

证券研究报告 / 金融工程研究报告

## 因子计算标准化流程中的若干问题及解决方案

--因子选股系列之十

### 报告摘要:

#### 去极值标准化流程中存在的问题:

- 1) 没有对尾分布进行检验, 对于如何判定极端值并没有明确定义, 在中位数/3sigma/分位数法中对规定边界以外的所有数据点拉回到同一取值丢失分布信息, 在多因子组合时可能由于尾部不保序的原因导致进入组合的标的出现差异;
- 2) 没有对不同极值处理方法的优缺点和适用性进行讨论, 而实际检验发现, 不同数据分布特征的数据在不同极值处理方法下得到的分布结果差异非常大;
- 3) 没有讨论数据正态转换的必要性: 通常来看, 对于非幂律分布的样本, 正确的正态转换可以有效将数据分布向正态收敛, 且不会降低因子表现, 可以纠正偏度和峰度, 降低误判为极端值的概率, 同时保留了原始数据的多峰或其他的分布特征。

**线性中性化流程中存在的问题:** 若线性中性化是有效的, 则残差应满足正态独立同分布, 但现实中大部分因子的中性化残差都未满足这一条件。一方面是由于因子不会和解释变量强线性, 否则可以直接使用相关因子替代; 第二方面是由于正交变量选择不足或者过度; 第三方面原因是很多因子在不同行业、不同市值域可能体现出分层线性差异, 而非全局线性, 使得回归系数取值有偏差; 第四方面也与去极值操作时将边界点拉平有关, 带来一定的拟合误差; 第五点与部分数据样本的过度肥尾有关, 由于被解释变量是幂律分布, 正交变量无法改变其分布特性, 使得线性回归的残差因子性质较差。

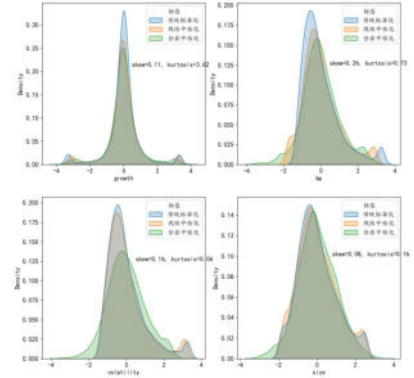
#### 提供以下规范化流程方案:

- 1) 对原始数据进行 Box-Cox 变换, 改善非幂律分布的样本数据的分布性质, 降低尾部数据错误判定为异常值的可能, 同时保留其他分布特征;
- 2) 计算左右尾经验累积分布结果, 并计算与幂律分布临界值的大小, 若某侧尾部的概率分布更高, 则进行分位数法截尾操作;
- 3) 使用 Barra 中的规范化方法计算其标准化值;
- 4) 市值因子的中性通过行业内标准化, 而非以行业作为哑变量执行回归处理; 其他因子行业内部与市值中性; 或进行行业哑变量的行业中性, 而后行业中性市值因子分层, 分别在层内进行市值回归处理。

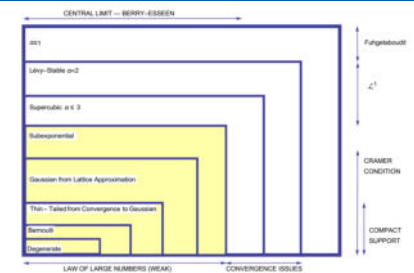
**处理结果对比:** 收益和 IC 表现变化有多种可能的, 如果此前的因子分布性质较好, 可以通过线性进行完善的风格剔除, 则传统与改进做法的收益表现差异不大; 如果此前的因子性质较差, 大概率导致全域线性风格中性化的风格剔除不完全, 则分域做法的收益表现可好可坏, 取决于此前未被剔除的风格 beta 对因子是加强作用还是削弱作用, 但整体来看, 可以降低风格 beta 的影响。多个因子测试显示, 基本面因子小幅改善, 量价因子较为有效, 各因子测试的分组线性性提高, 超额分年度收益更稳定, 受到市值因子影响更小。

**风险提示:** 市场变化、政策变化风险, 模型失效风险

### 常见因子不同各阶段密度函数



### 肥尾分布表



### 相关报告

- 《个股及基金风格收益与风险归因分析》  
--20160826
- 《因子模型理论与实践及因子检验方法--因子选股系列之一》  
--20220921
- 《基于高斯秩的估值因子改进》  
--20240918

证券分析师: 王琦

执业证书编号: S0550521100001

021-61002390 wangqi\_5636@nesc.cn

## 目 录

<b>1. 引言 .....</b>	<b>5</b>
<b>2. 因子计算标准化流程中的若干问题及解决方案 .....</b>	<b>5</b>
2.1. 数据预处理、去极值与标准化中的问题和解决方案.....	5
2.1.1. 问题 .....	5
2.1.2. 解决方案 .....	9
2.2. 中性化过程中的问题和解决方案.....	11
2.2.1. 问题 .....	11
2.2.2. 解决方案 .....	14
2.3. 总结.....	17
<b>3. 常见单因子测试对比 .....</b>	<b>17</b>
3.1. SUE .....	17
3.2. 单季度净利润增速.....	19
3.3. BP .....	20
3.4. 20 日波动率.....	22
3.5. SUE+低波复合结果对比.....	24
<b>4. JASON 复合策略的提升 .....</b>	<b>26</b>
4.1. JASON 组合对比表现 .....	26
4.2. 各单因子对比表现.....	29
<b>5. 风险提示.....</b>	<b>32</b>

