

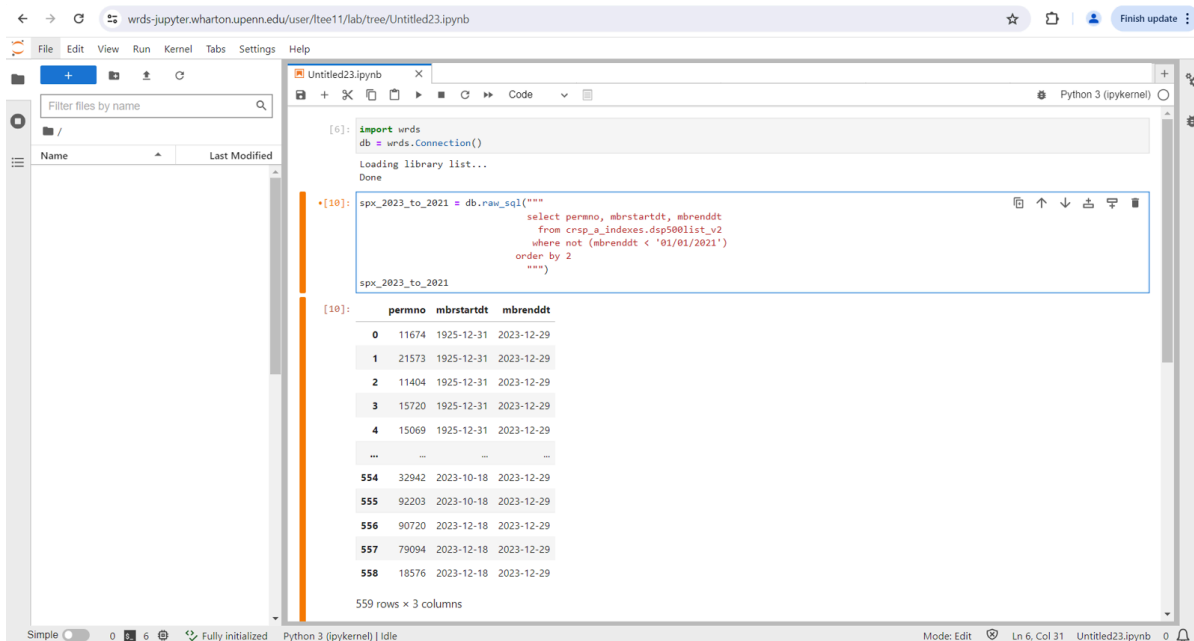
Notes and Thoughts on Retrieving Historical Members of the S&P 500 from WRDS

Tee Lip Hwe

Senior Librarian, SMU Libraries, Singapore Management University, lhtee@smu.edu.sg

Introduction

Retrieving all historical members or constituents of the S&P 500 can be done using Python coding on the CRSP dataset through an API connection with WRDS (Wharton Research Data Services (WRDS), n.d. -f; WRDS, n.d. -e).



```
[6]: import wrds
db = wrds.Connection()
Loading library list...
Done

[10]: spx_2023_to_2021 = db.raw_sql("""
select permno, mbrstartdt, mbrenddt
from crsp_a_indexes.dsp500list_v2
where not (mbrenddt < '01/01/2021')
order by 2
""")
spx_2023_to_2021

[10]:
```

	permno	mbrstartdt	mbrenddt
0	11674	1925-12-31	2023-12-29
1	21573	1925-12-31	2023-12-29
2	11404	1925-12-31	2023-12-29
3	15720	1925-12-31	2023-12-29
4	15069	1925-12-31	2023-12-29
...
554	32942	2023-10-18	2023-12-29
555	92203	2023-10-18	2023-12-29
556	90720	2023-12-18	2023-12-29
557	79094	2023-12-18	2023-12-29
558	18576	2023-12-18	2023-12-29

559 rows x 3 columns

Figure 1. Screenshot of WRDS' JupyterHub platform with original Python code by the author for accessing historical members of S&P 500 from CRSP dataset. Author's code.

