

Alexander Hillert

Introduction to the CRSP stock market database

Leibniz Institute for Financial Research SAFE
Research Data Center

Agenda

1. Overview

2. Key variables

a. Security identifiers

b. Security characteristics

c. Return, price, and volume data

3. Data selection and download

4. Conclusion

Overview (1)

Introduction

- CRSP = Center for Research in Security Prices.
- Our focus: stock market database.
- High quality stock market data for the U.S.
- Coverage
 - 1926 to today.
 - Annual data updates.
- Data access
 - Web browser or Python API.
 - Part of Wharton Research Data Services.
 - Available in the House of Finance Data Room.
- CRSP offers further databases like mutual funds.

Overview (2)

Introduction – continued

- Data frequency: daily and monthly.

Stock / Security Files

Stock / Security Files

» [Daily Stock File](#)

» [Stock Market Indexes](#)

» [Monthly Stock File](#)

» [U.S. Daily Event Study: Upload your own events](#)

» [Stock Header Info](#)

- Individual stock data vs. aggregate index data.

Agenda

1. Overview
2. Key variables
 - a. Security identifiers
 - b. Security characteristics
 - c. Return, price, and volume data
3. Data selection and download
4. Conclusion

Security identifiers (1)

CRSP's security identifiers

- PERMNO: unique security-level identifier.
- PERMCO: unique company-level identifier.
- One firm has one PERMCO but can have multiple PERMNOs.

date	comnam	permno	permco	ticker	shrcls
31-Jan-23	ALPHABET INC	14542	45483	GOOG	C
31-Jan-23	ALPHABET INC	90319	45483	GOOGL	A
28-Feb-23	ALPHABET INC	14542	45483	GOOG	C
28-Feb-23	ALPHABET INC	90319	45483	GOOGL	A
31-Mar-23	ALPHABET INC	14542	45483	GOOG	C
31-Mar-23	ALPHABET INC	90319	45483	GOOGL	A

- PERMNO and PERMCO are permanent IDs.

Security identifiers (2)

CRSP's security identifiers – continued

- PERMNO and PERMCO are permanent IDs.
- Do not change when there are minor events like, e.g., company name changes.

date	comnam	permno	permco	ticker	shrcls
30-Sep-15	GOOGLE INC	14542	45483	GOOG	C
30-Sep-15	GOOGLE INC	90319	45483	GOOGL	A
30-Oct-15	ALPHABET INC	14542	45483	GOOG	C
30-Oct-15	ALPHABET INC	90319	45483	GOOGL	A

- Can change after major events like, e.g., bankruptcy restructuring.

date	comnam	permno	permco	ticker
30-Apr-09	GENERAL MOTORS CORP	12079	20799	GM
29-May-09	GENERAL MOTORS CORP	12079	20799	GM
30-Jun-09	GENERAL MOTORS CORP	12079	20799	GM
30-Nov-10	GENERAL MOTORS CO	12369	53554	GM
31-Dec-10	GENERAL MOTORS CO	12369	53554	GM
31-Jan-11	GENERAL MOTORS CO	12369	53554	GM

Security identifiers (3)

Further security identifiers

- Ticker symbol.
- CUSIP
 - Variable “CUSIP” is today’s CUSIP. NOT historical data.
 - Variable “NCUSIP” is historical data.
 - Important difference when merging with other data sets.
- CUSIPs can change even for minor firm events.

date	comnam	permno	permco	ticker	shrcls	ncusip	cusip
30-Sep-15	GOOGLE INC	14542	45483	GOOG	C	38259P70	02079K10
30-Sep-15	GOOGLE INC	90319	45483	GOOGL	A	38259P50	02079K30
30-Oct-15	ALPHABET INC	14542	45483	GOOG	C	02079K10	02079K10
30-Oct-15	ALPHABET INC	90319	45483	GOOGL	A	02079K30	02079K30

Agenda

1. Overview
2. Key variables
 - a. Security identifiers
 - b. Security characteristics
 - c. Return, price, and volume data
3. Data selection and download
4. Conclusion

