

基于沪深 300 样本股的因子选股策略

股票池

从 baostock 下载数据，时间范围为 2013 年 7 月 1 日-2023 年 6 月 9 日，选取沪深 300 指数样本股。根据沪深 300 指数调样规则同步更换样本股。

理论基础

根据资本资产定价模型 CAPM 理论，资产的预期超额收益由一元线性模型决定： $E[R_i] - R_f = \beta_i (E[R_M] - R_f)$ 其中 $E[.]$ 是期望符号， R_i 为某资产 i 的收益率， R_f 为无风险收益率， R_M 为市场组合的预期收益率。式中 $\beta_i = \frac{cov(R_i, R_M)}{var(R_M)}$ 刻画了该资产收益对市场收益的敏感程度，它也被称为资产 i 对市场风险的暴露程度。CAPM 正是最简单的线性因子模型，它指出资产的预期超额收益率由市场组合的预期超额收益率和资产对市场风险的暴露大小决定，而市场组合也被称为市场因子。

不同资产的收益率并非由单一的市场因子决定，而是同时受到其他因子的影响。以此为契机，套利定价理论 (Arbitrage Pricing Theory, APT)，在 CAPM 的基础上做了进一步延伸，构建了线性多因子定价模型（简称多因子模型）。多因子模型假设资产 i 的预期超额收益由以下多元线性模型决定： $E[R_i^e] = \beta_i * \lambda$ 其 $E[R_i^e]$ 表示资产 i 的预期超额收益， β_i 是资产 i 的因子暴露 (factor exposure) 或称因子载荷 (factor loading)， λ 是因子预期收益 (factor expected return)，也常被称为因子溢价 (factor risk premium)

因子选股规则

剔除停牌，退市，测试区间内未编列入沪深 300 的股票。按照因子值的前 20%，20%-40%，40%-60%，60%-80%，后 20% 构建投资组合。入选的每只股票投资相同资金，每逢 5 个交易日重新计算因子，并进行调仓。收益率计算公式为 5 日前的收盘价/当日收盘价 - 1，收盘价为后复权，手续费为买入的万分之二。

因子分类

因子主要分为估值、盈利、成长、动量、波动率、流动性六大类因子，具体因子定义见下表：

大类因子	小类因子	因子解释
动量	RSTR_barra	Barra 因子: $\sum_{t=L}^{T+L} w_t [\ln(1 + r_t)]$, L=21, T=500, 半衰期 120 日
	RSTR_m24	$\sum_{t=L}^{T+L} w_t [\ln(1 + r_t)]$, L=1, T=240, 半衰期 120 日
	RSTR_m12	$\sum_{t=L}^{T+L} w_t [\ln(1 + r_t)]$, L=1, T=240, 半衰期 60 日
	RSTR_m6	$\sum_{t=L}^{T+L} w_t [\ln(1 + r_t)]$, L=1, T=120, 半衰期 30 日
	RSTR_m3	$\sum_{t=L}^{T+L} w_t [\ln(1 + r_t)]$, L=1, T=60, 半衰期 15 日
	RSTR_m1	$\sum_{t=L}^{T+L} w_t [\ln(1 + r_t)]$, L=1, T=20, 半衰期 5 日
	RS_1	最新收盘价/21 个交易日前收盘价
	RS_3	最新收盘价/63 个交易日前收盘价
	RS_6	最新收盘价/126 个交易日前收盘价
	RS_12	最新收盘价/252 个交易日前收盘价
	Alpha	alpha 系数: 个股收益率序列与沪深 300 指数收益率序列以半衰期指数加权, 得到 alpha 系数, 半衰期为 60 日
流动性	STOM	月度平均换手率: 最近一个月的交易量/流通股数
	STOQ	季度平均换手率: 最近一季度的交易量/流通股数
	STOS	半年平均换手率: 最近半年的交易量/流通股数
	STOA	年度平均换手率: 最近一年的交易量/流通股数
	STOM_barra	Barra 因子: 公式: $\ln\left(\sum_{t=1}^{T+1} \frac{V_t}{S_t}\right)$, V_t 为 t 日成交金额, S_t 为 t 日流动市值
	STOQ_barra	Barra 因子: 公式: $\ln\left[\sum_{t=1}^T \exp(STOM_t)\right]$, T=63 个交易日
	STOA_barra	Barra 因子: 公式: $\ln\left[\sum_{t=1}^T \exp(STOM_t)\right]$, T=244 个交易日
	Ins	机构持股比例: 机构持股变动/总股本
	ins_c	机构持股比例变动
	Size	总市值
	non-linear-size	Barra 因子: 中等市值: 将总市值的对数与总市值立方的对数回归得到残差, 再对残差做标准化处理
	marketcap	流通市值

