**Monthly price and volume data dictionary**

Data range: Data from January, 1990 to Dec, 2017

This new data set was downloaded to reduce computational costs of working with a very large Daily and Volume data set. Further, since the other important data sets expected to be joined start from only 1990, I did not download monthly data before 1990. At a later stage of the project, I may choose to focus exclusively on the daily price and volume data set.

The data dictionary for the monthly price and volume data set is very similar to the daily price and volume data set. However, it does not include number of trades and Opening Price columns. All other columns contain information aggregated monthly.

**Columns in the data set**

| **Name** | **Code** | **Comments** |
| --- | --- | --- |
| PERMNO | PERMNO | Primary key |
| date | date | From 1925 |
| Share code | SHRCD | Is it a stock? ADR? REIT? |
| SICCD | SIC code | Industry or sector. Provided by the commerce department |
| Company Name | COMNAM | Company name |
| Share class | SHRCLS | Class A? (voting rights) |
| Trading Symbol | TSYMBOL | Ticker |
| Trading Status | TRDSTAT | Trading? Halted? |
| Security status | SECSTAT | New issue? Regular? Bankrupt? |
| Permanent company number | PERMCO | Company code. Keep a mind a company can issue multiple stocks |
| Cusip | CUSIP | S&P’s company code. Not really a unique code. But we may need this later |
| New PERMNO | NWPERM | Applicable if a stock got merged into another stock |
| Bid or Low | BIDLO | Intraday low price |
| Ask or High | ASKHI | Intraday high price |
| Price | PRC | Closing price |
| Share Volume | VOL | Intraday volume |
| Holding Period Return | RET | Percentage change from previous day |
| Number of shares outstanding | SHROUT | Total no. of shares |
| Cumulative Factor to adjust price | CFACPR | To account for stock splits, dividends |
| Cumulative Factor to adjust shares | CFASHR | To account for stock splits and reverse stock splits etc. |
| Value weighted return (includes dividends) | vwretd | Market cap weighted portfolio of entire market |
| Equal weighted return (includes dividends) | ewretd | Equal weighted portfolio of the entire market |
| Return on S&P composite index | sprtrn | S&P500 and its predecessors return |

**Explanation of the variables in more detail (as provided by CRSP)**

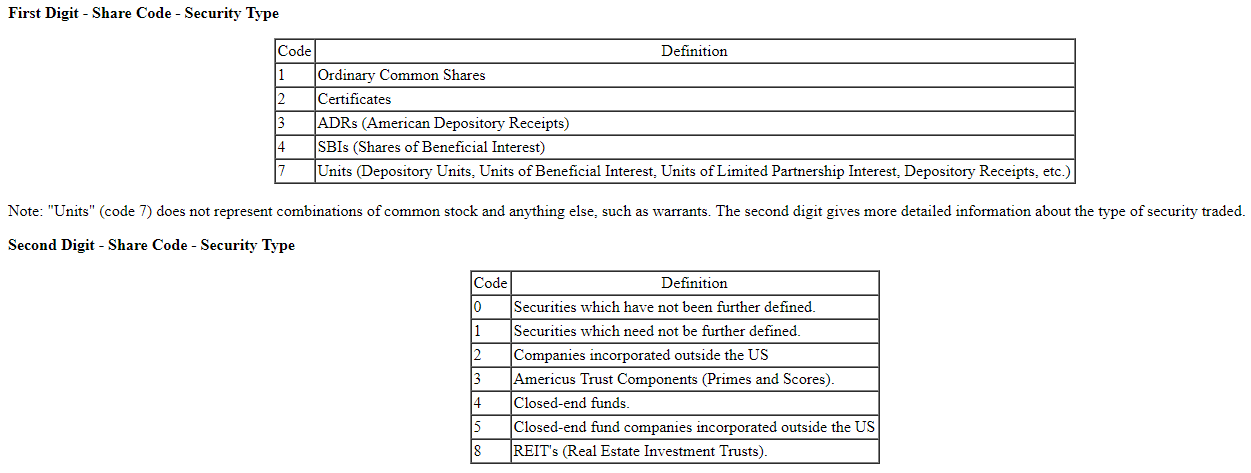
PERMNO

CRSP’s permanent issue identifier. One PERMNO belongs to only one PERMCO. One PERMCO can have one or more PERMNOs.

Date

Date in MMDDYYYY format

SHRCD



SICCD

SICCD is the Standard Industrial Classification code. SIC Code is used to group companies with similar products or services. The Standard Industrial Classification Manual contains descriptions of categories recognized by the US Government. SIC Code is an integer between 100 and 9999. The first two digits refer to a major group. The first three digits refer to an industry gro up. All four digits indicate an industry. Missing SIC Codes are set to 0. Most Nasdaq SIC codes have only the first three digits; CRSP has added a fourth digit of zero.