Security characteristics (1)

Share code

- CRSP includes not only U.S. stocks but also other securities like foreign stocks listed in the U.S., closed-end funds, and REITs.
- Share code (SHRCD) allows to select relevant securities.
 - Standard stock universe: ordinary U.S. stocks → share codes 10 and 11.

date	comnam	permno	permco	shrcl	ticker	shrcls
30-Oct-15	ALPHABET INC	14542	45483	11	GOOG	C
30-Oct-15	ALPHABET INC	90319	45483	11	GOOGL	A

- Foreign stocks listed in the U.S.: share code 12.

date	comnam	permno	permco	shrcl	ticker
30-Nov-23	DEUTSCHE BANK A G	89199	42291	12	DB
29-Dec-23	DEUTSCHE BANK A G	89199	42291	12	DB

Security characteristics (2)

Exchange code

- Provides information where stocks are traded.
- Standard universe includes NYSE (exchcd = 1), AMEX (exchcd = 2, starting in 1962), and Nasdaq (exchcd = 3, starting in 1973).

date	comnam	permno	exchcd	ticker
29-Dec-23	GENERAL MOTORS CO	12369	1	GM
29-Dec-23	ADAMS RESOURCES & ENERGY INC	61487	2	AE
29-Dec-23	ALPHABET INC	90319	3	GOOGL

- Other examples: Toronto (exchcd = 19), over-the-counter (exchcd = 20).
- Caution: same ticker at different exchanges at the same time.

date	comnam	permno	exchcd	ticker
30-Oct-64	RELIANCE MFG CO	20511	1	RMC
30-Oct-64	RAMCO ENTERPRISES INC	35713	2	RMC

Security characteristics (3)

Firm characteristics

- **Company name**
 - Historical information.
 - Can include abbreviations like “MGMT” (management) and “SVC” (service).
- **Industry classification**
 - Standard Industry Classification codes (SIC codes).
 - 4-digit number from 100 to 9999.
 - first two digits indicate sector, e.g., 6xxx is finance, insurance, real estate.
 - Historical information.
 - Popular [Fama and French \(1997\)](#) industries are based on SIC codes.
 - [Ken French's data library](#) → “Industry portfolios” → “Details”

Agenda

1. Overview
2. Key variables
 - a. Security identifiers
 - b. Security characteristics
 - c. Return, price, and volume data
3. Data selection and download
4. Conclusion

Return, price, and volume data (1)

Price data

- CRSP reports closing prices (PRC), opening prices (OPENPRC), highest and lowest prices/quotes (ASKHI and BIDLO).
- If no closing price is available, bid-ask average is used.
- Important: CRSP puts a minus sign before bid-ask averages to indicate that there are not actual closing prices.

date	comnam	permno	prc	exchcd
28-Feb-23	ELECTRO SENSORS INC	32062	4.58	3
31-Mar-23	ELECTRO SENSORS INC	32062	-4.57	3
28-Apr-23	ELECTRO SENSORS INC	32062	4.56	3
31-May-23	ELECTRO SENSORS INC	32062	-4.445	3
30-Jun-23	ELECTRO SENSORS INC	32062	4.44	3

Return, price, and volume data (2)

Return data

- CRSP provides holding period returns (RET) adjusted for distributions and stock splits.
- Example for stock split: NVIDIA's 1-to-4 stock split

date	comnam	permno	prc	ret	shrout	cfacpr
28-May-21	NVIDIA CORP	86580	649.78	0.0823	623000	4
30-Jun-21	NVIDIA CORP	86580	800.1	0.2316	623000	4
30-Jul-21	NVIDIA CORP	86580	194.99	-0.0252	2500000	1
31-Aug-21	NVIDIA CORP	86580	223.85	0.1482	2500000	1

$$\rightarrow \text{Return in July: } \frac{194.99 \cdot 4 - 800.1}{800.1} = -0.0252$$

- Returns excluding dividends are also available (RETX).