	MSM	一个月换手率变动: 最近 1 个月换手率/最近 1 年换手率
	MSQ	季度换手率变动: 最近 3 个月换手率/最近 1 年换手率
	MSS	半年换手率变动: 最近 6 个月换手率/最近 1 年换手率
波动率	DASTD	Barra 因子: 年度平均波动率: 累计日波动率以半衰期指数加权, 半衰期为 40 日
	CMRA	Barra 因子: 年度收益率波动
	HSIGMA	Barra 因子: sigma: 个股收益率序列与沪深 300 指数收益率序列以半衰期指数加权, 得到残差, 对残差求标准差得到 sigma, 半衰期为 60 日
	Beta	Barra 因子: 贝塔系数: 个股收益率序列与沪深 300 指数收益率序列以半衰期指数加权, 得到 beta 系数, 半衰期为 60 日
	yieldvol_1	月度日收益率波动率: 一个月日收益率标准差
	yieldvol_3	季度日收益率波动率: 三个月日收益率标准差
	yieldvol_6	半年日收益率波动率: 半年日收益率标准差
	high_low_1	月度股价波动: 最高价 / 最低价 (最近 一个月内存格)
	high_low_3	季度股价波动: 最高价 / 最低价 (近三个月内存格)
	high_low_6	半年股价波动: 最高价 / 最低价 (近六个月内价格)
	high_low_12	全年股价波动: 最高价 / 最低价 (近十二个月内价格)
	VOL_1	成交量月度波动率: 1 月波动率标准差
	VOL_3	成交量季度波动率: 3 月波动率标准差
	VOL_6	成交量半年波动率: 6 月波动率标准差
	VOL_12	成交量年度波动率: 12 月波动率标准差
成长	growth_ttm_or	营业收入 ttm 一年增长率
	growth_ttm_profit	净利润 ttm 一年增长率
	qfa_yoy sales_qq	单季度营业收入一年增长率
	qfa_yoy netprofit_qq	单季度归母净利润一年增长率
	qfa_yoy oc f_qq	单季度经营性现金流一年增长率
	qfa_yoy profit_qq	单季度净利润一年增长率
	qfa_roe_qq	单季度 roe 一年增长率
	yoy_profit_qq	净利润一年增长率
	yoy_growth_netprofit_qq	归母净利润一年增长率
	yoy_or_qq	营业收入一年增长率
	yoyroe_qq	roe 一年增长率
	yoyocf_qq	经营性现金流一年增长率

