

米筐科技因子风险模型数据

在量化交易中，因子数据是策略开发、风险预测、配置优化、风险敞口管理、业绩归因等步骤的核心部分。米筐科技为用户提供高质量的因子风险模型数据，帮助用户搭建完整、可靠的量化投研流程。相对于传统因子风险模型，米筐因子风险模型数据有如下改进：

- 基于申万行业分类定义行业因子，构建了 2014 年前后两套申万行业分类标准的映射¹，2014 年前提供两套行业暴露度（分别基于新旧行业分类标准）；
- 提供细分因子暴露度、个股对各指数贝塔等衍生因子数据，灵活应用于不同投研场景；
- 提供以常用指数（沪深 300、中证 500 及中证 800）成分股为股票池的因子收益率，便于挑选在特定指数有效的因子，构建指数增强策略（详见 3.2 节测试）；
- 特异风险模型中，对新股、复牌股、停牌股进行了优化处理，有效提升了风险预测精度和权重优化效果（详见 3.3 和 3.4 节测试）

表 1：数据详情

更新频率	日频率
增量更新时间	交易日早上 9 点前
数据缺失情况	无缺失值
数据种类	<ul style="list-style-type: none">• 个股暴露度、细分风格因子暴露度（2002 年 3 月 25 日 ~ 至今）• 个股对指数贝塔（2006 年 1 月 18 日 ~ 至今）• 个股特异收益（2002 年 3 月 26 日 ~ 至今）• 全市场因子收益率（2002 年 3 月 26 日 ~ 至今）• 以沪深 300、中证 500、及中证 800 指数成分股为股票池的因子收益率（2007 年 1 月 16 日 ~ 至今）• 日度因子协方差和特异风险（2006 年 6 月 28 日 ~ 至今）• 月度因子协方差和特异风险（2007 年 2 月 7 日 ~ 至今）• 季度因子协方差和特异风险（2009 年 3 月 29 日 ~ 至今）
因子数目	风格因子 10 个、行业因子 28 个、市场联动因子 1 个
行业分类方式	申万一级行业分类

¹ 申万在 2014 年 1 月 1 日曾对其行业分类标准进行过调整。

一 数据说明及数据 API.....	3
二 字段说明	7
三 测试结果.....	9
3.1 风格因子累积收益率和相关性	9
3.2 模型修正可决系数（ADJUSTED R-SQUARED）	10
3.3 偏差统计量（BIAS STATISTICS）测试.....	13
3.3.1 因子协方差 Bias Statistics.....	13
3.3.2 特异风险 Bias Statistics.....	15
3.4 风险最小化组合	18
3.4.1 测试组合构建及回测设置	19
3.4.2 每日调仓测试结果	19
3.4.3 每周调仓测试结果.....	21
3.4.4 每月调仓测试结果	23
3.4.5 每季度调仓测试结果.....	25
四 方法论介绍	27
4.1 多因子模型	27
4.2 因子暴露度	28
4.3 因子收益率	29
4.4 因子风险模型	30

一 数据说明及数据 API

表 1.1：数据说明

因子数据	解释
风格因子暴露度	个股对于特定风格因子的风险暴露。暴露度绝对值越大，则投资组合表现对因子表现变化越敏感。可用于投资组合风格评估、风险敞口管理等。部分风格因子由多个细分风格因子组合而成。暴露度的解释和计算规则详见本说明 4.2 部分。
细分风格因子暴露度	个股对细分风格因子的风险暴露。细分风格因子表示某一类风格中的细分风险维度，用户可根据实际投资研究需求，使用细分风格因子代替风格因子。
个股对指数贝塔	个股对指数（上证 50、沪深 300、中证 500 等）的原始贝塔值（未进行标准化，因此区别于贝塔因子），可用于指数跟踪或贝塔中性处理。
因子收益率	因子在给定股票池产生的超额收益。目前提供全市场、沪深 300、中证 500、中证 800 四个股票池的因子收益。用户可根据实际投资研究中所使用的股票池，选择相应的因子收益进行风格追踪和构建指数增强策略。显示和隐式因子收益率介绍见本说明 4.3 部分。
特异收益率	个股收益中无法被因子解释的部分。例如上市公司出现高管人员变动，可能引起股价剧烈波动，产生较大的特异收益。此时上市公司股价主要由消息面驱动，而与其所处行业、基本面、及市场行情相关性较低。
因子协方差	投资组合风险中能被因子解释的系统性风险部分。目前提供日度、月度、季度三套不同预测期限的风险模型，适用于不同调仓频率的交易策略。因子风险模型介绍见 4.4 部分，测试结果见 3.3.1 和 3.4 部分。
特异风险	投资组合风险中不能被因子解释的，与个股自身特殊因素相关的部分（见上述特异收益率解释）。目前提供日度、月度、季度三个不同预测期限的风险模型，适用于不同调仓频率的交易策略。特异风险测试结果见 3.3.2 和 3.4 部分。

1.1 获取一组股票的因子暴露度：

```
get_factor_exposure(order_book_ids, start_date, end_date, factors = None,  
industry_mapping = True )
```

参数如下：

参数	类型	解释
order_book_ids	str or list of str	证券代码（例如：‘600705.XSHG’）
start_date	str	开始日期（例如：‘2017-03-03’）
end_date	str	结束日期（例如：‘2017-03-20’）
factors	None or list of str	因子。默认获取全部风格因子暴露度（None）。风格因子、行业因子和市场联动字段名称见表 2.1 和 2.2
industry_mapping	boolean	是否按 2014 年后的申万行业分类标准计算行业暴露度。默认为 True。若取值为 False，则 2014 年前的行业暴露度按旧行业分类标准计算

返回 MultiIndex 的 pandas.DataFrame，index 第一个 level 为 order_book_id，第二个 level 为 date，columns 为因子字段名称。

1.2 获取一组股票的细分风格因子暴露度：

```
get_descriptor_exposure(order_book_ids, start_date, end_date, descriptors = None)
```

参数如下：

参数	类型	解释
order_book_ids	str or list of str	证券代码（例如：‘600705.XSHG’）
start_date	str	开始日期（例如：‘2017-03-03’）
end_date	str	结束日期（例如：‘2017-03-20’）
descriptors	None or list of str	细分风格因子。默认获取全部细分风格因子的暴露度（None）。细分风格因子字段名称见表 2.1

返回 MultiIndex 的 pandas.DataFrame，index 第一个 level 为 order_book_id，第二个 level 为 date，column 为细分风格因子字段名称。

1.3 获取一组股票的基准贝塔值：

```
get_stock_beta(order_book_ids, start_date, end_date, benchmark='000300.XSHG')
```

参数如下：

参数	类型	解释
order_book_ids	str or list of str	证券代码（例如：‘600705.XSHG’）
start_date	str	开始日期（例如：‘2017-03-03’）
end_date	str	结束日期（例如：‘2017-03-20’）
benchmark	str	基准指数。默认为沪深 300（‘000300.XSHG’），可选上证 50（‘000016.XSHG’）、中证 500（‘000905.XSHG’）、中证 800（‘000906.XSHG’）以及中证全指（‘000985.XSHG’）

返回 pandas.DataFrame，index 为日期，column 为个股的 order_book_id

1.4 获取因子收益率：

```
get_factor_return(start_date, end_date, factors=None, universe='whole_market', method='implicit')
```

参数如下：

参数	类型	解释
start_date	str	开始日期（例如：‘2017-03-03’）
end_date	str	结束日期（例如：‘2017-03-20’）
factors	None or list of str	因子。默认获取全部因子的因子收益率（None）。风格因子、行业因子和市场联动字段名称见表 2.1 和 2.2
universe	str or list	基准指数。默认为全市场（‘whole_market’），可选沪深 300（‘000300.XSHG’）、中证 500（‘000905.XSHG’）、中证 800（‘000906.XSHG’）
method	str	计算方法 ⁽¹⁾ 。默认为 ‘implicit’（隐式因子收益率），可选 ‘explicit’（显式风格因子收益率）
备注： (1) 由于隐式和显式收益率的计算方法差异（见本说明 4.3 部分），当 method 参数取值为 ‘implicit’，可返回三类因子（风格、行业、市场联动）的隐式因子收益率；而当 method 参数取值为 ‘explicit’，只可返回风格因子的显式因子收益率		

返回 pandas.DataFrame，index 为日期，column 为因子字段名称。

1.5 获取一组股票的特异收益率：

`get_specific_return(order_book_ids, start_date, end_date)`

参数如下：

参数	类型	解释
order_book_ids	str or list	证券代码（例如：‘600705.XSHG’）
start_date	str	开始日期（例如：‘2017-03-03’）
end_date	str	结束日期（例如：‘2017-03-20’）

返回 `pandas.DataFrame`，index 为日期，column 为个股的 `order_book_id`。

1.6 获取因子协方差：

`get_factor_covariance(date, horizon= ‘daily’)`

参数如下：

参数	类型	解释
date	str	开始日期（例如：‘2017-03-03’）
horizon	str	预测期限。默认为日度（‘daily’），可选月度（‘monthly’）或季度（‘quarterly’）

返回 `pandas.DataFrame`，index 和 column 均为因子名称。

1.7 获取一组股票的特异波动率：

`get_specific_risk(order_book_ids, start_date, end_date, horizon= ‘daily’)`

参数如下：

参数	类型	解释
order_book_ids	str or list	证券代码（例如：‘600705.XSHG’）
start_date	str	开始日期（例如：‘2017-03-03’）
end_date	str	结束日期（例如：‘2017-03-20’）
horizon	str	预测期限。默认为日度（‘daily’），可选月度（‘monthly’）或季度（‘quarterly’）

返回 `pandas.DataFrame`，index 为日期，column 为个股的 `order_book_id`。

二 字段说明

表 2.1 和 2.2 给出了第 1 部分数据 API 中风格因子、市场联动、行业因子的字段名称。表 2.3 给出了风格因子和市场联动因子的解释，图 2.1 展示了个股非线性市值和个股市值的非线性关系。

