



多因子框架

使用基础

数据清理、排序、分组、资金分配

评价指标计算与解读

TSMultiFactor()中的面向对象

组合因子





学习项目:
题目自拟,
2-3人/组,
不需要提交报告, 每组需现场完成一个15分钟左右的项目展示,
成绩基于业界评委打分

展示时间: 12月12日或19日下午,(需统计)





多因子框架

- 天软科技平台提供了一整套多因子框架, 极高的提升了效率
- 这套系统是在通用框架的基础上针对因子定制化
- 我们将关心的因子放入框架中, 调整好相关参数, 最终可以展现给我们通过这个因子选股策略的回测结果.

- 因子的选择与比例配置
- 评价解读

用户核心



- 因子值计算
- 极值处理、标准化
- 分组、资金分配
- 回测
- 评价指标计算

天软多因子框架
TSMultiFactor



FFactorArr的设置是用户使用多因子框架的核心

字段	类型	说明	注意事项
因子名称	String	自定义用于显示的名称	因子名称的唯一性
因子公式	String	TSL语言中的指标公式，可以解析的字符串	因子公式常见参数： stockid,endt,rdate,stockarr
因子方向	Integer	1: 正向因子 0: 反向因子	因子方向只有0或1
因子比例(%)	Numeric	代表该因子在所有因子中的 权重，百分数	因子比例的和必须等于100

范例:

```
FactorDef:= array((  
    ‘因子名称’:‘ROE’,  
    ‘因子公式’: ‘Last12MData(Rdate,9900100)’,  
    ‘因子方向’:1,  
    ‘因子比例(%)’: 100));
```

- 用ROE作为一个因子, 并且是唯一因子(单因子), 来一次作为预测投资的信号. 在这里的设定为看涨(因子方向为1), 意思是在这期观测到的企业的ROE越高下一期股票涨的可能性就越大.
- 问题: 可以多于一个因子吗? 可以合成一个新因子吗? 因子方向是否永远不变? 比例是否永远不变?



多因子框架:关键设定

- 回测框架是一个类(class):
 - `obj:= new TSMultiFactor()`
- 设定回测区间 (开始日FBegT与结束日FendT)
 - `obj.FBegT:=20220101t; //设置回测起始日期`
 - `obj.FendT:=20221231t; //设置回测截止日期`
- 基准代码.
 - `obj.FIndexId:='SH000016'; //设置基准为上证50, 任意基准只要有对应代码即可`
- 调仓周期.
 - `obj.Fcycle:=cy_month(); //按月调仓, 也可按天,周,季度....`

属性	类型	含义	默认值
FBegT	DateTime	开始日期	当前系统时间往前推 3 年
FEndT	DateTime	截止日期	当前时间
FIndexId	String	基准代码	SH000300
FCycle	String	调仓周期, 字符串	Cy_month()
FFactorArr	Array	因子库	





多因子框架: 回测概览

FactorDef:= array(('因子名称':'ROE','因子公式':'Last12MData(Rdate,9900100)','因子方向' :1,'因子比例(%)':100));
//设定因子基本信息: 名称,公式,方向和比例

obj:= new TSMultiFactor(); //继承多因子框架class

obj.FBegT:=20200101t; //设置回测起始日期
obj.FendT:=20201231t; //设置回测截止日期
obj.FIndexId:=‘SH000016’; //设置基准为上证50
obj.Fcycle:=cy_month(); //每月底最有一个交易日调仓
obj.FfactorArr:=FactorDef; //设置因子库

obj.BackTest(); //回测

return obj.GetGroupReturn(1,cy_month());

//返回各组 and 基准的每月累计收益
//股票池默认today() 的上证50
//交易成本为0.1%
//这里的截止日也是调仓日
• 注意: 一定要搞清楚每个信息的意义和背后的逻辑!

数组数据		运行信息							
🏆	开始日	截止日	交易点数	基准	第1组	第2组	第3组	第4组	第5组
0	2020-01-01	2020-01-23	1	-2.83	-0.10	-0.10	-0.10	-0.10	-0.10
1	2020-02-01	2020-02-28	20	-6.52	-3.25	-4.55	-3.52	-3.86	-0.73
2	2020-03-01	2020-03-31	22	-10.89	-8.82	-12.17	-8.62	-5.23	-8.28
3	2020-04-01	2020-04-30	21	-5.17	1.52	-4.02	-5.93	-2.24	-6.26
4	2020-05-01	2020-05-29	18	-7.00	-0.02	-7.39	-6.40	-6.30	-10.13
5	2020-06-01	2020-06-30	20	-2.51	9.12	-5.78	2.50	-4.47	-7.19
6	2020-07-01	2020-07-31	23	7.67	23.95	4.33	13.68	12.04	1.80
7	2020-08-01	2020-08-31	21	10.80	28.38	9.43	14.39	11.05	3.00
8	2020-09-01	2020-09-30	22	7.11	25.13	8.78	6.83	6.43	-4.92
9	2020-10-01	2020-10-30	16	8.70	28.26	9.31	9.53	8.86	-6.50
10	2020-11-01	2020-11-30	21	14.96	31.50	16.47	15.25	13.96	0.03
11	2020-12-01	2020-12-31	23	20.64	49.58	14.97	9.37	10.10	3.61



设定调仓期

- 调仓期和初始样本的选择
- 行情因子策略可为周或月.
- 财务因子策略集中在月或者季度.
- 信息更新频率与交易成本的trade-off.





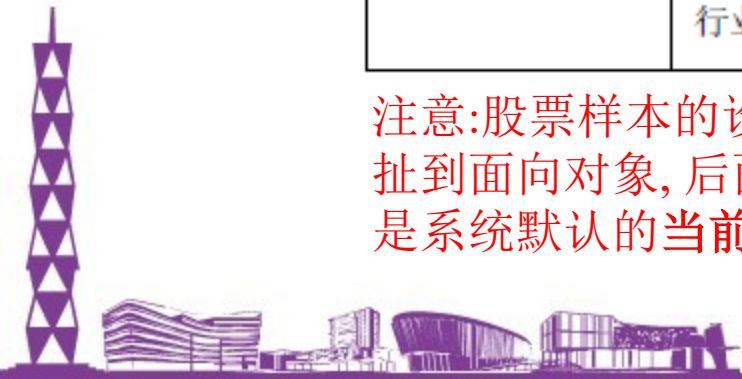
选择股票样本（域）

- 初始样本(股票池)的选择
 - 没有统一标准,根据策略目标来选择.
 - 注意: 股票池的基准时间应和历史数据时间配合, 不可按今天的成分股来判断历史的回报(新加入的成分股一般都是在历史中市值都是增长的)

分类	说明
市场板块	上证 A 股、深证 A 股、中小企业版、创业板等
行业板块	申万行业、证监会行业、中证行业等
指数成分	规模类：沪深 300、中证 800、上证 50 等 风格类：300 成长、300 价值等 主题类：中证红利、内地消费等 策略类：基本面 200、基本面 400 等 行业类：中证金融、中证信息等

- 样本剔除
 - 上市时间过短
 - ST股
 - 长期亏损

注意:股票样本的设定本质上市选择用来统计预测变量的数据池, 这个设定牵扯到面向对象, 后面我们会介绍如何在回测框架中使用和调整. 现阶段全部都是系统默认的当前时间的上证50.





因子选择和设置

- 可以是一个函数里面包含了计算对应因子的步骤.
- 范例1: Last12MData(RDate, 99000000) :
 - TTM从Rdate开始过去12个月的每股年化收益.
- 范例2: StockMarketValue(EndT):
 - 代表EndT日期的流通市值
- 注意: 必须提供一个变量日期,因为因子分析是回测,每个调仓期都会更新. **Rdate**特指报告期, **EndT**特指调仓日期, 是TSMultiFactor()这个类public参数.

与报告期有关的因子公式:

- 盈利类因子 (ROE)
- 债务类因子 (资产负债率)
- 成长类因子 (营业收入增长率)
- 经营能力类因子 (存货周转率)
- 分红和投资收益类因子 (股息支付率)
- 现金流类因子 (销售现金比率)
- 机构持股类因子 (第一大股东比例)

在这类财务因子中, 都必须包含报告期且需在天软平台定义为**Rdate**.

量价相关的因子公式

- 动量类因子 (最近20日涨幅)
- 规模类因子 (指定日流通股本,流通市值)
- 风险类因子 (最近1年Beta)

量价相关的因子日期设定相对多变.



- 因子方向可以通过统计相关性或者参考回归模型的系数方向.
- 因子方向是判断因子和股票收益的相关性.
- 因子方向1代表正相关或正序(做多),0代表负相关或逆序.
- 正或逆虽然重要,更关键是有“序”!
- 只要有“序”,我们就可以通过“序”预测.

