

# How to Deal With Some Commonly Encountered Financial Data Research Problems



**October 6, 2016**

**Todd Hines**  
**Assistant Economics, Finance and Data Librarian**  
**Princeton University**

# Finance Research



- **Most finance research papers do not document how they generated their dataset**
  - “We used Compustat and CRSP. . .”
- **This talk covers some common areas where it can be helpful to have a deeper understanding of the sources.**
  - **Theme:** Often not adequate to merely download the data and start using it. Researcher should spend some time understanding the data too.

# Finance Research



- **Finance database areas covered**
  - Header data
  - Linking data from different finance databases using company ids
  - Industry code data
  - Non-US fundamentals – issues
  - Building a corporate bond dataset
    - ✦ **Might not have enough time for this. Still wanted to include because it's important.**

# Header Data



- Not uncommon to see papers that look at headquarters and use Compustat or the non-CD-ROM version of Worldscope as their source.
- Problem
  - These databases use header data for that field

# Header Data (cont.)



- **What is header data?**
  - Historical values not retained
  - Only current value is available
  - In the past, storage space was expensive so this system was implemented to save storage space.
    - ✦ Storage space is now cheap, but this kind of data lives on.
  - Problem
    - ✦ Can provide very inaccurate historical picture
    - ✦ I'll show a headquarters example below.
  - When will I encounter header data? Most common for company id variables.
    - ✦ Name
    - ✦ Ticker
    - ✦ Industry code
    - ✦ Headquarters (address), etc.

# Header Data (cont.)



- **How to identify if a variable has header data**
  - Check database documentation (data manual)
  - Run an example of several companies where the variable chosen should have changed
    - ✦ **Headquarters example: Boeing changed their headquarters from Seattle to Chicago in the early 2000s. Compustat's data only shows Chicago.**

# Boeing Headquarters – from Compustat



Search results - G x Participants - Be x btn\_presentation x Compustat Mon x Compustat Fund x boeing ticker - G x Boeing Internatio x

https://wrds-sol2.wharton.upenn.edu/output/94502429a0eca1a2.htm

**Compustat Fundamentals Annual**

Global Company Key	Data Date	Data Year - Fiscal	Ticker Symbol	City	State Province	Company Name	Level of Consolidation - Company Annual Descriptor	Industry Format	Data Format	Population Source	ISO Currency Code	Active Inactive Status Marker
002285	19961231	1996	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	19971231	1997	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	19981231	1998	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	19991231	1999	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20001231	2000	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20011231	2001	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20021231	2002	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20031231	2003	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20041231	2004	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20051231	2005	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20061231	2006	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20071231	2007	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20081231	2008	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20091231	2009	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20101231	2010	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20111231	2011	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20121231	2012	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20131231	2013	BA	Chicago	IL	BOEING CO	C	INDL	STD	D	USD	A
002285	20141231	2014	BA	Chicago	IL	BOEING	C	INDL	STD	D	USD	A

# Header Data (cont.)



- **Solutions (cont.)**

- Try using a database that does not have header data for the variable you need
  - ✦ **Example – headquarters**
    - SEC EDGAR documents
    - Worldscope on CD-ROM
- Some databases have header and historical data for the same variables
  - ✦ **Examples – SIC codes in CRSP, SIC & NAICS codes in Compustat**



# Linking Finance Data Sets



- For more complete coverage see my co-presentation from the 2014 Beyond the Numbers conference: “Linking Financial Data Sets: Possible Problems and Solutions”
  - This presentation will only cover the highlights

# Linking Finance Data Sets



- **Very common occurrence in finance research**
  - Researcher is using more than one database/source. The different sources use different company identifiers.
- **This can cause a host of problems**
  - IDs can change over time. Examples of name and ticker changes.
    - ✦ Companies can change names, change tickers
      - Ticker “S” over the last twenty years
        - Part of the time was Sears’s ticker, part of the time was Sprint’s
      - Philip Morris
        - Philip Morris International (current stock ticker “PM”)
        - Successor company is actually Altria (ticker “MO”)
      - Kmart bought Sears. The merged company was called Sears, even though Sears was the acquired company.

