

FAMA-FRENCH HIGH FREQUENCY DATA DOCUMENTATION

Xinyu LIU, University of Chicago

02/17/2020

Portfolio_RCC.py

Input: In order to calculate factors, we need to first build our portfolio based on monthly (for Fama-French five factors) and daily (for Momentum) returns from CRSP and financial statement form Compustat.

Querying tool: This program access the aforementioned databases using *DataHub* developed by Yuxing. The portfolio construction part largely resembles WRDs demonstrations. The main body of this program is defined as a function *portfolio(start,end, comp)*.

Output: Each output file one year of daily return, market cap and portfolio classification information for each individual stock, and is always organized from July $year_{t-1}$ to June $year_t$. Find the time range setting session at the beginning of the code to adjust.

Path: /project2/dachxiu/hf_ff_project/Implmt_Code/Xinyu_test/outlier/Portfolio/
Current portfolios cover from 19970701 to 20190630

Listing 1: Time Range Setting

```
1 # Date range for portfolio construction
2 # Always set the startdate at least two years earlier than desired range ↔
  # to meet the construction requirement, also make the enddate slightly ↔
  # wider than what you want.
3 startdate = 20140101
4 enddate = 20191231
5
6 # The following parameters directly control the output range. For exmaple ↔
  # this setting below will form two years of portfolio:
7 # 20170701_0630_daily_all_RCC.csv
8 # and 20180701_0630_daily_all_RCC.csv
9 startyear = 2017
10 endyear = 2019
```

Save_TAQ.py

Input: This is a synthetic program that helps to generate daily level matching table between CRSP and TAQ, as well as to save daily TAQ file to the RCC server.

Querying tool: This program also access the aforementioned databases using *DataHub* developed by Yuxing. The daily matching table work is an authentic piece packaged in the function *Generating_Daily_Matching.matchingtable(date)*. The main body contains two for loops to save daily matching table and daily TAQ return files. Trading day information is obtained from the portfolio output created from the first program.

Output: Each output file is a single day record. Similar to the previous one, just find the time range setting session at the beginning of the code to adjust. Default setting will be from July $year_t$ to June $year_{t+n}$.

Path:

- Matching Table:

/project2/dachxiu/hf_ff_project/Implmt_Code/Xinyu_test/outlier/TAQmatch/

- TAQ daily Table:

/project2/dachxiu/hf_ff_project/Implmt_Code/Xinyu_test/outlier/TAQdata/

Current TAQ records cover from 19970701 to 20190630

Listing 2: Time Range Setting

```
1 # The following parameters directly control the output range. For exmaple ↔  
   this setting below will form hundreds of daily records within these ↔  
   two years of portfolio like:  
2 # taq_20180501.csv  
3 # and matchtable_20180501.csv  
4 startyear = 2017  
5 endyear = 2019
```

More on *Generating_Daily_Matching.py* : Merging CRSP daily return with TAQ can be troublesome, mostly because they use different tickers(symbols) as primary ID. Permno for CRSP is exclusive whereas symbols of companies may change. To join them together properly, the principle of matching table generating is that:

- First check if CRSP symbol exists, if so take it as primary reference to TAQ; if not, check CRSP ticker with the historical table provided by Yuxing, if they are the same for a given permno, then take this ticker as symbol. Note the none-trading firms with negative close price are also excluded at this step.

- Second screen all symbols and check suffix of 'WI', 'WD' and 'V', take off these suffix so that the mapping to TAQ will not be messed up. There is a detailed documentation I based my work

on from previous RAs.

- Third merge cleaned CRSP stock price with taq price on symbol. I have coded an optional system to record each cleaning step. This can be output by editing the textitreturn object.

Intraday_RCC_concat.py

Input: Now that we have all the inputs ready for the punch action. This program takes the portfolio, matching table and TAQ calculated from previous programs to generate factors following four main steps.

Querying tool:

- Step one imports portfolios and computes *Intraday* and *Overnight* returns.
- Step two merges portfolio with TAQ according to the daily matching table and replace the the opening and close price with CRSP daily records.
- Step three first clean misreported TAQ intraday records such as jumps and bounces, as well as BRKA case. Then it calculates the weight for later value-weight method use.
- Step four calculates intraday factors and output final results together with outlier records.

Output: From July $year_{t-1}$ to June $year_t$, outputs are created into 7 csv file due to memory limitation. Find the time range setting session at the beginning of the code to adjust.

Current portfolios cover from 19970701 to 20181231 due to the availability of TAQ intraday data.

Listing 3: Time Range Setting and Output Path

```
1 # The following parameters directly control the output range. One can also↵  
   parallel it by submitting mutiple tasks of different time range if ↵  
   calculation speed is a major concern:  
2 startyear=2007  
3 endyear=2013  
4  
5 steponeout='/project2/dachxiu/hf_ff_project/Implmt_Code/Xinyu_test/outlier↵  
   /Stepone'  
6 steptwoout='/project2/dachxiu/hf_ff_project/Implmt_Code/Xinyu_test/outlier↵  
   /Steptwo'  
7 stepthreeout='/project2/dachxiu/hf_ff_project/Implmt_Code/Xinyu_test/↵  
   outlier/Stepthree'  
8 stepfourout='/project2/dachxiu/hf_ff_project/Implmt_Code/Xinyu_test/↵  
   outlier/Stepfour'
```

Plotting with *plot_compare.py*: This is an example of comparing Python construction with Matlab and Ken-French library data. Resulting plot can be found at:

/project2/dachxiu/hf_ff_project/Implmt_Code/Xinyu_test/outlier/98-07-01_18-12-31mat_py_ff.pdf

Note: All these programs along with their sbatch codes are stored in the path mention above.
Please make sure you make your own copy and carefully adjust all relevant output paths before any execution!