表 2.1：风格因子和市场联动因子字段说明

备注：对于包含细分因子的风格因子，风格因子字段加粗标记，细分因子字段圆点标记

字段	说明
beta	贝塔
momentum	动量
size	规模
earning_yield <ul style="list-style-type: none"><i>trailing_earnings_to_price_ratio</i><i>cash_earnings_to_price_ratio</i><i>predicted_earning_to_price</i>	盈利率
residual_volatility <ul style="list-style-type: none"><i>daily_standard_deviation</i><i>cumulative_range</i><i>historical_sigma</i>	残余波动率
growth <ul style="list-style-type: none"><i>sales_growth</i><i>earnings_growth</i><i>short_term_predicted_earnings_growth</i><i>long_term_predicted_earnings_growth</i>	成长性
book_to_price	账面市值比
leverage <ul style="list-style-type: none"><i>market_leverage</i><i>debt_to_assets</i><i>book_leverage</i>	杠杆率
liquidity <ul style="list-style-type: none"><i>one_month_share_turnover</i><i>three_month_share_turnover</i>	流动性

• <i>twelve_month_share_turnover</i>	
non_linear_size	非线性市值
comovement	市场联动

表 2.2：行业因子字段说明

备注：申万在 2014 年 1 月 1 日曾对行业分类标准进行调整，下表为 2014 年后分类名称

1	农林牧渔	15	商业贸易
2	采掘	16	休闲服务
3	化工	17	综合
4	钢铁	18	建筑材料
5	有色金属	19	建筑装饰
6	电子	20	电气设备
7	家用电器	21	国防军工
8	食品饮料	22	计算机
9	纺织服装	23	传媒
10	轻工制造	24	通信
11	医药生物	25	银行
12	公用事业	26	非银金融
13	交通运输	27	汽车
14	房地产	28	机械设备

表 2.3：风格因子及市场联动因子解释

风格因子	解释
贝塔	个股/投资组合收益对基准收益的敏感度
动量	股票收益变化的总体趋势特征
规模	上市公司的市值规模
盈利率	上市公司的营收能力
残余波动率	个股残余收益的波动成都
成长性	上市公司的营收增长情况
账面市值比	上市公司的股东权益-市值比，反映其估值水平
杠杆率	上市企业负债占资产比例，反映企业的经营杠杆率
流动性	股票换手率，反映个股交易的活跃程度

非线性市值	反映中等市值股票和大/小市值股票的表现差异（图 2.1） ²
市场联动	反映市场整体涨落对个股/投资组合的影响

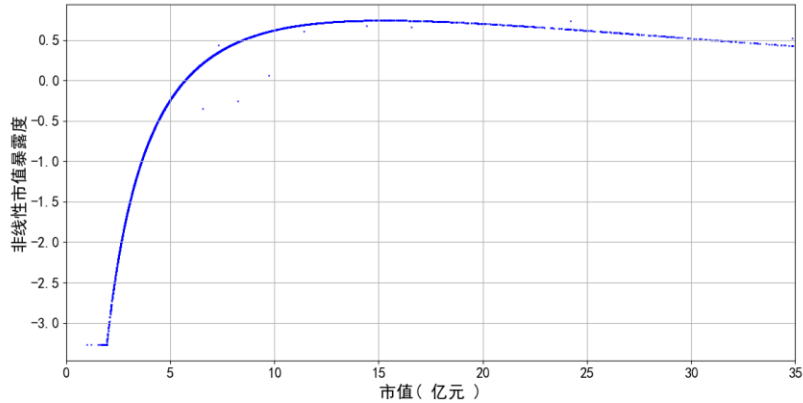


图 2.1：个股非线性市值因子暴露度随市值非线性变化，中等市值股票相对于小市值/大市值股票有更高的非线性市值因子暴露度

三 测试结果

3.1 风格因子累积收益率和相关性

图 3.1 展示 10 个风格因子（见第 2 部分表 2.2）从 2002 年到 2018 年的累计隐式收益率（显式和隐式因子收益率区别见 4.3 节说明）。从图中可以看出：

- 贝塔因子在 17 年间的累积收益率显著高于其它风格因子（即贝塔较高的股票长期平均表现好于贝塔较低的股票）；
- “盈利率”和“账面市值比”因子的累积收益率较高，证明在一个较长的投资时间期限内，价值选股在 A 股市场是有效的；
- “市值”和“非线性市值”等因子因子的累积收益率较低，这和 A 股市场小市值股票在过去多年能产生超额收益率的市场经验是吻合的（个股对非线性市值暴露随市值变化趋势见图 2.1）
- 在 2007 年 5 月和 2015 年 6 月，由于 A 股市场出现极端行情³，导致风格因子收益率出现了短期异常波动。

图 3.2 展示风格因子收益率 17 年间的相关系数，图中显示 10 个风格因子总体相关性较低，因此能够有效表示不同的风险维度。其中，贝塔因子和残余波

² 从图 2.1 可以看出，相对于小市值/大市值股票，中等市值股票有较高的非线性市值因子暴露度，因此非线性市值可用于区分中等市值和小市值/大市值股票

³ 2007 年 5 月 30 日，由于财政部在凌晨宣布提高股票交易印花税，导致股票市场连续大跌；2015 年 6 月 12 月，2014 年启动的牛市结束，A 股市场开始进入股灾阶段。

动率因子的相关性略高（0.47）。说明在特定时期，贝塔较高的个股（对市场波动较敏感）的个股，其残余收益率的波动通常也较大。图 3.1 中贝塔因子和波动率累积收益率走势差异明显，则说明这两个因子并不具有稳定的相关性。

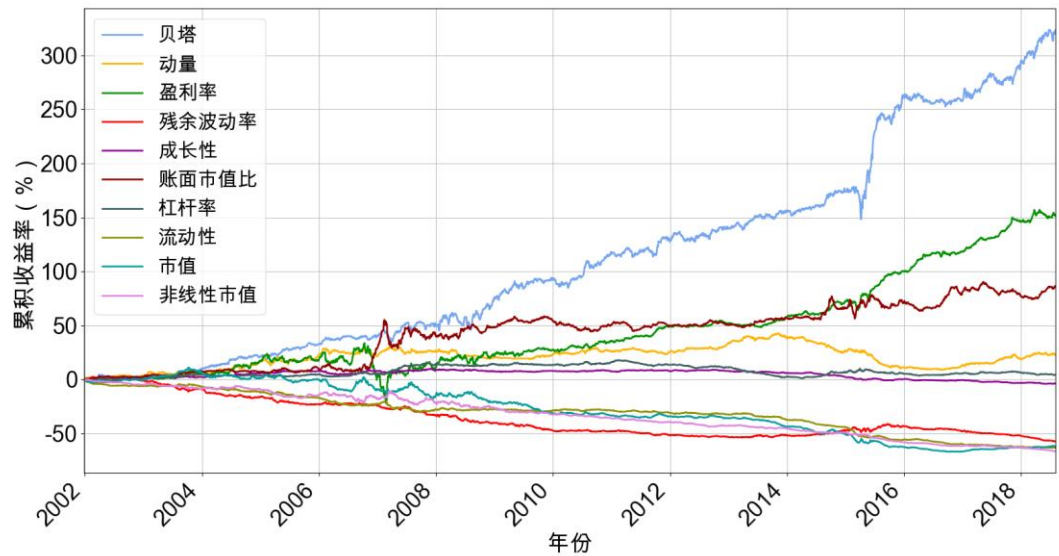


图 3.1：2002~2018 年全市场累积隐式因子收益率

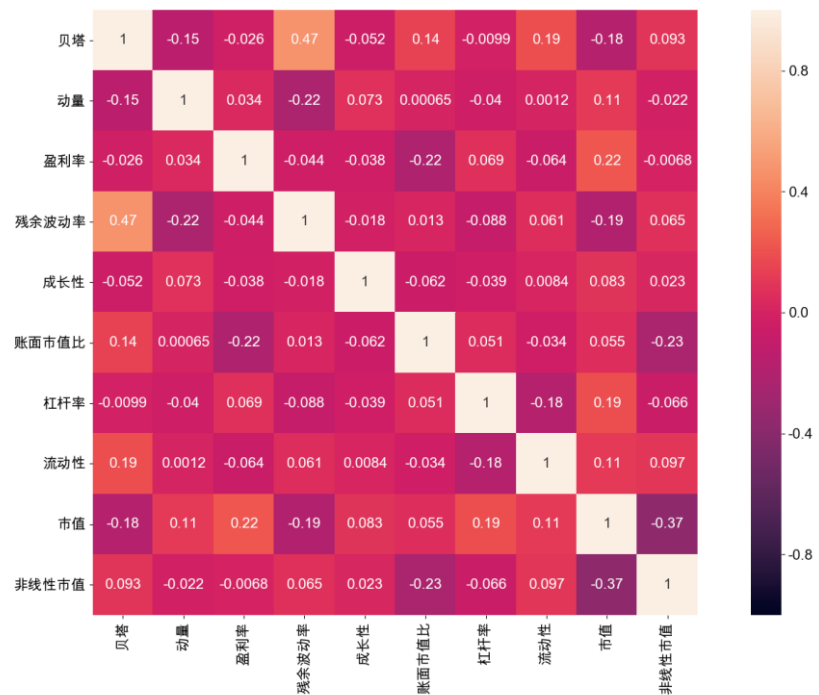


图 3.2：2002~2018 年全市场隐式因子收益率相关系数矩阵

3.2 模型修正可决系数（adjusted R-squared）

在量化投研中，因子的有效性依赖于股票池的选择。例如，同样是使用小市值因子选股，在全市场选择的可能是经营有问题的 ST 股；而在沪深 300 中选

择的则是市值相对偏小，但公司基本面良好的价值型股票。因此，若研究阶段所使用的股票池和实际投资的股票池存在较大差异，则研究阶段表现出色的因子，可能在实际投资中表现会出现明显衰减。

