## Data Fetching

Our target is to build the 6-factor model including market, Size (SMB), Value (HML), Profitability (RMW), Investment (CMA), and Momentum (UMD). We decided to use *tushare* which is an opensource financial data library to get the daily stock prices, index price and the financial report of the stocks.

Our stock universe is selected from the component stocks in CSI 300 index, and we also use CSI 300 index as the market. We filtered stocks in banks and insurance industries to focus on non-financial Sectors.

The time is from 2000-01-01 to 2024-12-31. The meta data of the stocks includes daily percentage changes using *pro.daily()*, and their financial indicators using *pro.fina\_indicator()* for each year. For each stock, in order to calculate the above factors, we fetch the indicators including market value (*t\_mv*), asset growth over the previous year(assets\_yoy), net profit to total assets (*npta*), PB value(*t\_pb*). Normally the annual financial report is announced before 31st Mar, so we update the portfolio for each factor every year on 1st April. Accordingly, we use the same strategy to get the index data by using the api *pro.index\_daily()* to get the daily returns for index 399300.SZ.

The risk-free rate is fetched using the api *pro.shibor()* to get the Shanghai Interbank Offered Rate (Shibor) daily one-year rate. The monthly rate is the average of the daily rate for each month. For the missing data, we use 3.5% as default rate.

## Data Processing

To calculate the factors, we use the following formulas:

, , , , , where is the current month. To get the monthly return, we accumulate the daily returns using the following equation:

, where is the trading days in the month $m$.

Similarly, we get the monthly returns for the index using the same strategy.

## Factor Constructing

Our data period is from 2000-01-01 to 2024-12-31 and the new financial indicators are available on each April. Thus, we build the factor portfolios at every April starting from year 2001. To make the procedure clear, we use size as an example to describe the steps for building the SMB factor each month.

1. If the month is April, we update the factor portfolio for size and rank the stocks for April according to size from small to large, then divide the stocks into small-cap (top50%) and large-cap (bottom 50%) to build the size portfolios.

2. For every month, we calculate the average monthly returns of the latest factor portfolios, namely and , as the top-portfolio monthly return and the bottom-portfolio monthly return.

3. The SMB is calculated just simply using .

For other factors, we use the same strategy but with minor adjustments, such as ranking order, top-bottom ratio, etc, separately. The following table shows the details of those parameters.

|  |  |  |
| --- | --- | --- |
| Factor | Order | Ratio |
| Size (SMB) | Ascending | 50% |
| Value (HML) | Descending | 30% |
| Profitability (RMW) | Descending | 30% |
| Investment (CMA) | Descending | 30% |
| Momentum (UMD) | Descending | 50% |

After the factor construction, for each stock at each month, we have the 6 factors including market, Size (SMB), Value (HML), Profitability (RMW), Investment (CMA), and Momentum (UMD). Finally, we generate a file includes about 272 stocks monthly data from April 2001 to Dec. 2024. In addition to the stock name and month date, each row of the file consists of the above 6 factors and the risk-free rate.