the first 6 digits of CUSIP represent the issuer**, it is possible to have multiple links. In this case, use tickers or company names as an additional check. Use the following SAS statement to create a 6-digit `CUSIP`:

```
CUSIP6 = substr(cusip, 1, 6);
```

If `CUSIP` **is not available** in the other database, but ticker symbols are provided, it is possible to **establish the link through the ticker**. Take note that **discarded tickers are often reused for newly listed companies**, so it is crucial to **use ticker effective dates** when linking by this method.

If both `CUSIP`s and tickers are not available, another option is to **link by company names**. Total assets, company addresses, and more can provide supporting evidence. Be aware that a company can have its **name spelled in a variety of different ways**. For example, IBM can be found as as "INTERNATIONAL BUSINESS MACHINE" or "INTERNATIONAL BUSINESS MACHS CO". In SAS, make use of the **spelling distance function (`SPEDIS`)** to quantify the difference between two company names.

<div align="right">Top of Section</div>

Database Notes

For additional help regarding the CRSP U.S. Stock product, [refer to the WRDS support page](#).

Comparing CRSP Stock and Index Products

The CRSP U.S. Stock product includes **minimal** information about stock indexes and the market as a whole. **Beta** and **index membership (index constituents)** are sometimes thought of as stock characteristics, but actually only exist within the context of the market. These items are included in the CRSP Index product.

Adjusting for Stock Splits and Other Corporate Actions

Returns are already adjusted for splits, but prices and shares outstanding are not. To adjust prices and shares outstanding, use the following code in SAS:

```
1   Adjusted prices = PRC / CFACPR
2   Adjusted shares = SHROUT * CFACSHR
```

CFACSHR is **not always equal** to CFACPR. This can be caused by less common distribution events, spin-offs, and rights.

Negative Stock Price (PRC)

If the closing price **is not available** for any given period, the number in the price field **is replaced with** a bid/ask average. **Bid/ask averages** have dashes placed in front of them. These do not incorrectly reflect negative prices; they **serve to distinguish** bid/ask averages from actual closing prices. If neither the price nor bid/ask average is available, **the field is set to zero**.

Missing Value Codes in SAS

When looking up a holding period return (RET), or delisting return (DLRET), SAS may return a missing value code. These are indicated with a decimal point followed by a letter (e.g. ".E", ".D"). The descriptions for the missing value codes in SAS can be found on the "Variable Description" page for CRSP Stock.

Share Code Security Type (SHRCD)

Security type code (SHRCD) has **two digits**, with each digit representing a specific piece of information. The **first** defines the security type; the **second** provides more detailed information about the type of security traded. For example, a SHRCD of 10 or 11 would represent U.S. common stocks.

Table 3: SHRCD First Digit

First Digit	Definition
1	Ordinary Common Shares
2	Certificates
3	ADRs (American Depository Receipts)
4	SBIs (Shares of Beneficial Interest)
7	Units (Depository Units, Units of Beneficial Interest, Units of Limited Partnership Interest, Depository Receipts, etc.)

Note: **Code 7 (Units) do not** represent combinations of common stock and anything else (e.g. warrants).

Table 3: SHRCD Second Digit

Second Digit	Definition
0	Securities which have not been further defined.
1	Securities which need not be further defined.
2	Companies incorporated outside the US
3	Americus Trust Components (Primes and Scores).
4	Closed-end funds.
5	Closed-end fund companies incorporated outside the US
8	REIT's (Real Estate Investment Trusts)

Dataset Differences in Web Queries & UNIX/SAS Access

When accessing Monthly Stock file (msf) or Daily Stock file (dsf) on the WRDS server through UNIX or SAS, a user will notice **some information is unavailable**, such as the share code (SHRCD) or delisting price (DLPRC). This information is available when using the web query form, however. This is because **WRDS web queries merge** the Monthly and Daily Stock file (msf/dsf) with the Event file (MSEALL/DSEALL) and Market Indexes Monthly (msi) on the fly. **Share Code and Share Class** are part of the Events table that only records changes, rather than creating an entry for every day or month. Please refer to the CRSP datalist to find out what is provided in each SAS dataset. Additionally, use the variable search tool to find out which dataset contains the item on WRDS.

