

量化多因子选股框架

——金融工程专题报告

2024年9月4日



核心观点

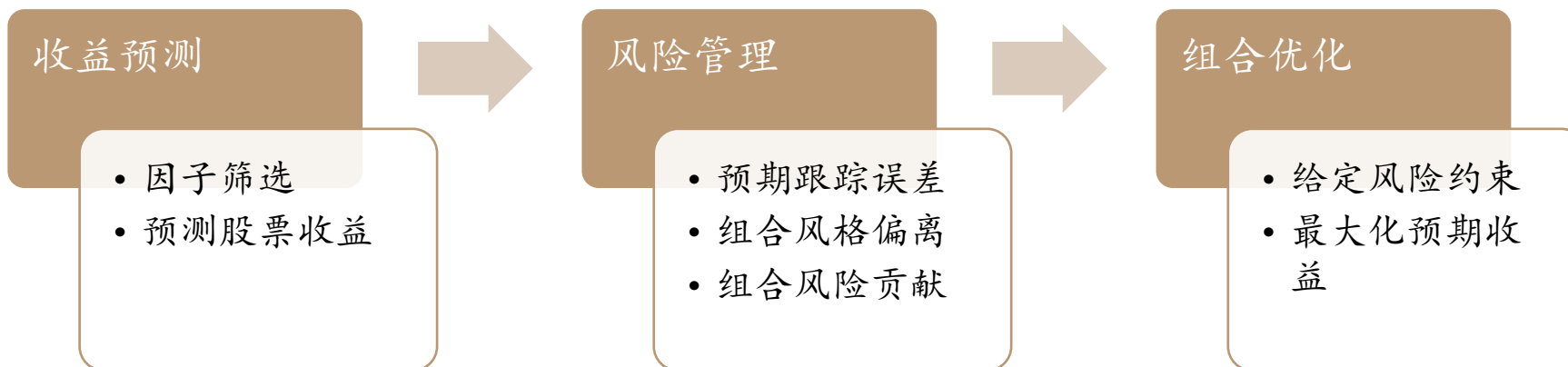
- **【核心结论】** 本报告详细介绍了量化多因子选股的流程与框架，涵盖因子预处理、评价及收益预测模型等关键步骤。
- **【报告亮点】** 系统性整理了从因子测试、因子筛选到组合构建的全过程，为高效构建量化投资组合提供了清晰路径。
- **【主要逻辑】**
 - **单因子测试**
 - ✓ 介绍单因子预处理的基本方法，主要包括去极值、标准化、中心化。
 - ✓ 因子检验的评价维度和评价方式，其中评价方式主要有ICIR法、分组测试法、双变量排序法和回归法。
 - **收益预测模型**
 - ✓ 对因子进行筛选，判断其是否能够纳入多因子组合。
 - ✓ 用一系列筛选得到的因子预测个股收益，这一步本质可理解为多因子的聚合。常见方法可分为线性和非线性，其中线性法主要有加权和回归两种方式，非线性主要包括神经网络、树模型等。
 - **组合构建**
 - ✓ 基于得到的个股收益预测，计算组合的净值曲线和收益风险特征。
 - ✓ 组合构建的主要步骤包括设定选股池与调仓周期、以及确定选股逻辑和权重。其中，选股逻辑和权重一般基于人工逻辑直接构建得到，或者利用优化算法在更复杂的选股约束下进行求解。
 - ✓ 我们基于上述框架，结合多因子复合打分构建了一个中证500指数增强策略，业绩表现优于基准。
- **风险提示：**数据 第三方，或有遗漏、滞后、误差；选股研究框架及筛选指标为西部证券自建，或有考虑不周之处；股票历史业绩不代表未来表现，研究框架也会随时间变化。



量化多因子选股基本框架

- 收益预测模型：通过历史估计给出股票未来的预期收益率。
- 风险管理模型：控制组合与业绩基准之间的跟踪误差。
- 组合权重优化：在给定的风险约束下，最大化组合的预期收益率。

图：多因子组合构建基本流程





目 录

CONTENTS

01

单因子测试

02

收益预测模型

03

组合构建



1 单因子测试

➤ 1.1 备选因子：基于原始指标或开发；挑选时需要注意因子是否**直接、直观且有意义**。

- 直接 (incisive)：通过因子能够明确对股票进行分类。
- 直观 (intuitive)：对股票的分类标准具有可以解释的逻辑性。
- 有意义 (interesting)：分类后的股票风险收益特征具有明显差异。

表：确定备选因子的3个I

备选因子	Incisive	Intuitive	Interesting
首字母	■	□	□
公司地点	■	■	□
股票盈利	■	■	■

参考文献：Active Portfolio Management, by Grinold and Kahn.



1 单因子测试

➤ 1.2 因子预处理

- 01 - 去极值
- 02 - 标准化
- 03 - 中性化



1 单因子测试

➤ 1.2 因子预处理

- **01 - 去极值：**因子数据中的极端值会影响分析结果，离群值会扭曲因子与收益率之间的相关性估计。
 - ✓ 去极值的基本逻辑：确定上下限，将超出范围的因子值修改为上下限。
 - ✓ 常见的去极值方法有：
 - ① MAD法：上下限为 $x_M \pm n * D_{MAD}$ ， x_M ：序列 x_i 的中位数， D_{MAD} ：序列 $|x_i - x_M|$ 的中位数
 - ② 3σ 法：上下限为 $x_M \pm n * \sigma$ ， x_M ：序列 x_i 的均值， σ ：序列 x_i 的标准差
 - ③ 百分位法：上限分位数和下限分位数（一般为97.5%和2.5%）。