基于上述原因，除全市场以外，米筐因子数据以三个常用指数（沪深 300、中证 500、中证 800）成分股作为股票池，进行横截面回归计算隐式因子收益率（显式和隐式因子收益率区别见 4.3 节说明）。在用户使用因子收益率监控市场风格或构建因子指数增强策略时，可以选择贴近实际投资场景的因子收益率，避免“研究-投资”的不一致性带来的因子效果衰减。

修正可决系数（Adjusted R-Squared）为评估回归模型解释力的重要指标。为评估米筐因子数据在不同股票池的解释力，在表 3.2.1 和图 3.3~3.6 给出了米筐因子数据在不同股票池的 Adjusted R-Squared。可以看出若以指数成分股为投资股票池，使用该股票池的因子收益率能够明显提高模型解释力。因此，用户在挑选因子构建指数增强策略的时候，应考虑选择相应股票池的因子收益率进行监控和研究。

表 3.1：米筐因子数据在各个股票池 Adjusted R-Squared 指标（2009~2018 年）

股票池	因子收益率类型	均值	标准差
全市场	全市场	0.29	0.056
沪深 300	全市场	0.34	0.074
	沪深 300	0.40	0.060
中证 500	全市场	0.29	0.064
	中证 500	0.32	0.054
中证 800	全市场	0.34	0.063
	中证 800	0.36	0.057

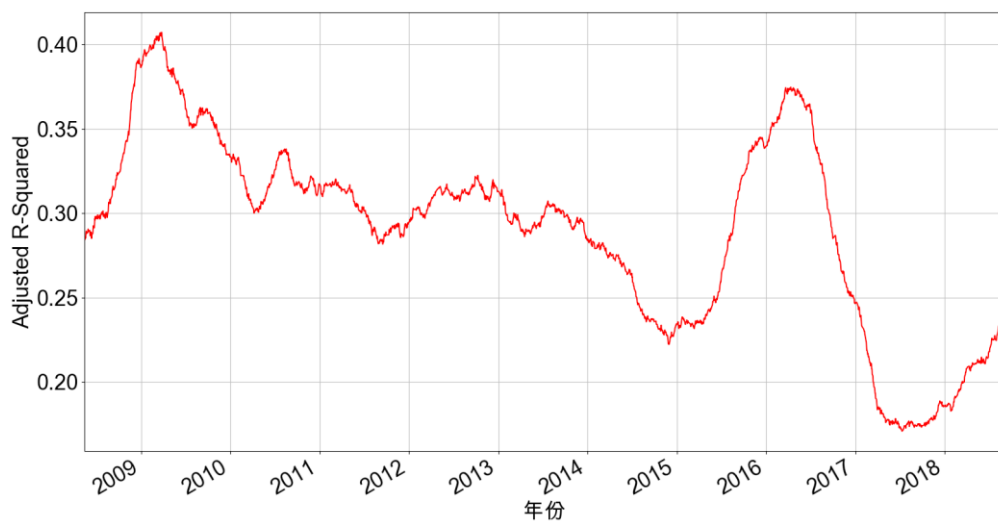


图 3.3: 以全市场为股票池，基于全市场因子收益率的滚动 Adjusted R-Squared

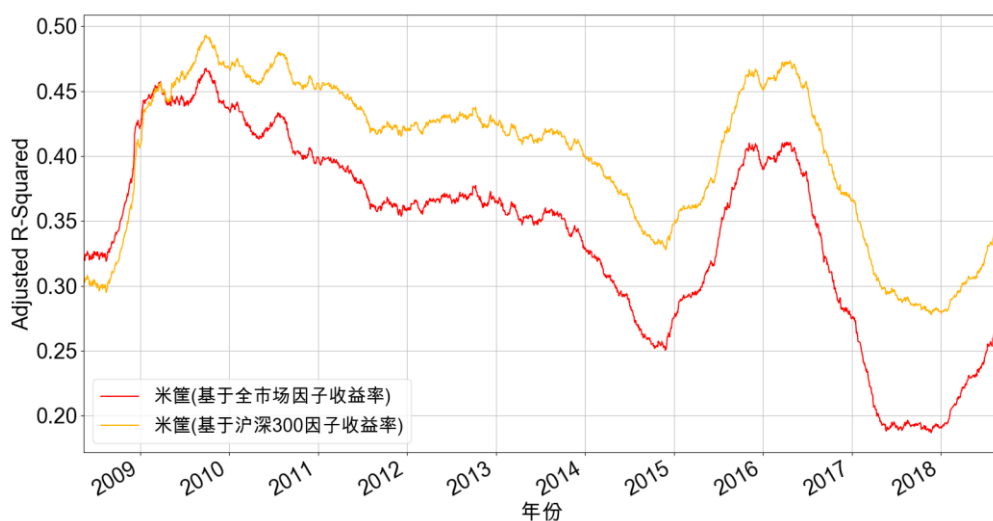


图 3.4: 以沪深 300 成分股为股票池，基于全市场因子收益率和基于沪深 300 因子收益率的滚动 Adjusted R-Squared

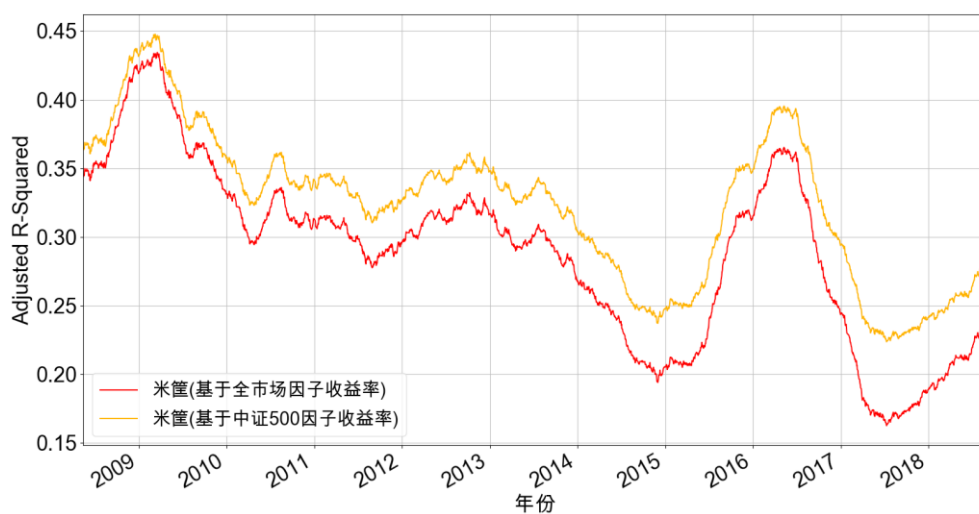


图 3.5：以中证 500 成分股为股票池，基于全市场因子收益率和基于中证 500 因子收益率的滚动 Adjusted R-Squared

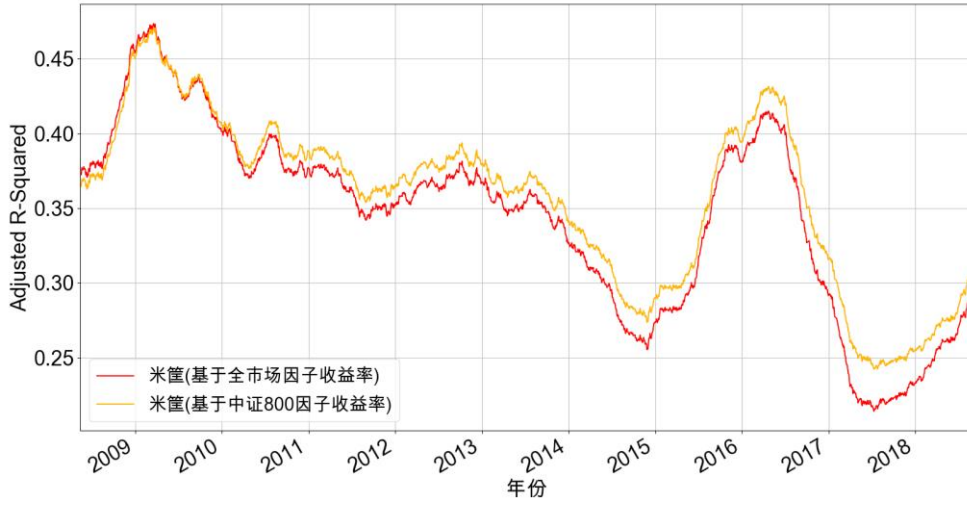


图 3.6：以中证 800 成分股为股票池，基于全市场因子收益率和基于中证 800 因子收益率的滚动 Adjusted R-Squared

3.3 偏差统计量（Bias Statistics）测试

偏差统计量（Bias Statistics）为评估风险模型样本外表现的重要指标。这一部分展示了因子协方差及特异风险的 Bias Statistics 测试结果。核心结论如下：

- 米筐因子协方差及特异风险测试表现均明显优于经验估计；
- 基于 A 股市场特点改进后的特异风险模型模型优于未改进的特异风险模型

3.3.1 因子协方差 Bias Statistics

对于因子协方差，Bias Statistics 定义如下：

$$b_{kt} = \frac{f_{kt}}{\sigma_{kt}} \quad (3.1)$$

$$bias\ statistics_k = \sqrt{\frac{1}{T-1} \cdot \sum_{t=1}^T (b_{kt} - \bar{b}_k)^2} \quad (3.2)$$

f_{kt} ， σ_{kt} 和 b_{nt} 分别为第 t 期因子 k 的累积收益率、预测因子收益波动率和标准化因子收益率（standardized factor return）； \bar{b}_k 为整个测试周期 b_{kt} 的均值； T 为总期数（日度测试一期为 1 个交易日，月度测试一期为 21 个交易日，季度测试一期为 63 个交易日）。

Bias Statistics 指标可理解为“实际收益波动率”和“模型预测收益波动率”的比值。因此，若风险模型的偏差统计量在 1 上下小幅度波动，即说明风险模型预测较为准确，没有系统性地低估/高估风险。

