

林晓明 执业证书编号：S0570516010001
研究员 0755-82080134
linxiaoming@htsc.com

陈烨 执业证书编号：S0570518080004
研究员 010-56793927
chenye@htsc.com

相关研究

- 1 《金工：动量增强因子在行业配置中的应用》2018.10
- 2 《金工：基于回归法的基金持股仓位测算》2018.10
- 3 《金工：周期视角下的因子投资时钟》2018.10

酌古御今：指数增强基金收益分析

指数增强基金分析系列之二

指数增强型基金的超额收益来源与行业、风格配置等多种因素相关

指数增强型基金作为主动投资与被动投资的有机结合，近年来受到投资者的广泛关注。本文旨在从实战经验出发，探秘现有指数增强型公募基金产品的收益来源。本文首先基于收益分解的直接证据，依次对比行业选择与行业内选股、指数成分股内与成分股外选股获益的相对难易。之后，基于收益相关分析，按照从单维到多维的顺序，挖掘行业、风格、持股集中度的配置特色与基金超额收益表现的关系。我们发现，指数增强基金的收益受到行业选择、行业内选股、风格配置等多种关键因素影响。

与行业选择相比，行业内选股是指数增强基金更大的超额收益来源

本文基于 Brison 绩效归因模型将指数增强型基金的超额收益率分解为行业选择收益、行业内选股收益和交互项。与行业选择相比，行业内选股更可能获得高期望值的超额收益率。但是二者在获益胜率和获益稳定性上没有明显区别。在单个行业内选股方面，轻工制造和机械行业内选股相对容易获取高期望、高胜率、稳定的超额收益率。

与指数成分股内选股相比，指数成分股外选股能提供更多超额收益

本文从指数成分股内、外选股角度出发，将指数增强型基金的超额收益率分解为指数成分股内选股收益和指数成分股外选股收益。与指数成分股内选股相比，指数成分股外选股更有可能获得高期望、高胜率、稳定的超额收益率。

主动进行行业配置、不同行业使用不同选股因子可能带来更多超额收益

在行业配置方面，相对指数的行业中性程度较低、行业配置稳定性较低、行业择时准确度较高的基金，其超额收益大概率较好。在持仓风格上，持仓风格的行业间一致性较低、持仓风格稳定性较低的基金，其超额收益大概率较好。换言之，主动进行行业配置、在不同行业内使用不同选股因子、主动进行风格择时可能会为指数增强基金带来更多超额收益。在持仓集中度上，持仓集中度较高的基金，其超额收益不一定较好，但是其跟踪误差大概率较大。

聚类结果进一步证实指数增强基金采取主动型行业、风格配置方案更佳

本文利用 k-means 聚类方法将指数增强型基金按照多维配置特色分为两类，考察基金类别间的配置特色与超额收益的相关关系。聚类结果表明，相对指数行业中性程度较低、持仓风格行业间一致性较低、并且持仓风格稳定性较低的基金，其超额收益大概率较好。也就是说，指数增强基金投资人应该更多地考虑采取主动型行业与风格配置方案：不严格限制行业中性、不同行业内区别性选股、并进行风格择时配置。

风险提示：本报告中的收益分析是基于历史数据，由此总结的规律在未来存在失效的可能。本报告仅使用基金的半年报和年报公布的持仓数据，时间跨越尺度较大，存在一定局限性。由于现有宽基指数的公募指数增强型基金数目有限，本报告只考察了基金配置特色与超额收益的线性相关关系，未讨论可能的非线性关系，敬请注意。

正文目录

指数增强型公募基金收益分析概览.....	5
收益分析：行业选择、行业内选股的获益难易对比	7
行业选择与行业内选股获益对比	7
行业选择与行业内选股收益的期望值	7
行业选择与行业内选股的获益胜率	8
行业选择与行业内选股的获益稳定性	9
行业间对比：相对容易获取超额收益率的行业	10
本节小结	11
收益分析：指数成分股内、外选股的获益难易对比	12
指数成分股内、外配股权重	12
指数成分股内、外选股获益对比	12
指数成分股内、外选股收益的期望值	13
指数成分股内、外选股的获益胜率	13
指数成分股内、外选股的获益稳定性	14
本节小结	14
配置特色：相对指数的行业中性程度、行业配置稳定性与行业择时准确度	15
相对指数的行业中性程度	15
非行业中性基金的整体行业偏离度稳定性	16
非行业中性基金的单行业超/低配上限	16
相对指数的行业中性程度与超额收益率	16
行业配置稳定性	18
行业配置稳定性与超额收益率	18
行业择时准确度	19
本节小结	21
配置特色：持仓风格的行业间一致性与持仓风格稳定性	22
持仓风格的行业间一致性	23
持仓风格的行业间一致性与超额收益率	25
持仓风格稳定性	27
持仓风格稳定性与超额收益率	28
本节小结	29
配置特色：持仓集中度	30
持仓集中度与超额收益率	30
持仓集中度与跟踪误差	30
本节小结	31
基于配置特色的基金聚类	32
指标集构建	32
k-means 聚类	33
本节小结	34
全文总结	35
附录一 持仓风格行业间差异度举例	36
附录二 风格偏好稳定性与超额收益率的关系	39
风险提示	40

图表目录

图表 1: 指数增强型基金收益分析框架	5
图表 2: 本文分析的 34 只指数增强型基金	6
图表 3: 行业选择和行业内选股收益期望值的分布直方图	8
图表 4: 行业选择和行业内选股收益期望值的关系图	8
图表 5: 行业选择与行业内选股收益占比的关系图	8
图表 6: 行业选择和行业内选股获益胜率分布直方图	9
图表 7: 行业选择和行业内选股获益胜率的关系图	9
图表 8: 行业选择与行业内选股获益稳定性的分布直方图	9
图表 9: 行业选择与行业内选股获益稳定性的关系图	9
图表 10: 一级行业的单行业内选股收益的期望值	11
图表 11: 一级行业的单行业内选股获益胜率	11
图表 12: 一级行业的单行业内选股获益稳定性	11
图表 13: 指数成分股内配股权重的分布直方图	12
图表 14: 指数成分股内、外选股收益期望值的分布直方图	13
图表 15: 指数成分股内、外选股收益期望值的关系图	13
图表 16: 指数成分股内、外选股获益胜率的分布直方图	13
图表 17: 指数成分股内、外选股获益胜率的关系图	13
图表 18: 指数成分股内、外选股获益稳定性的分布直方图	14
图表 19: 指数成分股内、外选股获益稳定性的关系图	14
图表 20: 整体行业偏离度的分布直方图	15
图表 21: 非行业中性基金的整体行业偏离度稳定性的分布直方图	16
图表 22: 非行业中性基金的单行业最大偏离度的分布直方图	16
图表 23: 年化超额收益率与整体行业偏离度的关系图	17
图表 24: 超额收益率排名百分位与整体行业偏离度的关系图	17
图表 25: 相对指数行业中性基金与行业非中性基金的超额收益率对比	17
图表 26: 行业权重前后期变动量的分布直方图	18
图表 27: 行业权重前后期相关系数的分布直方图	18
图表 28: 年化超额收益率与行业权重前后期变动量的关系图	19
图表 29: 超额收益率排名百分位与行业权重前后期变动量的关系图	19
图表 30: 年化超额收益率与行业权重前后期相关系数的关系图	19
图表 31: 超额收益率排名百分位与行业权重前后期相关系数的关系图	19
图表 32: 行业择时准确度的分布直方图	20
图表 33: 年化超额收益率与行业择时准确度的关系图	20
图表 34: 超额收益率排名百分位与行业择时准确度的关系图	20
图表 35: 本文使用的因子列表	22
图表 36: 自身因子偏好行业间差异度最大值 0.1624: 基金 A-2017 年报的自身因子偏好行业分布图	24
图表 37: 自身因子偏好行业间差异度最小值 0.0743: 基金 B-2017 半年报的自身因子偏好行业分布图	25
图表 38: 自身持仓风格的行业间差异度的分布直方图	25

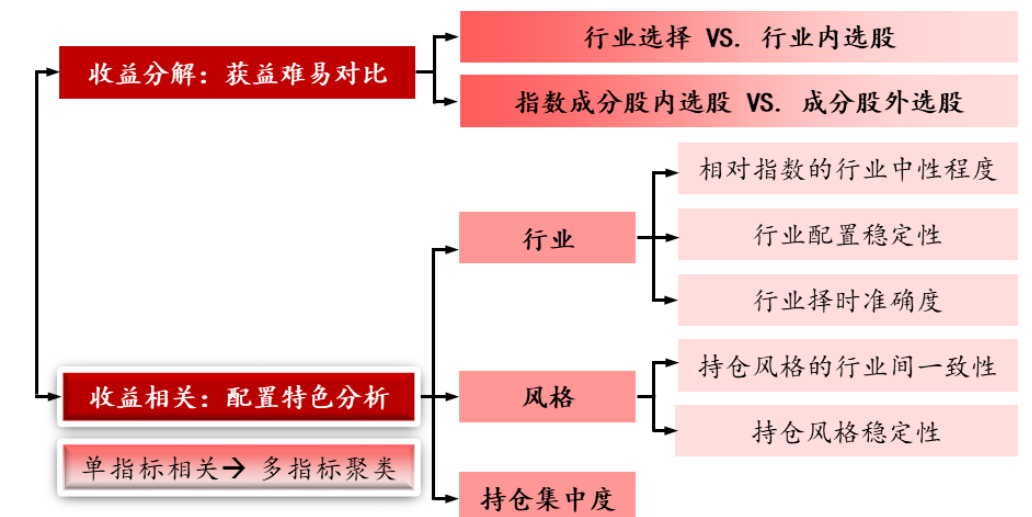
图表 39: 相对指数持仓风格的行业间差异度的分布直方图	25
图表 40: 不同指数增强基金自身因子偏好行业间差异度的分布直方图	26
图表 41: 不同指数增强基金相对基准的因子偏好行业间差异度分布图	26
图表 42: 年化超额收益率与因子偏好行业间差异度的关系图	26
图表 43: 超额收益率排名百分位与因子偏好行业间差异度的关系图	26
图表 44: 年化超额收益率与风格偏好行业间差异度的关系图	26
图表 45: 超额收益率排名百分位与风格偏好行业间差异度的关系图	26
图表 46: 持仓风格前后期变动量的分布直方图	28
图表 47: 持仓风格前后期相关系数的分布直方图	28
图表 48: 年化超额收益率与因子偏好前后期变动量的关系图	28
图表 49: 超额收益率排名百分位与因子偏好前后期变动量的关系图	28
图表 50: 年化超额收益率与因子偏好前后期相关系数的关系图	29
图表 51: 超额收益率排名百分位与因子偏好前后期相关系数的关系图	29
图表 52: 不同指数增强基金前 10 重仓股权重的分布直方图	30
图表 53: 不同指数增强基金前 20 重仓股权重的分布直方图	30
图表 54: 年化超额收益率与前 10 重仓股权重的关系图	31
图表 55: 年化超额收益率与前 20 重仓股权重的关系图	31
图表 56: 年化跟踪误差与前 10 重仓股权重的关系图	31
图表 57: 年化跟踪误差与前 20 重仓股权重的关系图	31
图表 58: 初始指标集内单个配置指标的拟合优度及方向	32
图表 59: 初始指标集的指标间相关关系矩阵	32
图表 60: 基金类别中心的指标值	34
图表 61: 基金类别的年化超额收益率	34
图表 62: 基金类别的超额收益率排名百分位	34
图表 63: 自身风格偏好行业间差异度最大值 0.1343: 基金 C-2016 半年报的自身风格偏好行业分布图	36
图表 64: 自身风格偏好行业间差异度最小值 0.061: 基金 D-2018 半年报的自身因子偏好行业分布图	36
图表 65: 相对指数因子偏好行业间差异度最大值 0.2522: 基金 E-2015 年报的相对指数因子偏好行业分布图	37
图表 66: 相对指数因子偏好行业间差异度最小值 0.0025: 基金 F-2016 年报的相对指数因子偏好行业分布图	37
图表 67: 相对指数风格偏好行业间差异度最大值 0.2695: 基金 G-2016 半年报的相对指数风格偏好行业分布图	37
图表 68: 相对指数风格偏好行业间差异度最小值 0.0025: 基金 F-2016 年报的相对指数风格偏好行业分布图	38
图表 69: 年化超额收益率与风格偏好前后期变动量的关系图	39
图表 70: 超额收益率排名百分位与风格偏好前后期变动量的关系图	39
图表 71: 年化超额收益率与风格偏好前后期相关系数的关系图	39
图表 72: 超额收益率排名百分位与风格偏好前后期相关系数的关系图	39