1 单因子测试

➤ 1.2 因子预处理

- **02 - 标准化**: 不同的因子数据往往具有不同的单位和数值范围, 这种差异会导致数值较大的因子在分析中占据过大的比重, 而数值较小的因子则会被相对削弱。因此, 通常需要对这些因子数据进行标准化处理, 以消除量纲的影响。

✓ 常见标准化方法有:

① Z-score

② Max-Min

③ 取对数

④ 排序标准化



1 单因子测试

➤ 1.2 因子预处理

- **03 - 中性化**：剔除因子数据中显著的风险暴露，因为若使用具有明显风险暴露的因子值作为选股条件，会导致筛选出的股票集中在特定的风险暴露上。
 - ✓ 中性化的基本逻辑：因子值作为 y ，风险因子作为 x ，进行线性回归，回归残差即为风险中性后的因子值。
 - ✓ 常见中性化方法有：
 - ① 行业中性
 - ② 市值中性
 - ③ 其他风险因子中性



1 单因子测试

➤ 1.3 因子检验

- 因子的评价维度：

- ✓ 因子单调性：因子单调性越好，收益区分能力越强。
- ✓ 因子有效性：因子对未来收益的解释能力。
- ✓ 因子稳定性：因子的有效性是否稳定、可持续
- ✓ 因子时效性：在过去有效的因子，现在或将来不一定有效。



1 单因子测试

➤ 1.3 因子检验

• 因子评价方式：

- ✓ 01 - IC、IR：常用于评价因子有效性和稳定性。
- ✓ 02 - 分组测试：将股票按照因子值大小分组，统计每组在下一周期的收益。
- ✓ 03 - 双变量排序：计算控制因子1下的因子2分组收益，以防止排序受其他因子的干扰。
- ✓ 04 - 回归法：将T期的因子暴露与T+1期股票收益率进行线性回归。



1 单因子测试

➤ 1.3 因子检验

✓ 01 - IC、IR：常用于评价因子有效性和稳定性。

- ◆ Pearson相关系数（IC）：当期因子值与下期股票收益之间的线性相关性。
- ◆ Spearman相关系数（RankIC）：当期因子与下期股票收益之间的序数相关性。
- ◆ ICIR信息比率（IC Information Ratio）：IC的均值与IC标准差之比，代表因子获取稳定Alpha的能力。

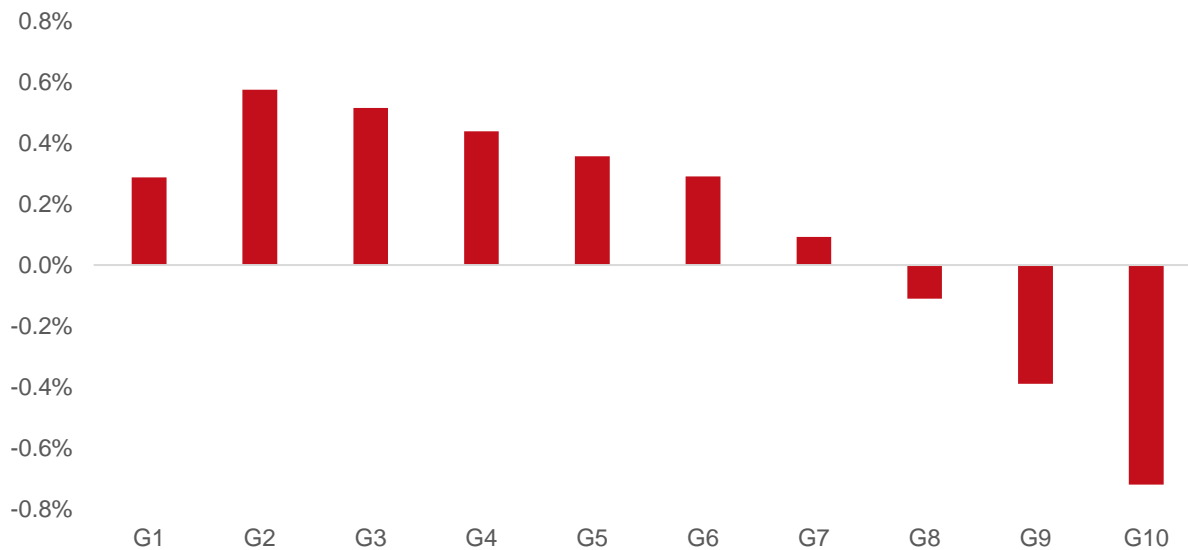


1 单因子测试

➤ 1.3 因子检验

- ✓ 02 - 分组测试：将股票按照因子值大小分组，统计每组在下一周期的收益。
 - ◆ 每组间的收益差异越大、单调性越明显，因子有效性越好。

图：PB因子分组月均收益（2018.01-2024.07）



注：横轴为t月按照股票的PB从小到大等分成10组，纵轴代表每个PB组合t+1月收益的均值



1 单因子测试

➤ 1.3 因子检验

✓ 03 - 双变量排序：计算控制因子1下的因子2分组收益，以防止排序受其他因子的干扰。

◆ （有条件）双变量排序流程：

1) 按因子值1将股票分为10组。

2) 在各因子值1分组中按因子值2分为10组。

3) 将不同因子值1分组内相同因子值2排序分组的收益加总，得到控制因子值1下的因子值2的分组收益。



1 单因子测试

➤ 1.3 因子检验

✓ 04 - 回归法：将因子T期的因子暴露与T+1期股票收益率进行截面线性回归。

$$r^{t+1} = X^t f^t + \varepsilon^t$$

- ◆ 回归系数 f^t ：因子当期的收益率（因子溢价）。
- ◆ 因子收益率显著度水平（t-test）：对于每一期的截面回归，都可以得到一个因子收益率 f^t ，对于 f^t 序列进行t检验，以观察因子收益率序列的方向一致性。

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/898065063127006124>