表 3.2 和图 3.7~3.9 展示了日度、月度、季度的米筐因子协方差和经验估计⁴（empirically estimated）因子协方差的滚动 Bias Statistics。可以看出米筐因子协方差能够有效消除经验估计中的系统性偏差（Bias Statistics 的均值更接近于 1），同时提供预测的准确性和稳定性（Bias Statistics 的标准差更小）。

表 3.2：经验因子协方差和米筐因子协方差 Bias Statistics 均值和标准差对比

		均值	标准差
日度	经验	0.949216	0.238186
	米筐	0.985416	0.015108
月度	经验	1.028358	0.320764
	米筐	1.004237	0.098769
季度	经验	1.192166	0.443940
	米筐	1.053151	0.173348

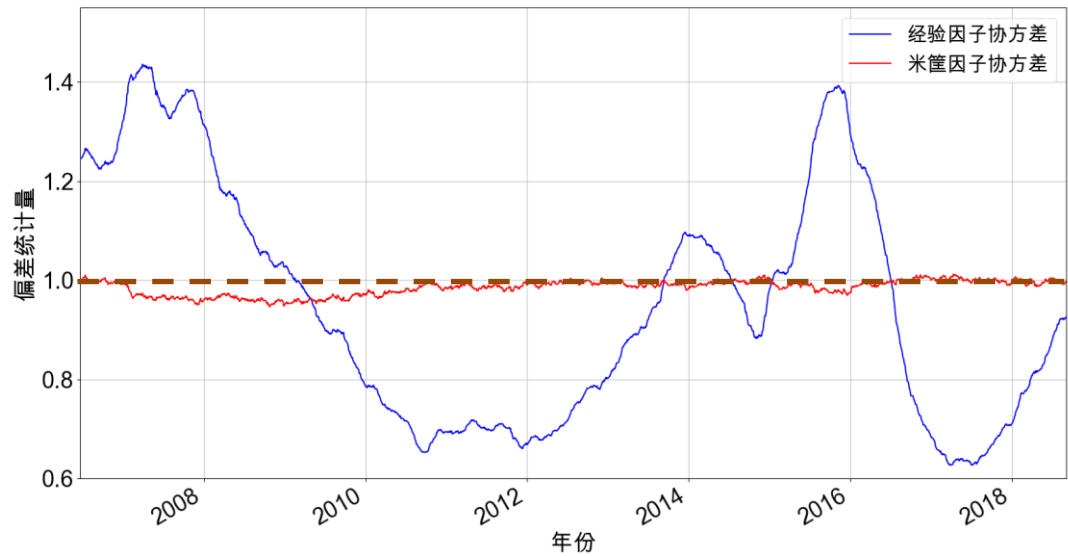


图 3.7：米筐日度因子协方差滚动 Bias Statistics

⁴ 即直接使用历史收益率数据计算协方差矩阵

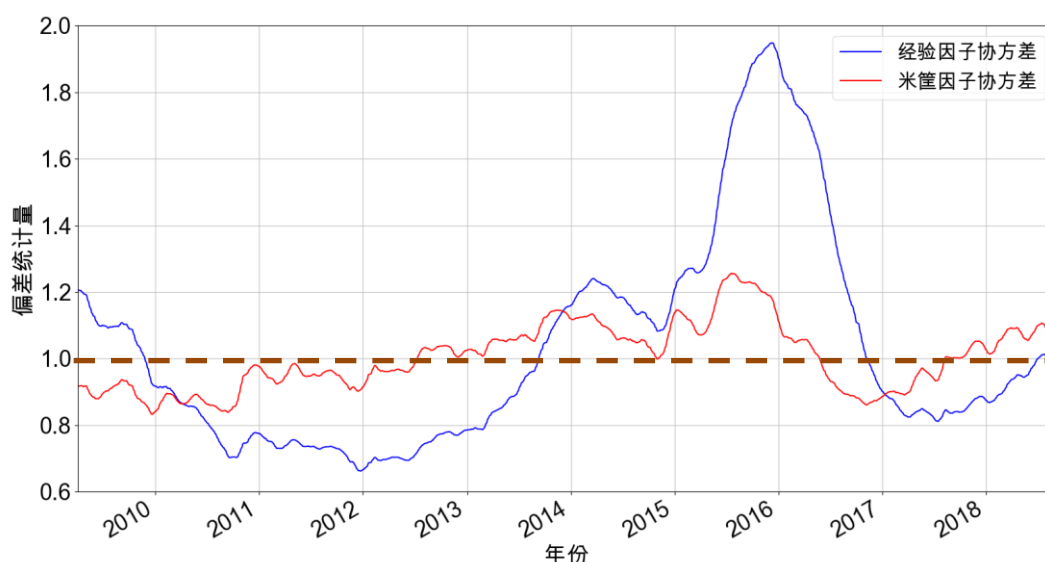


图 3.8: 米筐月度因子协方差滚动 Bias Statistics

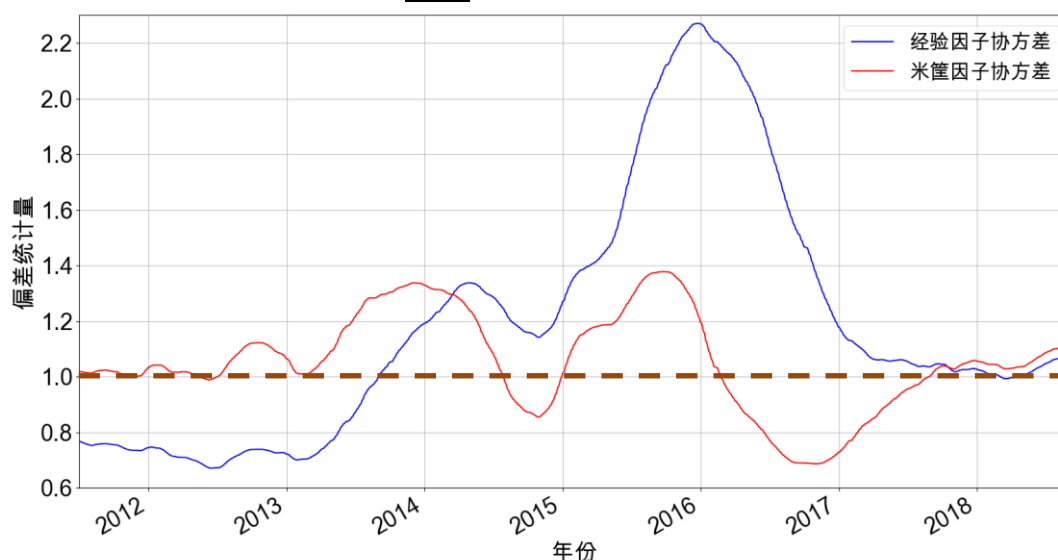


图 3.9: 米筐季度因子协方差滚动 Bias Statistics

3.3.2 特异风险 Bias Statistics

特异风险反映个股风险中不能被因子解释的部分。例如，上市公司高管人员临时变动所引起的股价波动，通常与上市公司基本面、所处行业、及市场整体行情无关，因此不能被因子所解释，从而归入到特异风险中。特异风险的预测通常较容易受噪声干扰——一个股的上市时间过短（导致历史特异收益率数据不足）、停牌（导致股票历史特异波动率过低）、市场对于消息面因素的过度反应等，都会造成特异风险估计的误差。

相对于欧美股票市场，A 股市场的特异风险建模有两点重要区别：（1）股票在上市/复牌后的第一个交易日，没有涨跌停板限制下股价的大幅波动⁵，在未来一段时间不可重现，因此该交易日的特异收益率应视为离群值；（2）上市公司把申请停牌视为规避市场风险的一种特殊手段⁶，导致市场停牌股数量在行情欠佳时显著增加。图 3.10 为近 10 年 A 股市场停牌股数目变化情况。其中显示 A 股市场的停牌股数目从 2014 年开始显著增加（尤其在 2015 年股灾期间），并长时间稳定在 200 个以上。因此，对上述股票上市/复牌第一个交易日股价的异常波动，以及股票停牌的恰当处理，是准确预测 A 股市场特异风险的关键。

基于上述 A 股市场的特点，米筐特异风险模型对传统特异风险模型中离群值及停牌股处理流程进行了改进。为测试特异风险模型的效果，特异风险的 Bias Statistics 定义如下：

$$b_{nt} = \frac{r_{nt}}{\sigma_{nt}} \tag{3.3}$$

$$bias\ statistics_n = \sqrt{\frac{1}{T-1} \cdot \sum_{t=1}^T (b_{nt} - \bar{b}_k)^2} \tag{3.4}$$

其中 r_{nt} ， σ_{nt} 和 b_{nt} 分别为第 n 个股票的累积特异收益率、预测特异风险和标准化特异收益率（standardized specific return）； \bar{b}_k 为整个测试周期 b_{kt} 的均值； T 为总期数（日度测试一期为 1 个交易日，月度测试一期为 21 个交易日，季度测试一期为 63 个交易日）。

表 3.3 和图 3.11~3.13 对比了经验估计、未改进特异风险模型、及改进后的米筐特异风险模型的 Bias Statistics 测试结果。其中月度特异风险模型的测试效果改善明显。此外，在 3.4 节，我们通过对比一系列风险最小化组合的回测表现，进一步证明了改进后的日度、月度、季度特异风险模型，均能够有效提升优化组合表现。

表 3.3：经验特异风险和米筐特异风险 Bias Statistics 均值和标准差对比

		均值	标准差
	经验	1.001665	0.106535

⁵ A 股上市公司上市的第一个交易日没有涨跌停板的限制。某些情况下（例如进行股权分置改革），上市公司停牌后复牌的第一个交易日也没有涨跌停板限制。

⁶ 例如，若某一上市公司股票持续下跌，则该上市公司大股东的股权质押会出现平仓风险。而停牌则成为避免股价继续下跌的一种临时性措施。

日度	米筐（未改进）	1.010809	0.029167
	米筐（改进后）	1.019603	0.029271
月度	经验	0.911546	0.198186
	米筐（未改进）	1.008305	0.097394
	米筐（改进后）	0.964666	0.059002
季度	经验	0.824596	0.186647
	米筐（未改进）	0.951106	0.138990
	米筐（改进后）	0.932783	0.122038