大类因子	因子名称	因子公式	因子方向	因子描述
盈利因子	净资产收益率(%)ROE	Last12MData(RDate,9900100)	1	净利润/平均净资产
	总资产报酬率(%)ROA	Last12MData(RDate,9900105)	1	利润总额/平均总资产
	销售毛利率(%)	Last12MData(RDate,9900103)	1	(营业收入-营业成本)/营业收入
	销售净利率(%)	Last12MData(RDate,9900102)	1	净利润/营业收入
成长因子	营业收入增长率(%)	Last12MData(RDate,9900600)	1	
	营业利润增长率(%)	Last12MData(RDate,9900602)	1	
	利润总额增长率(%)	Last12MData(RDate,9900603)	1	
	归属于母公司净利润增长(%)	Last12MData(RDate,9900604)	1	
经营能力	存货周转率	Last12MData(RDate,9900400)	1	营业成本/平均存货
	应收账款周转率	Last12MData(RDate,9900404)	1	营业成本/平均应收账款
	总资产周转率	Last12MData(RDate,9900416)	1	营业成本/平均总资产
	流动比率	Last12MData(RDate,9900200)	1	流动资产/流动负债

因子方向和比例

- 对于单因子,因子比例必须等于100.
- 对于多个因子,各个单因子比列的和必须是100.
- 设置比列没有统一标准,可根据自己主管判断或某一量化标准分配.

负债因子	速动比率	Last12MData(RDate,9900201)	1	(流动资产-存货)/流动负债
	资产负债率(%)	1/Last12MData(RDate,9900203)	1	总负债/总资产
规模因子	总市值	Ln(StockTotalValue(EndT))	1	
	流通市值	Ln(StockMarketValue(EndT))	1	
	总股本	Ln(StockTotalShares(EndT))	1	
	流通股本	Ln(StockHasNAShares(EndT))	1	
	总资产	Last12MData(RDate, 44059)	1	
估值因子	市盈率(TTM)	1/StockPE_VI(EndT)	1	归属净利润/总市值
	市净率(TTM)	1/StockPNA_VI(EndT)	1	归属净资产/总市值
	市销率(TTM)	1/StockPMI_VI(EndT)	1	营业收入/总市值
	市现率(TTM)	1/StockPCF_VI(EndT)	1	经营现金净额/总市值
动量因子	最近 1 个月涨幅(%)	StockZf2(20)	1	
	最近 3 个月涨幅(%)	StockZf2(60)	1	
	最近 6 个月涨幅(%)	StockZf2(120)	1	
	最近 12 个月涨幅(%)	StockZf2(240)	1	
其他因子	收盘价	StockClose(EndT)	1	
	最近 1 个月日均换手率	StockAveHsl2(10)	1	
	最近 3 个月日均换手率	StockAveHsl2(60)	1	
风险因子	标准差(%)	StockRisk("SH00300",IncYear(EndT,-1),EndT)	0	用最近 1 年数据, 计算标准差
	Beta	StockBeta("SH00300",IncYear(EndT,-1),EndT)	0	用最近 1 年数据, 计算Beta





多因子框架: 回测结果again

```
FactorDef:= array((  
'因子名称':'ROE' ,  
'因子公式':'Last12MData(Rdate,9900100)',  
'因子方向'1,  
'因子比例(%)':100));
```

```
obj:= new TSMultiFactor();
```

```
obj.FBegT:=20200101t;
```

```
obj.FendT:=20201231t;
```

```
obj.FIndexId:='SH000016';
```

```
obj.Fcycle:=cy_month();
```

```
obj.FfactorArr:=FactorDef;
```

```
obj.BackTest();
```

```
return obj.GetGroupReturn(1,cy_month());
```

//返回各组 and 基准的每月累计收益

//股票池默认today() 的上证50

//交易成本为0.1%

//截止日也是调仓日

至此为止我们完成了一次对ROE因子基于上证50股票池2020年的回测! 操作很简单, 但是要搞清楚每个信息的意义和背后的逻辑!

数组数据		运行信息							
序号	开始日	截止日	交易点数	基准	第1组	第2组	第3组	第4组	第5组
0	2020-01-01	2020-01-23	1	-2.83	-0.10	-0.10	-0.10	-0.10	-0.10
1	2020-02-01	2020-02-26	20	-6.52	-3.25	-4.55	-3.52	-3.86	-0.73
2	2020-03-01	2020-03-31	22	-10.89	-8.82	-12.17	-8.62	-5.23	-8.28
3	2020-04-01	2020-04-30	21	-5.17	1.52	-4.02	-5.93	-2.24	-6.26
4	2020-05-01	2020-05-29	18	-7.00	-0.02	-7.39	-6.40	-6.30	-10.13
5	2020-06-01	2020-06-30	20	-2.51	9.12	-5.78	2.50	-4.47	-7.19
6	2020-07-01	2020-07-31	23	7.67	23.95	4.33	13.68	12.04	1.80
7	2020-08-01	2020-08-31	21	10.80	28.38	9.43	14.39	11.05	3.00
8	2020-09-01	2020-09-30	22	7.11	25.13	8.78	6.83	6.43	-4.92
9	2020-10-01	2020-10-30	16	8.70	28.26	9.31	9.53	8.86	-6.50
10	2020-11-01	2020-11-30	21	14.96	31.50	16.47	15.25	13.96	0.03
11	2020-12-01	2020-12-31	23	20.64	49.58	14.97	9.37	10.10	3.61



多因子框架: 回测回报率解读

- 开始日: 计算回报率的起始时间.
- 截止日: 计算回报率的截止时间, 收集因子数据的时间点.
- 交易点数: 开始日与截止日被统计的交易日.
(为何第一行是1个交易点?)
- 基准: 用来判断是否跑赢大盘(这里设置的是上证50), (思考: 从机会成本的角度思考为何一定要跑赢大盘).
- 第1-5组: 将ROE因子值按大小排列, 第一组理应为最有价值. (如何分组? 如何排列? 如何分配自己? 为何第一行全都一样?, 每个时间段的每组股票是否会变化?)
- 基准和每组显示都是累计回报率

数组数据 运行信息									
🏆	开始日	截止日	交易点数	基准	第1组	第2组	第3组	第4组	第5组
0	2020-01-01	2020-01-23	1	-2.83	-0.10	-0.10	-0.10	-0.10	-0.10
1	2020-02-01	2020-02-28	20	-6.52	-3.25	-4.55	-3.52	-3.86	-0.73
2	2020-03-01	2020-03-31	22	-10.89	-8.82	-12.17	-8.62	-5.23	-8.28
3	2020-04-01	2020-04-30	21	-5.17	1.52	-4.02	-5.93	-2.24	-6.26
4	2020-05-01	2020-05-29	18	-7.00	-0.02	-7.39	-6.40	-6.30	-10.13
5	2020-06-01	2020-06-30	20	-2.51	9.12	-5.78	2.50	-4.47	-7.19
6	2020-07-01	2020-07-31	23	7.67	23.95	4.33	13.68	12.04	1.80
7	2020-08-01	2020-08-31	21	10.80	28.38	9.43	14.39	11.05	3.00
8	2020-09-01	2020-09-30	22	7.11	25.13	8.78	6.83	6.43	-4.92
9	2020-10-01	2020-10-30	16	8.70	28.26	9.31	9.53	8.86	-6.50
10	2020-11-01	2020-11-30	21	14.96	31.50	16.47	15.25	13.96	0.03
11	2020-12-01	2020-12-31	23	20.64	49.58	14.97	9.37	10.10	3.61



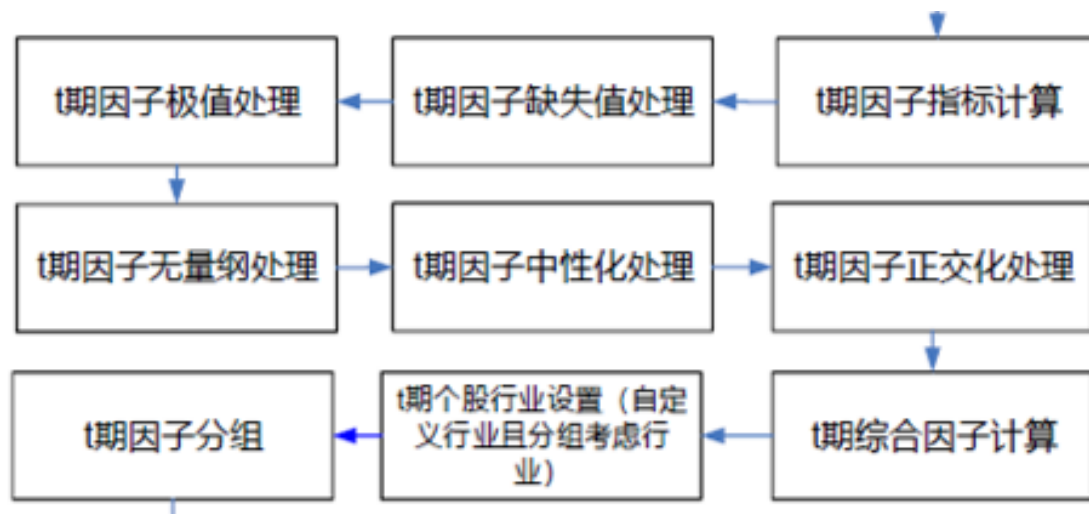
多因子框架

数据清理、排序、分组、资金分配



原始数据至可用因子

- 天软框架让我们选定参数(因子, 开始与截止日期, 调仓期, 股票池, 基准, 等等)然后输出结果.
- 结果很重要, 但是如何获得此结果更重要!
- 只有理解背后机制, 你才能知道你在干什么, 当回测结果不好时, 如何做对应调整!
- 下面我们介绍, 每一组的股票是如何筛选出?





