Data

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 decideted to use `tushare` which is an opensource fincial data library to get the daily stock prices, index price and the finnancial reprot of the stocks.

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

The time period 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 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:

$$size = t\_mv$$

$$value = \frac{1}{t\_pb}$$

$$profitability = npta$$

$$investment = assets\_yoy$$

$$momentum = \frac{1}{6}\sum_{j=m-6}^{m-1} r_j$$

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

$r_m = \exp(\sum_{i=1}^{m_{days}} log(1 + r_{m,i})) - 1$, m_{days} 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 porfolios 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. Rake the stocks for April according to size from small to large. 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 montly returns of the latest factor portfolios, namely $r_m^{p_t}$ and r_m^{pb}, as the top-portfolio monthly return and the bottom-portfolio monthly return.
3. The SMB is calculated just using $SMB = r_m^{pt} - r_m^{pb}$.

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

Factor	Order	Ratio
Size(SMB)	Assending	50%
Value(HML)	Dessending	30%
Profitability(RMW)	Dessending	30%
Investment(CMA)	Dessending	30%
Momentum(UMD)	Dessending	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 montly 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 risk free rate.