Removal of S&P Index Constituents Dataset from Compustat

Previously, the constituents of S&P 500 were retrievable via Compustat. In July 2020, the S&P Dow Jones Indices (SPDJI) constituent names data was removed from Compustat, due to SPDJI direct licensing (WRDS, n.d. -a; S&P Global Market Intelligence, n.d.). With the move, the constituents of the S&P 500, along with other S&P indices, such as the S&P 1500 Super Composite, are no longer available via Compustat.

About CRSP's S&P 500 Historical Member Listing Dataset

The historical members of S&P 500 (SPX) can be retrieved using the tables `crsp_a_indexes.dsp500list_v2` or `crsp_a_indexes.dsp500list`, both of which capture the daily membership of SPX (WRDS, n.d. -d; WRDS, n.d. -c).

The script in Figure 2 below retrieves the SPX members for 29 December 2023, which is the last trading day of 2023. The output shows there were 503 SPX members on 29 December 2023.

```
spx_2023_12_29 = db.raw_sql("""
    select count (*)
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '12/29/2023')
    --order by 3
    --limit 10
    """)
spx_2023_12_29
```

	count
0	503

Figure 2. Number of S&P 500 members on or as of 12/29/2023. Author's code.

The mnemonic variable 'mbrenddt' denotes 'Constituent Membership End Date' (WRDS, n.d. -d).

The SQL condition with its mnemonic variable "not (mbrenddt < '12/29/2023')" stands for "membership does not end on or before 28 December 2023". It therefore retrieves all members for one specific day, namely 29 December 2023 the last trading day of 2023. But why?

The above script works because the CRSP's SPX members data is updated annually in WRDS (WRDSs, n.d. -b). Hence, at the point of writing this article (in 2024 before it ends), we can retrieve the SPX members only up to the year 2023, or more specifically up to 29 December 2023 the last trading day of 2023, which is also the 'maximum allowed date' (for the year 2023) for the script execution to be meaningful or sound.

It is also for the above reason that all the stock entities that were members as of 29 December 2023 still (or at least till 29 December 2023 as we speak), would have the mnemonic variable 'mbrenddt' indicated as or defaulted to 12/29/2023, as shown in Figure 3. This practice was carried out to keep track and allow for the next members data update for the year 2024.

```
spx_2023_12_29 = db.raw_sql("""
    select mbrenddt, count(*)
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '12/29/2023')
    group by mbrenddt
    --order by 1
    """)
spx_2023_12_29
```

mbrenddt	count
0 2023-12-29	503

Figure 3. 'member end date' of all 503 members of S&P 500 on or as of 12/29/2023. Author's code.

In other words, SPX members with 'mbrenddt = 12/29/2023' could either be still active in 2024 or having their last day with SPX on 29 December 2023. Mathematically speaking, a majority of

them would still be SPX member in 2024. We would be able to verify (in CRSP itself) those members that cross over to the year 2024, when the annual update for 2024 is performed, in which case the 'mbrenddt' data point would be revised and updated accordingly based on the above-mentioned data update logic and practice.

That is to say, stock entities that remain as SPX members on the last trading day of 2024 will have their 'member end date' indicated as or defaulted to the last trading day of 2024, to make way for the next annual update for 2025. To leave a deep impression on the unique approach and rationale of this data update practice, I coined the catchphrase/slogan 'There is no tomorrow.'

Note: Considering that the data for SPX members is recorded daily (as indicated by 'dsp' in 'dsp500list_v2', which stands for 'daily stock price'), specifying non-trading days like '12/31/2023' or '12/30/2023' in the condition "not (mbrenddt < 'MM/DD/YYYY)" will result in retrieving no data. See Figure 4 below.

```
spx_2023_dec_31_or_30 = db.raw_sql("""
    select count (*)
      from crsp_a_indexes.dsp500list_v2
     where not (mbrenddt < '12/31/2023' or mbrenddt < '12/30/2023')
    --order by 3
    --limit 10
    """)
spx_2023_dec_31_or_30
```

count	
0	0

Figure 4. Number of S&P 500 members recorded on or as of 12/30/2023 or 12/31/2023 non-trading days. Author's code.