预测变量原始数据清理

- 我们以截止日为2020年10月30日的上证50成分股, 提取TTM ROE 为预测因子(正相关)来挑选ROE表现最突出的10只股票作为投资对象.
- 截止日为2020年10月30日, 部分上市公司可能已经公布了三季度报(20200930), 部分还未公布(用年中报).



原始数据

	代码	名称	截止日	报告期	ROE
0	SH600519	贵州茅台	2020-10-30	20200930	30.03
1	SH600031	三一重工	2020-10-30	20200930	27.01
2	SH600887	伊利股份	2020-10-30	20200930	25.80
3	SH601012	隆基绿能	2020-10-30	20200630	23.82
4	SH603160	汇顶科技	2020-10-30	20200930	22.91
5	SH600585	海螺水泥	2020-10-30	20200930	22.77

以上内容可以通过 `obj.GetGroupFactorValue()` 在多因子框架中获得.



预测变量原始数据清理

缺失值处理: (不处理, 剔除, 平均值, 中位数)

- 以当期所有个券对应因子值的中位数填充
- 上证50都是大蓝筹主要是国企, 很少不上报, 停牌, 所以这里无缺失.



	代码	名称	截止日	报告期	ROE	ROE_缺失值处理后
0	SH600519	贵州茅台	2020-10-30	20200930	30.03	30.03
1	SH600031	三一重工	2020-10-30	20200930	27.01	27.01
2	SH600887	伊利股份	2020-10-30	20200930	25.80	25.80
3	SH601012	隆基绿能	2020-10-30	20200630	23.82	23.82
4	SH603160	汇顶科技	2020-10-30	20200930	22.91	22.91
5	SH600585	海螺水泥	2020-10-30	20200930	22.77	22.77
6	SH600276	恒瑞医药	2020-10-30	20200930	20.72	20.72
7	SH601138	工业富联	2020-10-30	20200630	19.94	19.94

缺失值处理

- 以当期所有个股对应因子值的中位数填充.(不建议中平均值,why?)

预测变量原始数据清理

极值处理

- 在处理完缺失值后再进行极值处理.
- 将小概率事件剔除. 我们从数据中发掘的信息是期望一个长期并且可复制预测变量(因子), 贵州茅台的每股收益是否属于极值?
- 天软系统默认中位数去极值法. 也可选3倍标准差法或四分位法.
- 系统默认叫法为奇异处理



	代码	名称	截止日	报告期	ROE	ROE_缺失值处理后	ROE_奇异处理后
0	SH600519	贵州茅台	2020-10-30	20200930	30.03	30.03	27.10
1	SH600031	三一重工	2020-10-30	20200930	27.01	27.01	27.01
2	SH600887	伊利股份	2020-10-30	20200930	25.80	25.80	25.80
3	SH601012	隆基绿能	2020-10-30	20200630	23.82	23.82	23.82
4	SH603160	汇顶科技	2020-10-30	20200930	22.91	22.91	22.91
5	SH600585	海螺水泥	2020-10-30	20200930	22.77	22.77	22.77

注意: 只有茅台超标被赋予新值.

中位数去极值法

- $y_i = |x_i - x_{50}|$
- x_{50} 是原始数据的中位数
- z_{50} 是 y 的中位数
- x_i 是观测到的数据
- $\hat{x}_i = x_{50} \pm 5.2 * z_{50}$, 如果 x_i 超过这个范围, 那么被赋值 \hat{x}_i .

四分位去极值

$$G_U = x_{75} - x_{50}$$

$$G_L = x_{50} - x_{25}$$

$$\hat{x}_i = x_{75} + 1.5 * G_U$$

$$\hat{x}_i = x_{25} - 1.5 * G_L$$



预测变量原始数据清理

无量纲化处理

- 单因子一般使用的无量纲化默认方法为标准法, 除此之外还有极差标准化, 极差正规化, 百分比打分法(按降序打分最高是100最低是0). (详见天软多因子框架手册)
- 在多因子分析时, 每个因子差异较大(单位不同), 为了方便比较进行无量纲化处理.



	代码	名称	截止日	报告期	ROE	ROE_缺失值处理后	ROE_奇异处理后	ROE_标准化后
0	SH600519	贵州茅台	2020-10-30	20200930	30.03	30.03	27.10	2.30
1	SH600031	三一重工	2020-10-30	20200930	27.01	27.01	27.01	2.29
2	SH600887	伊利股份	2020-10-30	20200930	25.80	25.80	25.80	2.11
3	SH601012	隆基绿能	2020-10-30	20200630	23.82	23.82	23.82	1.80
4	SH603160	汇顶科技	2020-10-30	20200930	22.91	22.91	22.91	1.66
5	SH600585	海螺水泥	2020-10-30	20200930	22.77	22.77	22.77	1.64





预测变量原始数据清理

中性化处理和正交化

- 一般指对市值和行业中性化,就是在同行业内. 学界常用,业界多用后面提到的检验分组来达到同样目的.
- 对市直和行业中性化也可在分组中实施.
- 系统默认为不进行中性化处理和不进行正交化处理.
- 正交化一般在多因子中,目的是剔除各个因子之间的相关性.



代码	名称	截止日	报告期	ROE	ROE_缺失值处理后	ROE_奇异处理后	ROE_标准化后	ROE_中性化后	ROE_正交化后
0	SH600519	贵州茅台	2020-10-30	20200930	30.03	30.03	27.10	2.30	2.30
1	SH600031	三一重工	2020-10-30	20200930	27.01	27.01	27.01	2.29	2.29
2	SH600887	伊利股份	2020-10-30	20200930	25.80	25.80	25.80	2.11	2.11
3	SH601012	隆基绿能	2020-10-30	20200630	23.82	23.82	23.82	1.80	1.80
4	SH603160	汇顶科技	2020-10-30	20200930	22.91	22.91	22.91	1.66	1.66
5	SH600585	海螺水泥	2020-10-30	20200930	22.77	22.77	22.77	1.64	1.64



预测变量原始数据清理

综合打分

- 综合打分的目的是为了下一步排序方便**分组**, 所以值的意义被弱化, 从而强化了顺序 (相对关系).
- 单因子综合打分直接对上一步处理后的数据直接复制.
- 多因子需要按因子比例组合.
- 此案例直接复制标准化的的结果.



序号	代码	名称	截止日	报告期	ROE	ROE_缺失值处理后	ROE_奇异处理后	ROE_标准化后	ROE_中性化后	ROE_正交化后	综合因子
0	SH600519	贵州茅台	2020-10-30	20200930	30.03	30.03	27.10	2.30	2.30	2.30	2.30
1	SH600031	三一重工	2020-10-30	20200930	27.01	27.01	27.01	2.29	2.29	2.29	2.29
2	SH600887	伊利股份	2020-10-30	20200930	25.80	25.80	25.80	2.11	2.11	2.11	2.11
3	SH601012	隆基绿能	2020-10-30	20200630	23.82	23.82	23.82	1.80	1.80	1.80	1.80
4	SH603160	汇顶科技	2020-10-30	20200930	22.91	22.91	22.91	1.66	1.66	1.66	1.66
5	SH600585	海螺水泥	2020-10-30	20200930	22.77	22.77	22.77	1.64	1.64	1.64	1.64



数据清理相关参数设定

属性	类型	含义	默认值
FFillMissType	Integer	缺失值处理方法	0: 剔除
FAbnormalType	Integer	极值处理方法	0: 中位数法
FStandarizeType	Integer	标准化方法	1: Z 值法
FNeuType	Integer	中性化方法	0: 不进行中性化
FOrthType	Integer	正交化方法	0: 不进行正交化
FIndustryEval	String	行业表达式	'MF_Industry(6,1)'
FConstr	String	剔除条件	空字符串
FBuyRatio	Numeric	开仓费率(%)	0.1
FSellRatio	Numeric	平仓费率(%)	0.1

- 其他参数选择详见天软多因子框架说明书





排序

因子排序

- 根据综合打分以降序排列
- 为何前10支股票为第一组? 为何综合因子得分最高排第一? 如果因子方向为逆序呢? 第一组股票暗示什么?