- Shrout: number of shares outstanding (in 1000s).
- CFACPR: cumulative factor to adjust prices.

Return, price, and volume data (3)

Delisting return information

- Delisting return (DLRET) is the return after a stock delisted.
- For example, shareholders may receive a final distribution.
- Delisting return captures the difference between last closing price and final payment/value.

date	comnam	permno	prc	ret	dlret	dlstcd
31-Mar-09	GENERAL MOTORS CORP	12079	1.94	-0.1378		
30-Apr-09	GENERAL MOTORS CORP	12079	1.92	-0.0103		
29-May-09	GENERAL MOTORS CORP	12079	0.75	-0.6094		
30-Jun-09	GENERAL MOTORS CORP	12079			-0.1867	574

- Delisting code provides information on why the stock was delisted.
 - For example, 574 = “*Delisted by current exchange - bankruptcy, declared insolvent.*”
 - see CRSP manual for list of codes.

Return, price, and volume data (4)

Delisting return information – continued

- To avoid biased results, important to incorporate delisting information.
- Compound returns if both ret and dlret are available.

date	comnam	permno	prc	ret	dlret	dlstcd
30-May-08	COUNTRYWIDE FINANCIAL CORP	64565	5.26	-0.0640		
30-Jun-08	COUNTRYWIDE FINANCIAL CORP	64565	4.25	-0.1920	0.0209	231
31-Jul-08	COUNTRYWIDE FINANCIAL CORP	64565				

DLSTCD = 231: *“When merged, shareholders primarily receive common stock or ADRs.”*

- Use dlret if ret is not available (example on previous slide)
- Delisting return information is not always available.
→ [Shumway \(1997\)](#) and [Shumway and Warther \(1999\)](#) propose delisting corrections for NYSE/AMEX and Nasdaq stocks, respectively.

Return, price, and volume data (5)

Trading volume

- VOL: number of shares traded during a day/month.
 - Monthly file: units of 100 shares.
 - Daily file: units of one share.
 - Starts only in 1983 for Nasdaq stocks.

