# First Round: Stocks-per Resolution

The two issues occur because the “max\_daily\_stocks\_per” variable was based on the “stocks\_per” variable that simply noted the number of stocks pulled from the raw TAQ data. The actual construction makes a variety of adjustments to it, namely forward-filling missing data and creating composite prices, which affect the actual stocks per day. The details are explained below, but in short:

* By consolidating the multi-share class stocks and accurately reporting the forward filled data, I created the post-adjustment “stocks\_per\_full” variable, resulting in the graph below.
* Existing occurrences above 500 are due to spin-off and replacement technicalities
* Existing occurrences below 500 are due to TAQ data quality or because I combined multiple tracking stocks into one composite component.
* Overall, being off by 1-2 is generally because of our adjustments with creating composite prices for multiple share classes and tracking stocks and is intended OR missing TAQ data.

Chart

AI-generated content may be incorrect.

1. Occurrences of over 500:
   1. This is not a reflection of unmerged PERMCOs, rather that the SP500 has had more than 500 components since then. There are three distinct reasons
      1. Multiple share classes
         1. Companies like Alphabet with dual share classes inflate the number of components as both together count as one “component”.
      2. Spin-off finalizations
         1. When a company completes a spinoff, if one of the companies is to-be removed it may stay in the index for an additional day or so.
      3. Company replacement finalizations
         1. When a company is being dropped, it may overlap in the index with the added company for a day or so.
   2. Example day: 11/01/2016 has 508 stocks on the day. Below are the eight causes of the +8 extra stocks.
      1. Alphabet Class A and C (Multiple share classes)
      2. Discovery Class A and C (Multiple share classes)
      3. 21st Century Fox Class A and C (Multiple share classes)
      4. News Corp Class A and B (Multiple share classes)
      5. Under Armor Class A and C (Multiple share classes)
      6. Alcoa/Arconic (Spin-off finalizations)
         1. Alcoa split into Alcoa and Arconic on 11/01/2016. Arconic remained in the index while Alcoa was to be dropped, but Alcoa remained in the index temporarily on 11/01/2016.
      7. Yum!/Yum China (Spin-off finalizations)
         1. Yum China spun off from Yum! Brands on 11/01/2016. As Yum China was ineligible for the SP500 it was to be removed, but stayed in the index temporarily for a day until the removal was finalized.
      8. HCP/Quality Care Properties (Spin-off finalizations)
         1. [QCP spun off from HCP](https://content.edgar-online.com/ExternalLink/EDGAR/0001104659-16-153366.html?dest=A16-12302_15EX2D1_HTM&hash=9b31206fe5fc911333c2638909d6ff87049e4abcab93747d734f682286a9336b&utm_source=chatgpt.com#A16-12302_15EX2D1_HTM) on 11/01/2016. It was [to be removed from the main SP500](https://press.spglobal.com/2016-10-24-Quality-Care-Properties-Set-to-Join-the-S-P-MidCap-400-Guess-to-Join-S-P-SmallCap-600?utm_source=chatgpt.com), but stayed for a day until finalized.
   3. To further the point that this is not an issue of merging on the same PERMCO, the issue of not merging the PERMCOs properly doesn’t manifest if we are looking at just one day nor if we look at a single day’s stock count. This issue only appears if we look at the total stocks over many days (not stocks per day over many days) and the amount is much higher than (500 + dual share class companies + spin-offs resulting in permco changes + stock replacements).
2. Occurrences under 500
   1. The lower amount of “stocks\_per” pre-2003 reflects the less frequent trade data in the 1994-2003 TAQ dataset (the older monthly TAQ dataset). I went through the raw TAQ trade data and confirmed this was not due to my download and due to less frequent data points. We use forward-filling to resolve this issue.
      1. Ex: 25may1995 12:20:00 has a pre-adjustment “stocks\_per” of 473. [This folder](data_quality_test_files) contains a dataset of all the missing stocks data for that datetime from TAQ between 11:00:00-13:00:00. Note that all the missing stocks indeed have no trade data between 12:16:00-12:20:59.
   2. There are some extended occurrences of 499 stocks in early years even post adjustment. There are two causes to this.
      1. First, I believe this is due to a data issue on TAQ’s end where some companies are just not found in their database.
         1. Ex: Between 6/30/1993 – 9/20/1994, McCaw Cellular Communications (CUSIP: 579468109) was in the SP500 and the Hue dataset reports that. Looking through the TAQ master file for 1994-06 though, there is no entry for McCaw or its CUSIP. Until 09/20/1994, when it was replaced, the max stocks are 499. Afterwards, it rises to 500.
         2. [Sources folder](data_quality_test_files):
            1. McCaw addition/removal – nyu\_sp500\_changes\_1976\_2000
            2. Jun1994 TAQ Master – TAQ\_master\_199406
      2. Second, there are cases where the 500 stocks include multiple tracking stocks. Since I create composites of companies with multiple tracking stocks, our count drops to 499 or 498.
         1. Ex: On 11Nov2001, both tracking stocks of Sprint FON and PCS were part of the 500 stocks. Since I combine them into FON\_CP, we technically only have 499 stocks.
         2. [Sources folder](file:///\\arc\dfs\unix\if\research-gmsm\Kroner\SP500_intraday_replication\a01_create_SP500_weights\documentation\data_quality_test_files):
            1. 11Nov2001 stocks - tracking\_stock\_example.dta

# Second Round: Missing TAQ Data

Here we review the missing TAQ data noted in the step above and confirm why every case differs from 500.

* Get daily Hue components data, match it up with TAQ data to see which companies are missing
  + Get daily components out of intraday dataset
  + Merge with Hue
  + Adjust for our changes (tracking stocks + share changes)
  + Compare how many missing stocks per day to what our permco\_per tells us
    - Make sure to adjust the expected number for composite prices

1. **Adjusting TRTH Data** 🡪 After this adjustment, we no longer had entire days missing from TAQ when comparing TAQ vs HUE data daily component data.
   * There were some days fully dropped (ex. 8-10apr1996) because TRTH SP500 data was missing on those days and we dropped rows with TRTH SP500 in the code. Additionally, we needed to adjust for early market closes reflected by missing TRTH data on days like July 3rd.
   * Missing days
     + Forward filled TRTH data instead of dropping rows
   * Early market close days
     + Removed 1pm-4pm times as the market was closed then.