- 如果我们按照ROE正序选股投资, 那么第一组就是在2020-10-30截止日上证50里最新报告期最有增长潜力的10只股票(每组10只股票共5组).
- 这是2020-10-30截止日更新的排序, 下一个调仓期会重新反复一遍.

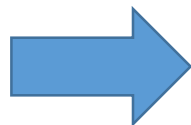
序号	代码	名称	截止日	报告期	ROE	ROE_缺失值处理后	ROE_奇异处理后	ROE_标准化后	ROE_中性化后	ROE_正交化后	综合因子	组名
0	SH600519	贵州茅台	2020-10-30	20200930	30.03	30.03	27.10	2.30	2.30	2.30	2.30	第1组
1	SH600031	三一重工	2020-10-30	20200930	27.01	27.01	27.01	2.29	2.29	2.29	2.29	第1组
2	SH600887	伊利股份	2020-10-30	20200930	25.80	25.80	25.80	2.11	2.11	2.11	2.11	第1组
3	SH601012	隆基绿能	2020-10-30	20200630	23.82	23.82	23.82	1.80	1.80	1.80	1.80	第1组
4	SH603160	汇顶科技	2020-10-30	20200930	22.91	22.91	22.91	1.66	1.66	1.66	1.66	第1组
5	SH600585	海螺水泥	2020-10-30	20200930	22.77	22.77	22.77	1.64	1.64	1.64	1.64	第1组
6	SH600276	恒瑞医药	2020-10-30	20200930	20.72	20.72	20.72	1.33	1.33	1.33	1.33	第1组
7	SH601138	工业富联	2020-10-30	20200630	19.94	19.94	19.94	1.21	1.21	1.21	1.21	第1组
8	SH601888	中国中免	2020-10-30	20200930	18.21	18.21	18.21	0.94	0.94	0.94	0.94	第1组
9	SH601318	中国平安	2020-10-30	20200930	17.32	17.32	17.32	0.80	0.80	0.80	0.80	第1组
10	SH600309	万华化学	2020-10-30	20200930	17.25	17.25	17.25	0.79	0.79	0.79	0.79	第2组
11	SH600048	保利发展	2020-10-30	20200930	17.11	17.11	17.11	0.77	0.77	0.77	0.77	第2组



排序

2020年10月30日调仓

名称	截止日	报告期	ROE	ROE_奇异处理后
贵州茅台	2020-10-30	20200930	30.03	27.10
三一重工	2020-10-30	20200930	27.01	27.01
伊利股份	2020-10-30	20200930	25.80	25.80
隆基绿能	2020-10-30	20200630	23.82	23.82
汇顶科技	2020-10-30	20200930	22.91	22.91
海螺水泥	2020-10-30	20200930	22.77	22.77
恒瑞医药	2020-10-30	20200930	20.72	20.72
工业富联	2020-10-30	20200630	19.94	19.94
中国中免	2020-10-30	20200930	18.21	18.21
中国平安	2020-10-30	20200930	17.32	17.32



2020年11月30日调仓

名称	截止日	报告期	ROE	ROE_奇异处理后
三一重工	2020-11-30	20200930	27.01	26.34
贵州茅台	2020-11-30	20200930	30.03	26.34
伊利股份	2020-11-30	20200930	25.80	25.80
隆基绿能	2020-11-30	20200930	24.61	24.61
汇顶科技	2020-11-30	20200930	22.91	22.91
海螺水泥	2020-11-30	20200930	22.77	22.77
恒瑞医药	2020-11-30	20200930	20.72	20.72
中国中免	2020-11-30	20200930	18.21	18.21
工业富联	2020-11-30	20200930	18.12	18.12
中国平安	2020-11-30	20200930	17.32	17.32



因子数据处理方式思考

- 如果是两个因子呢?
- 综合因子得分如何计算?
- 原始数据提炼因子数据中的方法是否合理? 可否有创新?





学界与业界分组

- 业界中的分组方法和学界资产定价较相似
- 前者为通过分组检验超额收益是否稳定. 后者主要过滤干扰项.
- 常见方法类似学界的排序法, 业界用三分位, 五分位(系统默认), 十分位等等都是以百分位排名分组.
- 在学术界遇到的分组问题业界也有, 所以在业界也有以因子值和行业排序的重排序分组.
- 比如电子行业市值相对于金融行业市值较大, 不考虑行业因素的话, 会导致投资标准虽然按市值, 其实背后是行业主导. (如同计量经济学中的内生性 endogeneity).





分组

范例:无行业考虑分组

- 以收盘价作为正向因子
- 奇异处理包括缺失和极值处理.
- 但因子所以无量纲化用的是百分比打分法
- 综合因子一般指多因子结合后的百分百打分,这里是单因子所以无变化.

代码	名称	收盘	奇异处理后	无量纲化后	综合因子	行业名称	排名
SH601318	中国平安	45.29	45.29	100.00	100.00	申万金融服务	1
SH601601	中国太保	22.50	22.50	90.90	90.90	申万金融服务	2
SH601628	中国人寿	21.40	21.40	81.80	81.80	申万金融服务	3
SH600585	海螺水泥	18.45	18.45	72.70	72.70	申万建筑建材	4
SH601166	兴业银行	16.69	16.69	63.60	63.60	申万金融服务	5
SH600036	招商银行	13.75	13.75	54.50	54.50	申万金融服务	6
SH600030	中信证券	13.36	13.36	45.40	45.40	申万金融服务	7
SH600068	葛洲坝	5.49	5.49	36.30	36.30	申万建筑建材	8
SH600010	包钢股份	5.40	5.40	27.20	27.20	申万钢铁	9
SH600019	宝钢股份	4.89	4.89	18.10	18.10	申万钢铁	10
SH601668	中国建筑	3.90	3.90	9.00	9.00	申万建筑建材	11
SH601669	中国电建	3.82	3.82	0.00	0.00	申万建筑建材	12

- 分三组!
- 可见申万钢铁全部都在第三组,那么第三组表现的好坏到底是因子决定的还是行业决定?
- 同样第一组也面临类似问题!
- 我们的投资收益是根据所设计的因子而得,通过分组将其他因素降到最低,常见方式是考虑行业因素.



考虑行业的第一种分组方式

范例: 12支股票共分三组需考虑因子值和行业分类.

- 同行业企业数被组数整除: 按行业内因子值排名依次入组. (金融服务行业每组依次排入2家)



名称	综合因子	行业名称	行业内排名	分组
中国平安	100.00	申万金融服务	1	1
中国太保	90.90	申万金融服务	2	1
中国人寿	81.80	申万金融服务	3	2
兴业银行	63.60	申万金融服务	4	2
招商银行	54.50	申万金融服务	5	3
中信证券	45.40	申万金融服务	6	3

- 同行业企业数被除有余数(建筑建材行业有4家),先按行业内因子值排名依次入组,余数企业自最后一组开始入组(排名第4的中国电建入住第3组).



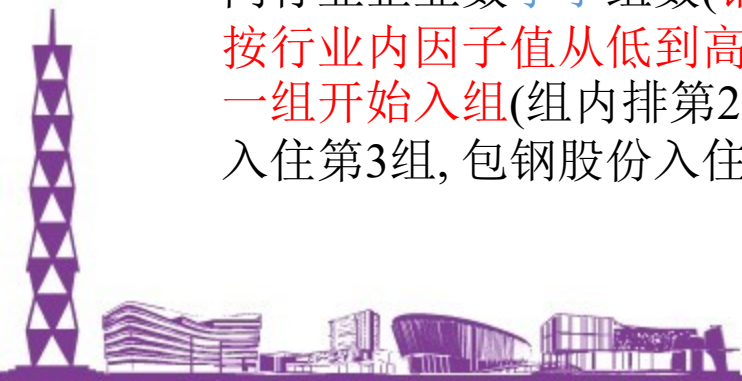
名称	综合因子	行业名称	行业内排名	分组
海螺水泥	72.70	申万建筑建材	1	1
葛洲坝	36.30	申万建筑建材	2	2
中国建筑	9.00	申万建筑建材	3	3
中国电建	0.00	申万建筑建材	4	3

- 同行业企业数小于组数(钢铁行业),按行业内因子值从低到高并从最后一组开始入组(组内排第2宝钢股份入住第3组,包钢股份入住第2组)



名称	综合因子	行业名称	行业内排名	分组
包钢股份	27.20	申万钢铁	1	2
宝钢股份	18.10	申万钢铁	2	3

注意:这里的分组如同一个动态的投资组合. 动态因为每个调仓期会变化.





考虑行业的第二种分组方式

- 前者的考虑行业分组导致组之间的行业分布极为不均衡从而缺乏代表性. 另一种考虑行业的分组方法如下.
- 对于同行业企业数小于组数情况, 每组都放入此行业所有企业, 如右图所示.