指数增强型公募基金收益分析概览

指数增强基金作为主动投资与被动投资的有机结合，具有策略相对透明、风格相对平稳、风险可控性较强、获取超额收益可能性较高的优点，近十年来投资热度不断攀升，基金数目逐年稳步递增。总体来看，指数增强型基金业绩表现不俗，但正所谓，龙生九子凤育九雏，如何在波动市场中获取高期望、稳定的超额收益率，各家基金各显神通。在这百花齐放的指数增强型基金市场中，是否存在普遍的获利规律，表现优秀的基金产品又有怎样的共同特点？本文作为华泰指数增强基金分析系列的第二篇，旨在从实战经验出发，治繁总要，挖掘现有指数增强型基金历史数据中的确定性规律，探秘名列前茅基金的取胜法宝。

本文的分析框架如图表 1 所示。我们将首先基于收益分解的直接证据，依次对比指数增强型基金行业选择、行业内选股的获益难易程度，和在指数成分股内、外选股的获益难易程度。之后从收益相关角度出发，剖析现有指数增强型基金在行业、风格、持股集中度上的配置特色，并逐一分析单个配置指标与基金超额收益率的相关关系。最后，基于多个配置指标将基金聚类，再次验证配置指标对基金超额收益率的影响，从而归纳总结，为指数增强策略投资人提供参考。

图表1： 指数增强型基金收益分析框架



资料来源：华泰证券研究所

我们在目前所有跟踪沪深 300、中证 500、中证 1000、上证 50 四个宽基指数的公募指数增强基金产品中，选取了成立日在 2017 年底之前的 34 只基金作为数据样本，列举在图表 2 中。

我们提取了每只基金在 2015-2017 年半年报、年报以及 2018 年半年报，共计 7 个截面期上公布的持仓明细，在每个截面期上，从基金持仓股票中剔除 IPO 不满 3 个月的个股，并将剩余个股的持仓权重归一化。此外，我们提取了 7 个截面期上标的指数的成分股明细，以及全部 A 股所属 29 个一级行业分类，用于后续数据分析。

我们将收益率的考察区间定义为截面期前一个月最后一个交易日至截面期后一个月最后一个交易日。我们提取截面期前后时间节点上，基金的复权单位净值、标的指数收盘价以及全部 A 股的复权收盘价，分别计算基金、标的指数和个股的区间收益率。例如，2018 年基金半年报对应的截面期为 2018.6.30，我们考察的收益率区间为 2018.5.31-2018.7.31，亦即我们假设基金半年报公布的详细持仓数据在截面期前后各一个月的时间内基本稳定不变，通过基金持仓股票在这段时间内的表现来分析基金的操作特色和收益来源。

若无特别说明，后文中的分析指标值为 7 个截面期上的均值。

图表2： 本文分析的 34 只指数增强型基金

基金代码	基金名称	现任基金经理	成立日期	基金规模	
				(亿元)	业绩比较基准
050002.OF	博时裕富沪深 300A	桂征辉	2003/8/26	57.59	沪深 300 指数*95%+银行同业存款利率*5%
200002.OF	长城久泰沪深 300	杨建华	2004/5/21	7.42	沪深 300 指数收益率*95%+银行同业存款利率*5%
310318.OF	申万菱信沪深 300	袁英杰	2004/11/29	4.14	沪深 300 指数收益率*95%+1.5%
450008.OF	国富沪深 300	张志强	2009/9/3	1.38	沪深 300 指数*95%+银行同业存款利率*5%
100038.OF	富国沪深 300	李笑薇,方旻	2009/12/16	46.28	沪深 300 指数收益率*95%+1.5%
162213.OF	泰达宏利沪深 300A	杨超	2010/4/23	1.88	沪深 300 指数收益率*95%+同业存款利率*5%
*166007.OF	中欧沪深 300A	曲径	2010/6/24	-	沪深 300 指数收益率*95%+银行活期存款利率*5%
*注：自 2018 年 10 月 8 日起已更名为“中欧互通精选混合型证券投资基金”，基金经理未变。					
163407.OF	兴全沪深 300	申庆	2010/11/2	15.74	沪深 300 指数*95%+同业存款利率*5%
519116.OF	浦银安盛沪深 300	陈士俊	2010/12/10	1.21	沪深 300 指数收益率*95%+1%
110030.OF	易方达沪深 300 量化	官泽帆	2012/7/5	10.36	沪深 300 指数收益率*95%+活期存款利率*5%
000312.OF	华安沪深 300 量化 A	许之彦,孙晨进	2013/9/27	5.40	沪深 300 指数收益率*95%+商业银行活期存款基准利率*5%
000311.OF	景顺长城沪深 300	黎海威	2013/10/29	84.76	沪深 300 指数收益率*95%+1.5%
000176.OF	嘉实沪深 300 增强	龙昌伦	2014/12/26	6.21	沪深 300 指数收益率*95%+银行活期存款利率*5%
001015.OF	华夏沪深 300 增强 A	宋洋	2015/2/10	3.56	沪深 300 指数收益率*95%+1.5%
002310.OF	创金合信沪深 300A	程志田	2015/12/31	1.38	沪深 300 指数收益率*95%+一年期人民币定期存款利率*5%
003015.OF	中金沪深 300A	魏李	2016/7/22	0.15	沪深 300 指数收益率*95%+银行活期存款利率*5%
003261.OF	安信沪深 300A	龙川	2016/10/12	0.36	沪深 300 指数收益率*95%+1.5%
003876.OF	华宝沪深 300	徐林明	2016/12/9	2.09	沪深 300 指数收益率*95%+1.5%
004190.OF	招商沪深 300A	王平	2017/2/10	0.43	沪深 300 指数收益率*95%+同期银行活期存款利率*5%
167601.OF	国金沪深 300	宫雪,艾翀	2017/9/1	0.26	沪深 300 指数收益率*95%+同期银行活期存款利率*5%
161017.OF	富国中证 500	徐幼华,李笑薇,方旻	2011/10/12	24.10	中证 500 指数收益率*95%+1%
000478.OF	建信中证 500A	叶乐天	2014/1/27	32.36	中证 500 指数收益率*95%+商业银行活期存款利率*5%
002311.OF	创金合信中证 500A	程志田	2015/12/31	1.52	中证 500 指数收益率*95%+一年期人民币定期存款利率*5%
002510.OF	申万菱信中证 500	俞斌,袁英杰	2016/4/21	0.55	中证 500 指数收益率*95%+银行活期存款利率*5%
003016.OF	中金中证 500A	魏李	2016/7/22	0.19	中证 500 指数收益率*95%+银行活期存款利率*5%
002906.OF	南方中证 500 增强 A	李佳亮	2016/11/23	2.05	中证 500 指数收益率*95%+1%
003986.OF	申万菱信中证 500 优选	袁英杰	2017/1/10	2.02	中证 500 指数收益率*95%+银行活期存款利率*5%
004192.OF	招商中证 500A	王平	2017/5/17	0.10	中证 500 指数收益率*95%+同期银行活期存款利率*5%
004945.OF	长信中证 500	宋海岸	2017/8/30	0.32	中证 500 指数收益率*95%+银行活期存款利率*5%
005062.OF	博时中证 500A	桂征辉	2017/9/26	1.15	中证 500 指数收益率*95%+银行活期存款利率*5%
003646.OF	创金合信中证 1000A	程志田	2016/12/22	0.18	中证 1000 指数收益率*95%+一年期人民币定期存款利率*5%
004194.OF	招商中证 1000A	王平	2017/3/3	0.09	中证 1000 指数收益率*95%+同期银行活期存款利率*5%
110003.OF	易方达上证 50 指数 A	张胜记	2004/3/22	116.52	上证 50 指数
399001.OF	中海上证 50	彭海平	2010/3/25	1.38	上证 50 指数*95%+一年期银行定期存款利率*5%

资料来源：Wind，华泰证券研究所

收益分析：行业选择、行业内选股的获益难易对比

花无百日红，尚有重开日。行业周期转，浮沉此彼时。行业的发展状况随着经济周期的变化而呈现此涨彼跌的波动趋势。在经济周期的某个阶段，某个行业可能整体发展较好，而其他某个行业可能整体表现较差。在一个完整的经济周期内，某些先行行业发展良好，可能预示着跟随行业在下一阶段内获益颇丰。倘若基金经理能够认知精准、择时准确，每次超配预期收益较好的行业，低配预期收益较差的行业，那么获取超额收益率就变得非常容易。但是实际上，市场风云变化，行业间相对发展趋势切换不定，判断失误很可能导致基金产品相对指数发生回撤。因此，不同的基金经理在行业配置上选用的策略不同。