图表目录

图 1: 对数市值去极值标准化分布对比(2024/08/30).....	6
图 2: 对数市值因子分位数法去极值标准化后数值范围 .....	7
图 3: 单季度净利润增速去极值标准化分布对比(2024Q2).....	7
图 4: 流通市值经 BOX-COX 转换前后密度分布.....	8
图 5: 净利润增速经 BOX-COX 转换前后密度分布.....	8
图 6: 肥尾分布表 .....	9
图 7: 流通市值-去极值标准化分布结果对比 .....	10
图 8: 单季度净利润增长率-去极值标准化分布结果对比 .....	11
图 9: 净利润增速因子与对数市值的标准化和行业中性化散点图 .....	12
图 10: 净利润增速因子市值中性化后散点图及前后密度函数对比 .....	12
图 11: BP 因子与对数市值不同处理阶段关系对比.....	13
图 12: 不同因子各处理阶段密度分布对比 .....	15
图 13: 净利润增速 v.s. 市值 分层秩相关性 .....	16
图 14: BP v.s. 市值 分层秩相关性 .....	16
图 15: 波动率 v.s. 市值 分层秩相关性 .....	16
图 16: SUE 净值-新版规范化处理 .....	18
图 17: SUE 净值-传统规范化处理 .....	18
图 18: SUE IC 和累计 IC-新版规范化处理 .....	18
图 19: SUE IC 和累计 IC-传统规范化处理 .....	18
图 20: 单季度净利润增速净值-新版规范化处理 .....	19
图 21: 单季度净利润增速净值-传统规范化处理 .....	19
图 22: 单季度净利润增速 IC 和累计 IC-新版 .....	19
图 23: 单季度净利润增速 IC 和累计 IC-传统 .....	19
图 24: BP-新版规范化处理 .....	21
图 25: BP-传统规范化处理 .....	21
图 26: BP-IC 和累计 IC-新版规范化处理 .....	21
图 27: BP-IC 和累计 IC-传统规范化处理 .....	21
图 28: 20 日波动率-新版规范化处理 .....	22
图 29: 20 日波动率-传统规范化处理 .....	22
图 30: 20 日动量 IC 和累计 IC-新版规范化处理 .....	23
图 31: 20 日动量 IC 和累计 IC-传统规范化处理 .....	23
图 32: SUE+低波-新版规范化处理 .....	24
图 33: SUE+低波-传统规范化处理 .....	24
图 34: SUE+低波 IC 和累计 IC-新版规范化处理 .....	24
图 35: SUE+低波 IC 和累计 IC-传统规范化处理 .....	24
图 36: JASON 净值-新版规范化处理 .....	26
图 37: JASON IC-新版规范化处理 .....	26
图 38: JASON 净值-传统规范化处理 .....	27
图 39: JASON IC-规范化处理 .....	27
图 40: J 净值-新版规范化 .....	29
图 41: J-IC-新版规范化 .....	29
图 42: J 净值-传统规范化 .....	29
图 43: J-IC-传统规范化 .....	29
图 44: A 净值-新版规范化 .....	30
图 45: A-IC-新版规范化 .....	30
图 46: A 净值-传统规范化 .....	30
图 47: A-IC-传统规范化 .....	30
图 48: S 净值-新版规范化 .....	30
图 49: S-IC-新版规范化 .....	30

图 50: S 净值-传统规范化.....	31
图 51: S-IC-传统规范化.....	31
图 52: O 净值-新版规范化.....	31
图 53: O-IC-新版规范化.....	31
图 54: O 净值-传统规范化.....	31
图 55: O-IC-传统规范化.....	31
图 56: N 净值-新版规范化.....	32
图 57: N-IC-新版规范化.....	32
图 58: N 净值-传统规范化.....	32
图 59: N-IC-传统规范化.....	32
表 1: SUE 风险收益指标-新版规范化处理.....	18
表 2: SUE 风险收益指标-传统规范化处理.....	19
表 3: 单季度净利润增速风险收益指标-新版规范化处理.....	20
表 4: 单季度净利润增速风险收益指标-传统规范化处理.....	20
表 5: BP 风险收益指标-新版规范化处理.....	21
表 6: BP 风险收益指标-传统规范化处理.....	22
表 7: 20 日波动率风险收益指标-新版规范化处理.....	23
表 8: 20 日波动率风险收益指标-传统规范化处理.....	23
表 9: SUE+低波风险收益指标-新版规范化处理.....	25
表 10: SUE+低波风险收益指标-传统规范化处理.....	25
表 11: JASON 风险收益指标-新版规范化处理.....	27
表 12: JASON 风险收益指标-传统规范化处理.....	27
表 13: JASON 风险收益排序-新版规范化处理.....	28
表 14: JASON 风险收益排序-传统规范化处理.....	28

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/147032121134006166>