代码	名称	收盘	奇异处理后	标准化后	综合因子	行业名称	行业内排名	分组
SH600010	包钢股份	5.40	5.40	45.40	45.40	申万钢铁	1	1
SH600019	宝钢股份	4.89	4.89	18.10	18.10	申万钢铁	2	1
SH600585	海螺水泥	18.45	18.45	81.80	81.80	申万建筑建材	1	1
SH601318	中国平安	45.29	45.29	100.00	100.00	申万金融服务	1	1
SH601601	中国太保	22.50	22.50	90.90	90.90	申万金融服务	2	1
SH600010	包钢股份	5.40	5.40	27.20	27.20	申万钢铁	1	2
SH600019	宝钢股份	4.89	4.89	18.10	18.10	申万钢铁	2	2
SH600068	葛洲坝	5.49	5.49	72.70	72.70	申万建筑建材	2	2
SH601628	中国人寿	21.40	21.40	88.40	88.40	申万金融服务	3	2
SH601166	兴业银行	16.69	16.69	80.50	80.50	申万金融服务	4	2
SH600010	包钢股份	5.40	5.40	27.20	27.20	申万钢铁	1	3
SH600019	宝钢股份	4.89	4.89	18.10	18.10	申万钢铁	2	3
SH601668	中国建筑	3.90	3.90	9.00	9.00	申万建筑建材	3	3
SH601669	中国电建	3.82	3.82	0.00	0.00	申万建筑建材	4	3
SH600036	招商银行	13.75	13.75	78.50	78.50	申万金融服务	5	3
SH600030	中信证券	13.36	13.36	78.20	78.20	申万金融服务	6	3



分组相关的参数

所属	属性	类型	含义	默认值
分组相关	FGroupType	Integer	分组方式	1: 百分位
	FGroups	String		五分位
	FGroupHy	Integer	分组时是否考虑行业因素	-1: 分组时不考虑行业，直接按综合得分排名分组。
资金分配相关	FIndustryMethod	Integer	行业资金分配方法	-1: 不进行行业中性处理
	FRateType	Integer	个股资金分配方法	4: 等权重
行业	FIndustryType	Integer	行业分类方式	1: 按申万一级行业来划分行业





资金分配

- 每组有了股票后，每只股票得多少资金被投资该如何选？
- 多因子框架中的资金占比**默认为均分**
- 问题：一组有10只股票，那么买入的话如何分配资金？每只各占10%？
- 回测的过程中，每组的股票都要在每个调仓期做多，组内的资金分配方法有多种选择。
- 不同方法侧重点不一，比如按市值可能导致银行股变为主导

其他方案：

- 是否可以考虑协方差的目标优化来分组？
- 如何区分到底是因子的贡献还是组内资金分配的结果？



最小方差投资组合优化：n只股票

- $\mathbf{w} = [w_1, w_2, \dots, w_n]'$ be the vector of portfolio weights (size $n \times 1$).
- Σ be the $n \times n$ variance-covariance matrix of asset returns.

Optimization Problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}'\Sigma\mathbf{w} \\ \text{subject to} \quad & \mathbf{1}'\mathbf{w} = 1 \end{aligned}$$

This is a **quadratic programming** problem with a linear equality constraint.

Closed-form Solution:

Let $\mathbf{1}$ be an $n \times 1$ vector of ones.

$$\mathbf{w}^* = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$$

Let the covariance matrix Σ is defined as:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

注意:

- $w > 0$ 意味着股票不可卖空（股票看跌）
- 总权重不可大于1





最小方差投资组合优化：n只股票

只要能准确估算出未来的 Σ 即可解决投资组合优化的问题！

挑战依旧存在，只是少了一点点儿：

- Σ 的估算是否准确？
- Σ 的估算方法很多，该用哪一种？（金融计量主要研究内容）
 - Sample Covariance
 - 基于时间序列的预测方法（RiskMetrics, BEKK等等）

Let the covariance matrix Σ is defined as:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$



Sample Covariance

用样本协方差估计的特点：

设有 N 只资产，其收益在观测区间 $t = 1, \dots, T$ 记为列向量 $\mathbf{r}_t \in \mathbb{R}^N$ 。

令 $\Sigma = \text{Cov}(\mathbf{r}_t)$ 表示真实但未知的协方差矩阵。目标是用历史样本对 Σ 进行估计。

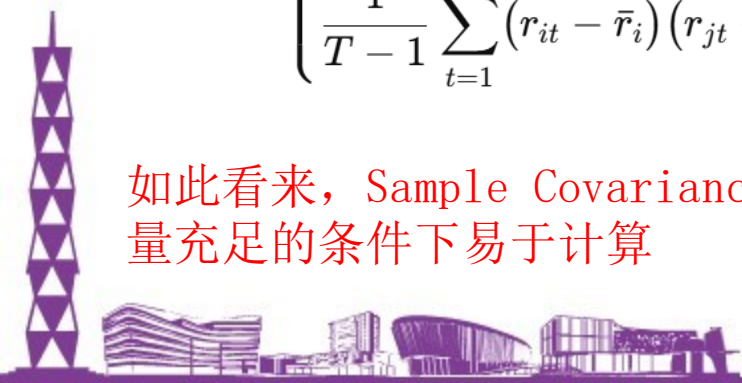
其中， \bar{r}_i ($i = 1, 2, \dots, N$) 为第*i*个资产收益的均值。

则资产收益协方差矩阵中的每个元素计算方式如下：

$$\hat{\sigma}_{ij} = \begin{cases} \frac{1}{T-1} \sum_{t=1}^T (r_{it} - \bar{r}_i)^2, & i = j, \\ \frac{1}{T-1} \sum_{t=1}^T (r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j), & i \neq j. \end{cases}$$

场景	Sample covariance 的具体症状
样本量要求高	当资产数量 <i>N</i> 接近或超过时间序列长度 <i>T</i> 时，样本协方差矩阵可能出现高维问题，导致估计误差增大。
权重均等分配	<p>样本协方差矩阵估计中对每个时间点赋予均等权重。</p> <p>「均等权重」意味着：无论是去年平静的市场，还是今年剧烈的股灾，每一天的价格波动对最终结果的贡献完全相同。</p> <p>但是真实的金融数据常呈现波动聚集（如市场危机期间波动率骤升），等权重意味着稀释近期高波动信息。</p>

如此看来，Sample Covariance的优势是在样本量充足的条件下易于计算



RiskMetrics model

“最新信息权重大，历史信息快速衰减”——用指数权数 (Exponentially Weighted Moving Average) 给过去的收益冲击打分，形成滚动协方差。

即离当前时间越近的数据权重越高，越远的数据权重越低，从而更灵敏地反映最新市场变化。

对应到 Σ_t 中的每个元素：

$$\sigma_{ij,t} = \begin{cases} \lambda \sigma_{ii,t-1} + (1 - \lambda) r_{i,t-1}^2, & \text{if } i = j \\ \lambda \sigma_{ij,t-1} + (1 - \lambda) r_{i,t-1} r_{j,t-1}, & \text{if } i \neq j \end{cases}$$

实现极简、实时更新、抓住波动簇集效应

$$\Sigma_t = \lambda \Sigma_{t-1} + (1 - \lambda) r_{t-1} r_{t-1}^\top, \quad 0 < \lambda < 1$$

- Σ_t : 当日（开盘前）风险矩阵
- λ : 衰减因子；日频数据常用 0.94（RiskMetrics™ 1996 建议）



资金分配中考虑行业

- 行业处理的方法多种多样。
- 也可在资金分配中考虑行业差异。
- 分组方式和组内资金分配都可直接或间接降低行业影响。
- 怎么分不是关键，关键是你的目标。





其他相关参数设定

所属	属性	类型	含义	默认值
回测相关	FpriceType	Integer	成交价类别	2: 收盘价
	FvolModType	Integer	成交量取整模式	0: 不取整
	FdividendType	Integer	是否分红再投资	0: 分红不再投资
	FAllotmentType	Integer	是否参与配股	0: 不自动参与配股
	FBacktestType	Integer	使用的回测方法	0: 表示使用回测框架清算