Shares outstanding

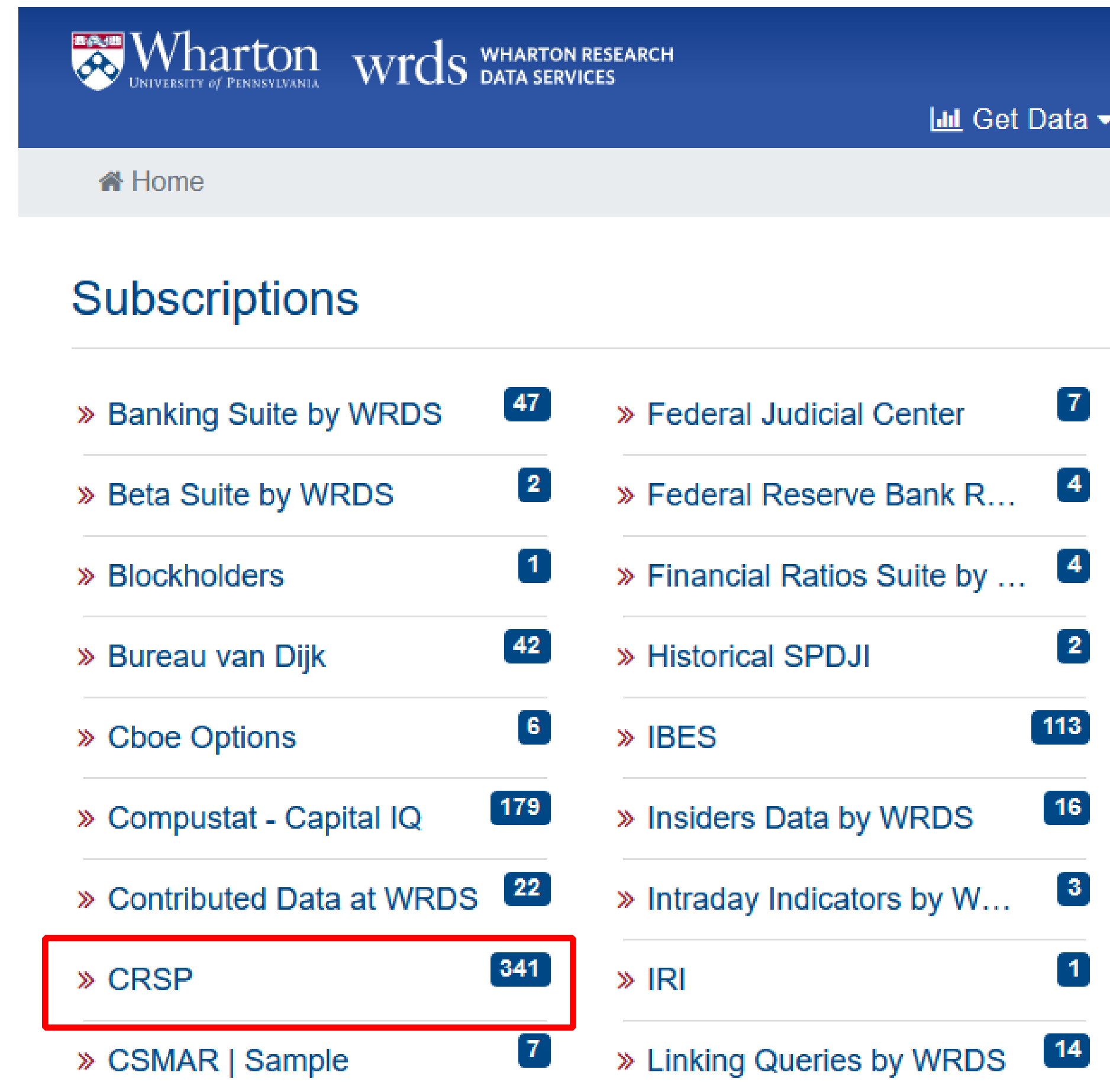
- SHROUT: number of publicly held shares in units of 1000 shares.
- To compute market cap, multiply SHROUT by the absolute value of price.

Agenda

1. Overview
2. Key variables
 - a. Security identifiers
 - b. Security characteristics
 - c. Return, price, and volume data
3. Data selection and download
4. Conclusion

Data selection and download (1)

CRSP in WRDS



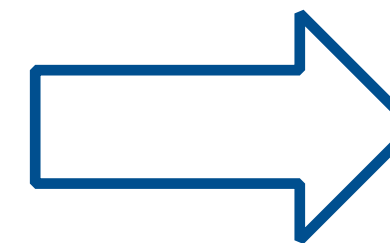
Wharton wrds WHARTON RESEARCH DATA SERVICES

Get Data

Home

Subscriptions

» Banking Suite by WRDS	47	» Federal Judicial Center	7
» Beta Suite by WRDS	2	» Federal Reserve Bank R...	4
» Blockholders	1	» Financial Ratios Suite by ...	4
» Bureau van Dijk	42	» Historical SPDJI	2
» Cboe Options	6	» IBES	113
» Compustat - Capital IQ	179	» Insiders Data by WRDS	16
» Contributed Data at WRDS	22	» Intraday Indicators by W...	3
» CRSP	341	» IRI	1
» CSMAR Sample	7	» Linking Queries by WRDS	14



CRSP

The Center for Research in Security Prices, LLC (CRSP) maintains the most comprehensive collection of security price, return, and volume data for the NYSE, AMEX and NASDAQ stock markets. Additional CRSP files provide stock indices, beta-based and cap-based portfolios, treasury bond and risk-free rates, mutual funds, and real estate data.

Annual Update

Databases in this section are updated once each year, in early February. Update schedules should not be confused with end-of-day, end-of-month, or end-of-quarter data such as stock prices.