Users accessing **SAS datasets through UNIX** to perform complex calculations can refer to the "crspmerge" SAS macro for merging CRSP Stock and Events data.

Event Tables

The stock price tables (dsf, msf) are time series, they have one row per month for each company. The **event tables** have one row per event - they only **note changes**.

There are 5 types of events:

- Names History (NAMES).
- Delisting Event Histories (DELIST).
- Distribution Events (DIST).
- NASDAQ Event Information Histories (NASDIN).
- Shares Event Histories (SHARES).

There are tables for end-of-day and end-of-month frequencies for each event type (e.g. dsenames, msenames) as well as tables of all events together (dse, mse). Additional tables dseall, mseall, and stocknames were created by WRDS and are described below.

The DATE variable in the Monthly Stock Events (mse) file for each event type **corresponds to different items** across the various files. In the msenames file it refers to the Name Date (NAMEDT), in the msedist file it is the ex-distribution date (EXDT), for msedelist it is the delisting date (DLSTDT), mseshares uses it for the shares observation date (SHRSDT) in , and in msenasdin it is the NASDAQ Traits Date (TRTSDT).

The following table shows a slice taken from the mse file for Microsoft.

Table 4: mse

Date	EVENT	COMNAM	PERMNO	PERMCO	TSYMBOL	CUSIP	NCUSIP	NAICS	HSICCD	SICCD	SHRCD	NAMEENDT
...	...	...	...	...	...	...	...	...	...	...	...	...
19860313	NAMES	MICROSOFT CORP	10107	8048	MSFT	59491810	59491810		7370	7370	11	20040609
...	...	...	...	...	...	...	...	...	...	...	...	...
20040528	NASDIN		10107	8048		59491810			7370			
20040604	NASDIN		10107	8048		59491810			7370			
20040610	NAMES	MICROSOFT CORP	10107	8048	MSFT	59491810	59491810	511210	7370	7370	11	20140829
20040610	SHARES		10107	8048		59491810			7370			
...	...	...	...	...	...	...	...	...	...	...	...	...
20041115	DIST		10107	8048		59491810			7370			
20041115	DIST		10107	8048		59491810			7370			

When NAMES occurs in the "Event Type" field, **there is a change** to one or more identifying fields: company name, NCUSIP, Tsymbol, NAICS and so on. Two SIC variables are present. HSICCD is the non-historical header version that is the same for the company's entire history. SICCD is the historic version that changes over time.

In the **example above for Microsoft**, there are two NAMES events between 1986 and 2004. The row added on June 10, 2004 **indicates** the NAMES event where a NAICS code was added.

The EVENT type is **not available** in mseall. This dataset is **generated by WRDS** and displays a summary of a company as of each date.

Table 5: mseall

Date	COMNAM	PERMNO	PERMCO	TSYMBOL	CUSIP	NCUSIP	NAICS	HSICCD	SICCD	SHRCD	NAMEENDT	DCLF
19860331	MICROSOFT CORP	10107	8048	MSFT	59491810	59491810		7370	7370	11	20040906	
...	...	...	...	...	...	...	...	...	...	...	...	...
20040528	MICROSOFT CORP	10107	8048	MSFT	59491810	59491810		7370	7370	11		
20040630	MICROSOFT CORP	10107	8048	MSFT	59491810	59491810	511210	7370	7370	11	20140829	
...	...	...	...	...	...	...	...	...	...	...	...	...
20041130	MICROSOFT CORP	10107	8048	MSFT	59491810	59491810	511210	7370	7370	11		2004(
20041130	MICROSOFT CORP	10107	8048	MSFT	59491810	59491810	511210	7370	7370	11		2004^

In the mseall file, items associated with a **one-time event**, such as dividend cash amount (divamt), will not be carried onto the next observation. If there are multiple one-time events within one month (e.g. see the month of November 2004 for Mircoosft), **multiple observations will appear** for the same date (msi.date).