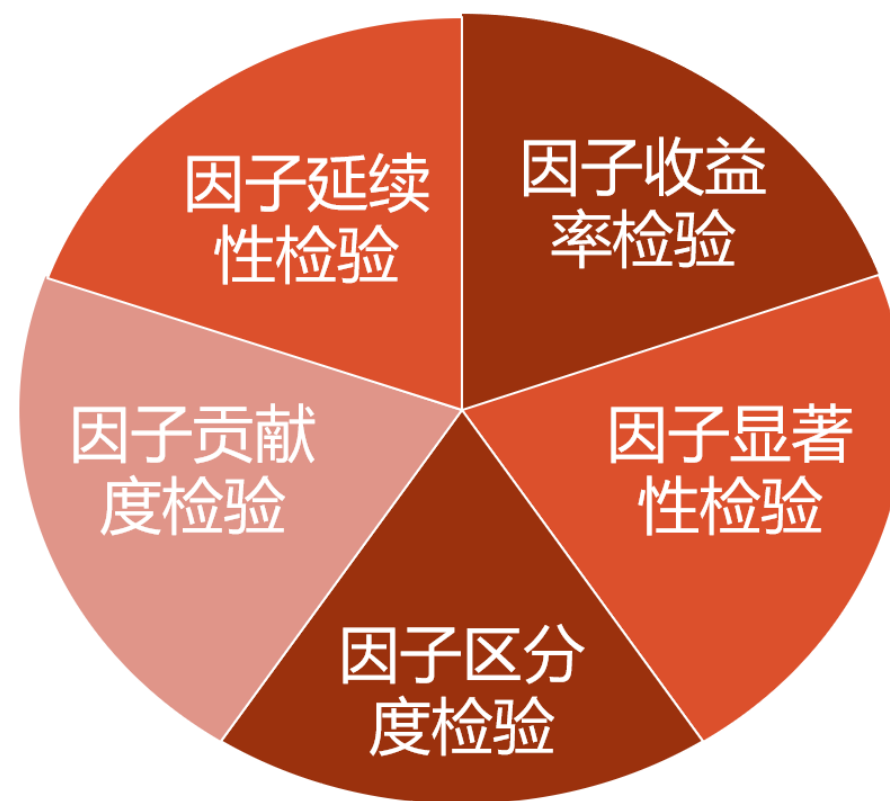
多因子框架

评价指标计算与解读





多因子框架: 回测评价





多因子框架: 回测评价

- 前面讲到都是设置因子相关参数,目的是建立一个定制化的回测环境.
- 设定的环境是否能带来超额回报需要通过因子评价.
 1. 收益和换手率分析: 展示各组的收益率(`GetGroupReturn()`)和换手率(`GetGroupTurnOver()`)
 2. 收益率检验缝隙: 展示平均收益,标准差,夏普率. (`GetTrailingReturn()`)
 3. 显著性检验:通过t检验方法进一步验证之前的结果. (`GetSignificanceStatistics`)
 4. 区分度检验: 通过对比第一组和最后一组的显著性差异来判断因子价值. (`GetLongShortStatistics()`).
 5. 延续性检验: 检验因子方向的持续性性. (`GetInformationCoefficient()`).
 6. 稳定性检验:通过自相关系数和股票成分在分组中的稳定性. (`GetSignalStability()`)
 7. 衰减性分析: 通过自相关系数检验T期与T+lag收益关系 $\text{lag}=1,2,\dots,N$.
 8. 贡献度分析: 通过最小二乘法(一般是线性回归)或主成分分析(PCA)分析每个因子的贡献.

上述的评价方法函数的具体参数选择详见天软多因子框架



回测评价:收益率

```
return obj.GetGroupReturn(pType=1,cy_month());
```

数组数据 运行信息

开始日	截止日	交易点数	基准	第1组	第2组	第3组	第4组	第5组
2020-01-01	2020-01-23	1	-2.83	-0.10	-0.10	-0.10	-0.10	-0.10
2020-02-01	2020-02-28	20	-6.52	-3.25	-4.55	-3.52	-3.86	-0.73
2020-03-01	2020-03-31	22	-10.89	-8.82	-12.17	-8.62	-5.23	-8.28
2020-04-01	2020-04-30	21	-5.17	1.52	-4.02	-5.93	-2.24	-6.26
2020-05-01	2020-05-29	18	-7.00	-0.02	-7.39	-6.40	-6.30	-10.13
2020-06-01	2020-06-30	20	-2.51	9.12	-5.78	2.50	-4.47	-7.19
2020-07-01	2020-07-31	23	7.67	23.95	4.33	13.68	12.04	1.80
2020-08-01	2020-08-31	21	10.80	28.38	9.43	14.39	11.05	3.00
2020-09-01	2020-09-30	22	7.11	25.13	8.78	6.83	6.43	-4.92
2020-10-01	2020-10-30	16	8.70	28.26	9.31	9.53	8.86	-6.50
2020-11-01	2020-11-30	21	14.96	31.50	16.47	15.25	13.96	0.03
2020-12-01	2020-12-31	23	20.64	49.58	14.97	9.37	10.10	3.61

pType详见天软多因子框架说明书.

0	收益率(%)
1	累计收益率(%)
2	超额收益率(%)
3	超额累计收益率(%)
4	第一组 vs 最后一组超额收益率(%),第一组 vs 最后一组累计超额收益(%)

回测评价:换手率

- 这里换手率的计算方式:是否有存在于'上期截止日'股票而不存于'截止日',被移除的股票除以上期截止日股票总数.
- 影响换手率的因素:新报告期ROE发布,成分股(上证50)调整(每年5月中下和11月中下),但指数真正剔除时间要到6月和12月!
- 某些股票不合规而剔除(停牌,上市时间过短).
- 上证50新进股票不影响,但剔除股票会!

return obj.GetGroupTurnOver(weight_type); 显示单位为百分比.

weight_type= 1 (个数法换手率) or 2 (权重法换手率)

数组数据		运行信息						
	截止日	上期截止日	基准	第1组	第2组	第3组	第4组	第5组
0	2020-02-28	2020-01-23	0.00	0.00	0.00	0.00	0.00	0.00
1	2020-03-31	2020-02-28	0.00	10.00	10.00	30.00	30.00	10.00
2	2020-04-30	2020-03-31	0.00	10.00	30.00	50.00	40.00	20.00
3	2020-05-29	2020-04-30	0.00	0.00	0.00	0.00	0.00	0.00
4	2020-06-30	2020-05-29	10.00	10.00	10.00	10.00	20.00	10.00
5	2020-07-31	2020-06-30	0.00	10.00	10.00	0.00	0.00	0.00
6	2020-08-31	2020-07-31	0.00	0.00	40.00	60.00	30.00	10.00
7	2020-09-30	2020-08-31	0.00	10.00	10.00	10.00	10.00	0.00
8	2020-10-30	2020-09-30	0.00	11.11	40.00	40.00	30.00	10.00
9	2020-11-30	2020-10-30	0.00	0.00	20.00	30.00	10.00	0.00
10	2020-12-31	2020-11-30	10.00	30.00	30.00	30.00	30.00	20.00

关于调整上证50、上证180、上证380等指数样本的公告

2021-05-28

☆ | ④ | A+ | A-

根据指数规则,经指数专家委员会审议,上海证券交易所与中证指数有限公司决定调整上证50、上证180、上证380、科创50等指数样本,于2021年6月11日收盘后生效。其中上证50指数更换5只样本,上证180指数更换18只样本,上证380指数更换38只样本,科创50指数更换5只样本。部分指数样本调整名单和备选名单见下表,其他上证指数调整名单见上海证券交易所网站 (www.sse.com.cn) 和中证指数公司网站 (www.csindex.com.cn)。

回测评价:换手率 (个数法)



数组数据 运行信息

id	截止日	上期截止日	基准	第1组	第2组	第3组	第4组	第5组
0	2020-02-28	2020-01-23	0.00	0.00	0.00	0.00	0.00	0.00
1	2020-03-31	2020-02-28	0.00	10.00	10.00	30.00	30.00	10.00
2	2020-04-30	2020-03-31	0.00	10.00	30.00	50.00	40.00	20.00
3	2020-05-29	2020-04-30	0.00	0.00	0.00	0.00	0.00	0.00
4	2020-06-30	2020-05-29	10.00	10.00	10.00	10.00	20.00	10.00
5	2020-07-31	2020-06-30	0.00	10.00	10.00	0.00	0.00	0.00
6	2020-08-31	2020-07-31	0.00	0.00	40.00	60.00	30.00	10.00
7	2020-09-30	2020-08-31	0.00	10.00	10.00	10.00	10.00	0.00
8	2020-10-30	2020-09-30	0.00	11.11	40.00	40.00	30.00	10.00
9	2020-11-30	2020-10-30	0.00	0.00	20.00	30.00	10.00	0.00
10	2020-12-31	2020-11-30	10.00	30.00	30.00	30.00	30.00	20.00

