# 论文标题

**摘要：**

**关键词：**

## 引言

简要背景知识 + 论文内容方法 + 结果

此处绘制流程图，概述一下

## 单因子测试与分析

### 建立回归模型

我们使用以下线性回归模型进行单因子显著性测试，对因子进行时间横面的回归：

股票 在第期的收益率；

: 股票 在因子 上的因子暴露值；

在第 期，模型回归运算所得的因子收益率；

回归运算所得到的股票在第期时的残差。

我们使用 auto-trader软件提取沪深300成分股的日频数据和因子数据，在每一个时间横截面上使用上述模型进行最小二乘回归（WLS）。使用当期股票因子暴露值和下一期股票收益率做线性拟合。在进行因子测试时，由于要求每月月初进行调仓，我们以上一个月的各股票的因子暴露值作为自变量（第 T 期），上一个月各股票月平均收益率为因变量（第 T + 1期），拟合得到一个线性模型，然后将本月月初的股票因子值输入到拟合完成的线性模型，输出的各股票本月平均收益率进行选股回测。

### 数据采集

我们根据点宽网的BP因子数据字典，使用auto-trader软件提取股票日频因子数据。

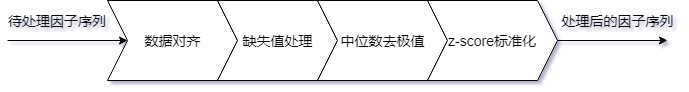
**样本范围：**沪深300成分股，剔除停牌的等不正常股票。

**样本期：**2016年1月1日至2018年9月30日，按月进行提取。

**因子选择**：我们具体测试了基础科目衍生类、质量类、收益风险类、情绪类、成长类、常用技术指标类、动量类、价值类、每股指标类、模式识别类、特色技术指标、行业与分析师类共十二大类因子。我们为每一类的因子随机挑选了10~20个因子进行测试。由于挑选因子较多以及本文篇幅有限，部分表现较优的因子定义将会在回测部分进行显示。

### 数据预处理

在使用线性模型进行因子回归的之前，我们需要对提取的各类别因子数据进行预处理步骤。大致流程如下：



**数据对齐：**上市公司财报的报告期和报告发布日期存在一定延迟，为了避免未来信息的干扰，在进行因子数据提取的时候，需要对日期进行修正，保证因子数据为当时所能获取的最新财报数据。

**缺失值：**提取得到的因子数据可能由于技术等原因出现缺失，为了避免缺失值影响拟合的效果，导致收益结果较差，对缺失值直接采取了移除的处理。

**中位数去极值法：**为了避免因子数据中存在个别极端值对回归产生较大的影响，我们使用“中位数去极值法”，将超过上下限的极端值用上下限继续替代，公式计算如下：

: 为待处理的原始因子序列；

: 为序列的中位数；

: 为处理后的因子序列；

: 为因子序列 中位数。

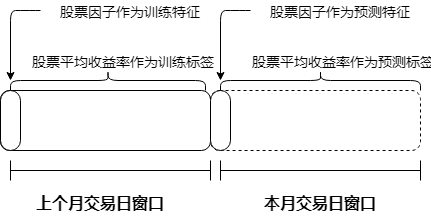
**z-score标准化：**由于各个因子的单位和量纲不同，为了使得因子之间存在可比性，需要对其进行z-score标准化处理，即对序列中的每一个因子，减去因子序列的均值，然后除以因子序列的标准差，使得因子序列近似成为一个符合均值为0，标准差为1的正态分布序列。

### 单因子回测

我们使用auto-trader软件获取特定单因子数据，每月初进行调仓，根据因子回归结果挑选较优或者排除较差的股票，经过回测得到测试报告。单因子整体测试思路如下：

1. 在月初第一天，获得前21天的股票K线数据和因子数据（近似为上一个月的数据），若未满足，则跳过该交易日；
2. 对上一个月初的因子数据进行预处理，计算上一个月股票池中各股票的平均收益率；
3. 使用单变量线性回归模型对处理后的因子数据和股票平均收益率进行拟合训练；
4. 获取本月初因子数据作为预测样本特征，输入训练后的线性回归模型得到本月各股票的预测平均收益率；
5. 视平均收益率前20%的股票为强势股，平均收益率后60%的股票为弱势股，每次调仓时，若持有弱势股则平仓；对未持仓的强势股，则以均等分配的可用资金进行开仓。

以一个月所有交易日为一个时间窗口单位，回归过程图示如下：



按照上述思路，从各大类因子中随机挑选小类因子进行因子回测，得到每一个因子相应的回测报告数据，主要涉及策略的年化夏普率(以月为单位，等于夏普率×，年化收益率，最大回撤率，信息比率等指标。在进行因子筛选时，主要考虑因素为年化夏普率，优先筛选每一类中年化夏普率最高的因子，当存在较多表现较优的因子时，综合各方面指标进行筛选。同时为了解决因子的多重共线性问题，我们测试了各大类因子间各因子历史序列的相关性，得到相关性矩阵，辨别出相关性较高的因子，对于这一类因子，采用因子剔除的方式进行处理。

以下展示各大类中表现较优因子的回测结果，并辅以简要的分析。

#### 基础科目衍生类因子

我们从BP因子数据字典中随机抽取了12个基础科目衍生类的因子进行测试，包括净运营资本( NetWorkingCapital )、毛利(GrossProfit)、企业自由现金流量(FCFF)、带息债务(IntDebt)、运营资本(WorkingCapital)、固定资产合计(TotalFixedAssets)、净利息费用(NetIntExpense)、价值变动净收益(ValueChgProfit)、留存收益(RetrainedEarnings)、财务费用(FinanExpenseTTM)、管理费用(AdminExpenseTTM)、净利润(NetProfitTTM)。经过历史选股回测之后，以年化夏普率为核心指标筛选出表现较优的6个因子（较差因子为了显示的简洁性，不予展示，下同），其各项指标如下：（结果保留两位小数，下同）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 因子ID | 年化夏普率 | 年化收益率 | 最大回撤率 | 信息比率 |
| GrossProfit | 144.04% | 7.50% | 10.27% | 78.33% |
| WorkingCapital | 59.16% | 4.57% | 22.86% | 54.80% |
| TotalFixedAssets | 92.33% | 5.61% | 13.38% | 63.15% |
| NetIntExpense | 216.74% | 11.91% | 13.67% | 105.11% |
| FinanExpenseTTM | 257.37% | **13.36%** | 12.96% | **115.41%** |
| AdminExpenseTTM | **279.01%** | 12.38% | **10.23%** | 113.61% |

表1：基础科目衍生类因子回归测试结果

从上述数据可看出管理费用因子（AdminExpenseTTM）经过回测后具有较高的年化夏普率和较小的最大回撤率，同时年化收益率和信息比率都有着较优的表现；而净利息费用（NetIntExpense）和财务费用（FinanExpenseTTM）的各项指标结果呈现较优，都可作为候选的因子。

从相关性矩阵可以看出，

---- FinanExpenseTTM AdminExpenseTTM

#### 4.2质量类因子

#### 4.3情绪类因子

#### 4.4成长类因子

#### 4.5价值类因子

#### 4.6每股指标类因子

#### 4.7行业分析师类因子

#### 4.8特色技术指标类因子

### 因子筛选总结

## 三、收益模型建立、对比与筛选

## 四、风险控制