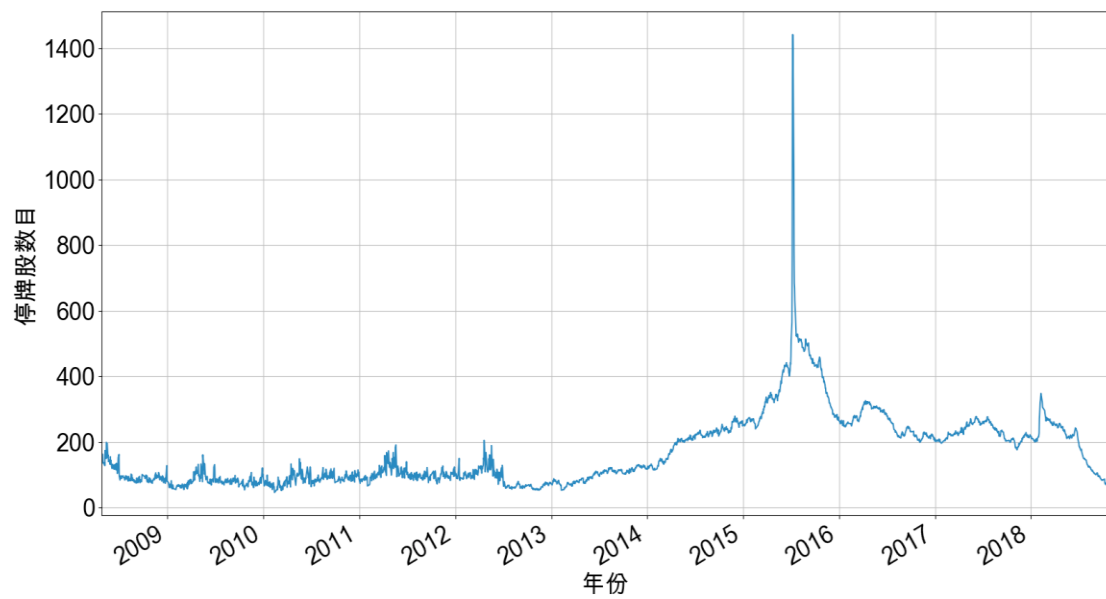


图 3.10：A 股近 10 年停牌股数量变化

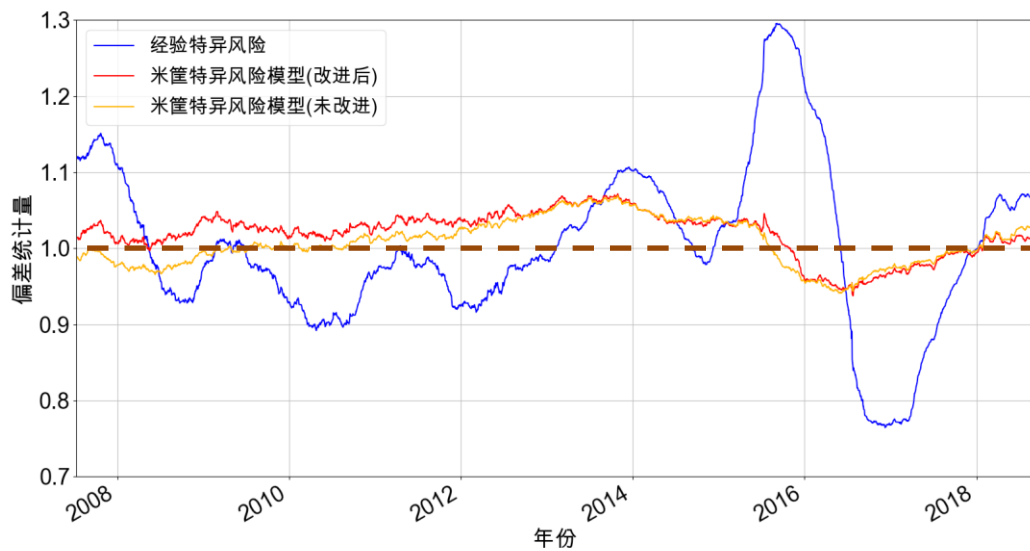


图 3.11：米筐日度特异风险 Bias Statistics 测试结果

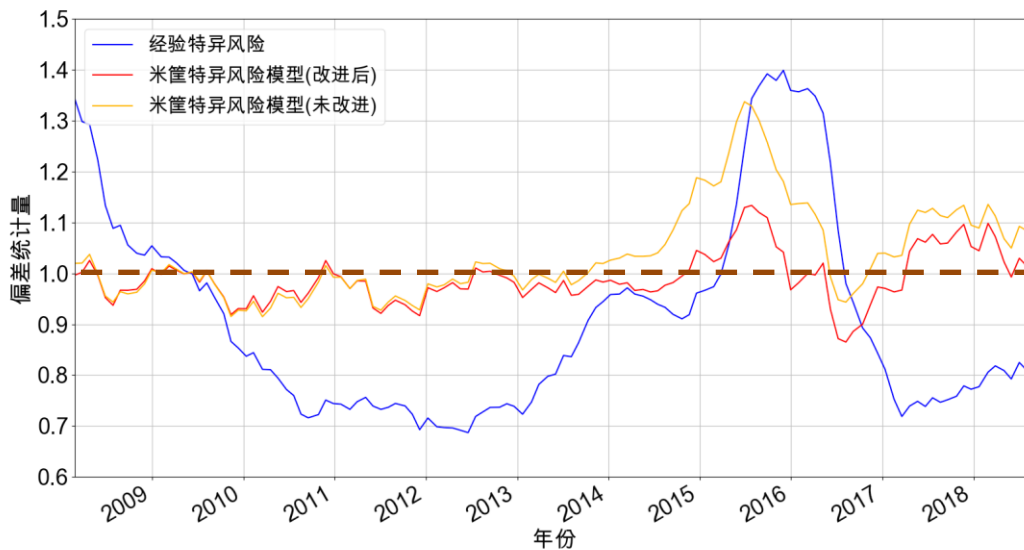


图 3.12：米筐月度特异风险 Bias Statistics 测试结果

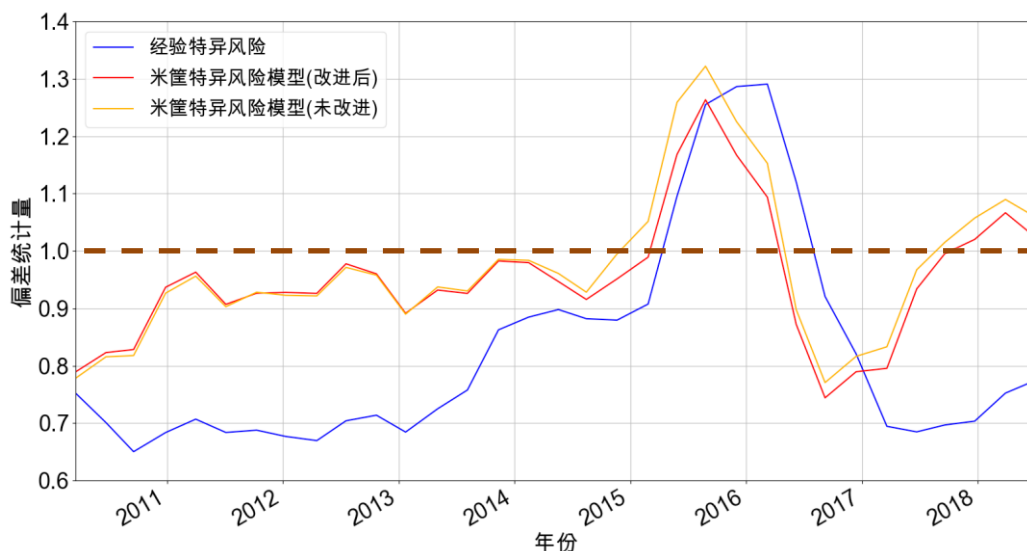


图 3.13：米筐季度特异风险 Bias Statistics 测试结果

3.4 风险最小化组合

为进一步评估米筐因子风险模型效果，这一部分按市值和行业对全市场股票进行筛选，构建 5 组共计 98 个优化测试组合进行回测。这一节核心结论如下：

- 因子风险模型在各组测试中表现均明显优于经验估计，同时改进后的因子模型略优于未改进的因子模型（特异风险估计的改进见 3.3.2 节讨论）；
- 投资组合包含的股票数目越多，因子风险模型的降维降噪效果越明显，模型的测试表现越好；

3.4.1 测试组合构建及回测设置

回测时间段为 2014 年 1 月 1 日至 2018 年 8 月 31 日。在每一期调仓中，使用不同的风险估计方法（经验估计和多因子风险模型），构建风险最小化测试组合。测试组合构建规则如下：

- 测试组合 1：每一期在全市场中选择市值最大的 600 只股票，从大到小排序分为 30 个组合，每个组合包含 20 个股票
- 测试组合 2：每一期在全市场中选择市值最大的 1200 只股票，从大到小排序分为 20 个组合，每个组合包含 60 个股票
- 测试组合 3：每一期在全市场中选择市值最大的 1000 只股票，从大到小排序分为 10 个组合，每个组合包含 100 个股票
- 测试组合 4：每一期以申万一级行业分类的 28 个行业作为股票组合
- 测试组合 5：每一期在 28 个行业中，选择每个行业中市值排名前 1、2、3、4、5 的股票（分别包含 28、56、84、112、140 个股票），以及市值排名后 1、2、3、4、5 的股票，构建共计 10 个测试组合

为全面评估日度、月度、季度风险模型效果，回测中考虑四种调仓频率：

- 每日调仓：每一个交易日按上述测试组合构建规则更新持仓，并按照风险最小化权重进行调仓；
- 每周调仓：除改为每周一调仓外，其它测试条件和上述相同；
- 每月调仓：除改为每月首个交易日调仓外，其它测试条件和上述相同；
- 季度调仓：除改为每季度首个交易日调仓外，其它测试条件和上述相同