- 这里换手率的计算方式:是否有存在于'上期截止日'股票而不存于'截止日',被移除的股票除以上期截止日股票总数.
- 权重分结果类似因为组内等权重

id	2020-01-23	2020-02-28	2020-03-31	2020-04-30	2020-05-29	2020-06-30	2020-07-31	2020-08-31	2020-09-30	2020-10-30	2020-11-30	2020-12-31
0	贵州茅台	贵州茅台	贵州茅台	华夏幸福	华夏幸福	汇顶科技	贵州茅台	汇顶科技	汇顶科技	贵州茅台	三一重工	海天味业
1	伊利股份	伊利股份	伊利股份	贵州茅台	贵州茅台	贵州茅台	汇顶科技	三一重工	三一重工	三一重工	贵州茅台	贵州茅台
2	华夏幸福	华夏幸福	华夏幸福	海螺水泥	海螺水泥	海螺水泥	海螺水泥	伊利股份	伊利股份	伊利股份	伊利股份	恒生电子
3	海螺水泥	海螺水泥	三一重工	隆基绿能	隆基绿能	隆基绿能	隆基绿能	贵州茅台	贵州茅台	隆基绿能	隆基绿能	三一重工
4	中国平安	三一重工	海螺水泥	三一重工	三一重工	三一重工	三一重工	海螺水泥	海螺水泥	汇顶科技	汇顶科技	伊利股份
5	三一重工	中国中免	万华化学	伊利股份	伊利股份	伊利股份	伊利股份	隆基绿能	隆基绿能	海螺水泥	海螺水泥	隆基绿能
6	中国中免	中国平安	中国中免	恒瑞医药	恒瑞医药	恒瑞医药	恒瑞医药	恒瑞医药	恒瑞医药	恒瑞医药	恒瑞医药	汇顶科技
7	工业富联	工业富联	中国平安	万华化学	万华化学	万华化学	万华化学	工业富联	工业富联	工业富联	中国中免	海螺水泥
8	恒瑞医药	恒瑞医药	恒瑞医药	工业富联	工业富联	工业富联	工业富联	万华化学	万华化学	中国中免	工业富联	恒瑞医药
9	海尔智家	海尔智家	工业富联	中国平安	中国平安	中国平安	保利发展	保利发展	保利发展	中国平安	中国平安	韦尔股份



回测评价:收益率检验

- 此处夏普比率不考虑无风险利率!
- 这里的胜率（按月为单位计算）是分组回报率为正数的比例. 比如第一组的63.64%来自7/11, 11个调仓日里7次第一组回报率高于0.
- 最大回撤%(日线)是局部高点对应之后的最低点的跌幅, 取最大的展示. 依据每组的日收盘价计算回报率.
- 平均收益是根据11个调仓期平均计算.

return obj.**GetTrailingReturn**(cy_month())

id	组名	整体				
		平均收益(%)	标准差(%)	夏普比率	胜率(%)	最大回撤%(日线)
0	第1组	3.95	7.01	0.56	63.64	-16.96
1	第2组	1.44	5.89	0.24	54.55	-18.51
2	第3组	0.98	5.93	0.17	54.55	-15.21
3	第4组	1.05	6.25	0.17	45.45	-12.73
4	第5组	0.47	5.54	0.09	54.55	-17.72





obj.GetGroupReturn(0,cy_month());

开始日	截止日	交易点数	基准	第1组	第2组	第3组	第4组	第5组
2020-01-01	2020-01-23	1	-2.83	-0.10	-0.10	-0.10	-0.10	-0.10
2020-02-01	2020-02-28	20	-3.80	-3.15	-4.45	-3.42	-3.77	-0.63
2020-03-01	2020-03-31	22	-4.67	-5.76	-7.98	-5.28	-1.43	-7.60
2020-04-01	2020-04-30	21	6.42	11.34	9.27	2.94	3.16	2.20
2020-05-01	2020-05-29	18	-1.93	-1.51	-3.50	-0.50	-4.16	-4.12
2020-06-01	2020-06-30	20	4.82	9.14	1.73	9.51	1.96	3.27
2020-07-01	2020-07-31	23	10.44	13.59	10.73	10.91	17.29	9.69
2020-08-01	2020-08-31	21	2.91	3.57	4.89	0.62	-0.89	1.18
2020-09-01	2020-09-30	22	-3.33	-2.53	-0.59	-6.61	-4.16	-7.69
2020-10-01	2020-10-30	16	1.48	2.51	0.48	2.54	2.29	-1.67
2020-11-01	2020-11-30	21	5.76	2.52	6.56	5.22	4.68	6.98
2020-12-01	2020-12-31	23	4.94	13.74	-1.29	-5.10	-3.38	3.59

- 这里的胜率（按月为单位计算）是分组回报率为正数的比例. 比如第一组的63.64%来自7/11, 11个调仓期里7次第一组回报率高于0.

组名	整体				
	平均收益(%)	标准差(%)	夏普比率	胜率(%)	最大回撤%(日线)
第1组	3.95	7.01	0.56	63.64	-16.96
第2组	1.44	5.89	0.24	54.55	-18.51
第3组	0.98	5.93	0.17	54.55	-15.21
第4组	1.05	6.25	0.17	45.45	-12.73
第5组	0.47	5.54	0.09	54.55	-17.72



回测评价:因子显著检验

- “超额”指针对基准的同时段回报率。超额概率是指每个调仓周转超过基准的比例. 比如第一组有9/11次超过基准.
- 跟踪误差= $\sqrt{\text{var}(r_{\text{group}} - r_{\text{bench}})}$, bench是基准.
- 信息比率: IR=超额收益/跟踪误差, 一个单位的跟踪误差有多少额外超额收益补偿. 类似夏普比率的表达方式.
- 如果基准的波动较大,那么波动较小的分组组合也会生成较大的跟踪误差,此时IR的所代表的的信息就会产生误差.
- 超额概率,t-stat,p-value (%)源自统一检测过程,假设检验.
- 使用角度我保留意见, 回忆一下假设检验的基础, 样本和总体.

```
return obj.GetSignificanceStatistics(cy_month());
```

组号	组名	整体					
		超额均值(%)	跟踪误差(%)	信息比率	超额概率(%)	T-Stat	P-Value
0	第1组	1.86	3.27	0.57	81.82	1.88	95.54
1	第2组	-0.65	2.81	-0.23	45.45	-0.77	22.87
2	第3组	-1.11	3.76	-0.30	45.45	-0.98	17.49
3	第4组	-1.04	3.94	-0.26	36.36	-0.88	20.07
4	第5组	-1.62	2.25	-0.72	18.18	-2.40	1.88

H_0 :第i组回报率-基准回报率 ≥ 0

H_1 :第i组回报率-基准回报率 < 0

p-value 是t-stat左侧的概率

t-stat越大说明超额均值越大

如果超额均值是负数如何理解?



回测评价:因子区分度检验

- 这部分和学界的方法相似,目的是提炼出因子ROE对回报率的贡献.
- 标准差根据每个调仓期的收益差计算得出.
- 胜率是第一组收益率大于最后一组的次数比例(共计9/11个调仓期第一组回报率大于最后一组).
- t-stat检验和p-value根据如下标准

H_0 :第1组回报率-最后组回报率 ≥ 0

H_1 :第1组回报率-最后组回报率 < 0

return obj.**GetLongShortStatistics**(cy_month())

序号	项目	整体					
		平均收益(%)	标准差(%)	夏普比率	胜率(%)	T-Stat	P-Value
0	第一组vs最后一组	3.48	4.35	0.80	81.82	2.65	98.78

序号	开始日	截止日	交易点数	基准	第1组	第2组	第3组	第4组	第5组
0	2020-01-01	2020-01-23	1	-2.83	-0.10	-0.10	-0.10	-0.10	-0.10
1	2020-02-01	2020-02-28	20	-3.80	-3.15	-4.45	-3.42	-3.77	-0.63
2	2020-03-01	2020-03-31	22	-4.67	-5.76	-7.98	-5.28	-1.43	-7.60
3	2020-04-01	2020-04-30	21	6.42	11.34	9.27	2.94	3.16	2.20
4	2020-05-01	2020-05-29	18	-1.93	-1.51	-3.50	-0.50	-4.16	-4.12
5	2020-06-01	2020-06-30	20	4.82	9.14	1.73	9.51	1.96	3.27
6	2020-07-01	2020-07-31	23	10.44	13.59	10.73	10.91	17.29	9.69
7	2020-08-01	2020-08-31	21	2.91	3.57	4.89	0.62	-0.89	1.18
8	2020-09-01	2020-09-30	22	-3.33	-2.53	-0.59	-6.61	-4.16	-7.69
9	2020-10-01	2020-10-30	16	1.48	2.51	0.48	2.54	2.29	-1.67
10	2020-11-01	2020-11-30	21	5.76	2.52	6.56	5.22	4.68	6.98
11	2020-12-01	2020-12-31	23	4.94	13.74	-1.29	-5.10	-3.38	3.59



因子延续性检验

- 检验t期因子值与t+1期的股票回报率的相关系数.
- 范例: 2020-01-23期的各公司的因子值与2020-01-24至2020-02-28对应公司的股票回报率做相关系数计算.
- 如果IC值持续保持同一方向说明因子延续性较强.
- p-value是针对 $H_0: IC=0, H_1: IC \neq 0$. (Two Tails)
- 如果 $IC > 0$ 或 $IC < 0$ 并且持续同一方向,对于时期的p-value越小说明越显著不同于0或是越好.
- 根据t检验:

$$t = \frac{IC\sqrt{n-2}}{\sqrt{1-IC^2}}$$

n指周期内的交易日天数

return obj.**GetInformationCoefficient()**;