本节将从行业角度对超额收益率进行分解归因，首先整体对比行业选择和行业内选股获益的相对难易。之后，将行业内选股维度进一步拆解，比较不同行业的行业内选股获益情况，尝试挖掘相对容易获取超额收益率的行业。

行业选择与行业内选股获益对比

1986 年，Brinson 等人提出了 Brinson 绩效归因模型，该模型将投资组合的收益分解为四个部分——基准收益、资产配置收益、个股选择收益和交互项。作为绩效归因方法中重要的理论基石，该模型至今仍被广泛使用。我们基于单期 Brinson 绩效归因模型，从行业角度出发，将基金相对标的指数的超额收益率拆分为行业选择收益、行业内选股收益和交互项三个部分，即：

$$\text{超额收益率} = \text{行业选择收益} + \text{行业内选股收益} + \text{交互项}$$

令第 t 个截面期上，基金在行业 i 上的配股权重为 $w_{t,i}^P \in [0, 1]$ ，且 $\sum_{i=1}^{29} w_{t,i}^P = 1$ 。标的指数在行业 i 上的配股权重为 $w_{t,i}^B \in [0, 1]$ ，且 $\sum_{i=1}^{29} w_{t,i}^B = 1$ 。标的指数的收益率为 r_t^B 。若满仓投资于行业 i ，即将行业 i 内的持仓个股权重等比例扩充至权重和为 1，基金在行业 i 上可获得的收益率为 $r_{t,i}^P$ ，标的指数在行业 i 上可获得的收益率为 $r_{t,i}^B$ 。注意，如果 $w_{t,i}^P = 0$ ，则 $r_{t,i}^P = \text{NaN}$ ，NaN 表示该数据无效，在后续处理中被舍弃。同理，如果 $w_{t,i}^B = 0$ ，则 $r_{t,i}^B = \text{NaN}$ 。那么，在第 t 个截面期上，行业选择收益 AR_t 表示为：

$$AR_t = \sum_{i=1}^{29} (w_{t,i}^P - w_{t,i}^B) \cdot (r_{t,i}^B - r_t^B)$$

行业内选股收益 SR_t 表示为：

$$SR_t = \sum_{i=1}^{29} w_{t,i}^B \cdot (r_{t,i}^P - r_{t,i}^B)$$

在本小节中，我们将从期望值、胜率和稳定性三个指标上，依次考察行业选择和行业内选股获得超额收益率的相对难易。

行业选择与行业内选股收益的期望值

期望值定义为行业选择收益和行业内选股收益在截面期时间序列上的均值。这里为了方便理解，二者均采用区间收益率形式，而非年化收益率形式。

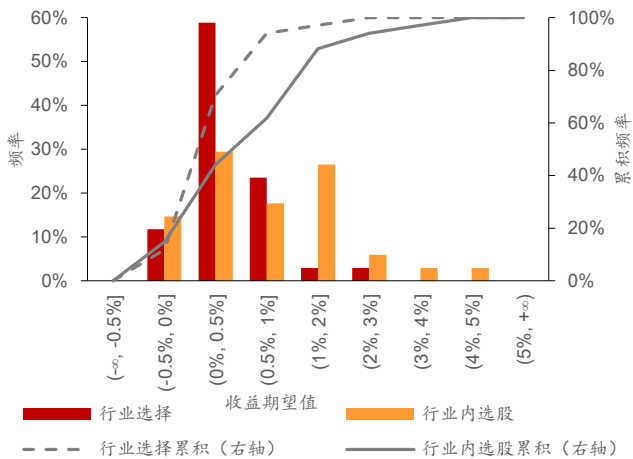
图表 3 为基金在行业选择和行业内选股收益期望值上的分布直方图。在行业选择收益的期望值上，约 82% 基金为正值，只有约 6% 基金高于 1%。而在行业内选股收益的期望值上，约 85% 基金为正值，约 38% 基金高于 1%，约 6% 基金高于 3%。二者对比可得，与行业选择相比，通过行业内选股获得高期望超额收益率的基金比例更大。

此外，对于所有基金，我们以行业选择收益的期望值为 x 轴，以行业内选股收益的期望值为 y 轴，绘制了二者的相对关系图，见图表 4。图中，灰色虚线表示 $y=x$ 的趋势线。由图可知，大部分基金都位于灰色虚线的上方，说明其行业内选股收益期望值高于行业选择。

此外，由于行业选择收益和行业内选股收益之和并不等于基金相对指数的超额收益率，因此，除了收益之外，我们还分析了收益占比。收益占比定义为行业选择收益和行业内选股收益分别占基金相对指数超额收益率的比例，并对截面期时间序列取均值。我们以行业选择收益占比为 x 轴，行业内选股收益占比为 y 轴，将二者的关系绘制在图表 5 中。由于存在某些基金相对指数超额收益率过小的情况，导致收益占比存在极端值。为了方便观察，我们在图表 5 的小图内提供了收益占比区间为[-25%, 100%]的局部放大图。收益占比与收益期望值的趋势一致：大部分基金都位于灰色虚线上方，说明其行业内选股收益占比大于行业选择。

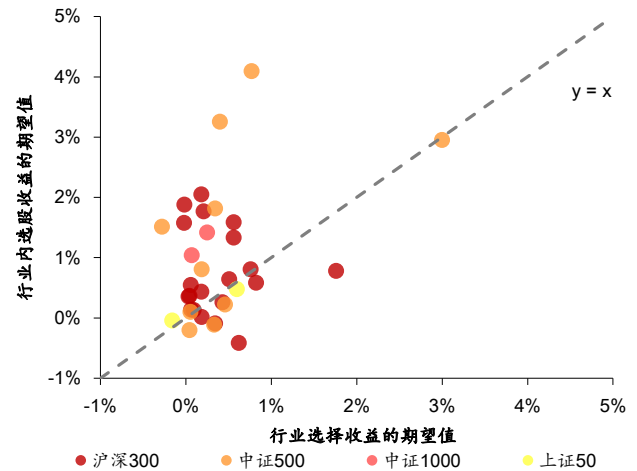
综上，与行业选择相比，行业内选股更有可能获取高期望值的超额收益率。

图表3： 行业选择和行业内选股收益期望值的分布直方图



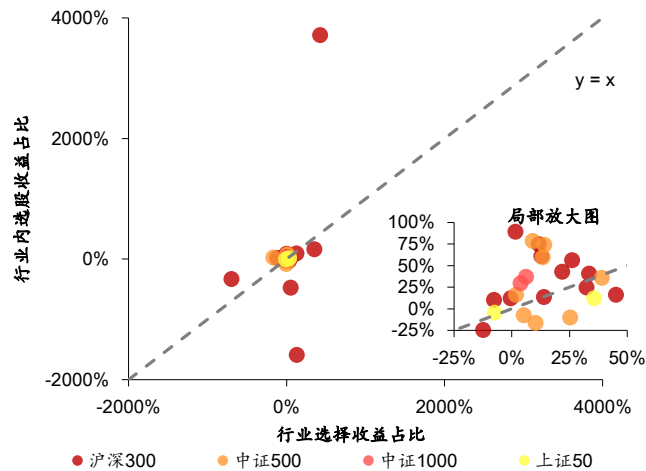
资料来源：Wind，华泰证券研究所

图表4： 行业选择和行业内选股收益期望值的关系图



资料来源：Wind，华泰证券研究所

图表5： 行业选择与行业内选股收益占比的关系图



资料来源：Wind，华泰证券研究所

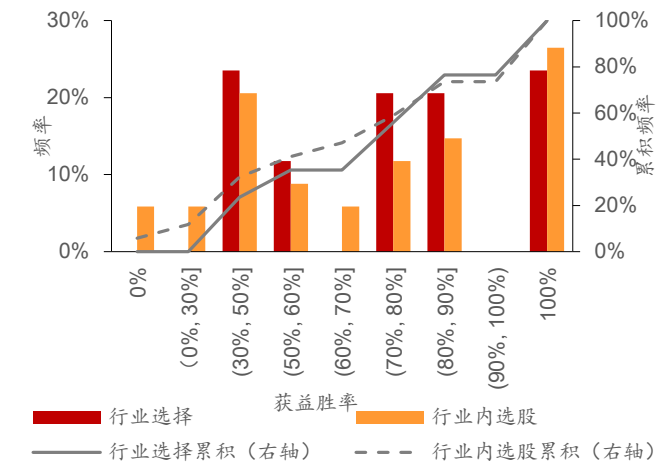
行业选择与行业内选股的获胜胜率

胜率定义为行业选择收益、行业内选股收益在截面期时间序列上为正值的频率（正值数目除以有效截面期数目）。基金在行业选择、行业内选股获胜胜率上的分布直方图如图表 6 所示。在行业选择获胜胜率上，所有基金都大于 30%，约 76% 基金大于 50%，约 24% 基金为 100%。在行业内选股获胜胜率上，约 6% 基金为 0%，约 68% 基金大于 50%，约 26% 基金 100%。

图表 7 为行业选择获益胜率与行业内选股获益胜率的关系图。从图中可以看出，基金较为均匀地分布在灰色虚线的两侧。

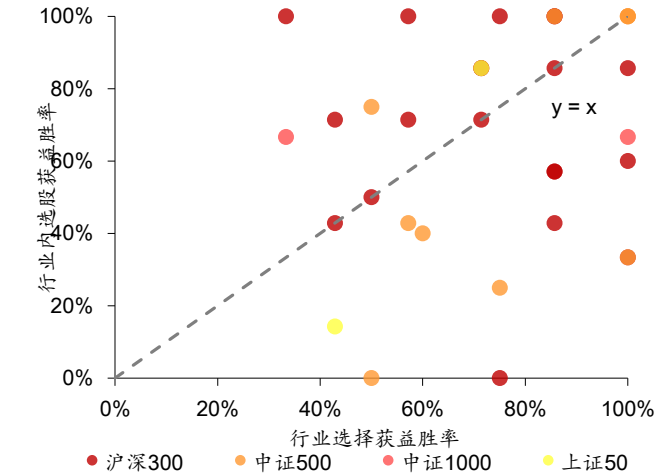
综上，在获益胜率上，行业选择与行业内选股无明显区别。

图表 6：行业选择和行业内选股获益胜率的分布直方图



资料来源：Wind，华泰证券研究所

图表 7：行业选择和行业内选股获益胜率的关系图



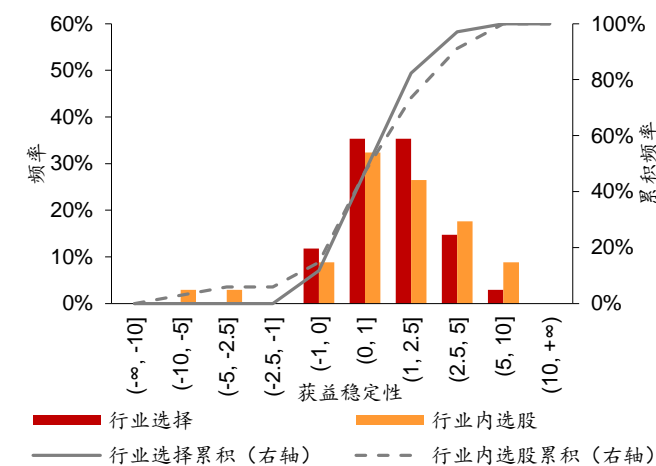
资料来源：Wind，华泰证券研究所

行业选择与行业内选股的获益稳定性

稳定性定义为行业选择收益与行业内选股收益在截面期时间序列上的 t 值。t 值可以反映样本值是否与原假设值存在显著差异。一般地，t 值大于 2.5 可认为差异显著，即有稳定的正值超额收益率。由于单个基金至多有 7 个截面期数据，数据量较少，统计效力较弱。但是，我们仍可以从 t 值上定性评价行业选择与行业内选股获益的稳定性。t 值的计算公式如下，这里，样本数目为有效截面期数目，原假设值为 0：