# Example of a Company with Name and ID Changes



## CRSP Names File

PERMNO	DATE	NAMEENDT	COMNAM	NCUSIP	TICKER
22840	19460710	19540223	CONSOLIDATED GROCERS CORP		
22840	19540224	19620701	CONSOLIDATED FOODS CORP		
22840	19620702	19680101	CONSOLIDATED FOODS CORP		CFD
22840	19680102	19850401	CONSOLIDATED FOODS CORP	20921910	CFD
22840	19850402	20010823	SARA LEE CORP	80311110	SLE
22840	20010824	20020101	SARA LEE CORP	80311110	SLE
22840	20020102	20040609	SARA LEE CORP	80311110	SLE
22840	20040610	20110207	SARA LEE CORP	80311110	SLE
22840	20110208	20120628	SARA LEE CORP	80311110	SLE
22840	20120629	20131231	HILLSHIRE BRANDS CO	43258910	HSB

# Linking Finance Data Sets



- **Company IDs can be header data**
  - ✦ **In Compustat, the company name, ticker & CUSIP are all header data**
    - They don't save the historical values

# Compustat Record

<u>GVKEY</u>	<u>FIRST DATE</u>	<u>LAST DATE</u>	<u>COMPANY NAME</u>	<u>TICKER</u>	<u>CUSIP FULL</u>
9411	1/1/1950	12/31/2013	HILLSHIRE BRANDS CO	HSH	432589109

# Linking Finance Data Sets



- Another issue: databases often use different company ids: CUSIP, ticker, etc.
  - ✦ **Solutions**
    - **Can get complex**
      - Need to have an understanding of all the databases being used
      - Know whether data are historical or header data
    - ✦ **My 2014 presentation shows how to link M&A data (from SDC) to stock data (from CRSP) to fundamentals data (from Compustat)**

# Industry Code Data



- **Finance research using industry codes**
  - SIC (Standard Industrial Classification)
  - Most heavily used in finance research (still)
  - Hasn't been updated since 1987
  - Not as many industries as other industry code classifications (NAICS, GICS, etc.)
- **Some issues when using industry code data**
  - Different database providers use different sources for industry data
    - ✦ See Melinda Stasch's presentation from the 2014 conference: "Vendors' Methodologies for Assigning Industry Codes"
    - ✦ Consequence – the same company can have different industry codes in different databases
      - Zynga: CRSP (SIC code 9999). Compustat (SIC code 7372 – Prepackaged software)

# Industry Code Data (cont.)



- **Some issues when using industry code data (cont.)**
  - **Can be header data**
    - ✦ CRSP has both a header SIC variable (HSICCD) and an historical SIC variable (SICH)
    - ✦ Compustat has both header SIC & NAICS and the historical values (SICH and NAICSH).



## Weyerhaeuser entry in CRSP

PERMNO	Date of Observation.	Names Ending Date	Company Name	Exchange Ticker Symbol	Header SIC Code	SIC Code
39917	19631206	19680101	WEYERHAEUSER CO	WY	2499	2631
39917	19680102	20010228	WEYERHAEUSER CO	WY	2499	2631
39917	20010301	20020101	WEYERHAEUSER CO	WY	2499	2621
39917	20020102	20040208	WEYERHAEUSER CO	WY	2499	2621
39917	20040209	20040609	WEYERHAEUSER CO	WY	2499	2421
39917	20040610	20050808	WEYERHAEUSER CO	WY	2499	2421
39917	20050809	20100103	WEYERHAEUSER CO	WY	2499	2411
39917	20100104	20110612	WEYERHAEUSER CO	WY	2499	2411
39917	20110613	20140108	WEYERHAEUSER CO	WY	2499	6798
39917	20140109	20151231	WEYERHAEUSER CO	WY	2499	2499

# Non-US Fundamentals



- **Historically most finance research has centered on the United States. Compustat North America is the most heavily used resource**
- **More and more interest in non-US companies**
- **Topics covered**
  - Primary research sources
  - Length of backfile
  - Problem with Worldscope's 1980 backfile
  - Permanent id numbers – not as prevalent for non-US companies

# Non-US Fundamentals (cont.)



- **Most heavily used non-US fundamentals databases. All available in WRDS.**
  - **Worldscope – most popular**
    - ✦ **Factset version and Thomson Worldscope version in WRDS**
  - **Compustat Global**
  - **Osiris (from Bureau van Dijk)**
- **WRDS produced a research paper comparing the three**
  - **International Accounting Databases on WRDS: Comparative Analysis.**  
Rui Dai. WRDS Working Paper (2011). <http://wrds-web.wharton.upenn.edu/wrds/research/applications/guides/pdf/March%2016%202012%20international.pdf>
- **Data don't go back nearly as far as for the US**
  - **Generally back to the 80s and 90s depending on the database and the country. For lesser developed countries, coverage may start later than that.**
    - ✦ **Subscription databases don't tend to add data until there is a market demand. Until their customers will pay for it.**