	growth_roe_qq_3	roe 三年增长率
	growth_netprofit_qq_3	归母净利润三年增长率
	growth_or_qq_3	营业收入三年增长率
	growth_profit_qq_3	净利润三年增长率
	growth_ocf_qq_3	经营性现金流三年增长率
	growth_profit_qq_5	净利润五年增长率
	growth_ocf_qq_5	经营性现金流五年增长率
	growth_roe_qq_5	roe 五年增长率
	SGRO	Barra 因子：过去 5 年企业营业收入复合增长率
	EGRO_5	Barra 因子：过去 5 年企业归属母公司净利润复合增长率
	EGIB	Barra 因子：未来 3 年企业一致预期净利润增长率
	EGIB_S	Barra 因子：未来 1 年企业一致预期净利润增长率
盈利 收益率	CETOP	Barra 因子：个股现金收益比股票价格
	roe_q	当季净资产收益率
	roe_ttm	滚动 ROE
	roa_q	当季资产收益率、资产回报率
	roa_ttm	滚动 ROA
	qfa_grossprofitmargin	当季毛利率
	grossprofitmargin_ttm	滚动毛利率
	profitmargin_q	当季扣非后利润率
	profitmargin_ttm	滚动扣非后利润率
	assetturn_q	当季资产周转率
	assetturn_ttm	滚动资产周转率
	operationcashflowratio_q	当季经营活动现金净流比率
	operationcashflowratio_ttm	滚动经营活动现金净流比率
	ROIC	投入资本回报率
	EBIT2EV	投资收益率：息税前利润/投入资本
	CASHROIC	现金 ROIC
	FreeCashflowYield	自由现金流比率：经营活动产生的净现金流-构建其他长期资产支付的现金/总市值
	sales2EV	营业收益率：营业收入_TTM/总市值+非流动负债
	cashflow1	经营活动净现金流/总市值
	cashflow2	经营活动净现金流/营业收入
	cashflow3	经营活动净现金流/营业收入净收益
	invturn_qq	存货周转率：存货成本/平均存货余额
	arturn_qq	应收账款周转率：当期销售净收入/平均应收账款
	faturm_qq	固定资产周转率：销售收入/平均固定资产
	assetturnover_ttm	滚动总资产周转率：营业收入 ttm / [(期初资产总额+期末资产总额) / 2]
	assetturn_qq	总资产周转率：营业总收入 / [(期初资产总额+期末资产总额) / 2]
	longdebtworkingcapital_qq	长期债务与营运资金比率：长期债务/营运资本
	finaexpensetogr_qq	财务费用比率：财务费用/主营业务收入
	gctogr_qq	营业费用比率：营业费用/主营业务收入
价值	ETOP	Barra 因子：历史 EP 值：利用过去 12 个月个股净利润除以当前市值。
	BTOP	Barra 因子：历史 BP 值：普通股权益价值/市值
	epcut	市盈率（扣除非经营性损益部分，即公司经营性盈利与市值之比值）
	bp_if	最近公告日 BP
	ncfp	净现金市值比：净现金流 / 总市值
	ocfp	营业现金流比率：经营性现金流 / 总市值
	dividendyield	股息率：过去一年分红/总市值
	stop1	营收市值比，市销率 PS（TTM）倒数
	stop2	营收市值比，市销率 PS（LYR）倒数
	ep_rel	相对 PE；PE/行业 PE
	bp_rel	相对 PB；PB/行业 PB
	PEG	市盈率相对盈利增长比率
	EPIBS	Barra 因子：预期 EP 值

因子评价方法

IC(Information Coefficient) 信息系数，表示所选股票的因子值与股票下期收益率的截面相关系数，

通过 IC 值可以判断因子值对下期收益率的预测能力。信息系数的绝对值越大，该因子越有效。IC 为负表示因子值越小越好，IC 为正表示因子值越大越好。IC 的计算方法是：计算全部股票在调仓周期期初排名和调仓周期期末收益排名的线性相关度。IC 越大的因子，选股能力就越强。

IC 最大值为 1，表示该因子选股 100%准确，对应的是排名分最高的股票，选出来的股票在下一个调仓周期中，涨幅最大；相反，如果 IC 值为-1，则代表排名分最高的股票，在下一个调仓周期中，跌幅最大，是一个完全反向的指标。最无用的 IC 值是 0 或者接近 0 的值，这代表该因子对于股票没有任何的预测能力。当 IC 的绝对值大于 0.02 时，因子的选股能力较强。

IR 即信息比率 (Information Ratio)，是超额收益的均值与标准差之比,代表因子获取稳定 Alpha 的能力。整个回测时段由多个调仓周期组成，每一个周期都会计算出一个不同的 IC 值，IR 等于多个调仓周期的 IC 均值除以这些 IC 的标准方差。所以 IR 兼顾了因子的选股能力（由 IC 代表）和因子选股能力的稳定性（由 IC 的标准方差的倒数代表），当 IR 大于 0.5 时因子稳定获取超额收益能力较强。

数据处理方法

标准化：计算 IC 值时，将同期不同股票因子进行归一化处理， $feature_{normalized}[i] = \frac{feature[i] - \mu}{\sigma}$ ， μ 为因子序列的均值， σ 为因子序列的标准差，得到一个新的近似服从 $N(0,1)$ 分布的序列，这样做可以让不同因子的暴露度之间具有可比性。

去极值化：设第 T 期某因子在所有个股上的暴露度序列为 D_i ， D_M 为该序列中位数， D_{M1} 为序列 $|D_i - D_M|$ 的中位数，则将序列 D_i 中所有大于 $D_M + 3D_{M1}$ 的数重设为 $D_M + 3D_{M1}$ ，将序列 D_i 中所有小于 $D_M - 3D_{M1}$ 的数重设为 $D_M - 3D_{M1}$ 。