3.4.2 每日调仓测试结果

表 3.4 对比了基于经验估计、未改进日度因子风险模型、改进后日度因子风险模型（特异风险估计的改进见 3.3.2 节讨论）构建的优化测试组合在每日调仓回测中的表现。其中改进后日度因子风险模型各项指标均为最优。图 3.14 展示了优化测试组合的“风险-收益”分布，其中显示在经历了 2015 年“牛市-股灾”极端行情的情况下，基于经验估计得到的优化测试组合表现差异较大，且部分组合出现过大的波动（图右侧红色散点）；而因子风险模型则有效地把优化测

试组合的年化波动率控制在 45% 以内，且优化测试组合整体呈现出更为显著的有效前沿方差最小化组合（minimum variance portfolio on efficient frontier）特征。

表 3.5 和图 3.15 进一步对比了各组优化测试组合的年化夏普率均值，其中改进后日度因子风险模型在各组优化测试组合中的年化夏普率均为最高，证明了该模型具有较好的通用性。从测试组合 1~3 中所包含的股票数目依次递增（见本部分开头测试组合构建规则说明），而经验估计和米筐因子风险模型的表现差距随之增大。理论上，因子风险模型是一种降维降噪的风险预测方法（见 4.4 节方法论介绍）。投资组合包含的股票数目越多，降维效果越明显，模型的预期表现越好。因此该测试结果和模型的理论表现是吻合的。值得注意的是，当投资组合仅包含 20 个股票的情况下，因子模型实际上是一种升维处理（米筐因子风险模型共计使用 39 个因子），但由于因子风险模型能够降低风险估计中的噪声，因此在实际测试中也比经验估计取得了更好的效果。

表 3.4：每日调仓优化测试组合各回测指标均值

	年化波动率（%）	年化收益率（%）	最大回撤（%）
经验	35.54	10.01	63.11
米筐（未改进）	28.75	13.21	57.24
米筐（改进后）	28.66	14.14	56.24

表 3.5：每日调仓各组优化测试组合年化夏普率均值

	经验	米筐（未改进）	米筐（改进后）
测试组合 1	0.4187	0.4140	0.4313
测试组合 2	0.2042	0.4095	0.4362
测试组合 3	-0.1693	0.3651	0.3937
测试组合 4	0.2624	0.3932	0.4158
测试组合 5	0.6689	0.6740	0.6886

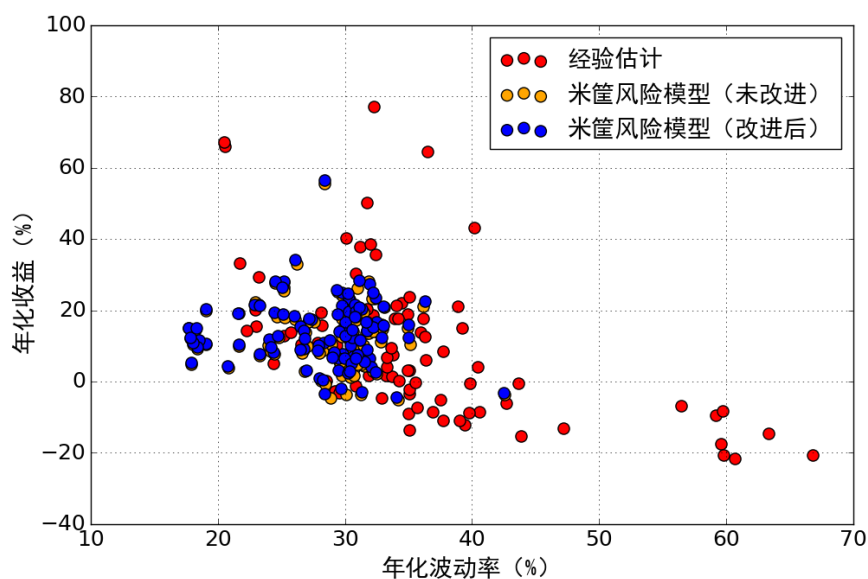


图 3.14: 每日调仓优化测试组合“风险-收益”分布

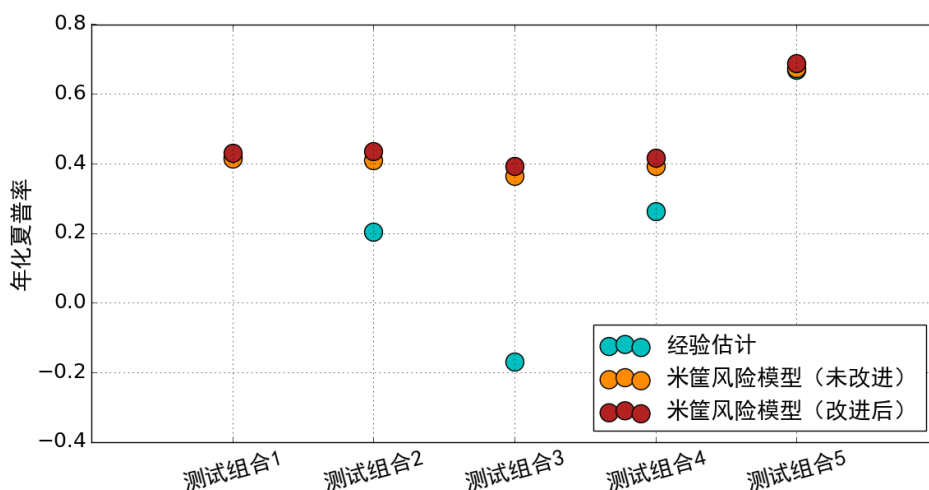


图 3.15: 每日调仓各组风险最小化组合年化夏普率均值

3.4.3 每周调仓测试结果

表 3.6 和图 3.16 对比了 5 种风险估计方法（基于经验估计、未改进的日度和月度因子风险模型、改进后日度和月度因子风险模型）构建的优化测试组合在每周调仓回测中的表现。其中可以看出：

- 因子风险模型各项指标均优于经验估计，改进后的因子模型略优于未改进的因子模型；
- 在每周调仓回测中，基于日度因子风险模型得到的优化测试组合年化波动率更低，而月度因子风险模型在提高收益和控制回撤方面效果更好。

表 3.7 和图 3.17 中对比了 3 种风险估计方法（经验估计、改进后日度和月度因子风险模型）得到的各组优化测试组合的年化夏普率均值，其中：

- 在优化测试组合 1~4 中，改进后因子风险模型年化夏普率均为最优；而在测试组合 5 中，经验估计的年化夏普率优于因子风险模型。测试组合 5 的测试结果不同于测试组合 1~4，是经验估计的不稳定性造成的。例如，在构建风险最小化组合的时候，经验估计可能会明显低估未来一段时间表现较好的股票的风险，导致其权重大幅提高，此时经验估计的误差，反而能进一步提高组合收益。在每日、每月、每季度调仓回测中，基于经验估计得到的优化测试组合 5 表现均不如基于因子风险模型，说明在每周调仓的情况下，经验估计在测试组合 5 的回测表现优于因子风险模型具有偶然性；
- 测试组合 1~3 的测试结果变化规律和每日调仓部分测试结果类似，投资组合包含的股票数目越多，因子模型降维降噪的效果越明显，测试效果越好。

表 3.6: 每周调仓优化测试组合各回测指标均值

	年化波动率 (%)	年化收益率 (%)	最大回撤 (%)
经验	29.18	14.63	55.51
米筐（日度，未改进）	21.32	14.75	49.87
米筐（日度，改进后）	21.22	15.14	49.55
米筐（月度，未改进）	21.98	19.09	47.10
米筐（月度，改进后）	21.86	19.58	46.66

表 3.7: 每周调仓优化测试组合年化夏普率均值

	经验	米筐（日度，改进后）	米筐（月度，改进后）
测试组合 1	0.4996	0.4703	0.6064
测试组合 2	0.4388	0.5718	0.7565
测试组合 3	0.0835	0.5758	0.7059
测试组合 4	0.1816	0.4967	0.5161
测试组合 5	0.8994	0.7516	0.6443