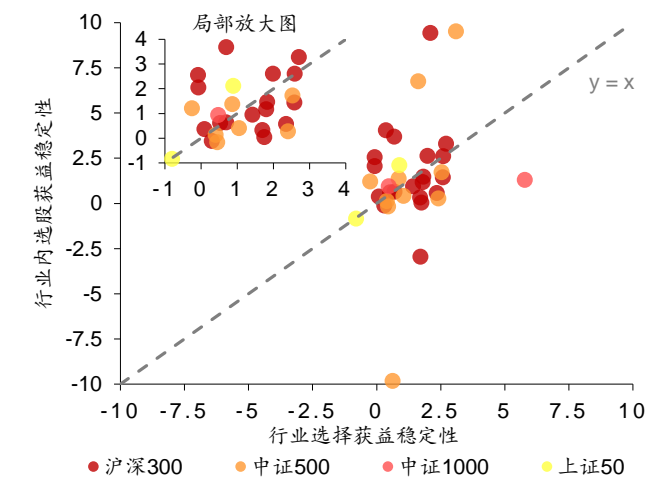
$$t \text{ 值} = \frac{\text{样本均值} - \text{原假设值}}{\sqrt{\frac{\text{样本方差}}{\text{样本数目} - 1}}}$$

图表 8：行业选择与行业内选股获益稳定性的分布直方图



资料来源：Wind，华泰证券研究所

图表 9：行业选择与行业内选股获益稳定性的关系图



资料来源：Wind，华泰证券研究所

图表 8 为基金在行业选择、行业内选股获益稳定性上的分布直方图。在行业选择获益稳定性上，约 88% 基金 t 值为正，仅有约 18% 基金 t 值大于 2.5。在行业内选股获益稳定性上，约 85% 基金 t 值为正，仅有约 26% 基金 t 值大于 2.5。这说明只有小部分基金可以较为稳定地通过行业选择或行业内选股获取超额收益。整体分布上，行业选择与行业内选股无明显区别。

同样，对于所有基金，我们以行业选择获益稳定性为 x 轴，以行业内选股获益稳定性为 y 轴，在图表 9 中绘制了二者的相对关系，并小图内提供了获益稳定性区间为 $[-1, 4]$ 的局部放大图。由图可知，基金较为均匀地分布在灰色虚线两侧。

综上，在获益胜率上，行业选择与行业内选股无明显区别。

行业间对比：相对容易获取超额收益率的行业

在上一小节中，我们对所有行业的行业内选股获益进行了整体分析，在本小节中，我们将其拆解到单行业维度，对比不同行业的单行业内选股获益情况，找寻相对容易获取超额收益率的行业。

在上一小节中，我们定义了第 t 个截面期上，若满仓投资于行业 i ，即将行业 i 内的持仓个股权重等比例扩充至权重和为 1，基金在行业 i 上可获得的收益率为 $r_{t,i}^P$ ，标的指数在行业 i 上可获得的收益率为 $r_{t,i}^B$ 。那么，在第 t 个截面期上，行业 i 的单行业内选股收益 $SR_{t,i}$ 表示为：

$$SR_{t,i} = r_{t,i}^P - r_{t,i}^B$$

在本小节中，我们将延续前文的分析思路，从期望值、胜率和稳定性三个指标上，对比不同行业在行业内选股获益上的相对难易。

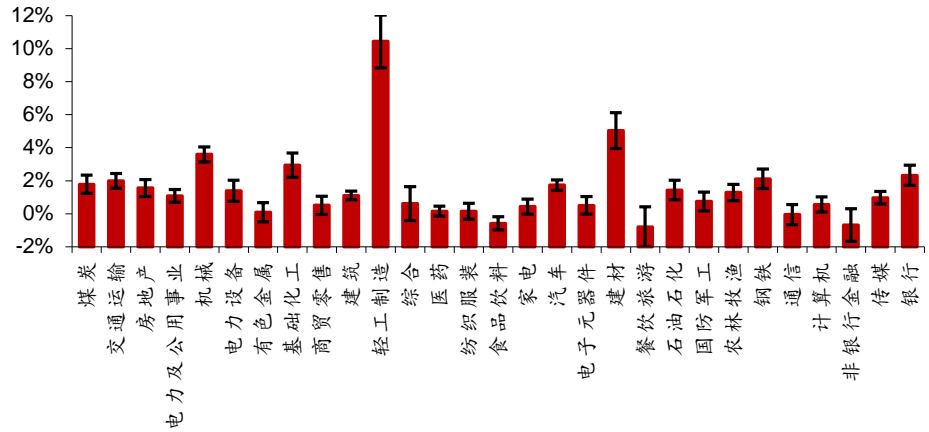
一级行业的单行业内选股收益的期望值、胜率、稳定性分别如图表 10-12 所示。图中柱形表示所有基金的指标均值，误差线表示所有基金的指标标准误。其中，标准误公式为：

$$\text{标准误} = \frac{\text{样本标准差}}{\sqrt{\text{样本数目}}}$$

这里，样本数目为指标值不是无效数据 (NaN) 的基金数目，标准误可以避免样本数目不同对标准差的影响。

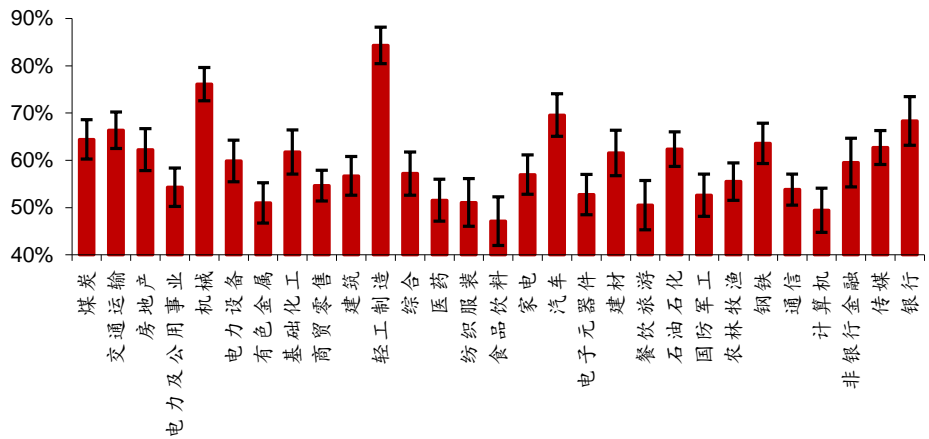
从图中可以看出，单行业内选股收益期望值最大的前五个行业依次为轻工制造、建材、机械、基础化工和银行；单行业内选股获益胜率最大的前五个行业依次为轻工制造、机械、汽车、银行和交通运输；单行业内选股收益稳定性最高的前 5 个行业依次为建筑、机械、轻工制造、汽车和交通运输。综合以上三个指标可得，在一级行业内，相对容易获取高期望、高胜率、稳定的超额收益率的行业为轻工制造和机械。

图表10： 一级行业的单行业内选股收益的期望值



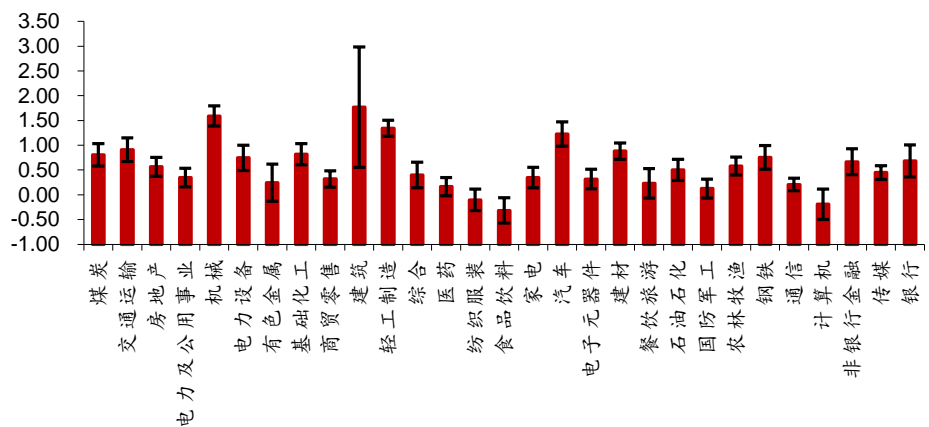
资料来源：Wind，华泰证券研究所

图表11： 一级行业的单行业内选股获胜胜率



资料来源：Wind，华泰证券研究所

图表12： 一级行业的单行业内选股获益稳定性



资料来源：Wind，华泰证券研究所

本节小结

与行业选择相比，行业内选股获取高期望值的超额收益率的概率更大。但是二者在获胜胜率和获益稳定性上没有明显区别。

在 29 个一级行业内，相对容易获取高期望、高胜率、稳定的超额收益率的行业为轻工制造和机械。

收益分析：指数成分股内、外选股的获益难易对比

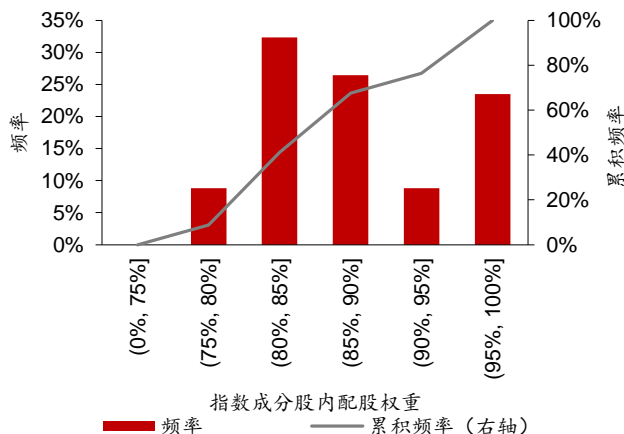
指数增强型基金的合同明确规定了基金的股票资产中投资于标的指数成分股资产的比例。一般地，基金股票资产中投资于标的成分股和备选成分股的资产不低于非现金基金资产的 80%。在此规定下，为了获取超额收益率，基金经理可能选择在指数成分股内选股，并调配成分股的权重以期获得成分内选股超额收益；也可能选择利用接近 20% 的投资比例，在指数成分股外挑选看好的个股。那么，从历史经验上来看，哪种选择更有可能产生高期望值、高胜率、稳定的超额收益率呢？