» Stock / Security Files	5	» Stock / Events	5
» Stock - Version 2 (CIZ)	9	» CRSP/Compustat Merged	10
» Index / Stock File Indexes	6	» Index / Portfolio Assignments	3

Data selection and download (2)

CRSP in WRDS – continued

- Monthly files are much smaller and easier to handle.
- Recommendation: use daily files only if needed.



The screenshot shows the Wharton WRDS website interface. At the top, there is a blue header with the Wharton University of Pennsylvania logo and the text 'wrds WHARTON RESEARCH DATA SERVICES'. Below the header is a navigation menu with links: Home, Get Data, CRSP, Annual Update, and a partially visible link starting with 'S'. The main content area features the heading 'Stock / Security Files' in a large, bold font. Below this heading, there is a sub-heading 'Stock / Security Files' and two buttons with red double arrows pointing right, labeled 'Daily Stock File' and 'Monthly Stock File'.

Data selection and download (3)

Step 1: select sample period

- Earliest start date: December 1925.

CRSP Monthly Stock

Query Form	Variable Descriptions	Manuals and Overviews	Knowledge Base	Data Preview
------------	-----------------------	-----------------------	----------------	--------------

CRSP Monthly Stock

You have 6 saved queries for this Query Form.

Step 1: Choose your date range.

Date Variable:

2019-01	to	2023-12
---------	----	---------

- You can save queries at the end of the form.
- To rerun them, click here.

Data selection and download (4)

Step 2: select security sample

- There are different options:
 - Enter specific company IDs.
 - Upload a list with IDs.
 - Download the entire database.
- Recommendation: obtain the entire data set.

Step 2: Apply your company codes.

What format are your company codes?

Autocomplete

ticker PERMNO (permno) PERMCO (permco) CUSIP Header (cusip)

ncusip siccd Standard Industrial Classification Code (hsiccd)

Select an option for entering your company codes:

Search Name or Ticker Code List Name

Save this code list to Saved Codes

Please enter company codes separated by a space.

Example: IBM MSFT AAPL

[Code Lookup: CRSP Stock (Annual)]

-----Select Saved Codes List-----

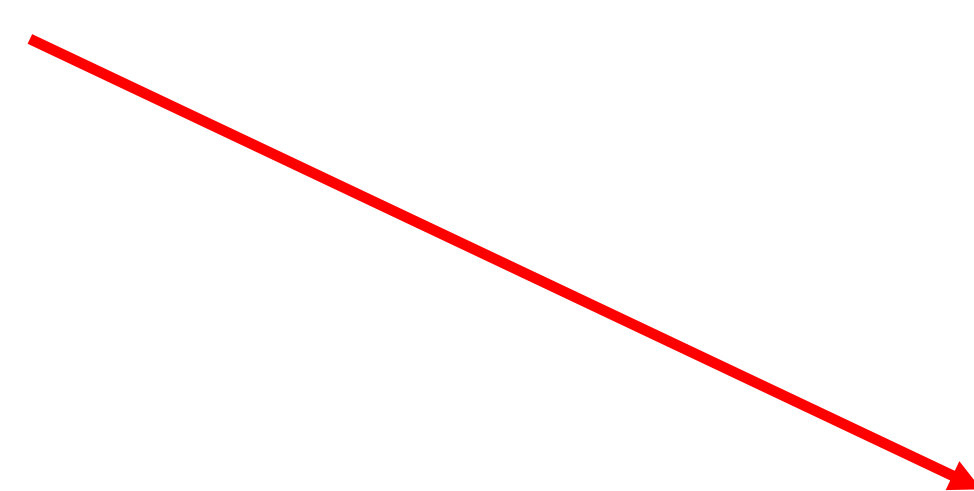
Choose from your saved code lists.

Browse... Company Codes Upload File

Upload a plain text file (.txt), having one code per line.

Search the entire database

This method allows you to search the entire database of records. Please be aware that this method can take a very long time to run because it is dependent upon the size of the database.