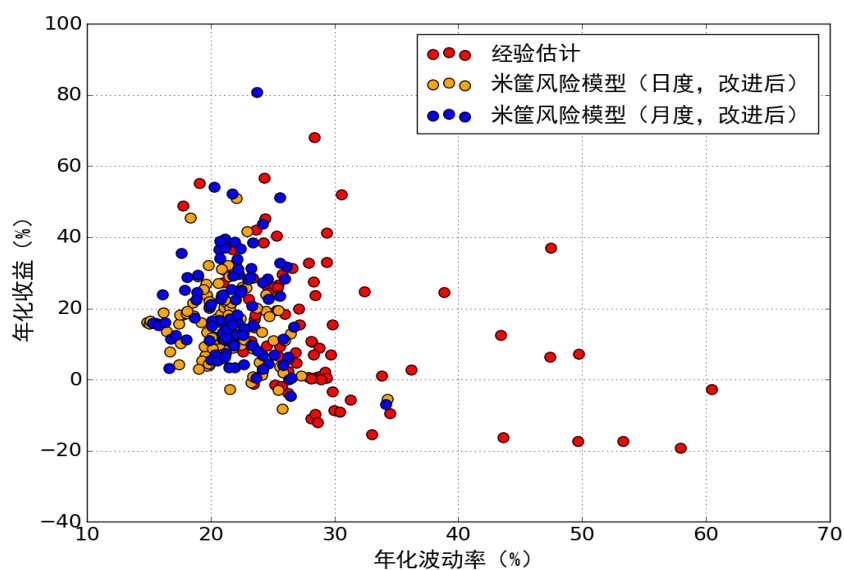


图 3.16: 每周调仓优化测试组合“风险-收益”分布

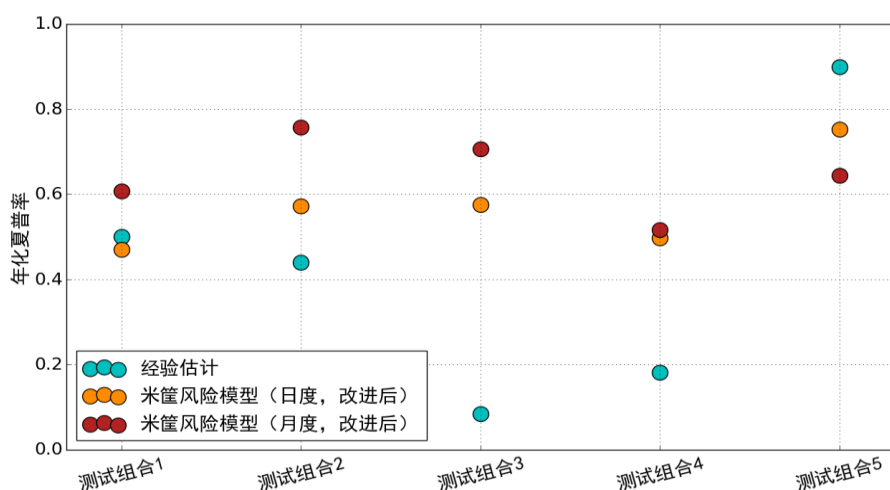


图 3.17: 每周调仓各组风险最小化组合年化夏普率均值

3.4.4 每月调仓测试结果

表 3.8 和 3.9，以及图 3.18 和 3.19 对比了 3 种风险估计方法（基于经验估计、未改进月度因子风险模型、改进后月度因子风险模型）构建的优化测试组合在每周调仓回测中的表现。其中可以看出：

- 改进后的因子模型的各项回测指标均为最优；
- 在所有优化测试组中，改进后因子风险模型年化夏普率均为最优；
- 测试组合 1~3 的测试结果和每日调仓及每周调仓部分测试结果类似，投资组合包含的股票数目越多，因子模型降维降噪的效果越明显，测试效果越好。

表 3.8：每月调仓风险最小化组合各指标均值

	年化波动率（%）	年化收益率（%）	最大回撤（%）
经验	37.32	10.05	62.69
米筐（未改进）	29.55	21.53	52.28
米筐（改进后）	29.48	21.94	51.93

表 3.9：每月调仓各组测试组合年化夏普率均值

	经验	米筐（未改进）	米筐（改进后）
测试组合 1	0.4608	0.5781	0.5835
测试组合 2	0.2456	0.6499	0.6609
测试组合 3	0.1009	0.6163	0.6295
测试组合 4	0.3576	0.4763	0.4829
测试组合 5	0.3111	0.7522	0.7592

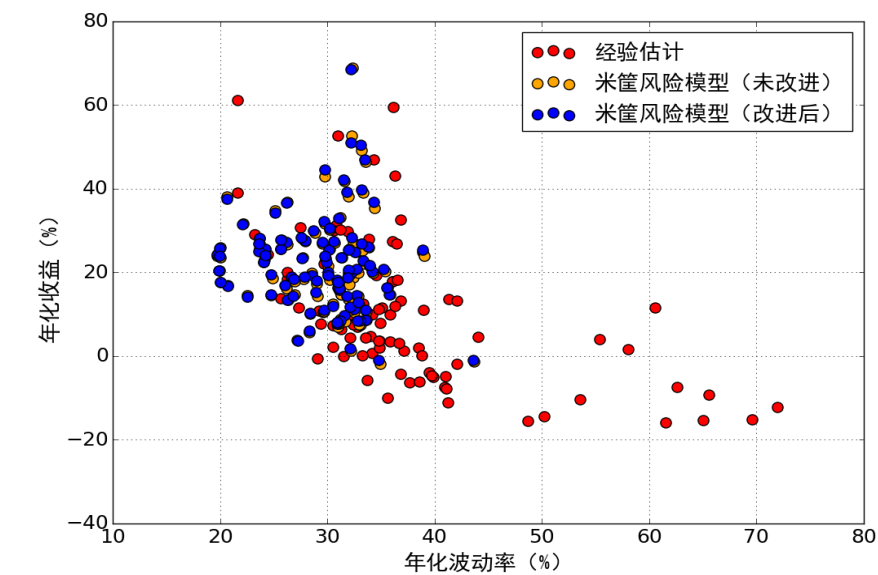


图 3.18：每月调仓风险最小化组合“风险-收益”分布

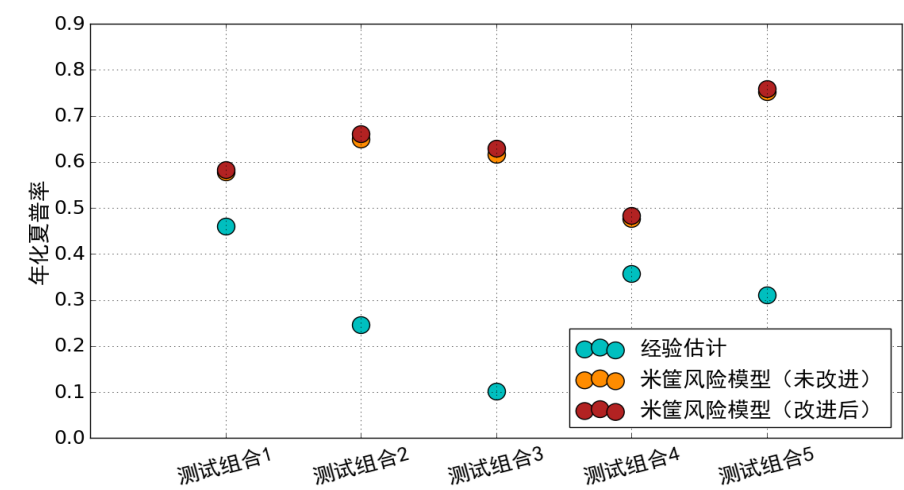


图 3.19：每月调仓各组风险最小化组合年化夏普率均值

3.4.5 每季度调仓测试结果

表 3.10 和 3.11，以及图 3.20 和 3.21 对比了 3 种风险估计方法（基于经验估计、未改进季度因子风险模型、改进后季度因子风险模型）构建的优化测试组合在每周调仓回测中的表现。其中可以看出：

- 改进后的因子模型的各项回测指标均为最优；
- 在所有优化测试组中，改进后因子风险模型年化夏普率均为最优；
- 测试组合 1~3 的测试结果和前述类似，投资组合包含越多，测试效果越好。

表 3.10：每季度调仓风险最小化组合各指标均值

	年化波动率（%）	年化收益率（%）	最大回撤（%）
经验	38.96	20.59	59.49
米筐（未改进）	30.32	25.37	51.31
米筐（改进后）	30.24	25.58	51.27

表 3.11：每季度调仓各组测试组合年化夏普率均值

	经验	米筐（未改进）	米筐（改进后）
测试组合 1	0.5629	0.5936	0.5950
测试组合 2	0.5649	0.7170	0.7207
测试组合 3	0.1002	0.7602	0.7633
测试组合 4	0.4532	0.5707	0.5734
测试组合 5	0.5790	0.7065	0.7111

综上，在 3.4.2~3.4.5 节中，我们给出了在不同调仓频率下（每日、每周、每月、每季度），基于米筐因子风险模型在一系列优化测试组合的回测结果。可以看出米筐因子风险模型在不同的条件下，均能有效地降低策略的风险和提高策略的收益，证明了其具有良好的适用性和稳定性。而由于因子风险模型能够实现降维降噪的效果，因此其在较大的投资组合中表现更为出色。此外，我们通过对改进前后的因子风险模型，证明了我们基于 A 股市场的特点，对特异风险估计的改进（见 3.3.2 节），能够有效提升优化测试组合的表现。