本节将从指数成分股内、外选股角度出发，首先分析指数增强型基金在指数成分股内、外配股的倾向性，之后对超额收益率进行分解归因，对比指数成分股内、外选股获益的相对难易。

指数成分股内、外配股权重

对于每个截面期，我们基于基金的持仓明细和标的指数的成分股明细，计算得到每个基金的指数成分股内配股权重，并对截面期取均值。图表 13 绘制了基金成分股内配股权重的分布直方图。由图可知，除了 3 只（约 9%）基金的指数成分股内配股权重处于 (75%, 80%] 区间之外，其余基金的成分股内配股权重均位于 80% 以上。约 32% 基金成分股外配股倾向较为明显，其成分股内配股权重刚刚超过规定的 80% 下限，位于 (80%, 85%] 区间内。而约 24% 基金在指数成分股内配股倾向较为明显，成分股内配股权重高于 95%。

图表13： 指数成分股内配股权重的分布直方图



资料来源：Wind，华泰证券研究所

指数成分股内、外选股获益对比

我们从指数成分股内外选股角度进行收益归因，将基金超额收益率分解为指数成分股内选股所得超额收益率和指数成分股外选股所得超额收益率。在每个截面期上，对于基金持仓股票中属于指数成分股的所有个股，将个股收益率按照基金在该个股上相对指数的配股权重差值加权求和，得到基金在指数成分股内选股所得超额收益率。将基金相对指数的超额收益率减去基金在指数成分股内选股所得超额收益率，得到基金在指数成分股外选股所得超额收益率。

接下来，我们将延续上一节的分析思路，从期望值、胜率和稳定性三个指标上，依次考察指数成分股内、外选股获得超额收益率的相对难易。

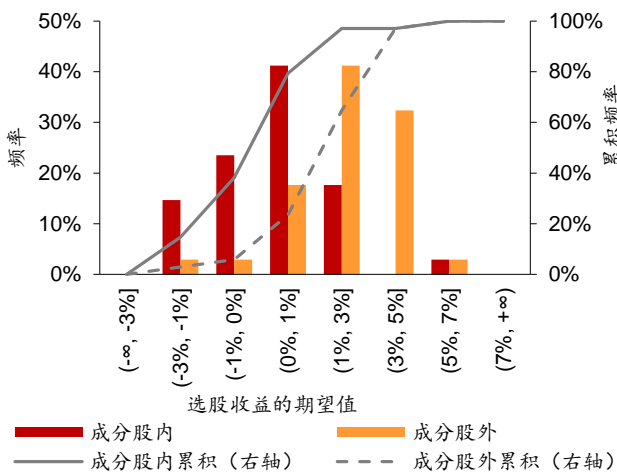
指数成分股内、外选股收益的期望值

期望值同上一小节，即成分股内、外选股所得超额收益率在截面期时间序列上的均值。基金在成分股内、外选股收益期望值上的分布如图表 14 所示。从图中可以看出，与指数成分股内选股相比，成分股外选股的收益期望值为正的基金比例更高，且期望值的均值、中位数更大。

同样，我们分别以成分股内、外选股的收益期望值为 x 、 y 轴，绘制了二者的关系图，如图表 15 所示。由图可知，大部分基金的数据点都位于表征 $y=x$ 的灰色虚线的上方，说明其指数成分股外选股所得超额收益率大于成分股内选股所得超额收益率。

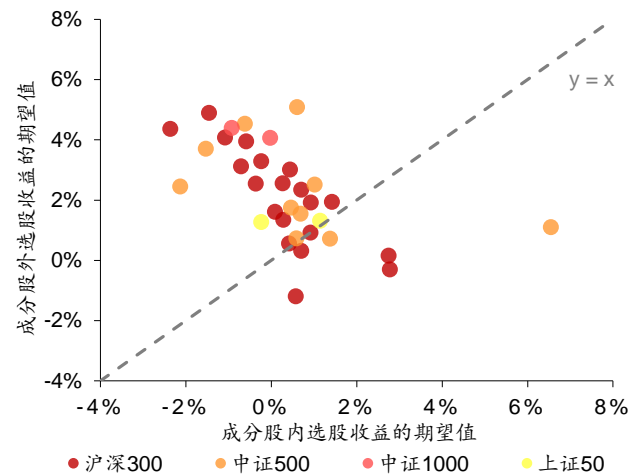
综上，与成分股内选股相比，指数成分股外选股更有可能获取高期望值的超额收益率。

图表14： 指数成分股内、外选股收益期望值的分布直方图



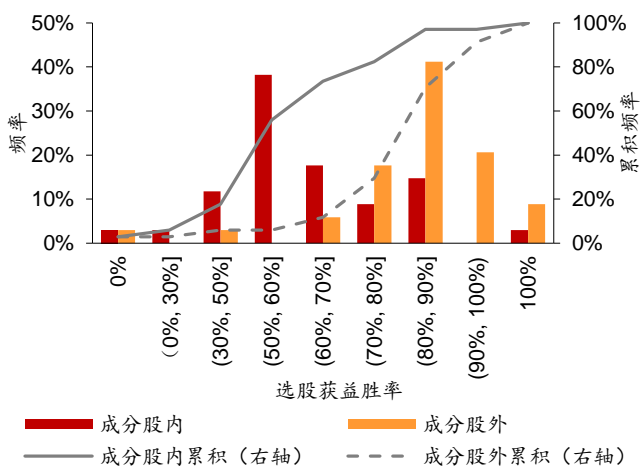
资料来源：Wind，华泰证券研究所

图表15： 指数成分股内、外选股收益期望值的关系图



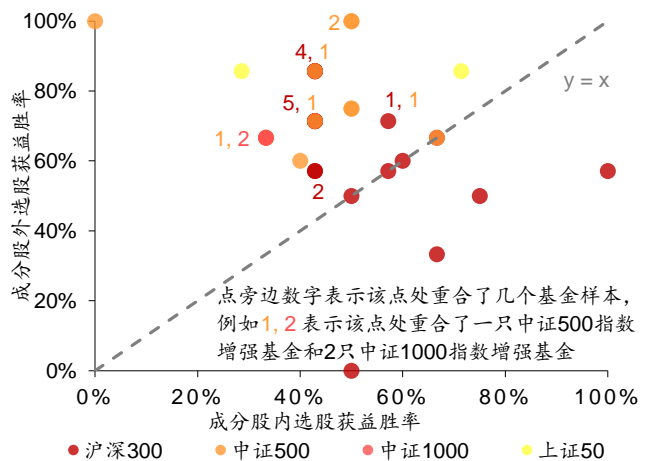
资料来源：Wind，华泰证券研究所

图表16： 指数成分股内、外选股获胜率的分布直方图



资料来源：Wind，华泰证券研究所

图表17： 指数成分股内、外选股获胜率的关系图



资料来源：Wind，华泰证券研究所

指数成分股内、外选股的获胜率

胜率定义同上一节，即成分股内、外选股所得超额收益率在截面期时间序列上为正值的频率（正值数目除以有效截面期数目）。基金在成分股内、外选股获胜率上的分布直方图如图表 16 所示。从图中可以看出，与成分股内选股相比，成分股外选股获胜率的分布整体更靠近 100%，说明成分外选股获得正值超额收益的频率更高。

同样，对于所有基金，我们以成分股内选股获益胜率为 x 轴，以成分股外选股获益胜率为 y 轴，绘制了二者的相对关系图，见图表 17。图中，灰色虚线表示 $y=x$ 的趋势线，数字表示重叠在一个点上的基金个数，数字的颜色表示基金跟踪的标的指数，不同指数间用逗号隔开，例如图中“5, 1”表示该数据点上重叠了 5 只沪深 300（红色）指数增强型基金和 1 只中证 500（橙色）指数增强型基金。由图可知，大部分基金的数据点都位于灰色虚线的上方，说明其成分股外选股获益胜率高于成分股内选股获益胜率。

综上，与成分股内选股相比，指数成分股外选股获得正值超额收益率的概率更大。

指数成分股内、外选股的获益稳定性

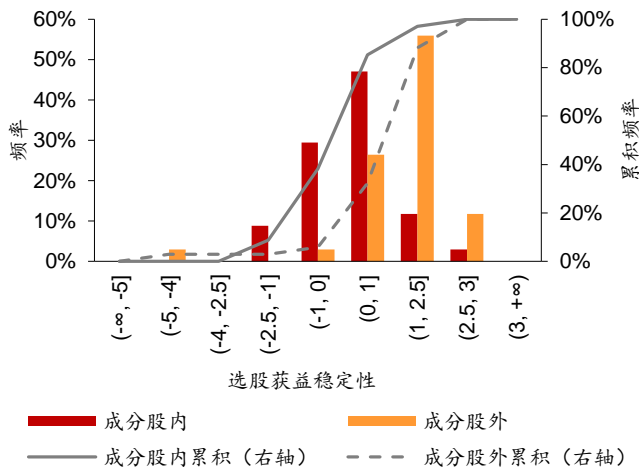
稳定性定义同上一节，即成分股内、外选股所得超额收益率在截面期时间序列上的 t 值。这里，原假设值仍为 0， t 值大于 2.5 可以认为获得了稳定的正值超额收益率。

图表 18 为基金在成分股内、外选股获益稳定性上的分布直方图。在成分股外选股获益稳定性上，约 94% 基金的 t 值为正，约 12% 基金的 t 值大于 2.5。而在成分股内选股获益稳定性上，仅有 62% 基金的 t 值为正，仅有约 3% 基金的 t 值大于 2.5。从整体分布上看，成分股外比成分股内更靠近正无穷一侧。

同样，对于所有基金，我们以成分股内选股获益稳定性为 x 轴，以成分股外选股获益稳定性为 y 轴，绘制了二者的关系图，如图表 19 所示。由图可知，大部分基金的数据点都位于灰色虚线的上方，说明其成分股外选股获益稳定性高于成分股内选股获益稳定性。