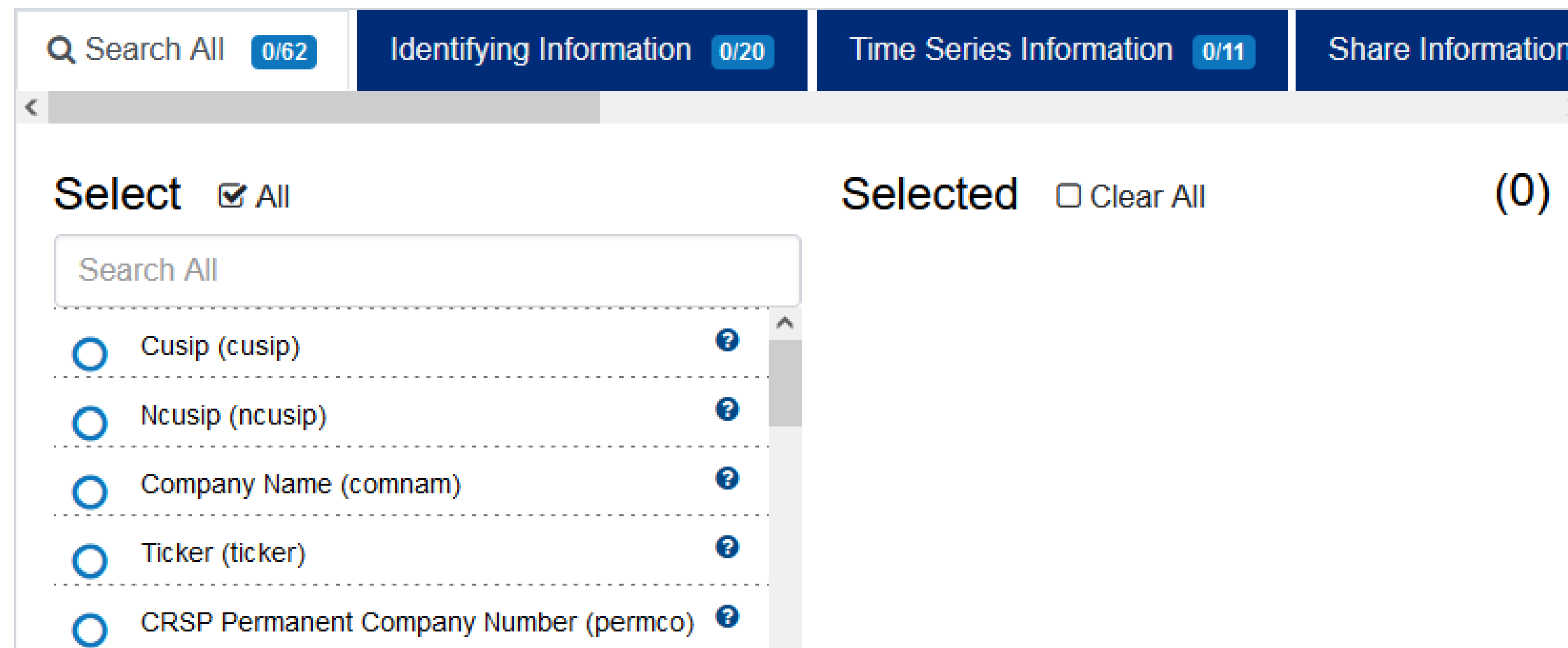


Data selection and download (5)

Step 3: select variables of interest

Step 3: Choose query variables.

How does this work?



The screenshot shows a web interface for selecting variables. At the top, there are four tabs: "Search All" (0/62), "Identifying Information" (0/20), "Time Series Information" (0/11), and "Share Information". Below the tabs, there is a "Select" section with a checked "All" checkbox and a "Selected" section with an unchecked "Clear All" checkbox and "(0)" items. A search bar labeled "Search All" is present. Below the search bar, a list of variables is shown, each with a radio button and a help icon:

- Cusip (cusip) ?
- Ncusip (ncusip) ?
- Company Name (comnam) ?
- Ticker (ticker) ?
- CRSP Permanent Company Number (permco) ?

No variables are currently selected. To set conditions on this query, Variables step.

