



Search WRDS

Get DataAnalyticsClassroomResearchSupport

[Home](#) / [Support](#) / [Manuals and Overviews](#) / [CRSP](#) / [CRSP/Compustat Merged \(CCM\)](#) / [WRDS Overview of CRSP/COMPUSTAT Merged \(CCM\)](#)

WRDS Overview of CRSP/COMPUSTAT Merged (CCM)

The CRSP/Compustat Merged database (CCM) provides CRSP and Compustat subscribers with a convenient way to link CRSP market and corporate action data with Compustat fundamental data.

Overview

The CCM database offers a variety of benefits including:

- **Comprehensive research:** Many of the links can be easily recreated by comparing identifiers such as **CUSIP**, ticker, and company name. Through in-depth research CRSP adds to CCM additional links that are complicated due to mergers & acquisitions, multiple securities, cross-listed securities, and more.
- **Multiple identifiers:** The CCM database may be accessed through either CRSP or Compustat identifiers, (i.e. **PERMNO**, **PERMCO**, **GVKEY**).

Top of Section

Coverage

CCM includes Standard & Poor's Compustat data, reformatted into CRSP's proprietary CRSPAccess database format, along with link information. Even though CCM is a CRSP product, it does not include CRSP's stock market data. It may be added through a separate step. Please refer to the Database Notes section of this document for more on this topic.

WRDS converts some of the CCM tables from CRSPAccess format to SAS format. This conversion is only done for the CCM tables that contain link information. A full copy of the Compustat data is provided by CRSP, which WRDS does not use. The CCM web queries join the link tables with WRDS's existing Compustat data, coming directly from Standard and Poors.

Top of Section

Dataset Organization

File Locations

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

Table 1: Datafile Locations

Product	Primary SAS Libname	Alternate SAS Libname	Unix Location
CCM - Annual Update	crsp	crspa	/wrds/crsp/sasdata/a_ccm
CCM - Quarterly Update	crsp	crspq	/wrds/crsp/sasdata/q_ccm
CCM - Monthly Update	crsp	crspm	/wrds/crsp/sasdata/m_ccm

Datasets Created By WRDS

No datasets are created by WRDS for this database. WRDS makes extensive use of the Link History table, which is provided directly by CRSP. Following are descriptions of the original CCM data tables.

CCM Datasets

Link History Table (CCMXPFLNKHIST)

The link history table (CCMXPFLNKHIST) is **the main table is used for WRDS CCM web queries**. It provides **access to all Compustat records**, regardless of whether or not the securities are in the CRSP universe. All Compustat data—including indexes, Canadian records, and off-exchange ranges not directly linked to CRSP Data—can be accessed using **GVKEY**, **GVKEY.IID**, and **GVKEYX**.

The table below displays a small slice of the data taken from CCMXPFLNKHIST. The prefix "L" is added to **LIID**, **LPERMNO** and **LPERMCO** in linking tables. Despite the addition of a prefix, The variables have the same meaning as they would normally.

- **LINKPRIM** clarifies the link's relationship to Compustat's marked primary security within the related range. "P" indicates a primary link marker, as identified by Compustat in monthly security data. "C" indicates a primary link marker, as identified by CRSP to resolve ranges of overlapping or missing primary markers from Compustat in order to produce one primary security throughout the company history. "J" indicates a joiner secondary issue of a company, identified by Compustat in monthly security data.
- **LIID** is the Compustat security identifier. When company data exists but no Compustat security can be properly identified within a range, it is assigned an "X" suffix.

- **LINKTYPE** is the link type code describing the connection between CRSP and Compustat data. Special attention should be paid to this variable to avoid double counting caused by multiple links.
- **LPERMNO** is missing if no link exists for a given **GVKEY**. It is up to the user to decide on how to handle overlaps and soft links.
- **LINKDT** is a calendar date marking the first effective date of the current link. If the link was valid before CRSP's earliest record, **LINKDT** is set to be SAS missing code ".B".
- **LINKENDDT** is the last effective date of the link record. It uses the SAS missing code ".E" if a link is still valid.