The stocknames file is a cross between dseall and dsenames. It has only **the most important** identification variables, eliminating much of the noise of dseall. It adds an effective date range for the identifiers (NAMEDT, NAMEENDDT) for each set of NCUSIP, COMNAM, TICKER, and EXCHCD variables and a date range for price data (ST_DATE, END_DATE). CUSIP and HEXCD are header variables and SICCD, SHRCD, and SHRCLS reflect the status of name start date (namedt). stocknames is **the most popular** event file.

The example below shows slices taken from stocknames file for Microsoft and Dell. In 2003, Dell changed its CUSIP from 24702510 to 24702R10; the two rows in stocknames reflect this change.

COMNAM	PERMNO	PERMCO	NCUSIP	NAMEDT	NAMEENDT	ST_DATE	END_DATE
MICROSOFT CORP	10107	8048	59491810	19860313	20140829	19860331	20140829
DELL COMPUTER CORP	11081	9833	24702510	19880622	20030721	19880630	20131031
DELL COMPUTER CORP	11081	9833	24702R10	20030722	20131029	19880630	20131031

In CRSP the variable BEGDAT from table dsfhdr is often used as **a rough estimate** of the first date of trading after the initial public offering (IPO). The document "WRDS Guide to IPO Databases and Research" discusses this topic in further detail.

SIC Codes

The Standard Industrial Classification code (SIC) **is used to group companies** with similar products or services. The SIC code is **an integer** between 100 and 9999. The **first 2** digits refer to a major industry group, the **third** digit identifies an industry group and the **fourth** digit indicates the industry.

In CRSP, the SIC code variable (SICCD) **contains the historical SIC information**. The `HSICCD` variable contains the header or most recent SIC codes. `HSICCD` is available in the stock files, however, you will have to look in the event files for `SICCD`.

People often report that there are differences between the SIC codes in CRSP, Compustat, and other databases. This is due to CRSP and Compustat **obtaining SIC codes from different sources**. Compustat assigns SIC codes by analyzing a company's 10K and annual report.

In the December 2009 stock database, **CRSP removed SIC codes** provided by Mergent from the Stock Databases and replaced them with SIC codes from Interactive Data Corporation (IDC). **Mergent was the primary source** of SIC codes for NYSE, NYSE MKT, and ARCA securities between August 24, 2001 (200108240) and 2009. **IDC** has always been a continuous alternate source of SIC Codes.

SIC codes **can be useful for** rough groupings of industries. Beyond that, they should be **used with caution**, as they are **not assigned or reviewed with a strict procedure** by any government agency. Moreover, most large companies belong to **multiple** SIC codes and can **change** over time. After the initial SIC code assignment when a company goes public, government agencies **do not refer** to the code or the company again - and **quite often a company will report its initial SIC code** forever. There have been cases in which companies have **obsolete SIC codes** from the 1972 numbering scheme in SEC filings from the early 1990s

Calculations

Market Cap

In CRSP the market capitalization as the product of price and shares outstanding is **computed** as follows:

```
1  mktcap = abs(prc)*shrout
```

Market to Book Ratios

Please refer to the [Market-to-Book (M/B) Ratio](#) or [Daniel, Grinblatt, Titman, and Wermers (1997) Characteristic-Based Benchmarks](#) applications on WRDS.

Top of Section

Top

Table of Contents

Wharton Research Data Services

About WRDS	3 Ways to use WRDS
WRDS FAQs	WRDS Account Types
WRDS News	Terms of Use
Account Preferences	Sample Data
Info / Support Request	Conference Calendar
Privacy Policy	Impactful Research