▼ Activate Conditional Statement Builder

- There is the option to use the “**Conditional statement builder**” to restrict the sample.
- E.g., you could require $EXCHCD = 1$ or $EXCHCD = 2$.

- Permno will always be included.
- Set of variables depends on project. Including basic information like SHRCD and EXCHCD is certainly helpful.

Data selection and download (6)

Step 4: select data format

- Select preferred data format for your software.
- To speed up the download, use a zip file.
- Save your query in case you need to run it again.
- Submit form.

Step 4: Select query output.

How does this work?

Select the desired format of the output file. For large data requests, select a compression type to expedite downloads. If you enter your email address, you will receive an email that contains a URL to the output file when the data request is finished processing.

Output Format

- comma-delimited text (*.csv)
- Excel spreadsheet (*.xlsx)
- tab-delimited text (*.txt)
- HTML table (*.htm)
- SAS Windows_64 dataset (*.sas7bdat)
- STATA file (*.dta)

Compression Type

- Uncompressed
- zip (*.zip)
- gzip (*.gz)

Date Format

- YYYY-MM-DD. (e.g. 1984-07-25)
- MM/DD/YYYY. (e.g. 07/25/1984)
- DD/MM/YYYY. (e.g. 25/07/1984)
- YYYYMMDD. (e.g. 19840725)

E-Mail Address *(Optional)*

Custom Field *(Optional)*

Save This Query *(Optional)*

Notes on this Query *(Optional)*

Data selection and download (7)

Step 5: download the data

- After submitting the query, a new tab will open.
- Depending on the amount of data, it can take a few seconds or several minutes to create your data set.
- Download the file when it is available.

Query 8941717, for crsp_a_stock

Status:	Success	Rerun				
Product:	crsp_a_stock					
Results:	<table><tr><td>Result Size: 10.6 MiB</td><td>Result Count: 522,899 Rows</td></tr><tr><td colspan="2">Output Files: Download .zip Output Results are also in your home directory at: ~/web_query_output/. See WRDS Cloud: Access Web Query Output for more details.</td></tr></table>		Result Size: 10.6 MiB	Result Count: 522,899 Rows	Output Files: Download .zip Output Results are also in your home directory at: ~/web_query_output/ . See WRDS Cloud: Access Web Query Output for more details.	
Result Size: 10.6 MiB	Result Count: 522,899 Rows					
Output Files: Download .zip Output Results are also in your home directory at: ~/web_query_output/ . See WRDS Cloud: Access Web Query Output for more details.						
Timing:	<table><tr><td>Elapsed Time: 18 seconds</td><td>Submitted: 2024-09-28 10:21</td></tr></table>		Elapsed Time: 18 seconds	Submitted: 2024-09-28 10:21		
Elapsed Time: 18 seconds	Submitted: 2024-09-28 10:21					



Agenda

1. Overview
2. Key variables
 - a. Security identifiers
 - b. Security characteristics
 - c. Return, price, and volume data
3. Data selection and download
4. Conclusion

Conclusion

CRSP stock market data

- High quality U.S. stock market data.
- Standard for U.S. empirical stock market research.
- Easy and convenient to use.
- Be aware of pitfalls like “negative” stock prices.
- Further information: CRSP data manual.

- For U.S. stock market data, please do NOT use
 - Refinitiv Workspace/Datastream because of data quality (see [Ince and Porter \(2006\)](#)).
 - Bloomberg because of data download restrictions.

References

- Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 43(2), 153-193.
- Ince, O. S., & Porter, R. B. (2006). Individual equity return data from Thomson Datastream: Handle with care!. *Journal of Financial Research*, 29(4), 463-479.
- Shumway, T. (1997). The delisting bias in CRSP data. *The Journal of Finance*, 52(1), 327-340.
- Shumway, T., & Warther, V. A. (1999). The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *The Journal of Finance*, 54(6), 2361-2379.

Thank you very much for watching!

Questions and feedback are very welcome!

Contact us at datacenter@safe-frankfurt.de.