Leibniz Institute for Financial Research SAFE
Theodor-W.-Adorno-Platz 3 I 60629 Frankfurt am Main I Germany

Information on the preprocessed CRSP data set

Alexander Hillert
SAFE Research Datacenter*

Last Update: March 15, 2025

Contents

1 Contact

If you encounter any difficulties or just want general information, do not hesitate to contact us.

SAFE Research Datacenter: datacenter@safe-frankfurt.de

More information about the SAFE Research Datacenter, and further guides can be found here, and here.

*datacenter@safe-frankfurt.de

2 Introduction

The monthly stock data from the Center for Research in Security Prices (CRSP) is one of the most popular data sets for empirical research on the U.S. stock market. While the CRSP database is known for its good data quality, the data still requires some preprocessing. Most importantly, delisting returns need to be taken into account to measure investors' returns correctly. This preprocessed data set combines CRSP's delisting return with the "normal" CRSP holding period returns. When delisting returns are missing, the correction procedure of Shumway (1997) and Shumway and Warther (1999) is used. The preprocessing steps are explained below.
The preprocessed file is based on the CRSP data from February 2025 and includes the period from December 1925 (start of CRSP) until December 2024 (most recent available period).
The preprocessed file is a csv file named

CRSP_monthly_1926_2024_adj_delisting.csv

It uses comma as delimiter. String variables are enclosed by double quotes. It has been saved as a zip file to improve efficiency.

3 Data processing steps

Note that below, variable names are indicated in brackets.

- Restrict the sample to ordinary shares of U.S. firms, i.e., keep securities with share codes [shrcd] 10 and 11.
- Create a delisting return variable [delisting_return]:
  - Use CRSP's delisting return [dlret] if available.
  - Otherwise, follow the procedure of Shumway (1997) and Shumway and Warther (1999), i.e.,
    - For NYSE and Amex stocks, impute -30% delisting return for the relevant delisting codes [dlstcd].
    - For Nasdaq stocks, impute -55% delisting return for the relevant delisting codes [dlstcd].
- Combine CRSP holding period returns [ret] with the delisting return, i.e., compute the compounded return if both are available and use either of the two if only one is available. The resulting variable [return_adj] measures investors return.

4 CRSP variables

Below is the list of CRSP variables included in the data set. It starts with the firm/stock ID and then continues with the actual data points like prices and volume.
Most variables are obtained directly from CRSP. If variables are self-computed, it is indicated.

4.1 Firm/stock identifiers

- **Comnam**: historical company name from CRSP.
  This variable contains abbreviations like "Mgmt" for "Management" or "Intl" for "International".

- **Cusip**: the security's most recent identifier from the Committee on Uniform Security Identification Procedures.
- **Ncusip**: the security' historical identifier from the Committee on Uniform Security Identification Procedures.
- **Permco**: unique firm-level identifier in the CRSP stock database.
  One company (=a single permco) can have multiple share classes outstanding (=multiple permnos). A prominent example is Alphabet Inc. (the parent company of Google). Alphabet (permco== 45483) has permno==14542 (class C shares) and permno== 90319 (class A shares) outstanding.
- **Permno**: unique security-level (=stock-level) identifier in the CRSP stock database.
- **Ticker**: the stock's historical trading symbol.

4.2 Dates and time identifiers

- **Date**: the date when the information was recorded. In the monthly CRSP data, it is the last trading day of the month. CRSP's point-in-time data (e.g., share price, number of shares outstanding) represent the information on that day.
- **Month_id** (self-computed): numerical time identifier on the monthly frequency based on the date from CRSP (January 1960 is month_id==0, February 1960 is month_id==1, January 1961 is month_id==12, February 1961 is month_id==13, October 2023 is month_id==765; the variable has been constructed using Stata's mofd()-function).
  This variable is helpful to define a panel data set (e.g., in Stata "xtset permno month_id").
- **Year**: the year of the date [Date].

4.3 Accounting and stock market variables

- **delisting_return** (self-computed): comprehensive return based on CRSP's delisting return [dlret] and the Shumway (1997) and the Shumway and Warther (1999) correction.
- **Dlprc**: the value of the liquidation payment that shareholders receive.
- **Dlret**: the delisting return from CRSP. It is based on the liquidation payment that shareholders receive. If, for example, a stock delists at a price of $10 per share and shareholders receive a liquidation payment of $4 per share, the delisting return will be -60%.
  The delisting return can be positive. For example, if the company is acquired and the acquirer pays a premium over the share price.
- **Dlstcd**: a code that indicates the reason for the delisting like, for example, bankruptcy, merger, the firm is going private, or the firm is acquired.
- **Exchcd**: the code for the exchange the stock is listed at
  - Exchcd==1: NYSE.
  - Exchcd==2: Amex.
  - Exchcd==3: Nasdaq.

- **Market_cap** (self-computed): a stock's market capitalization in million USD. It is calculated as the number of shares outstanding times the closing price (i.e., shrout * abs(prc) / 1000).

- **Prc**: closing price of a stock on the given date. In the monthly data, it is the closing price on the last trading day of the month.
  Negative numbers indicate that the price is not an actual closing price but the average of the bid and the ask. Recommendation: use the absolute value.

- **Ret**: the holding period return. In CRSP daily (monthly), it the return from yesterday's (last month's last trading day's) closing price to today (this month's last trading day's) closing price. The variable is adjusted for dividends and stock splits.

- **Return_adj** (self-computed): combines CRSP holding period return [ret] with the comprehensive delisting return [delisting_return]. Use this variable to measure shareholders' return. If you are using this variable no further return adjustment is needed.

- **Shrcd**: the share code from CRSP. The data set is filtered for codes 10 and 11 (see above unter "preprocessing steps").

- **Shrcls**: the share class of the stock. For example, for Alphabet, there are class A (permno== 90319) and class C (permno==14542) shares.

- **Shrout**: number of shares outstanding in thousands of shares.

- **Siccd**: historical Standard Industry Classification Code
  Based on this variable, on can create Fama and French (1997) industry groups (go to https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html and scroll down to "Industry Portfolios").

- **Vol**: the trading volume of a stock on the day (daily CRSP data) or during the month (monthly CRSP data). It is recorded in the number of shares traded. In the monthly CRSP files, volume is expressed in units of hundreds (100).

5 Further information

For further information about the CRSP stock market data, please see CRSP's official database manual (available on WRDS) and/or the Data Center's introductory video on CRSP available on the SAFE webpage.

References

Shumway, Tyler (1997) "The delisting bias in CRSP data," *The Journal of Finance*, 52 (1), 327–340.

Shumway, Tyler and Vincent A Warther (1999) "The delisting bias in CRSP's Nasdaq data and its implications for the size effect," *The Journal of Finance*, 54 (6), 2361–2379.