In addition to its availability in the SAS data format, the link history table can be also obtained using the `ccm_print` command with option `/li`.

Table 2: Slice of the CCMXPF_LNKHIST

GVKEY	LINKPRIM	LIID	LINKTYPE	LPERMNO	LPERMCO	LINKDT	LINKENDDT
COMPUSTAT global company key	Primary link marker	Security- level identifier	Link type code	Historical CRSP PERMNO link to COMPUSTAT record	Historical CRSP PERMCO link to COMPUSTAT record	First effective date of link	Last effective date of link
010411	P	01	LC	63773	5230	19811215	.E
010411	P	01	NR			19741129	19811214
010411	J	07	LC	90655	5230	20050516	20120131
010411	J	06	NR			20050429	20060131
010411	J	02	NR			19940331	.E
010411	J	04	NR			20020131	20060131

Link Used Table (CCMXPF_LNKUSED)

The link used table (CCMXPF_LNKUSED) has **all records of GVKEYs with a match to a PERMNO identifier**, even those not used. This table is **CRSP-centric** (i.e. starting with **PERMNO**), compared to the link history table which is Compustat-centric (i.e. starting with **GVKEY**). This table allows users searching by **PERMNO** to find all possible **GVKEYs** that have ever been linked, including those not used during a certain period.

Within this table, the item **USEDFLAG** marks whether the link is used in building the composite record. Please refer to the Database Notes section for further information.

In addition to the SAS data format, the link used history can be also obtained using the `ccm_print` command with the `/lu` option.

Link Range Table (CCMXPF_LNKRNG)

CRSP generates the link range table with information on the fiscal periods associated with each used link for each time series calendar frequency and keyset. This shows ranges in each of the fiscal and calendar calendars available in the CCM.

In addition to the SAS data format, CCMXPF_LNKRNG can be also obtained with the `ccm_print` command, using the `/lr` option.

Compustat Company Header History (COMPHIST)

COMPHIST contains **historical Compustat header information** (e.g. SIC code, CIK number, company name) as well as the effective date range of its historical description. Whenever a company has any descriptive information change, a new record will be added to the table.

In addition to the SAS data format, COMPHIST can be obtained with the `ccm_print` command, using the `/ch` option.

Compustat Company Header Information (COMPHEAD)

Compustat Company Header Information table (COMPHEAD) has **the latest Compustat header information**. It contains one entry per Compustat company. In addition to the SAS data format, the COMPHEAD table can be obtained with the `ccm_print` command and `/co` option.

Company Master Structure (COMPMASTER)

The master structure contains CCM Company identification and range data.

In addition to the SAS data format, COMPMASTER can be also obtained from the `ccm_print` command with `/ma` option.

Top of Section

Database Notes

Ambiguity

The CCM database includes very detailed coverage of the links between companies in the CRSP and Compustat databases. The level of detail can be both useful and confusing. Below are three common sources of confusion.

Differences in Coverage. CRSP covers stock prices on the four most popular US stock exchanges. This is a smaller set of companies than can be found in Compustat, which covers all public and some private companies. CRSP has a longer history of prices, and there are times when a company's stock may stop trading - causing gaps in the linkage.

Surviving Company Discrepancies. When companies merge, there may be discrepancies between CRSP and Compustat over which is the surviving company. In these cases, a researcher could disagree with the choice in the CCM dataset, and may need to resolve the disagreement or ambiguity about a which is the surviving company.

Multiple Identifiers Assigned. CRSP's primary identifier relates to securities while Compustat's relates to companies. Since companies can have more than one equity security, a single **GVKEY** can be link to multiple **PERMNO**. The Link Type and Used Flag fields attempt to address this issue.