# Non-US Fundamentals (cont.)



- **Worldscope**

- **Oldest**

- ✦ Started around 1986/87
- ✦ For many years, until 2005, was available as a CD-ROM product, so it was widely held

- **Problem with the backfile**

- ✦ Data can go back to 1980
- ✦ Problem: only firms that were around in 1986/87 were taken back to 1980
- ✦ **Survivorship bias!!!!**

# Non-US Fundamentals (cont.)



- **Company IDs**
  - All three databases only have header data for many IDs.
    - ✦ Yuck.
  - Compustat Global uses a permanent id – GVKEY. Osiris has BvDEP ID.
  - Most commonly used IDs
    - ✦ ISIN
    - ✦ SEDOL
    - ✦ CUSIP – not as heavily used for Non-US finance research
    - ✦ Ticker – very bad id. Dealing with multiple exchanges all over the world.

# Non-US Fundamentals (cont.)



- **File that shows historical ISINs and CUSIPs for non-US companies**
  - **Capital IQ Identifiers file in WRDS**
    - ✦ “Complete” coverage only goes back to around the late 90s
    - ✦ Provides historical ISINs and CUSIPs associated with a company. Doesn’t have historical SEDOLs.
    - ✦ Very useful for keeping track of non-US firms and their IDs over time.

# Corporate Bond Research



- Historically it has been harder to get corporate bond data to use for a research paper
  - Bond trading is over-the-counter
    - ✦ In US, data for most markets trades only since around 2006 (TRACE database)
  - Companies generally only have one primary stock
    - ✦ Can have many bonds out at one time
- Increasingly, the tools are available to most researchers to build bond datasets

# Corporate Bond Research (cont.)



- **Some important points**
  - As mentioned, companies can have many different bonds out at the same time



# United Technologies Corp. Bonds



Capital IQ - Identifiers x Watchlist x United Technologies To x united technologies ut x

finra-markets.morningstar.com/BondCenter/WatchList.jsp

Investors > Market Data > Watchlist

Last Updated: 10/03/2016

## Watchlist

Add Symbol or Bond CUSIP to Watchlist:

### Bonds

Symbol	CUSIP	Type	IssuerName	Coupon	Maturity	Moody's	S&P	Fitch	Last	Chg	Chg%	Yield	Date
UTX GT	913017BK4	Corporate Bond	UNITED TECHNOLOGIES CORP	6.050	08/01/2038	A3	A-	A-	134.72	1.011	0.76	3.575	09/30/2016 X
UTX GW	913017BP3	Corporate Bond	UNITED TECHNOLOGIES CORP	6.125	07/15/2038	A3	A-	A-	137.27	0.772	0.57	3.634	09/27/2016 X
UTX GZ	913017BS7	Corporate Bond	UNITED TECHNOLOGIES CORP	5.700	04/15/2040	A3	A-	A-	132.64	-0.162	-0.12	3.626	09/29/2016 X
UTX HA	913017BR9	Corporate Bond	UNITED TECHNOLOGIES CORP	4.500	04/15/2020	A3	A-	A-	110.81	-0.040	-0.04	1.350	10/03/2016 X
UTX3860 971	913017BU2	Corporate Bond	UNITED TECHNOLOGIES CORP	1.800	06/01/2017	A3	A-	A-	100.49	0.213	0.21	1.047	09/30/2016 X
UTX3860 973	913017BV0	Corporate Bond	UNITED TECHNOLOGIES CORP	3.100	06/01/2022	A3	A-	A-	107.36	0.000	0.00	1.727	09/30/2016 X
UTX3860 975	913017BT5	Corporate Bond	UNITED TECHNOLOGIES CORP	4.500	06/01/2042	A3	A-	A-	116.00	0.020	0.02	3.545	09/30/2016 X
UTX4240 120	913017CA5	Corporate Bond	UNITED TECHNOLOGIES CORP	4.150	05/15/2045	A3	A-	A-	111.36	-0.231	-0.21	3.511	10/03/2016 X
UTX4240 122	913017BZ1	Corporate Bond	UNITED TECHNOLOGIES CORP	1.776	05/04/2018	Baa1	BBB+	BBB+	100.55	-0.293	-0.29	1.428	09/30/2016 X

### Equities&Funds

List - Group

Symbol	Name	Last	Chg	Chg%	Vol (mil)
--------	------	------	-----	------	-----------

© 2016 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "Moody's"). All rights reserved. Moody's ratings and other information ("Moody's information") are proprietary to Moody's and/or its licensors and are protected by copyright and other intellectual property laws. Moody's information is licensed to Distributor by Moody's. Click moody's.com for complete legal terms and conditions governing use of Moody's information.