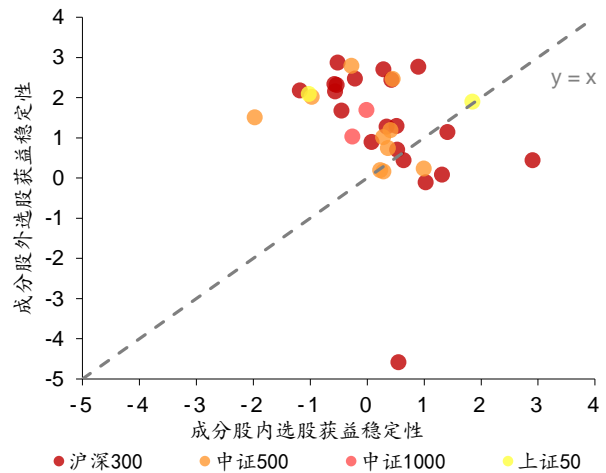
综上，与成分股内选股相比，指数成分股外选股获得正值超额收益率的稳定性更高。

图表 18：指数成分股内、外选股获益稳定性的分布直方图



资料来源：Wind，华泰证券研究所

图表 19：指数成分股内、外选股获益稳定性的关系图



资料来源：Wind，华泰证券研究所

本节小结

从配股权重上看，绝大多数指数增强型基金的指数成分股内配股权重都高于 80%，但不同基金在成分股内、外配股的倾向性不同：有些基金倾向于在指数成分股内配股，而有些基金倾向于利用规定比例以外的剩余配股额度，在指数成分股外选股。

从超额收益率上看，与指数成分内选股相比，指数成分股外选股获得的超额收益率在期望值、胜率、稳定性三个指标上均表现更好。说明对于指数增强型基金，指数成分股外选股相比于成分股内选股，更容易获得高期望、高胜率、稳定的超额收益率。

配置特色：相对指数的行业中性程度、行业配置稳定性与行业择时准确度

从本节开始，我们将从收益相关的分析角度出发，依次剖析指数增强型基金在行业、风格和持仓集中度上的配置特色，逐一探究单个配置指标与基金超额收益表现的相关关系。最后，汇总单个指标的分析结果，基于多个配置指标将基金聚类，再次验证基金类别的配置特色与类别超额收益表现的关系。

本节关注行业配置特色。在本节中，我们将依次考察以下三个配置指标——相对指数的行业中性程度、行业配置稳定性和行业择时准确度。

相对指数的行业中性程度

多数基于多因子模型的量化指数增强策略会选择相对基准行业中性度的设置，以控制跟踪误差。然而，仍有一些指数增强策略会选择相对基准非行业中性度的设置，以追求行业配置上可能获取的超额收益。

我们使用在指数增强系列首篇报告中定义的“行业偏离度”指标来衡量指数增强型基金相对基准的行业中性程度。令持仓权重之和为 1，“单行业偏离度”定义为某一行业在基金持仓中的权重占比与该行业在标的指数中的权重占比的差值的绝对值。“整体行业偏离度”定义为 29 个单行业偏离度求和，再除以 2。此外，为了表征基金在单行业上相对指数超配或低配的阈值上限，我们定义了“单行业最大偏离度”指标，即单行业偏离度的最大值。公式如下：

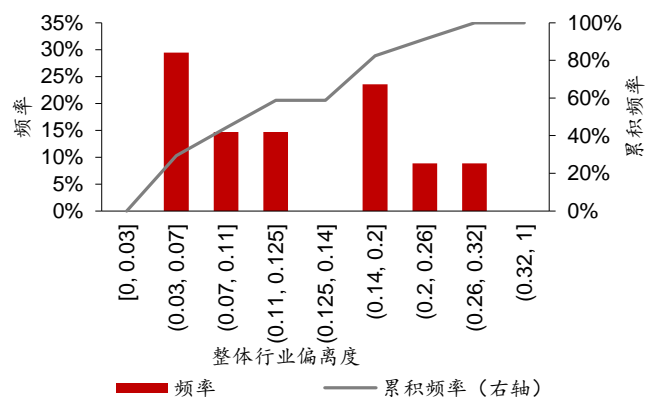
$$\text{单行业偏离度} = |\text{行业}i\text{在基金持仓中的权重占比} - \text{行业}i\text{在标的指数中的权重占比}|$$

$$\text{整体行业偏离度} = \frac{1}{2} \sum_{i=1}^{29} \text{行业}i\text{的单行业偏离度}$$

$$\text{单行业最大偏离度} = \max_{i=1}^{29} (\text{行业}i\text{的单行业偏离度})$$

整体行业偏离度的定义方式类似单边换手率，取值范围为[0, 1]。越接近 0 表示基金的行业配置权重与标的指数越相近，基金相对指数的行业中性程度越高。

图表20： 整体行业偏离度的分布直方图



资料来源：Wind，华泰证券研究所

我们基于 7 个截面期的基金持仓明细、标的指数成分股明细和全部 A 股所属一级行业分类，计算基金和标的指数在一级行业上的配股权重，进而得到每个截面期上的整体行业偏离度，并对截面期序列取均值。基金的整体行业偏离度的分布直方图如图表 20 所示。

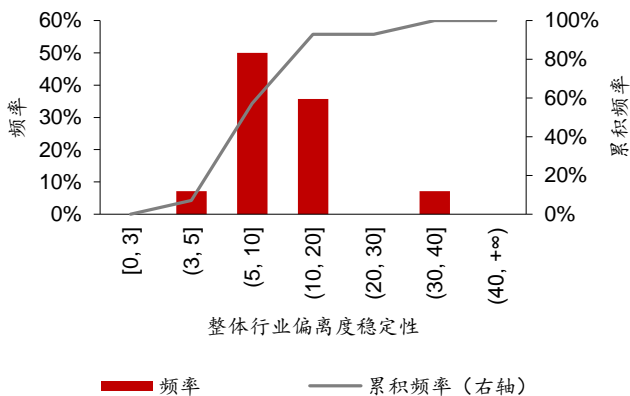
由图可见，基金按整体行业偏离度均值分布可以大致分为两类，其中 20 只（约 60%）基金的整体行业偏离度均值不高于 0.1250，即对于 29 个一级行业，基金在每个行业的平均超/低配权重小于 0.0087，剩余的 14 只（约 40%）基金的整体行业偏离度均值大于 0.1400，即基金在每个行业的平均超/低配权重大于 0.0096。我们以此空区间为界，将整体行业偏离度均值不高于 0.1250 的 20 只基金归为相对指数行业中性，将整体行业偏离度均值大于 0.1400 的 14 只基金归为相对指数非行业中性。

非行业中性基金的整体行业偏离度稳定性

对于 14 只非行业中性基金，我们计算整体行业偏离度在截面期时间序列上的 t 值以衡量其稳定性。这里，原假设值仍设为 0，t 值大于 2.5 可以认为基金的行业权重稳定地区别于指数，即基金相对指数的非行业中性配置在时间序列上稳定。

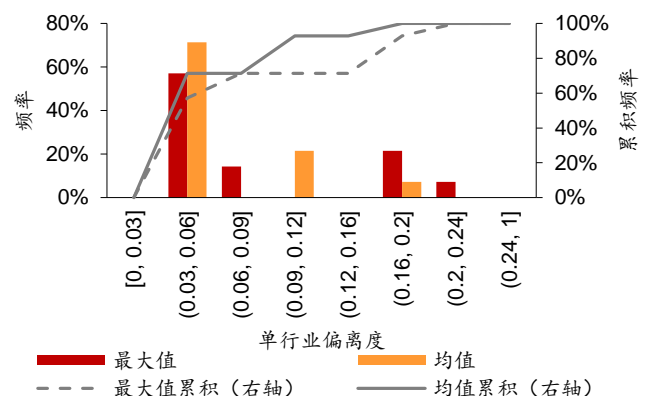
图表 21 为非行业中性基金在整体行业偏离度稳定性上的分布直方图。由图可知，所有非行业中性基金的 t 值都大于 3.0，因此，非行业中性基金在行业权重配置上稳定地区别于指数。

图表21： 非行业中性基金的整体行业偏离度稳定性的分布直方图



资料来源：Wind，华泰证券研究所

图表22： 非行业中性基金的单行业最大偏离度的分布直方图



资料来源：Wind，华泰证券研究所

非行业中性基金的单行业超/低配上限

进一步，对于非行业中性基金，我们考察其单行业超/低配是否存在阈值上限。将基金的单行业最大偏离度，对截面期时间序列取最大值和均值，绘制基金的单行业最大偏离度的分布直方图，如图表 22 所示。约 70% 基金的单行业最大偏离度均值不高于 0.06，最大值不高于 0.09，即基金相对标的指数的单行业超/低配权重上限不超过 0.09。在所有非行业中性基金中，单行业偏离度程度最大的基金的单行业超/低配权重上限为 0.23。

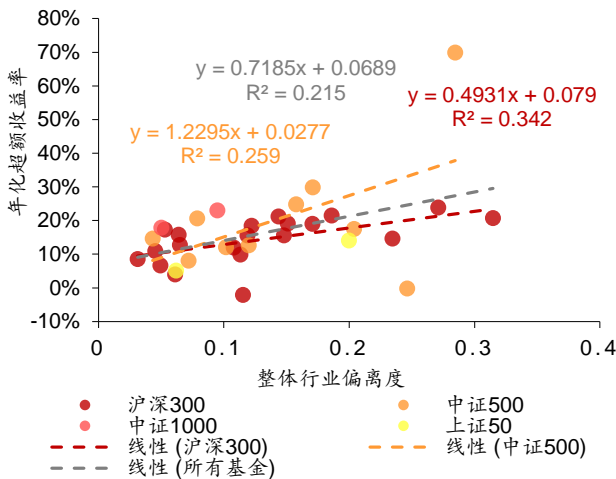
相对指数的行业中性程度与超额收益率

选择相对指数行业中性的基金经理普遍认为通过行业配置获取超额收益率的难度较大、风险较高、性价比较低。那么实际表现是否如此呢？相对指数行业中性的基金与非行业中性的基金，是否真的在超额收益率水平上相差无几呢？

图表 23 为基金年化超额收益率与整体行业偏离度的关系，虚线为分别在沪深 300 基金池、中证 500 基金池和所有基金池内的线性拟合结果。由于中证 1000 基金池和上证 50 基金池内基金数目过少，各只有两个，故不再单独分析。从图中可以看出，较大的整体行业偏离度，大概率对应较大的年化超额收益率。