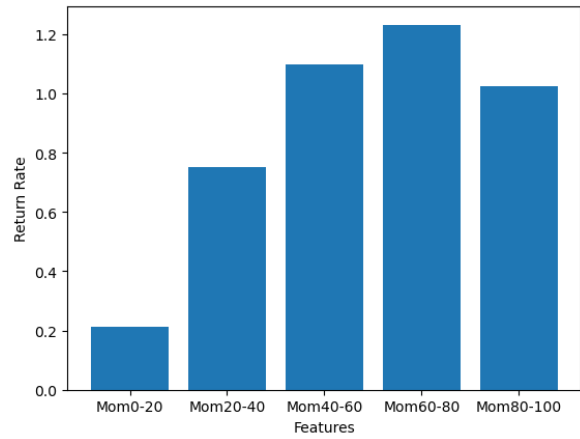
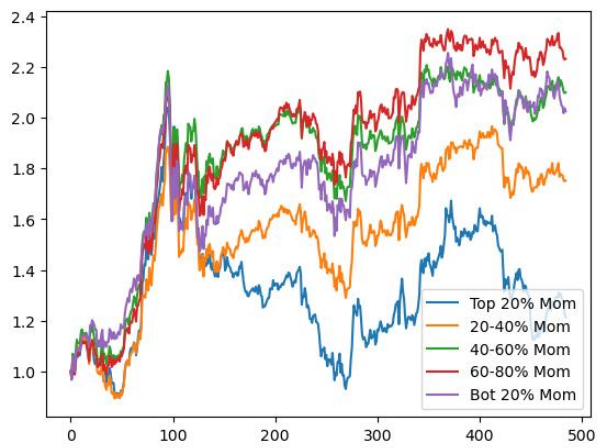
缺失值处理：得到新的因子暴露度序列后，将因子暴露度缺失的值排除于因子评价计算及收益率计算。

有效单因子测试

5 日动量因子:

计算公式: $\frac{close_price_i}{close_price_{i-5}}$, IC 值: -0.038 IR 值: 0.208

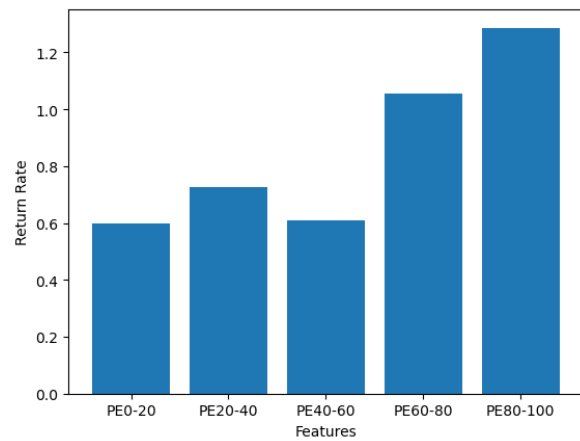
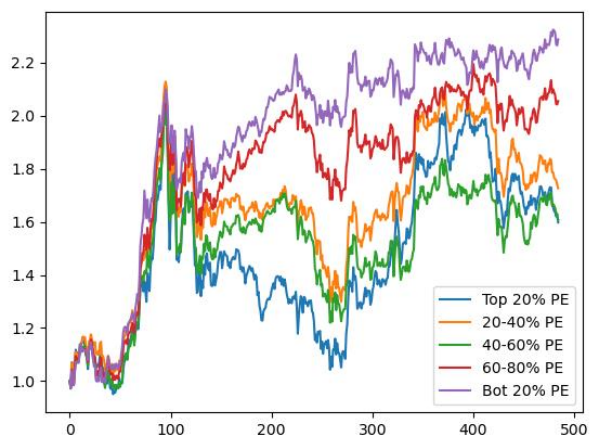
分层测试结果:



滚动市盈率因子:

计算公式: $peTTM_i$, IC 值: -0.025 IR 值: 0.125

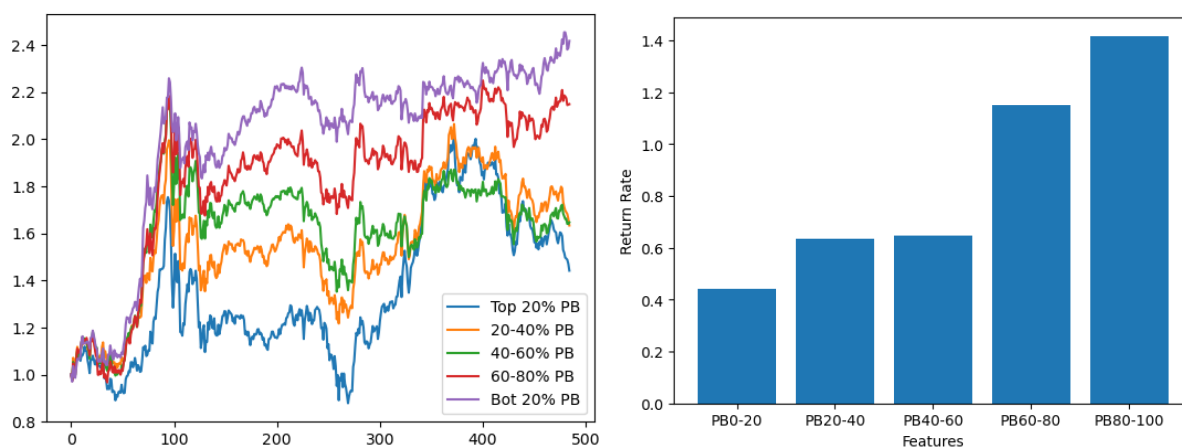
分层测试结果:



市净率因子:

计算公式: $pbMRQ_i$, IC 值: -0.026 IR 值: 0.106

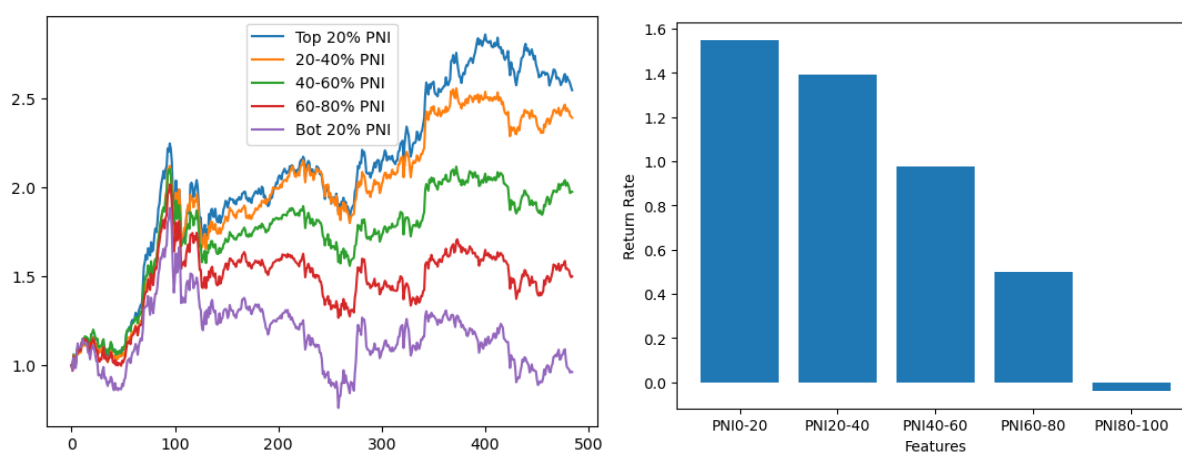
分层测试结果:



归属母公司股东净利润同比增长率因子:

计算公式: $YOYPNI_i$, 以财报统计的季度的最后一天为准, IC 值: 0.026 IR 值: 0.180

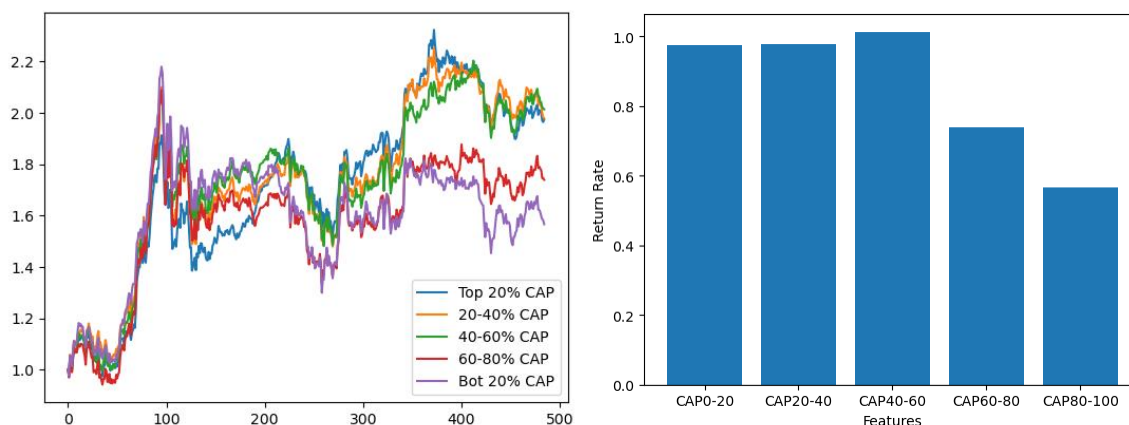
分层测试结果:



总流通市值因子:

计算公式: $\frac{close_i * volume_i}{turn_i}$, IC 值: 0.016 IR 值: 0.126

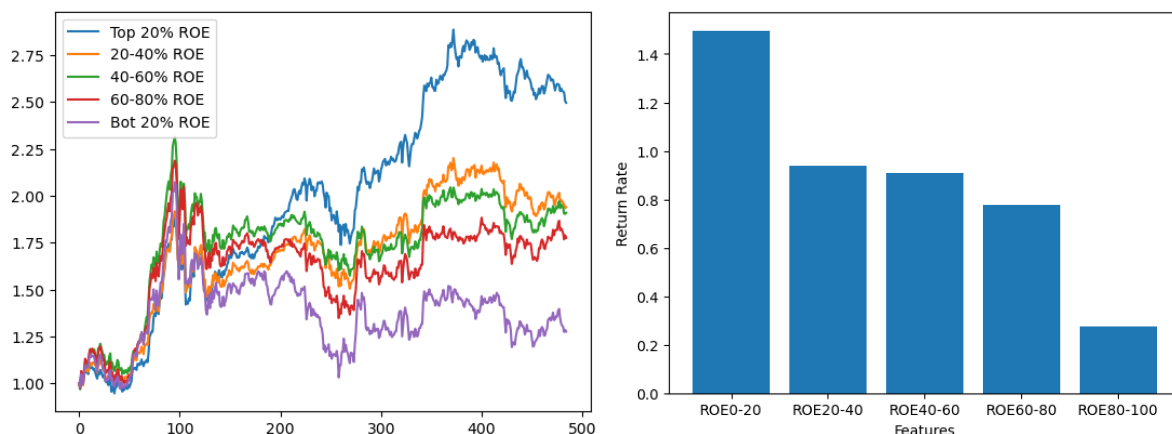
分层测试结果:



净资产收益率(平均)因子:

计算公式: $roeAvg_i$, IC 值: 0.027 IR 值: 0.144

分层测试结果:



多因子组合测试

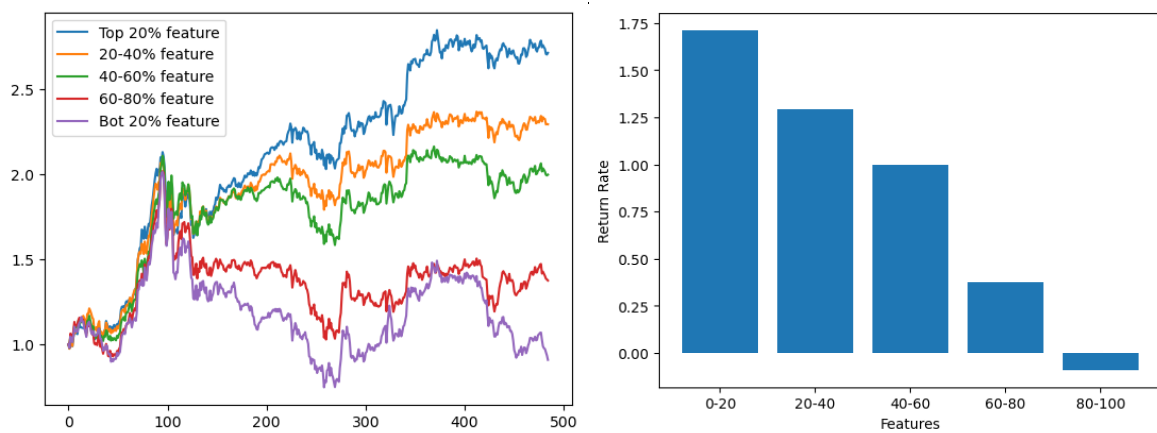
在六个有效因子中，滚动市盈率因子和市净率因子同属价值因子，且市净率因子比滚动市盈率因子显著，故去除滚动市盈率因子。