POWERED BY MORNINGSTAR

Quotes delayed at least 15 minutes. Market and fundamental data provided by Morningstar, Inc. Copyright © 2016 Morningstar, Inc. All Right Reserved. The information contained herein: (1) is proprietary to Morningstar and/or its content providers, (2) may not be copied or distributed, and (3) is not warranted to be accurate, complete or timely.

# Corporate Bond Research (cont.)



- **Some important points (cont.)**
  - Bonds may have been acquired in a merger
    - ✦ Example – United Technologies Corp bought BF Goodrich in 2011. BF Goodrich had bonds outstanding which were “taken over” by UTC.
      - Because those bonds were originally issued by BF Goodrich, they have BF Goodrich’s name on the bonds

# BF Goodrich Bonds Now “Owned” by United Technologies Corp



Capital IQ - Identifiers x Watchlist x United Technologies To x united technologies ut x

finra-markets.morningstar.com/BondCenter/WatchList.jsp

Bonds

Mutual Funds

Watchlist

Feedback

FAQ

Subscribe to Notices

**Watchlist**

Add Symbol or Bond CUSIP to Watchlist:  **ADD**

**Bonds**

Symbol	CUSIP	Type	IssuerName	Coupon	Maturity	Moody's	S&P	Fitch	Last	Chg	Chg%	Yield	Date
UTX GT	913017BK4	Corporate Bond	UNITED TECHNOLOGIES CORP	6.050	06/01/2036	A3	A-	A-	134.72	1.011	0.76	3.575	09/30/2016 X
UTX GW	913017BP3	Corporate Bond	UNITED TECHNOLOGIES CORP	6.125	07/15/2038	A3	A-	A-	137.27	0.772	0.57	3.634	09/27/2016 X
UTX GZ	913017BS7	Corporate Bond	UNITED TECHNOLOGIES CORP	5.700	04/15/2040	A3	A-	A-	132.64	-0.162	-0.12	3.628	09/29/2016 X
UTX HA	913017BR9	Corporate Bond	UNITED TECHNOLOGIES CORP	4.500	04/15/2020	A3	A-	A-	110.81	-0.040	-0.04	1.350	10/03/2016 X
UTX3668 604	382388AU0	Corporate Bond	GOODRICH CORP	6.800	07/01/2036		A-	A-	134.55	0.000	0.00		02/13/2015 X
UTX3676 999	382388AW6	Corporate Bond	GOODRICH CORP	4.875	03/01/2020		A-	A-	110.86	0.000	0.00		09/16/2016 X
UTX3685 624	382388AL0	Corporate Bond	GOODRICH CORP	7.000	04/15/2038		A-	A-	135.94	0.000	0.00		08/25/2016 X
UTX3860 971	913017BU2	Corporate Bond	UNITED TECHNOLOGIES CORP	1.800	06/01/2017	A3	A-	A-	100.49	0.213	0.21	1.047	09/30/2016 X
UTX3860 973	913017BV0	Corporate Bond	UNITED TECHNOLOGIES CORP	3.100	06/01/2022	A3	A-	A-	107.36	0.000	0.00	1.727	09/30/2016 X
UTX3860 975	913017BT5	Corporate Bond	UNITED TECHNOLOGIES CORP	4.500	06/01/2042	A3	A-	A-	116.00	0.020	0.02	3.545	09/30/2016 X
UTX4240 120	913017CA5	Corporate Bond	UNITED TECHNOLOGIES CORP	4.150	05/15/2045	A3	A-	A-	111.36	-0.231	-0.21	3.511	10/03/2016 X
UTX4240 122	913017BZ1	Corporate Bond	UNITED TECHNOLOGIES CORP	1.778	05/04/2018	Baa1	BBB+	BBB+	100.55	-0.293	-0.29	1.428	09/30/2016 X

**Equities&Funds**

List - Group

Symbol	Name	Last	Chg	Chg%	Vol (mil)
--------	------	------	-----	------	-----------

© 2016 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "Moody's"). All rights reserved. Moody's ratings and other information ("Moody's Information") are proprietary to Moody's and/or its licensors and are protected by copyright and other intellectual property laws. Moody's Information is licensed to Distributor by Moody's. Click moody's.com for complete legal terms and conditions governing use of Moody's Information.