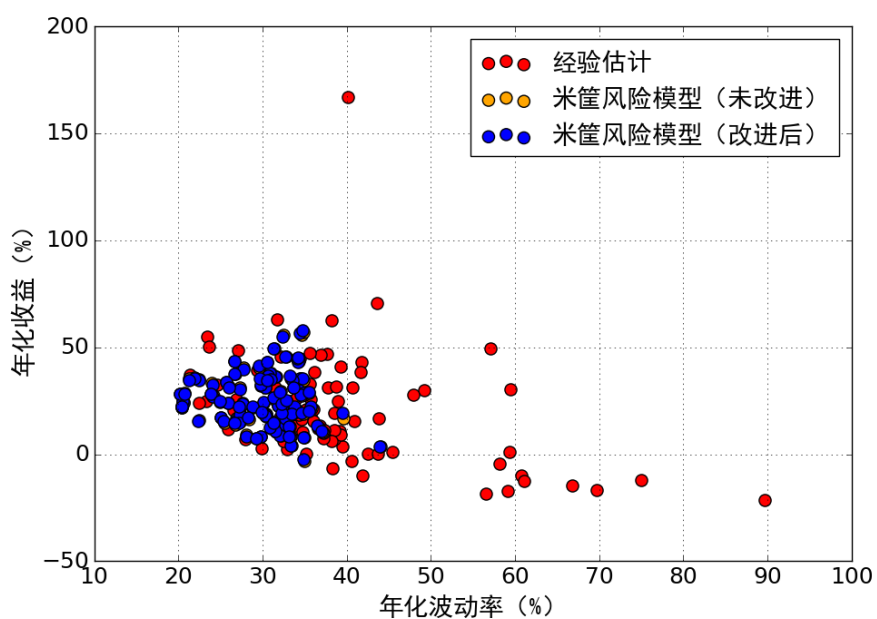


图 3.20: 每季度调仓风险最小化组合“风险-收益”分布

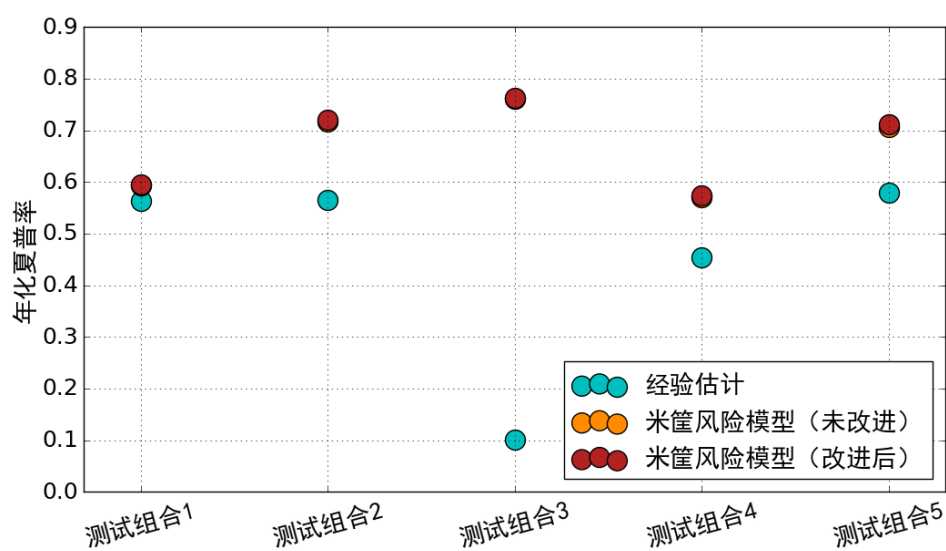


图 3.21: 每季度调仓各测试组合年化夏普率均值

四 方法论介绍

4.1 多因子模型

多因子模型认为股票组合的收益来源于股票市场上多个共同因素和特殊因素的共同作用（图 4.1），其通用表达式如下：

$$r_p = \sum_{k \in style} X_k f_k + \sum_{l \in industry} X_l f_l + X_c f_c + \mu \quad (4.1)$$

其中， r_p 为股票组合的收益率； X_k ， X_l 和 X_c 分别为股票组合对风格因子 k ，行业因子 l ，市场联动因子（comovement） c 的暴露度； f_k ， f_l 和 f_c 分别为风格因子 k ，行业因子 l ，市场联动因子 c 的因子收益； μ 为股票组合的特异收益。

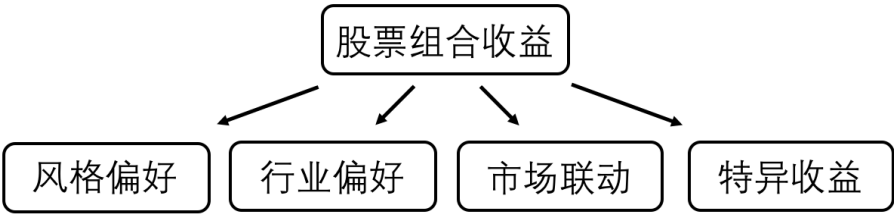


图4.1：股票组合收益来源

由上述分解可以看出，股票组合收益有四个收益来源：风格偏好、行业偏好、市场联动、以及特异收益；

- 风格偏好。风格为影响股票收益的共同风格因素（包括市值、盈利性、成长性等）。例如，若投资组合偏好于持有小市值股票，同时小市值股票表现显著优于大市值股票，则小市值的风格偏好能为投资组合带来超额收益；
- 行业偏好。例如，若投资组合偏好于持有“医药生物”行业，同时该行业总体表现优于其它行业，此时超配“医药生物”能为投资组合带来超额收益；
- 市场联动。市场联动表示股票市场整体涨落为投资组合带来的收益影响。对于任意两个满仓的股票组合，其获得的市场收益及承担的市场整体波动风险相同；而对于“股票+现金”的组合，随着组合中的股票仓位比例的增加，市场联动对组合收益影响越大；
- 特异收益。特异收益表示影响个股的特殊因素（例如上市公司管理层变动，地方政府政策优惠等），因而区别于影响股价的共同因素（风格、所属行业、市场整体波动）。

4.2 因子暴露度

- 因子对投资组合收益的影响由因子暴露度和因子收益率两部分决定，其中：
- 因子暴露度是投资组合主动押注的部分，表示投资组合自身偏好带来的收益和风险（例如重仓小市值股票，或重仓某一行业股票等）；
 - 因子收益是市场行情变化引起的投资组合被动风险变化。因子收益由不同风格，不同行业的股票表现差异所决定

根据表达式4.1，组合对某一因子 k 的暴露度为其收益率对因子收益率 f_k 计算偏导数：

$$X_k = \frac{\partial r_p}{\partial f_k} \tag{4.2}$$

因此股票组合对某一因子的暴露度绝对值越大，则投资组合对因子表现越敏感。在实际投资研究中，一个优秀的因子选股模型应该能够对因子未来的收益给出较为准确的预测，从而使得投资者能够基于因子表现对暴露度作出相应的调整（即当某一因子的因子收益预测为正，则选择对该因子正暴露；而当某一因子的因子收益为负，则选择对该因子负暴露）。表 4.1中提供了三类因子的因子暴露度计算规则说明。

表4.1：暴露度计算规则说明

因子	计算规则
风格因子	风格因子暴露度根据第四部分风格因子定义进行计算后，再进行市值加权标准化处理： <div>$X' = \frac{X - \bar{X}}{\sigma(X)}$</div> 其中 X 为原始暴露度； \bar{X} 为暴露度的市值加权横截面均值； $\sigma(X)$ 为暴露度的横截面标准差。经市值加权标准化处理后，能够保证市值加权的全市场指数 ⁽¹⁾ 的风格暴露度为 0。即全市场指数只承担市场整体波动风险，而没有任何特定的风格/行业风险暴露。
行业因子	若个股属于某一行业，则该个股对行业暴露度为 1，对其它行业暴露度为 0。
市场联动	投资组合对于市场联动因子的暴露度等于股票市值占比。例如，对于股票满仓组合，其暴露度为 1；对于股票市值占比为 50% 的组合，其暴露度为 0.5。
备注：（1）全市场指数也是CAPM模型中的自变量，代表市场的系统性收益和风险。	

4.3 因子收益率

隐式因子收益率和显式因子收益率区别如表 4.2 所示。在表 4.3 中，我们给出了一个特殊情景下，隐式因子收益率和显式因子收益率存在差异的实例解释。基于表 4.2 和表 4.3，可以看出对于风格因子，显式收益率的计算和意义较为直观，而隐式收益率则剔除了行业因素的影响，更能反映风格因子带来的真实超额收益率。用户可根据自己的投资研究习惯以及市场经验，选择合适的因子收益率进行市场风格追踪。

表4.2： 隐式和显式因子收益率对比

	隐式因子收益率	显式因子收益率
计算方式	以个股收益率作为因变量，因子暴露度（风格、行业、市场联动）作为自变量，进行横截面多元回归，得到回归系数即为因子收益率	做多因子暴露度大的 50% 股票，同时做空因子暴露度小的 50% 股票，再进行贝塔中性和市值中性处理，该多空组合收益即为因子收益率
是否考虑行业影响	是	否
使用场景	<ul style="list-style-type: none">• 市场风格追踪• 组合持仓分析	<ul style="list-style-type: none">• 市场风格追踪• 组合净值分析

表 4.3： 特定情景下市值因子显式/隐式因子收益率结果差异

情景	近一段时间银行业（大市值股票为主）表现远好于其它行业，导致大市值股票表现优于小市值股票
显式市值因子收益	显著为正。因为显式因子收益对应做多大市值股票，做空小市值股票的多空组合收益。
隐式市值因子收益	不显著为正。因为隐式收益计算中剔除了行业因素的影响，此时超额收益部分体现在银行业的因子收益中

4.4 因子风险模型

多因子风险模型认为股票组合的风险可以分解为因子部分和特异部分。其中，因子部分代表影响全市场股票风险的共同因素（风格、所属行业、市场整体涨落）；特异部分则代表影响个股风险的特殊因素（例如上市公司高管人员临时变动，或出现负面新闻等）。此外，影响某一股票的特殊因素，通常不会对其它股票造成影响（反之，若某一因素对相当数量的股票均存在稳定的，显著的影响，则应该成为因子），因此因子风险模型引入了两个假设：

- 因子收益率和个股特异收益率不相关；
- 不同个股的特异收益率不相关。

基于上述假设，可得投资组合收益波动率 $\sigma(r_p)$ 的因子分解表达式如下：

$$\sigma(r_p) = \sqrt{X_p^T \cdot \Sigma_f \cdot X_p + w_p^T \cdot \Delta_\mu \cdot w_p} \quad (4.3)$$

其中 X_p 和 w_p 分别为投资组合的因子暴露度矩阵和个股权重， X_p^T 和 w_p^T 分别为 X_p 和 w_p 的转置； Σ_f 为因子协方差矩阵； Δ_μ 为特异风险对角矩阵。

理论上，相对于经验估计（即直接使用历史收益率数据计算协方差矩阵），因子风险模型有如下优点：

- 实现了风险预测问题的降维降噪⁷；
- 降低股票上市时间过短（造成历史收益率数据不足）、停牌（导致个股的历史波动率过低）、数据缺失等边际情况对风险预测的干扰；
- 需要估计的变量数目减少，加快风险预测和权重优化的计算速度

在 3.4 节，我们通过测试不同一系列风险最小化组合，通过实证方式证明了因子模型在不同条件下，效果均显著优于经验估计。而且当投资组合包含的股票数目较多的情况下，因子风险模型的效果远好于经验估计，这与因子风险模型能够实现降维降噪的特点是吻合的。

⁷ 例如对于沪深 300 组合，其风险预测需要对一个 300 乘 300 的收益协方差矩阵进行估计；通过因子分解，则只需要对一个 39 乘 39 的因子协方差矩阵进行估计，从而实现了问题的降维处理；此外，问题的降维处理减少了需要进行估计的变量数目。在收益率序列长度一定的情况下，需要估计的变量数目减少，能够有效降低抽样误差（sampling errors）的影响，从而达到降噪的效果。