S&P 500 Member's Demographic

As shown in the output of the following script in Figure 5 which retrieves the 503 SPX members for 29 December 2023, there were 15 entities that became a member of SPX in 2023, 16 entities in 2022, 15 entities in 2021, and so on, as shown in the field 'member_start_date' or 'mbrstartdt'.

```
spx_2023_12_29 = db.raw_sql("""
    select substr(cast(mbrstartdt as VARCHAR(11)),1,4) as member_start_date, count (*)
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '12/29/2023')
    group by member_start_date
    order by 1 desc
    """)
spx_2023_12_29
```

	member_start_date	count
0	2023	15
1	2022	16
2	2021	15
3	2020	16
4	2019	21
...
58	1951	1
59	1949	1
60	1948	1
61	1944	2
62	1925	8

63 rows × 2 columns

Figure 5. Distribution of 'member start date' for all the 503 members of S&P 500 on or as of 29 December 2023. Author's code.

From Figure 5 we can identify the oldest members of SPX (as of 29 December 2023). We can see that there are 8 stock entities (out of 503) where the 'member start date' began in 1925.

Given that the earliest data available date for CRSP is 31 December 1925 (WRDS, n.d. -d), it is important to recognise that a stock entity with a 'member start date' of 31 December 1925 would be either: (i) a member of the SPX (or its predecessors) beginning 31 December 1925, or (ii) potentially a member the SPX (or its predecessors) before 31 December 1925 but the date was defaulted to 31 December 1925, or (iii) possibly having no recognisable 'member start date' and hence the date was defaulted to 31 December 1925.

From Britannica Money, the S&P 500 Index (SPX), was formerly called the Composite Index and later Standard & Poor's Composite Index. It was launched on a small scale in 1923, began tracking 90 stocks in 1926, and expanded to 500 in 1957 (Encyclopaedia Britannica, n.d.).

Retrieving All S&P 500 Members for 2023 or Multiple Years

It follows from the same data logic above that in order to retrieve all SPX members for the whole year of 2023, researchers will need "not (mbrenddt < '01/01/2023')", i.e. "membership does not end on or before 31 December 2022". This would allow the retrieval of any stock entity that has been a member of the SPX at any time in 2023, totaling 521. See Figure 6.

Academic BRASS

```

spx_2023 = db.raw_sql("""
                select mbrenddt, count (*)
                from crsp_a_indexes.dsp500list_v2
                where not (mbrenddt < '01/01/2023')
                group by mbrenddt
                order by 1
                """)
spx_2023

```

	mbrenddt	count
0	2023-01-04	1
1	2023-03-14	2
2	2023-03-17	1
3	2023-05-03	1
4	2023-06-16	1
5	2023-07-05	2
6	2023-08-24	1
7	2023-09-15	2
8	2023-10-02	2
9	2023-10-17	2
10	2023-12-15	3
11	2023-12-29	503

Figure 6. All 521 constituents of S&P 500 that were members at any time in 2023 (grouped by 'member end date'). Author's code.

One can further deduce that in order to retrieve all the constituents of SPX that were members at any time during the 3-year period between 2021 and 2023, the SQL condition would therefore be "not (mbrenddt < '01/01/2021')". This would retrieve a total of 559 constituents. See Figure 7.

```

spx_2023_to_2021 = db.raw_sql("""
                select permno, mbrstartdt, mbrenddt
                from crsp_a_indexes.dsp500list_v2
                where not (mbrenddt < '01/01/2021')
                order by 3, 2
                """)
spx_2023_to_2021

```

	permno	mbrstartdt	mbrenddt
0	75100	2000-06-21	2021-01-06
1	92239	2016-02-22	2021-01-20
2	16538	2017-01-17	2021-02-11
3	27983	1963-09-05	2021-03-19
4	30940	2008-10-02	2021-03-19
...
554	92203	2023-10-18	2023-12-29
555	32942	2023-10-18	2023-12-29
556	79094	2023-12-18	2023-12-29
557	18576	2023-12-18	2023-12-29
558	90720	2023-12-18	2023-12-29

559 rows × 3 columns

Figure 7. All 559 constituents of S&P 500 that were members at any time during the 3-year period between 2021 and 2023. Author's code.

What's Next

The discussion and data extraction above have been conducted with a focus on upholding survivorship impartiality.

With 'permno', the permanent stock-level identifier from CRSP, researchers can proceed to set their desired SPX members universe to retrieve their desired data items from CRSP, or perform data linking with other datasets such as Compustat, etc.