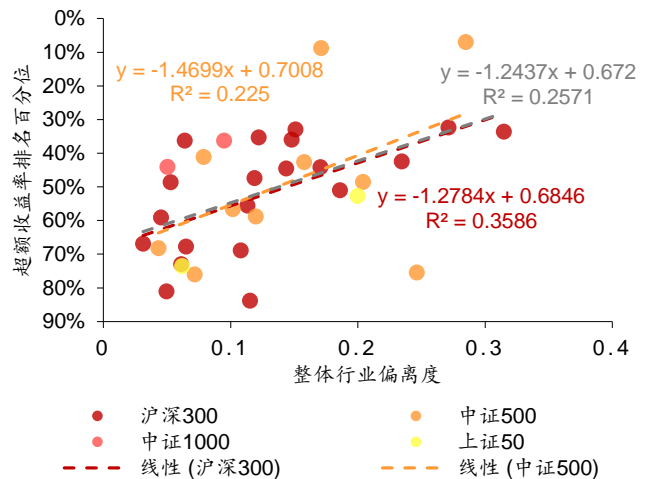
进一步，为了避免年化超额收益率在单个截面期上的极端值对均值的影响，我们考察基金的超额收益率排名百分位，将其作为基金业绩表现优劣的补充指标。对于每个截面期，我们在所有基金池内，计算每个基金的超额收益率由高到低的排名百分位。超额收益率越高，排名百分位值越小。之后，对截面期序列取均值，得到每个基金的超额收益率排名百分位，取值范围为[0%，100%]。基金超额收益率排名百分位与整体行业偏离度的关系如图表 24 所示。图表 24 和 23 的趋势一致，即较大的整体行业偏离度，大概率对应较靠前的超额收益率排名。

图表23： 年化超额收益率与整体行业偏离度的关系图



资料来源：Wind，华泰证券研究所

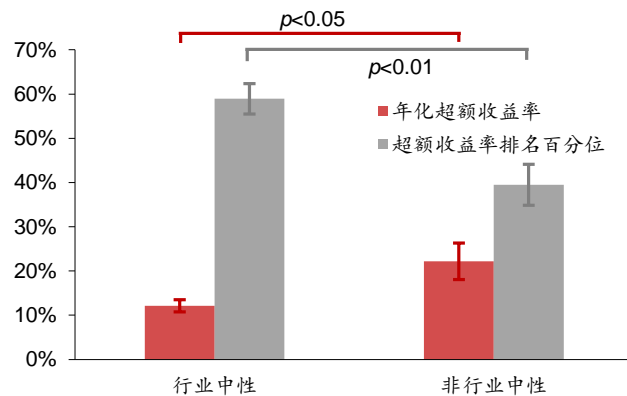
图表24： 超额收益率排名百分位与整体行业偏离度的关系图



资料来源：Wind，华泰证券研究所

在前文中，我们按照整体行业偏离度，将基金分为两类——相对指数行业中性 and 相对指数非行业中性。接下来，我们探究这两类基金在年化超额收益率和超额收益率排名百分位上是否存在显著差异。结果如图表 25 所示。其中，柱形表示基金类别的均值，误差线表示基金类别的标准误。

图表25： 相对指数行业中性基金与行业非中性基金的超额收益率对比



资料来源：Wind，华泰证券研究所

两类基金的年化超额收益率和超额收益率排名百分位都通过了类别间等方差 Levene 检验 ($F=1.644$, $p=0.209$; $F=0.034$, $p=0.855$)。由图可知，行业中性基金的年化超额收益率显著小于非行业中性基金 ($t_{32}=-2.657$, $p<0.05$)，并且行业中性基金的超额收益率排名百分位显著落后于非行业中性基金 ($t_{32}=3.228$, $p<0.01$)。因此，与行业中性基金相比，非行业中性基金获取超额收益率的能力更强。

综上，与相对指数行业中性基金的配置策略相比，相对指数非行业中性基金的配置策略更有可能获得高期望值的超额收益率。

行业配置稳定性

在上一小节中，我们站在单个截面期的角度上，探讨基金相对指数的行业中性程度。在本小节中，我们将从截面期时间序列的角度出发，探讨基金行业配置的稳定性。

我们选用两个指标来表征基金行业配置的稳定性——行业权重前后变动量和行业权重前后相关系数。公式如下：令第 t 个截面期上的基金行业权重向量为

$$\mathbf{w}_t = [w_{t,1}, w_{t,2}, \dots, w_{t,29}]^T, \quad \sum_{i=1}^{29} w_{t,i} = 1$$

则，第 t 期的行业权重前后变动量为

$$\Delta_t = \frac{1}{2} \sum_{i=1}^{29} |w_{t,i} - w_{t-1,i}|, \quad t > 2$$

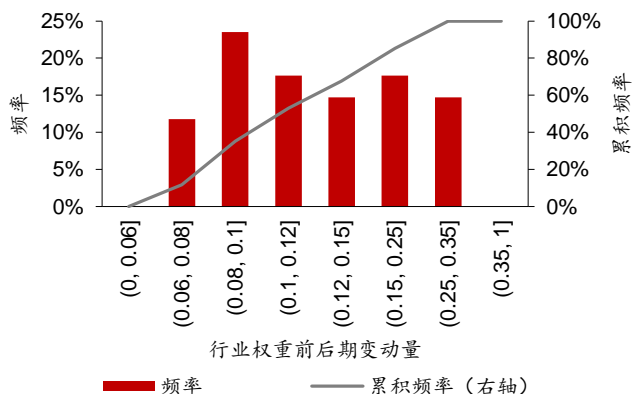
第 t 期的行业权重前后相关系数为

$$\text{Corr}_t = \frac{\text{Cov}(\mathbf{w}_t, \mathbf{w}_{t-1})}{\sqrt{\text{Var}(\mathbf{w}_t)} \cdot \sqrt{\text{Var}(\mathbf{w}_{t-1})}}, \quad t > 2$$

行业权重前后变动量同样类似单边换手率，取值范围为 $[0, 1]$ ，值越靠近 0，代表行业配置稳定性越高。行业权重前后相关系数的取值范围为 $[-1, 1]$ ，负值表示前、后期行业权重呈负相关，正值表示前、后期行业权重呈正相关，零值表示二者不相关。值越靠近 1，代表行业配置稳定性越高。

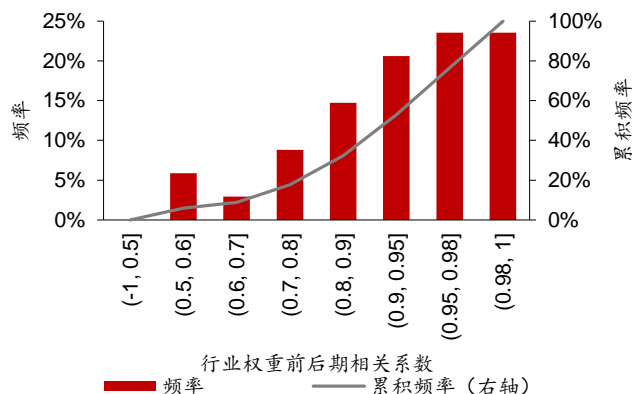
我们计算每个基金的行业权重前后变动量和前后相关系数，并对截面期时间序列取均值。基金在这两个指标上的分布直方图分别如图 26、27 所示。在行业权重前后变动量上，所有指数增强型基金都不大于 0.35，其中，85% 的基金不大于 0.25，53% 的基金不大于 0.12。在行业权重前后相关系数上，所有指数增强型基金都大于 0.50，其中，82% 的基金大于 0.80，47% 的基金大于 0.90。因此，整体来看，大多数指数增强型基金的行业配置在时间序列上偏好稳定。

图表26：行业权重前后变动量的分布直方图



资料来源：Wind，华泰证券研究所

图表27：行业权重前后相关系数的分布直方图



资料来源：Wind，华泰证券研究所

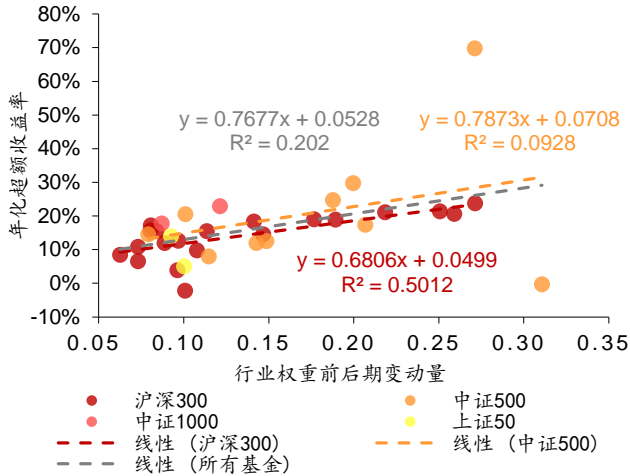
行业配置稳定性与超额收益率

面对风谲云诡的行业状况，是淡看风云变化，我自岿然不动，还是选择临机应变，主动审时调整？虽然大多数指数增强型基金的行业配置较为稳定，偏好以不变应万变，但仍有部分基金选择相对频繁地调整行业配置。那么，基金在行业配置稳定性上的不同是否会导致其在超额收益率上的差异呢？

我们延用在上一小节中定义的两个指标——年化超额收益率和超额收益率排名百分位——来表征基金业绩表现的优劣。图 28-29、30-31 分别为行业权重前后变动量和行业权重前后相关系数与这两个超额收益率指标的关系。虽然线性拟合的拟合效果在不同基

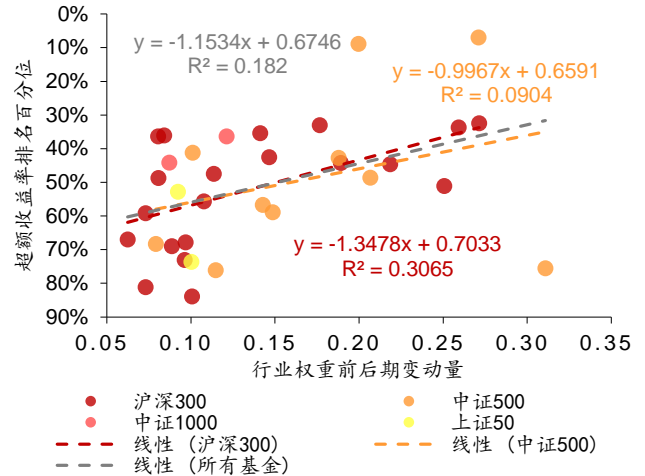
金池中不太一致，即在中证 500 基金池内较弱，而在沪深 300 基金池和所有基金池内较强，但是整体来看，较大的行业权重前后期变动量、较小的行业权重前后期相关系数，即较低的行业配置稳定性，大概率对应较高的超额收益率。