数组数据		运行信息			
	因子截止日	检验开始日	检验截止日	IC	P-value
0	2020-01-23	2020-01-24	2020-02-28	-0.01	0.96
1	2020-02-28	2020-02-29	2020-03-31	0.01	0.95
2	2020-03-31	2020-04-01	2020-04-30	0.46	0.00
3	2020-04-30	2020-05-01	2020-05-29	0.17	0.23
4	2020-05-29	2020-05-30	2020-06-30	0.16	0.26
5	2020-06-30	2020-07-01	2020-07-31	0.00	1.00
6	2020-07-31	2020-08-01	2020-08-31	0.12	0.40
7	2020-08-31	2020-09-01	2020-09-30	0.28	0.05
8	2020-09-30	2020-10-01	2020-10-30	0.13	0.37
9	2020-10-30	2020-10-31	2020-11-30	-0.15	0.30
10	2020-11-30	2020-12-01	2020-12-31	0.37	0.01



回测评价:因子稳定性检验

- 通过前后截止日对比因子值的变化对股票所在组的影响.
- 第一组的企业是否持续输出并且稳定输入因子值让其长期保持在第一组.
- 如果按此标准, 越稳定的因子会造就较低的调仓成本.

return obj.GetSignalStability()

序号	截止日	上期截止日	间隔期数	自相关系数	买入衰减(%)	买入反转(%)	卖出衰减(%)	卖出反转(%)
0	2020-02-28	2020-01-23	1	1.00	100.00	0.00	100.00	0.00
1	2020-03-31	2020-02-28	1	0.97	90.00	0.00	90.00	0.00
2	2020-04-30	2020-03-31	1	0.92	90.00	0.00	80.00	0.00
3	2020-05-29	2020-04-30	1	1.00	100.00	0.00	100.00	0.00
4	2020-06-30	2020-05-29	1	1.00	90.00	0.00	90.00	0.00
5	2020-07-31	2020-06-30	1	1.00	90.00	0.00	100.00	0.00
6	2020-08-31	2020-07-31	1	0.94	100.00	0.00	90.00	0.00
7	2020-09-30	2020-08-31	1	1.00	90.00	0.00	100.00	0.00
8	2020-10-30	2020-09-30	1	0.97	88.89	0.00	90.00	0.00
9	2020-11-30	2020-10-30	1	0.99	100.00	0.00	100.00	0.00
10	2020-12-31	2020-11-30	1	1.00	70.00	0.00	80.00	0.00

评价指标	说明	备注
自相关系数	前后期样本的因子暴露相关系数	越大说明信号越稳定
买入衰减(%)	T 期 Top 组的股票成分, 在 T+1 期继续在 Top 组中的比例。	该值越大说明买入组成分越稳定
买入反转(%)	T 期 Top 组的股票成分, 在 T+1 期转变为在 Bottom 组中的比例。	越小越好
卖出衰减(%)	T 期 Bottom 组的股票成分, 在 T+1 期继续在 Bottom 组中的比例	该值越大说明卖出组成分越稳定
卖出反转(%)	T 期 Bottom 组的股票成分, 在 T+1 期不在 Bottom 组, 而是在买入组, 即 Top 组的比例, 即完全由卖出信号反转为买入信号。	越小越好






回测评价:衰减分析

- 每个调仓期计算所有股票因子值和对应下一期的股票回报率为IC. 共计11个IC.
- 随着间隔(lag)增加,可用于计算IC的调仓期会减少(同为样本数).
- 显著比率(%)是根据每个调仓期计算的IC>0的比列.
- 随着间隔增加, 如果IC开始趋近于0意味着因子有效性(可预测性)衰减.

```
return obj.GetDecay_InformationCoefficient(FBegT, FendT, 1, 5)
```

	间隔期数	IC	IC_Std	IC_IR	显著比率(%)	样本数
0	1	0.14	0.18	0.79	100.00	11
1	2	0.19	0.16	1.23	100.00	10
2	3	0.24	0.11	2.24	100.00	9
3	4	0.28	0.07	3.83	100.00	8
4	5	0.29	0.06	5.00	100.00	7

$$IC = \text{corr}(f_t, R_{t+\text{lag}})$$



因子贡献度检验(单因子)

- 对t期因子值和t+1期对应股票回报率做线性回归
- 根据系数正负推荐因子方向.

return obj.GetLinearEvaluationResult(0 or 1)

期	截止日	回归系数		R方	T值		P值		推荐因子配置比例(%)
		ROE	常数项		ROE	常数项	ROE	常数项	ROE
0	2020-01-23	-0.05	-3.08	0.00	-0.05	-3.16	0.96	0.00	-100.00
1	2020-02-28	0.08	-5.57	0.00	0.07	-4.95	0.95	0.00	100.00
2	2020-03-31	3.57	5.85	0.21	3.58	5.92	0.00	0.00	100.00
3	2020-04-30	0.93	-2.76	0.03	1.20	-3.61	0.23	0.00	100.00
4	2020-05-29	1.65	5.14	0.03	1.14	3.59	0.26	0.00	100.00
5	2020-06-30	0.00	12.40	0.00	0.00	5.39	1.00	0.00	-100.00
6	2020-07-31	0.87	1.91	0.02	0.86	1.91	0.40	0.06	100.00
7	2020-08-31	1.76	-4.30	0.08	2.03	-5.02	0.05	0.00	100.00
8	2020-09-30	0.93	1.26	0.02	0.91	1.25	0.37	0.22	100.00
9	2020-10-30	-1.21	5.22	0.02	-1.06	4.60	0.30	0.00	-100.00
10	2020-11-30	4.94	1.57	0.14	2.79	0.90	0.01	0.37	100.00

期	项目	回归系数		R方	T值		P值		推荐因子配置比例(%)
		ROE	常数项		ROE	常数项	ROE	常数项	ROE
0	整体	1.22	1.60	0.01	2.72	3.60	0.01	0.00	100.00



因子贡献度检验(多因子)

- 多因子面临的最大问题就是如何分配各个因子之间的权重.
- 额外增加一个因子:
- FactorDef:= array(('因子名称':'ROE','因子公式':'Last12MData(Rdate,9900100)','因子方向':1,'因子比例(%)':50),
('因子名称':'销售毛利率','因子公式':'Last12MData(Rdate,9900103)','因子方向':1,'因子比例(%)':50));

日期	截止日	回归系数			R方	T值			P值			推荐因子配置比例(%)	
		ROE	销售毛利率	常数项		ROE	销售毛利率	常数项	ROE	销售毛利率	常数项	ROE	销售毛利率
0	2020-01-23	-0.56	1.79	-3.08	0.06	-0.56	1.78	-3.23	0.58	0.08	0.00	-23.93	76.07
1	2020-02-28	-0.91	3.39	-5.57	0.17	-0.83	3.09	-5.37	0.41	0.00	0.00	-21.18	78.82
2	2020-03-31	3.56	0.03	5.85	0.21	3.37	0.02	5.86	0.00	0.98	0.00	99.30	0.70
3	2020-04-30	0.30	1.93	-2.76	0.14	0.39	2.49	-3.80	0.70	0.02	0.00	13.61	86.39
4	2020-05-29	0.75	2.77	5.14	0.09	0.51	1.86	3.68	0.62	0.07	0.00	21.35	78.65
5	2020-06-30	-0.16	0.50	12.40	0.00	-0.07	0.20	5.34	0.95	0.84	0.00	-24.68	75.32
6	2020-07-31	1.53	-1.94	1.91	0.08	1.46	-1.86	1.96	0.15	0.07	0.06	43.97	-56.03
7	2020-08-31	2.16	-1.15	-4.30	0.11	2.35	-1.25	-5.04	0.02	0.22	0.00	65.35	-34.65
8	2020-09-30	1.16	-0.68	1.26	0.03	1.07	-0.63	1.24	0.29	0.53	0.22	62.92	-37.08
9	2020-10-30	-0.54	-2.02	5.22	0.08	-0.45	-1.69	4.69	0.66	0.10	0.00	-20.97	-79.03
10	2020-11-30	4.28	1.73	1.57	0.15	2.23	0.90	0.90	0.03	0.37	0.37	71.22	28.78

日期	项目	回归系数			R方	T值			P值			推荐因子配置比例(%)	
		ROE	销售毛利率	常数项		ROE	销售毛利率	常数项	ROE	销售毛利率	常数项	ROE	销售毛利率
0	整体	1.02	0.62	1.60	0.02	2.15	1.29	3.60	0.03	0.20	0.00	62.48	37.52