References

Encyclopaedia Britannica. (n.d.). *S&P 500*. Britannica Money. Retrieved April 28, 2024, from <https://www.britannica.com/money/SandP-500>

S&P Global Market Intelligence. (n.d.). *S&P Dow Jones indices constituent removal*. S&P Global Market Intelligence. Retrieved April 28, 2024, from <https://pages.marketintelligence.spglobal.com/SP-Dow-Jones-Indices-Constituent-Removal.html>

Wharton Research Data Services. (n.d. -a). *Compustat index constituents removed*. Wharton Research Data Services. Retrieved April 28, 2024, from <https://wrds-www.wharton.upenn.edu/pages/support/support-articles/compustat/north-america/compustat-index-constituents-removed/>

Wharton Research Data Services. (n.d. -b). *CRSP A indexes*. Wharton Research Data Services. Retrieved April 28, 2024, from https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_indexes/

Wharton Research Data Services. (n.d. -c). *CRSP S&P 500 list*. Wharton Research Data Services. Retrieved April 28, 2024, from https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_indexes/dsp500list/

Wharton Research Data Services. (n.d. -d). *CRSP S&P 500 list (v2)*. Wharton Research Data Services. Retrieved April 28, 2024, from https://wrds-www.wharton.upenn.edu/data-dictionary/crsp_a_indexes/dsp500list_v2/

Wharton Research Data Services. (n.d. -e). *Wharton Research Data Services (WRDS)*. Wharton Research Data Services. Retrieved April 28, 2024, from <https://wrds-www.wharton.upenn.edu/>

Wharton Research Data Services. (n.d. -f). *WRDS Jupyter platform*. Wharton Research Data Services. Retrieved April 28, 2024, from <https://wrds-jupyter.wharton.upenn.edu/>

Appendix A

Author's Code

Code Block 1: Establishing WRDS connection and accessing historical members of S&P 500 from CRSP dataset

```
import wrds
db = wrds.Connection()
spx_2023_to_2021 = db.raw_sql("""
    select permno, mbrstartdt, mbrenddt
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '01/01/2021')
    order by 2
    """)
spx_2023_to_2021
```

Code Block 2: Number of S&P 500 members on or as of 12/29/2023

```
spx_2023_12_29 = db.raw_sql("""
    select count (*)
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '12/29/2023')
    --order by 3
    --limit 10
    """)
spx_2023_12_29
```

Code Block 3: 'member end date' of all 503 members of S&P 500 on or as of 12/29/2023

```
spx_2023_12_29 = db.raw_sql("""
    select mbrenddt, count (*)
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '12/29/2023')
    group by mbrenddt
    --order by 1
    """)
spx_2023_12_29
```

Code Block 4: Number of S&P 500 members recorded on or as of 12/30/2023 or 12/31/2023 non-trading days

```
spx_2023_dec_31_or_30 = db.raw_sql("""
    select count (*)
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '12/31/2023' or mbrenddt < '12/30/2023')
    --group by 3
    --limit 10
    """)
spx_2023_dec_31_or_30
```

L. Tee

Code Block 5: Distribution of 'member start date' for all the 503 members of S&P 500 on or as of 29 December 2023

```
spx_2023_12_29 = db.raw_sql("""
    select substr(cast(mbrstartdt as VARCHAR(11)),1,4) as member_start_date,
count (*)
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '12/29/2023')
    group by member_start_date
    order by 1 desc
    """)
spx_2023_12_29
```

Code Block 6: All 521 constituents of S&P 500 that were members at any time in 2023 (grouped by 'member end date')

```
spx_2023 = db.raw_sql("""
    select mbrenddt, count (*)
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '01/01/2023')
    group by mbrenddt
    order by 1
    """)
spx_2023
```

Code Block 7: All 559 constituents of S&P 500 that were members at any time during the 3-year period between 2021 and 2023

```
spx_2023_to_2021 = db.raw_sql("""
    select permno, mbrstartdt, mbrenddt
    from crsp_a_indexes.dsp500list_v2
    where not (mbrenddt < '01/01/2021')
    order by 3, 2
    """)
spx_2023_to_2021
```