The North American Industry Classification System (NAICS) is replacing the U.S. Standard Industrial Classification (SIC) system. NAICS was developed jointly by the U.S., Canada, and Mexico to provide new comparability in statistics about business activity across North America

COMNAM

CRSP allocates a 32 character name description field for all securities. Preference is given to the spellings and abbreviations provided in Standard & Poor's CUSIP Directory. In cases where all name sources provide descriptions in excess of 32 characters, CRSP furnishes its own abbreviations.

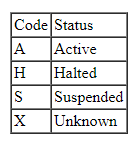
SHRCLS

SHRCLS describes the class of share and is generally blank. Any letter that identifies the class of stock (e.g., "A" for class A common) is contained in this field, left justified, and padded with three blank spaces.

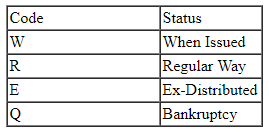
TSYMBOL

Trading Ticker Symbol is the trading symbol listed by exchanges and consolidated quote systems. It includes all temporary values, share classes, and share type suffixes. There is no punctuation (no periods) in the Trading Ticker Symbol. N.B. this field includes data starting on 20020102 for NYSE/AMEX, and 19821101 for NASDAQ.

TRDSTAT



SECSTAT



PERMCO

PERMCO is a unique permanent identifier 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. The mnemonics PERMCO and COMPNO are interchangeable. If the value in this field is less than 20,000, CRSP has used the Nasdaq-assigned Company Number, for when the company had an issue trading on The Nasdaq Stock Market(SM). A value over 20,000 indicates that the number was assigned by CRSP.

CUSIP

CUSIP is the latest eight-character CUSIP identifier for the security through the end of the file. CUSIP identifiers are supplied to CRSP by the CUSIP Service Bureau, Standard & Poor's, a division of McGraw-Hill, Inc., American Bankers Associate database, Copyright 1997. See Appendix A.6 for more CUSIP copyright information.

CUSIP identifiers were first assigned in 1968 as integers and expanded in 1984 to include alphabetic characters. The first six characters (including leading zeroes) identify the issuer, while the last two characters identify the issue. CUSIP issuer identifiers are assigned to maintain an approximately alphabetical sequence. The CUSIP identifier may change for a security if its name or capital structure changes. No header or historical CUSIPs are reused on our files. For securities no longer in existence or that were never assigned an official CUSIP identifier, CRSP has assigned a dummy CUSIP identifier for use in this field in accordance with the rules published in the CUSIP Directory.

There are two potential dummy CUSIPs which are assigned by CRSP. One, \*\*\*99\*9\*, (containing a 9 in the 4th, 5th and 7th character positions) represents a CRSP assigned CUSIP with a dummy issuer number (the first 6 character positions) and a dummy issue number (the last 2 character positions). The other, \*\*\*\*\*\*9\*, (containing a 9 in the 7th character position) represents a CRSP-assigned CUSIP with a real issuer number but a dummy issue number.

For example:

A CUSIP such as 12399099 or 12345699 is assigned by CRSP, and an identifier such as 12345610 is assigned by the CUSIP Agency. Securities actively traded on an international basis, domiciled outside the United States and Canada, will be identified by a CINS (CUSIP International Numbering System) number. CINS numbers employ the same Issuer (6 characters)/Issue (2 characters) 8-character identifier system set by the CUSIP Numbering System. It is important to note that the first portion of a CINS code is always represented by an alpha character, signifying the issuer's country code (domicile) or geographic region. See the current CUSIP Directory for more information. See Appendix A.1 in the Stock File User's Guide for a list of CINS country codes.

NWPERM

NWPERM is assigned when an issue ceases trading as a result of a merger or exchange where shareholders receive stock in the acquiring company. NWPERM is the permno of the acquiring company's stock. It acts as a forward pointer, allowing the user to trace the ongoing history of surviving companies. NWPERM may identify an issue that exists on a different CRSP stock file.

It is set to zero if there is no new primary security applicable, the issue is unknown, or the delisting does not represent a merger or exchange. The distribution history arrays contain an itemized record of all types of payments to shareholders in an exchange or merger. See Acquiring PERMNO for companies associated with individual payments.

BIDLO

Daily: *Bid or Low Price* is the lowest trading price during the day, or the closing bid price on days when the closing price is not available. The field is set to zero if no *Bid or Low Price* is available.

Daily trading prices for the NASDAQ National Market securities were first reported November 1, 1982. Daily trading prices for the NASDAQ SmallCap Market were first reported June 15, 1992. Therefore, *Bid or Low Price* for NASDAQ securities is always an ask before these days.

Monthly: Monthly files contain the lowest daily *Price or Bid/Ask Average* during the month. Closing price values are positive, bid/ask averages negative. The negative sign is a symbol used to differentiate between price and bid/ask average. The bid/ask average does not have a negative value, and the sign is ignored for ranking. The field is set to zero when no ask or bid/ask average is available.

