* 2019-08-04
  + In progress:
    - The fundamental data is updated quarterly, e.g. on datekey/reportperiod 2019-04-25 report earnings of 2019-03-31
    - This creates extra complexity of our quarterly rebalancing as it would introduce look-forward bias when we just rebalance at quarter end – the information we based on is actually available after the quarter end
    - Also we found that some previous earning records are missing from the our data (AMZN 2018 Q3). Have to back fill the gaps:
      * Re-get all fundamental data from Quandl, join with our data
    - Daily price data is ok.
  + Done:
    - Updated price and fundamental data
    - IC functionality: corr between alpha rank vs realized return rank
  + To-do:
    - Check data: use monthly (rank, rebalance) also update mechanism of raw data
    - Improve the rebalance logic to remove the look-forward bias: analyze the distribution of earning report dates, and rebalance after those dates.
  + **Action item:**
    - **Update fundamental data** 
      * **Turns out the data always use month-end date as calendar date. And we adjusted the date to biz date to get the data. This may lead some quarter end data are missing. We changed the code and re-getting all the data now.**
      * **After save the newly refreshed data, have to think about the refresh strategy**
* **2019-09-22**
  + We found that different company has different definitions of quarter-end/fisical year and reporting dates. So it is not realistic to rank their fundamental data on the same date. We decided to take the following approach – At each rebalance date, given a list of universe (SP500), for each company, use their latest record to that date, and then do the cross-sectional ranking (Done)
  + This introduce the need of the “lineup/universe” at each rebalance date (quarter end presumably). Our current sp500 univ data has the date column which indicates the time the universe changes. Hence given a rebalance date, we also need to get the most recent updated universe **no late than** that date (Done)
  + Then we need to update the rankZscore function (Done)
  + In the file generalBackTester.R, since we’ve changed our ranked\_univ function to only one single (rebalance) date, we have to re-write the logic to adapt the new back-testing logic
    - First, we need a series of rebalance dates, we use quarter end dates (3/31,6/30,9/30,12/31) -- Done
    - For the function constructPortQtly, previously we build portfolio along the period in this function. Now since our information are all based on single rebalance date, we need to totally revamp this function – Done
* **2019-11-22**
  + Factor Case Study Framework
    - Historical IC Analysis – pb DONE
      * Need to change to monthly and fwd one month return. Calculate monthly ic and average ic and standard deviation of ic
    - Factor Portfolio Performance Analysis - DONE
  + Factor Momentum – TO DO

• 2020-07-19

* + Add the update of SP all universe to the refresh data given it also only dates back 20 years – Done
  + Add detailed monthly stock level factor zscore and fwd return, analyze timeseries correlation and see if there are sector rotation – TO DO
  + Add historical market event/ business cycle –TO DO
  + Analyze other factor historical ICs – TO DO

2020-08-02

* + Add the update of SP all universe to the refresh data given it also only dates back 20 years – Done
  + Add detailed monthly stock level factor zscore and fwd return, analyze timeseries correlation and see if there are sector rotation – Done
  + need to interpret more from the chart – zscore sector average to see what sector the signal is favoring; add growth/value historical index return and compare to the sector level ICs and SP500 whole universe ICs **– In progress, more analysis on median IC for each sector, market-cap weighted return, to see which sector is favored. More to come next week**
  + Add historical market event/ business cycle –TO DO
  + Analyze other factor historical ICs – TO DO

2020-08-30

* + need to interpret more from the chart – zscore sector average to see what sector the signal is favoring; add growth/value historical index return and compare to the sector level ICs and SP500 whole universe ICs– In progress, more analysis on median IC for each sector, market-cap weighted return, to see which sector is favored. More to come next week – Completed
  + Add historical market event/ business cycle –TO DO
  + Analyze other factor historical ICs – TO DO

2020-09-13

* + WIP, the white-paper of pb factor. Working on the z-rank section, things to do next time is
    1. Deeper dive in high rank sectors.
    2. Sectors of volatile sectors
    3. IC and L/S portfolio

2020-09-27

* **Action item – the Communication Services sector is not updated. Pending feedback from SHARADAR. Pausing for Consumer Cyclical investigation**
* **TODO – We assume a high leverage would inflate the PB ratio. Thus, we should proceed the P/EV ratio as our next factor. Since PB ratio is the only factor we have looked so far, we are not able to interpret too much information from it. We should move on other factors then come back to fill the other sectors in the report (IC, portfolio etc.)**

2020-10-25

* We agreed on other than filling the leftover last time, we should focus on setting up the research framework so it can be easily leveraged for the analysis of other factors. The above items thus still pending.
* We looked at the top-level IC time series and observed a 10-year growth rally. **TODO: We can conduct some additional research to unfold forces that make/break this trend**

2020-11-22

* We conducted some analysis on the correlation of PB IC against some macro factors including GDP growth rate and treasury term premium. No material relationship has been found. Next session we would brainstorm other potential reasons for the growth rally or moving on to the next factor which leveraging the existing analysis framework.