鉴于不同因子的缺失项有所不同，如果将所有出现缺失值的个股都从回归票池中剔除，当数据源质量不佳时可能会造成票池大幅减少，回归结果不可信。所以这里折中处理，将因子暴露度缺失的地方设为新序列的均值，即设为零，可视作当存在缺失值时我们认为此个股的因子值与全市场平均情况相同，即持中性看法。

如果该单因子的 IC 值为负，则多因子矩阵的值减去该单因子矩阵的值，反之则累加。

计算公式: $\frac{close_i * volume_i}{turn_i} + YOYPNI_i + roeAvg_i - pbMRQ_i - \frac{close_price_i}{close_price_{i-5}}$, IC 值: 0.055 IR 值: 0.363

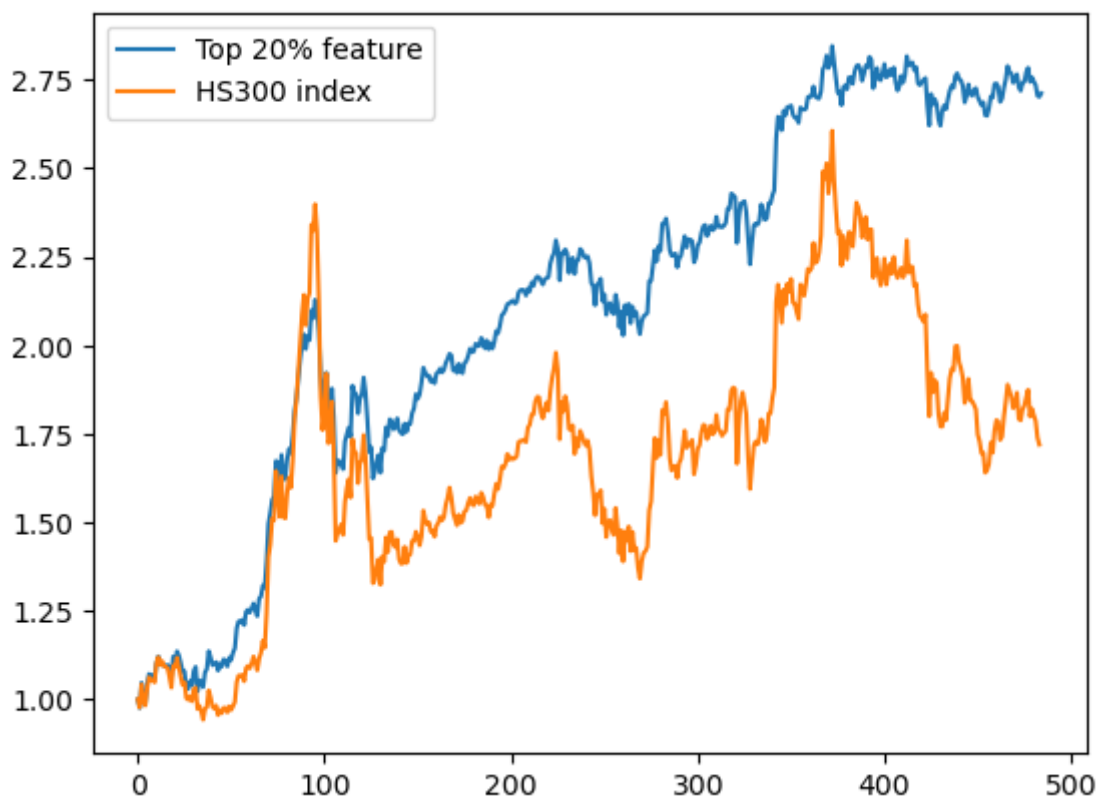
分层测试结果：



沪深 300 指数同期收益率为 72.00%，最大回撤为 44.80%

前 20%因子投资组合的十年总收益率为 171.12%，最大回撤为 23.69%，夏普比率为 0.81

将前 20%因子投资组合收益与同期沪深 300 指数收益比较：



买入因子前 20%股票同时卖空因子后 20%股票，构建投资组合。十年总收益率为 180.42%，最大回撤为 19.82%，夏普比率为 1.09

组合收益与同期沪深 300 指数收益比较：