ASKHI

Daily: *Ask or High Price* is the highest trading price during the day, or the closing ask price on days when the closing price is not available. The field is set to zero if no *Ask or High Price* is available. If the Price or Bid/Ask Average is negative, this field contains the closing ask. If positive, the field contains the highest trade.

Daily trading prices for the NASDAQ National Market securities were first reported November 1, 1982. Daily trading prices for the NASDAQ SmallCap Market were first reported June 15, 1992. Therefore, *Ask or High Price* for NASDAQ securities is always an ask before these days.

Monthly: Monthly files contain the highest daily *Price or Bid/Ask Average* during the month. The field is set to zero when no *Price or Bid/Ask Average* was available during the month a value is found for incomplete months. If *Price or Bid/Ask Average* contains any bid/ask averages, these will be marked with a negative symbol. The absolute value of *Price or Bid/Ask Average* is used to select the highest, but the sign is preserved if a bid/ask average is selected.

PRC

Prc is the closing price or the negative bid/ask average for a trading day. If the closing price is not available on any given trading day, the number in the price field has a negative sign to indicate that it is a bid/ask average and not an actual closing price. Please note that in this field the negative sign is a symbol and that the value of the bid/ask average is not negative.

If neither closing price nor bid/ask average is available on a date, prc is set to zero. In a monthly database, prc is the price on the last trading date of the month. The price series begins the first month-end after the security begins trading and ends the last complete month of trading.

If the security of a company is included in the Composite Pricing network, the closing price while listed on NYSE or AMEX on a trading date is the last trading price for that day on the exchange that the security last traded.

Similarly, highs, lows, and volumes include trades on all exchanges on which that security traded. For example, if a stock trades on both the NYSE and the PACX (Pacific Stock Exchange), and the last trade occurs on the PACX, the closing price on that day represents the closing price on the PACX, not the NYSE. Price data for Nasdaq securities comes directly from the NASD with the close of the day at 4:00 p.m. Eastern Time. Automated trades after hours on Nasdaq are counted on the next trading date, although the volumes are applied to the current date. Daily trading prices for The Nasdaq National Market securities were first reported November 1, 1982. Daily trading prices for The Nasdaq Small Cap Market were first reported June 15, 1992. prc for Nasdaq securities is always a negative bid/ask average before this time. All prices are raw prices as they were reported at the time of trading.

VOL

In monthly files, VOL is the sum of the trading volumes during that month. In daily files, VOL is the total number of shares of a stock sold on day I. It is expressed in units of one share, for daily data, and on hundred shares for monthly data. Our data source for NYSE/AMEX reports the number rounded to the nearest hundred. For example, 12,345 shares traded will be reported on the Nasdaq Stock Exchange as 12,345 and on the NYSE or AMEX exchanges as 12,300. Volume is set to -99 if the value is missing. A volume of zero usually indicates that there were no trades during the time period and is usually paired with bid/ask quotes in price fields.

NYSE/AMEX volumes are the sum of volumes on all exchanges where that security traded that day.

RET

A return is the change in the total value of an investment in a common stock over some period of time per dollar of initial investment. RET(I) is the return for a sale on day I. It is based on a purchase on the most recent time previous to I when the se curity had a valid price. Usually, this time is I - 1. Returns are calculated as follows:

For time t (a holding period), let:

t? = time of last available price < t

r(t) = return on purchase at t?, sale at t

p(t) = last sale price or closing bid/ask average at time t

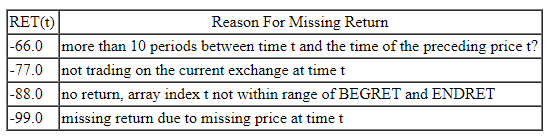
d(t) = cash adjustment for t

f(t) = price adjustment factor for t

p(t?) = last sale price or closing bid/ask average at time of last available price < t.

then r(t) = [(p(t)f(t)+d(t))/p(t')]-1

t? is usually one period before t, but t? can be up to ten periods before t if there are no valid prices2 in the interval. A series of special return codes specify the reason a return is missing.



SHROUT

SHROUT is the number of publicly held shares, recorded in thousands

CFACPR

Adjusted for stock splits and dividends

CFASHR

Adjusted for stock splits and reverse stock splits etc.

Vwretd

VWRETD indices contain either the daily or monthly returns, including all distributions, on a value-weighted market portfolio (excluding American Depository Receipts (ADRs)).

ewretd

EWRETD indices contain, depending on your subscription, daily or monthly returns, including all distributions, on an equally-weighted market portfolio (including ADRs).

Sprtrn

SPRTRN is the return on the Standard & Poor's Composite Index defined as:

(SPINDX(t)/SPINDX(t-1)) - 1

where SPINDX is the level of the Standard & Poor's 500 Composite Index (prior to March 1957, 90-stock index) at the end of the trading day or month.