2021-02-21

* We extended our chart of PB IC to 2021-01-31. And re-apply our IC framework to PE. Not surprisingly, PE almost tell an identical story with PB
* In next session, we agree to move on to another dimension other than growth/value. PEV could be a good choice as it contains leverage information

2021-03-06

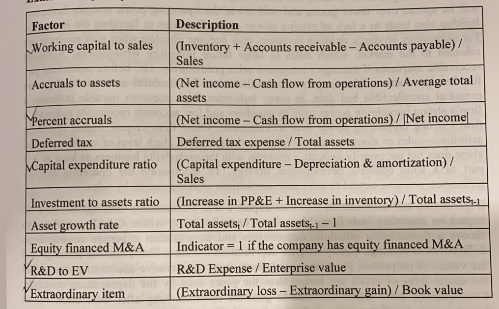
* We discussed the definition of momentum factor, agreed to start with the simpler version. Rolling return could be a good start. If not satisfied, we can move on to the EMA series. For that, we can also use the version w/o the fixed look back window. This would only require one time calculation for each stock.
* We looked at the IC chart of EV multiple (EV/EBITDA). Since it is still a valuation measure, it tells a close story as PB/PE. We observed though during 2005-2015, EV multiple IC exhibits more volatility. We think this might be related to the impact of the leverage, which could be our next topic.
* Fwd PE(data source?), quality can be studied in the future as well.
* Size is the easiest. Should be done first

2021-03-13

* TODO: Per Sharadar’s email, Upon Mar 29th, check the price table, a new col closeAdj should be added which accounts for dividend
* We looked at the factor size(Marketcap). **On average over 20 years period, small cap shows a slight premium about 0.02 IC. Though large cap has outperformed since 2017, small cap started to re-rally since the second half of 2020.**
* **Since small cap exhibits more seasonality, it might be helpful to study the timing of it. (e.g. sensitivity to the inst rate/tax)**
* NLSIZE from Barra, can be something to do if want to be fancy
* Next: quality (ROA avg+sd), mom, fwd-pe

2021-03-27

* We checked factor dividend yield. As a value factor, it basically shows the opposite as the growth factors
* For quality, we found ROA, ROE, ROIC, ROS are all only reported annually. To calculate those quarterly numbers by ourselves, ROA for example, we need to calculate the average of assets during the period, which requires looking back at each rebalance date.
* **TODO:** The above-mentioned transformation can be generalized for derived factors including MOM as well. It is a time-series function of certain columns from the fundamental table.
* **TODO:** It can be first simplified to transformations only applied to various columns at the rebalance date. This includes the following checked factors:



2021-04-03

* We looked at the newly adjusted price. The adjustment takes care of dividend, stock split, M&A etc. to ensure the total return of holding the stock can be achieved by the adjusted price. Details can be found at <https://blog.quandl.com/guide-to-stock-price-calculation>.
* To comply with the change, we need to update the way we calculated the returns using the adjusted price, for the sake of back-testing. However, when constructing any price-related factors, we should use the unadjusted price to avoid forward-looking bias.
* Action items related:
  + Re-load the past 20 yrs adjusted price. For historical data beyond 20 years which we saved at local, use the un-adjusted price as our best effort. Use closeadj from the
  + Update the return calculation function to use the adjusted close price.
  + Confirm the pe/pb ratio still using the un-adjusted price to avoid forward-looking bias
* Also, we create a new file calculateDerivedCol file to calculated derived factors based on raw columns from the fundamental table. The logic is to mutate on the fundamental\_dt\_all table and save it to another RData file. When calculating IC, just need to load the new RData file for the extra cols. Note, we still save the raw RData to update new data.
* We looked at the factor rd\_ratio = RnD/EV. Will dig into it next time

2021-04-23

* We updated Growth & Size IC, made a comparison to the return. **Had a very good discussion about the inability to use Factor (IC) to predict the future performance. As IC, has a look-forward bias by definition. More generally, any quant signal assumes the market will persist as the period when the signal is generated, which implicitly becomes a momentum factor more or less. This makes it hard to detect any turning/rotation point of the market. Also, a turning point becomes a turning point only after the turning is completed. While in the middle of it, we can not exactly tell whether it is only a temporary market correction or an official market rotation.**
* **TODO:** rd\_ratio, quality, factor correlations

2021-05-01

* We looked at rd\_ratio and capex\_ratio. Turns out rd\_ratio is highly oriented to IT and health-care, making it very close to a growth factor. On the other hand, capex\_ratio favors Cyclicals, defensives. Making more sense as a quality factor.
* TODO: ROA as quality factor

2021-05-15

* We finished the calculation of assets avg.
* TODO: Implement the function to handle missing values in asset etc. Idea is to interpolate the missing values

2021-06-13

* Note for the same reportperiod/calendardate, there may be multple records for the same company due to revision etc. E.g. AIRO at 2000-03-31. To avoid look-fwd bias, we decided to use the earliest reported observation. However, when the required data is missing at the first record, using those aviable from the later datekeys in an early to late order. After than, we then interpolate any remaining missing values

2021-06-20

* We finished the part of duplicated dates and interpolations(na.approx).
* Only extrapolation left

2021-08-08

* Extrapolation done
* General interpolation function done. Could use it for all fields.
* Need to clean up the NAs in assets derived

2021-11-07

* Figured out the cause of NA after interpolation (Duplicated datekey)
* Revenue still have NAs to calculate ROA
* Generalize the asset function to a normalization function