Link Boundaries

The link between two companies might end during a fiscal year. Researchers typically choose one of the following 3 rules to determine whether to include a fiscal period as a valid match (from most strict to least):

- Only if the entire period is within the link date range.
- If the period end date is within the link date range.
- If any part of the period is within the link date range.

Link Type Classifications

Primary Link Types

We often refer to the links with link types—which account for 41% of the links in CCM—of "LC", "LU" or "LS" as primary.

- **LC:** Link research complete (after extensive research by CRSP). Standard connection between databases.
- **LU:** Link is unresearched by CRSP. It is established by comparing the Compustat and historical CRSP **CUSIPs**. LU represents the most popular link type.
- **LS:** Link valid for this security only. Other CRSP **PERMNOs** with the same **PERMCO** will link to other **GVKEYs**. LS links mainly relate to ETFs where a single CRSP **PERMCO** links to multiple Compustat **GVKEYs**. In Compustat, even though they may belong to the same investment company (e.g. ISHARES), ETFs are presented with different **GVKEYs** and CRSP flags this situation.

Secondary Link Types

Secondary link types currently account for 2% of the links. Secondary links bring in unnecessary duplicates or matches that are not completely verified. Instead, using primary link types is highly recommended.

Secondary link types are classified as follows:

- **LX:** Link to a security that trades on foreign exchange not included in CRSP data. For instance, a Canadian firm with **GVKEY 001186** and **CUSIP 008474108** may trade on both the New York (NYSE) and Toronto (TSX) stock exchanges. Compustat contains price data for company **CUSIP 008474108** with NYSE **IID 01** and TSX **IID 01C**, while CRSP only collects information of **CUSIP 008474108** that trades on the NYSE. The linking table will reflect both matches, with the one trading on the NYSE as **LINKPRIM=P** and **LINKTYPE=LC**, and the other trading on TSX as **LINKPRIM=N** and **LINKTYPE=LX**.
- **LD:** Duplicate link to a security. Two **GVKEYs** map to a single **PERMNO (PERMCO)** during the same period, and this link should not be used. Almost all of these cases happened before 1990.
- **LN:** Primary link exists but Compustat does not have prices. Prices are used to check the accuracy of the link. For linktype LN there is no price information available even on a quarterly or annual basis. The user will have to decide whether or not to include these links.

Non-matching Link Types

57% of the links within CCM do not have matches. This is to be expected because of the different coverage of the two databases.

Unavailable link types are classified as follows:

- **NR:** No link available; confirmed by research.
- **NU:** No link available; not yet confirmed.

Top of Section

USEDFLAG

The **USEDFLAG** item notes if a link was **used** to to build a composite GVKEY record corresponding to the PERMNO. It originates from the CCMXPF_LNKUSED table and is therefore CRSP-centric. It may be helpful in cases of multiple securities to establish a connection between all securities involved, even those not directly linked.

For example, a company called "TELEPHONE & DATA SYSTEMS INC" (**PERMCO=5230**) has two equity issues traded on stock exchanges with overlapping time periods. These are **PERMNO=63773** (1981/12/31 to present; dates taken from CRSP monthly file with non-missing price information) and **PERMNO=90655** (2005/05/31 to 2012/01/31). The information is extracted from the CCMXPF_LNKUSED table, appearing with the prefix "U" for renaming purposes.

For the period 1981/12/15 to present (mid-month dates may originate from the CRSP names file), one of the observations in Table 3 below indirectly links to **PERMNO=90655**, which has a **USEDFLAG** of -1. Similarly, for the period 2005/05/16 to 2012/01/31, one of the two observations indirectly links to **PERMNO=63773**, as indicated by **USEDFLAG** being set to -1.

The fifth & sixth observations in this example indicate that Compustat contains price information for issue **IID=06 (CUSIP=879433852;** cross referenced using Compustat's monthly price file), but CRSP does not have a corresponding **PERMNO** or **PERMCO** for this issue. However those securities can be vaguely related with **APERMNO=63773, 90655**. In this scenario, **USEDFLAG** is set to -1. For **IID=02** and **IID=04**, a similar reason holds and flagged as "-1".