POWERED BY MORNINGSTAR

Quotes delayed at least 15 minutes. Market and fundamental data provided by Morningstar, Inc. Copyright © 2016 Morningstar, Inc. All Right Reserved. The information contained herein (1) is proprietary to Morningstar and/or its content providers; (2) may not be copied or distributed; and (3) is not warranted to be accurate, complete or timely. Neither Morningstar nor its content providers are responsible for any damages or losses arising from any use of this information. Past performance is no guarantee of future performance.

# Corporate Bond Research (cont.)



- **Some important points (cont.)**
  - Bonds can be issued by a subsidiary. Often will not have the parent company's name on the bond
    - ✦ **Example – Dayton Power & Light (a local utility) is owned by AES Corp (an S&P 500 constituent).**

# Dayton Power & Light Bonds – subsidiary of AES Corp



Capital IQ - Identifiers x Watchlist x

finra-markets.morningstar.com/BondCenter/WatchList.jsp

Market Data Investors > Market Data > Watchlist

Company Information Last Updated: 10/03/2016

Equities and Options

Bonds

Mutual Funds

**Watchlist**

Add Symbol or Bond CUSIP to Watchlist:  **ADD**

**Bonds**

Symbol	CUSIP	Type	IssuerName	Coupon	Maturity	Moody's	S&P	Fitch	Last	Chg	Chg%	Yield	Date
AES IR	00130HBN4	Corporate Bond	AES CORP	8.000	08/01/2020	Ba3	BB	BB	117.34	0.000	0.00	2.962	09/29/2016 X
AES3718 106	240019BC2	Corporate Bond	Dayton Power & Light	8.150	01/15/2026								X
AES3718 107	240019BF5	Corporate Bond	Dayton Power & Light	7.875	02/15/2024								X
AES3718 108	240019BJ7	Corporate Bond	Dayton Power & Light	7.875	02/15/2024								X
AES3874 654	00130HBS3	Corporate Bond	AES CORP	7.375	07/01/2021	Ba3	BB	BB	115.25	-0.250	-0.22		10/03/2016 X
AES4104 361	00130HBU8	Corporate Bond	AES CORP	5.500	03/15/2024	Ba3	BB	BB	104.56	0.562	0.54		09/30/2016 X
AES4126 894	00130HBV6	Corporate Bond	AES CORP	3.842	06/01/2019	Ba3	BB	BB	99.813	0.000	0.00		09/27/2016 X
AES4139 750	240019BS7	Corporate Bond	Dayton Power & Light	1.875	09/15/2016	WR	BBB-	NR	100.05	0.000	0.00	1.002	08/22/2016 X
AES4229 312	00130HBW4	Corporate Bond	AES CORP	5.500	04/15/2025	Ba3	BB	BB	103.02	-0.060	-0.06	4.952	10/03/2016 X
AES4363 141	00130HBX2	Corporate Bond	AES CORP	6.000	05/15/2026	Ba3	BB	BB	105.25	0.750	0.72	5.158	09/22/2016 X

**Equities&Funds**

List - Group

Symbol	Name	Last	Chg	Chg%	Vol (mil)
--------	------	------	-----	------	-----------

© 2016 Moody's Corporation, Moody's Investors Service, Inc., Moody's Analytics, Inc. and/or their licensors and affiliates (collectively, "Moody's"). All rights reserved. Moody's ratings and other information ("Moody's Information") are proprietary to Moody's and/or its licensors and are protected by copyright and other intellectual property laws. Moody's Information is licensed to Distributor by Moody's. Click moody's.com for complete legal terms and conditions governing use of Moody's Information.

POWERED BY MORNINGSTAR

Quotes delayed at least 15 minutes. Market and fundamental data provided by Morningstar, Inc. Copyright © 2016 Morningstar, Inc. All Right Reserved. The information contained herein: (1) is proprietary to Morningstar and/or its content providers; (2) may not be copied or distributed; and (3) is not warranted to be accurate, complete or timely. Neither Morningstar nor its content providers are responsible for any damages or losses arising from any use of this information. Past performance is no guarantee of future results.

INDUSTRY PROFESSIONALS INVESTORS ABOUT FINRA