Unless you are looking for all possible cases that have ever linked, it is best to avoid links marked as **USEDFLAG=-1**.

Table 3: USEDFLAG Example

UGVKEY	ULINKPRIM	UIID	ULINKTYPE	APERMNO	ULINKID	UPERMNO	UPERMCO	USEDFLAG	LINKDT
--------	-----------	------	-----------	---------	---------	---------	---------	----------	--------

UGVKEY	ULINKPRIM	UIID	ULINKTYPE	APERMNO	ULINKID	UPERMNO	UPERMCO	USEDFLAG	LINKDT
COMPUSTAT global company key	Primary link marker	Security- level identifier	Link type code	CRSP PERMNO	Unique ID per link associated with PERMNO	Historical CRSP PERMNO link to COMPUSTAT record	Historical CRSP PERMCO link to COMPUSTAT record	Flag marking whether link is used in building composite record	First effective date of link
010411	P	01	LC	63773	1	63773	5230	1	19811215
010411	P	01	LC	90655	1	63773	5230	-1	19811215
010411	J	07	LC	63773	5	90655	5230	-1	20050516
010411	J	07	LC	90655	5	90655	5230	1	20050516
010411	J	06	NR	63773	4			-1	20050429
010411	J	06	NR	90655	4			-1	20050429
010411	P	01	NR	63773	0			-1	19741129
010411	P	01	NR	90655	0			-1	19741129
010411	J	02	NR	63773	2			-1	19940331
010411	J	02	NR	90655	2			-1	19940331
010411	J	04	NR	63773	3			-1	20020131
010411	J	04	NR	90655	3			-1	20020131

Top of Section

Linking Without CCM

While performing the link between Compustat's accounting data and CRSP's pricing data, there are several important issues to consider. Compustat and CRSP have different universes, with slightly different coverage.

	CRSP	COMPUSTAT
Coverage	Stock market data from major stock exchanges (NYSE, AMEX, NASDAQ)	Accounting data for public, OTC, and private companies
Primary Identifiers	PERMNO, PERMCO	GVKEY
Secondary Identifiers	Header & Historical 8-digit CUSIP	Header 9-digit CUSIP
Additional Identifiers	Ticker Symbol, Company Name (Header & Historical)	Ticker Symbol, Company Name (Header only)
Securities	A firm can have multiple securities	
Inclusion Policy	All data after IPO	Market Capitalization thresholds are \$100M, with a threshold of \$25M for client requests. This represents the average market cap using the price-close monthly for the previous three months. Additionally, the security must have trading volume over the last three months.
Specialities		Red-Herring Companies
Historical Coverage	1925–present	1950–present

Merging by CUSIP

WRDS recommends starting with the SAS sample program [merge_funda_crsp_bycusip.sas](#) to merge CRSP and Compustat Xpressfeed databases by **CUSIP**.

Merging with Stock Data

While performing the link between Compustat's accounting data and CRSP's pricing data, there are several important issues to consider:

- **Different calendar/fiscal regime:** Compustat data needs to be converted from fiscal to calendar regime. When merging both sets, it is important to ask the question, "When did the market know about the accounting information?".
- **Different frequencies:** Compustat data is annual or quarterly, while CRSP data is monthly or daily.

Top of Section

Top

Table of Contents

- » [Overview](#)
- » [Coverage](#)
- » [Dataset Organization](#)
- » [Database Notes](#)
- » [USEDFLAG](#)
- » [Linking Without CCM](#)



Wharton Research Data Services

- About WRDS

WRDS FAQs

WRDS News

Account Preferences

Info / Support Request

Privacy Policy
- 3 Ways to use WRDS

WRDS Account Types

Terms of Use

Sample Data

Conference Calendar

Impactful Research

Unless otherwise noted, all material is © 1993-2020, The Wharton School, University of Pennsylvania. All rights reserved.