图表28： 年化超额收益率与行业权重前后期变动量的关系图



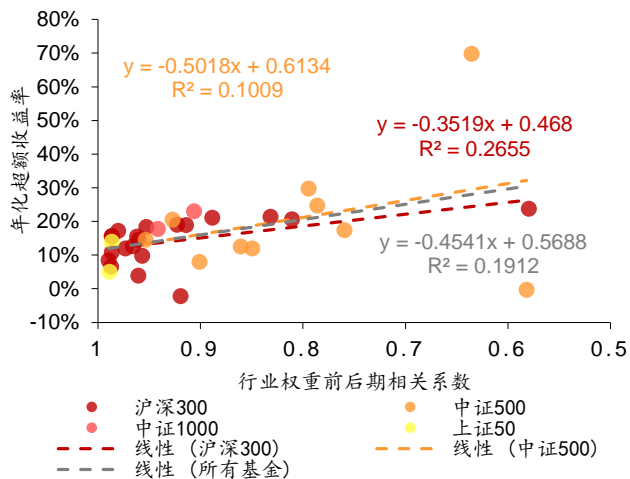
资料来源：Wind，华泰证券研究所

图表29： 超额收益率排名百分位与行业权重前后期变动量的关系图



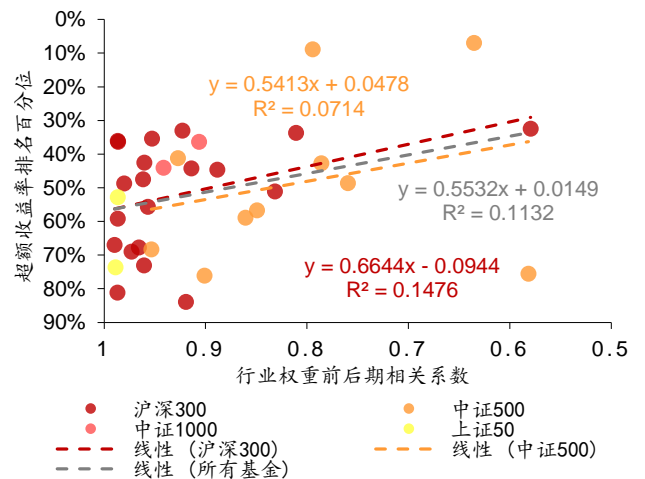
资料来源：Wind，华泰证券研究所

图表30： 年化超额收益率与行业权重前后期相关系数的关系图



资料来源：Wind，华泰证券研究所

图表31： 超额收益率排名百分位与行业权重前后期相关系数的关系图



资料来源：Wind，华泰证券研究所

行业择时准确度

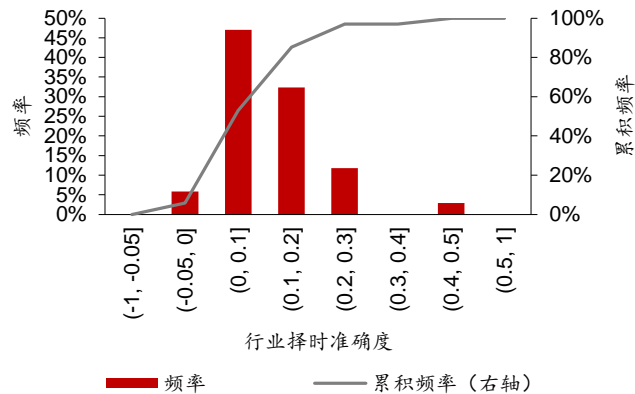
行业择时准确度定义为基金相对指数的行业权重差值与指数行业超额收益率的 Spearman 秩相关系数。令第 t 个截面期上，基金在行业 i 上的配股权重为 $w_{t,i}^P \in [0, 1]$ ，且 $\sum_{i=1}^{29} w_{t,i}^P = 1$ 。标的指数在行业 i 上的配股权重为 $w_{t,i}^B \in [0, 1]$ ，且 $\sum_{i=1}^{29} w_{t,i}^B = 1$ 。标的指数的收益率为 r_t^B 。若满仓投资于行业 i ，即将行业 i 内的持仓个股权重等比例扩充至权重和为 1，基金在行业 i 上可获得的收益率为 $r_{t,i}^P$ ，标的指数在行业 i 上可获得的收益率为 $r_{t,i}^B$ 。注意，如果 $w_{t,i}^P = 0$ ，则 $r_{t,i}^P = \text{NaN}$ ，NaN 表示该数据无效，在后续处理中被舍弃。同理，如果 $w_{t,i}^B = 0$ ，则 $r_{t,i}^B = \text{NaN}$ 。那么，第 t 个截面期的行业择时准确度 R_{spe_t} 表示为：

$$R_{\text{spe}_t} = \text{spe}(w_{t,i}^P - w_{t,i}^B, r_{t,i}^B - r_t^B)$$

行业择时准确度的取值范围为[-1, 1]。正值代表基金相对指数超配了收益高于指数平均收益的行业，低配了收益低于指数平均收益的行业；负值表示基金相对指数低配了收益高于指数平均收益的行业，高配了收益低于指数平均收益的行业；零值表示基金相对指数的行业权重差值与指数的行业超额收益率无关。行业择时准确度的值越接近 1，代表基金相对指数的行业权重差值与指数行业超额收益率的正相关程度越高，基金的行业择时能力越强。

图表 32 为基金在行业择时准确度上的分布直方图。由图可知，约 94% 基金的行业择时准确度均为正值，约 90% 的基金位于 (0.0, 0.3] 区间内，说明大部分指数增强型基金都具有一定的行业择时能力。特别地，一只基金一枝独秀，独领风骚，行业择时准确度大于 0.4。

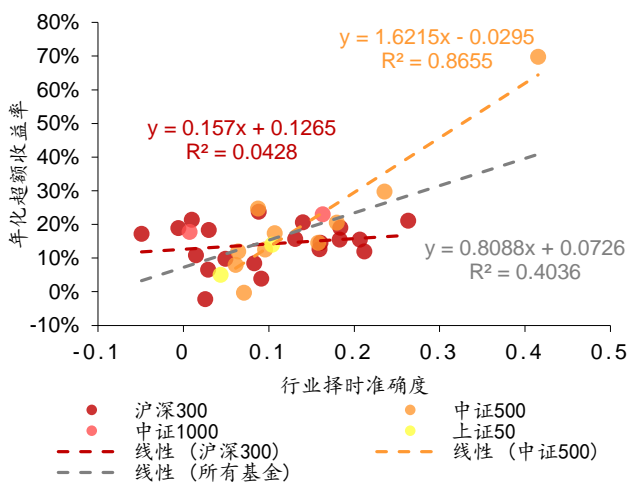
图表 32： 行业择时准确度的分布直方图



资料来源：Wind，华泰证券研究所

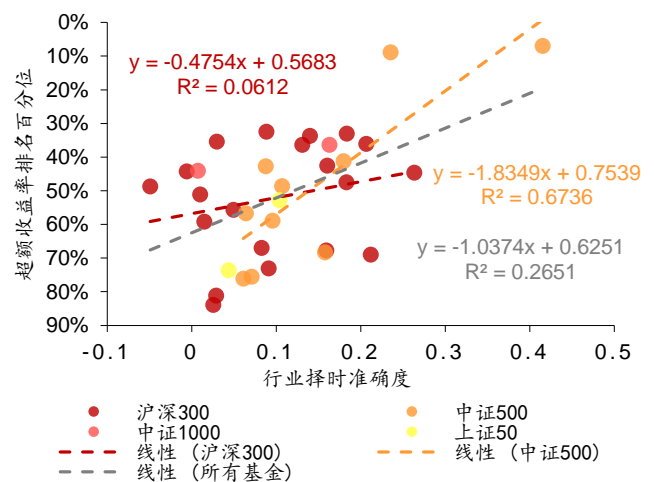
由于行业择时准确度的定义与超额收益率直接挂钩，因此，在预期上，行业择时准确度越高，基金相对标的指数的超额收益率也应该越大。那么，实际表现是否与预期一致呢？我们继续使用年化超额收益率和超额收益率排名百分位两个指标来表征基金的超额收益率水平，结果如图表 33-34 所示。从图中可以看出，基金超额收益率与其行业择时准确度的正相关关系在沪深 300 基金池内程度较弱，而在中证 500 基金池和所有基金池中程度较强。特别地，在行业择时准确度上表现出众（高于 0.4）的那只基金在超额收益率上同样独占鳌头：年化超额收益率高达 70%，超额收益率排名第一。综上可得，良好的行业择时准确度，可以为超额收益率的获取添砖加瓦、保驾护航。

图表 33： 年化超额收益率与行业择时准确度的关系图



资料来源：Wind，华泰证券研究所

图表 34： 超额收益率排名百分位与行业择时准确度的关系图



资料来源：Wind，华泰证券研究所

本节小结

我们按照基金相对指数的整体行业偏离度，将基金分为两类——相对指数非行业中性 and 相对指数行业中性。我们发现，相对指数非行业中性基金在超额收益表现上显著优于行业中性基金。此外，非行业中性基金的非行业中性配置特色在时间序列上较为稳定，在单行业上存在一定的超/低配阈值。整体来看，相对指数的行业中性程度较低的基金，其超额收益率大概率较高、超额收益率排名较靠前。

行业配置稳定性较低的基金，其超额收益率大概率较高、超额收益率排名较靠前。

行业择时准确度较高的基金，其超额收益率可能较高，超额收益率排名较靠前。但是此相关关系在沪深 300 基金池内程度较弱。

配置特色：持仓风格的行业间一致性与持仓风格稳定性

在指数增强系列首篇报告《指数增强方法汇总及实例——量化多因子模型构建指数增强组合实证》中，我们介绍了以多因子模型为主的量化指数增强方法。在本节中，我们将基于华泰金工多因子模型研究成果，从持仓风格角度出发，依次探讨指数增强型基金在持仓风格的行业间一致性和持仓风格稳定性上的配置特色，并对比不同配置特色的基金在超额收益率表现上的差异。

多因子模型是目前量化选股方法中理论研究最为丰富的主流模型之一，在投资市场中应用广泛。经典的多因子模型表达式为：

$$\tilde{r}_j = \sum_{k=1}^K X_{jk} * \tilde{f}_k + \tilde{u}_j$$

\tilde{r}_j : 股票 j 的收益率

X_{jk} : 股票 j 在因子 k 上的因子暴露（因子载荷）

\tilde{f}_k : 因子 k 的因子收益

\tilde{u}_j : 股票 j 的残差收益率

多因子模型本质是将对 N 只股票的收益-风险预测转变成对于 K 个因子的收益-风险预测，将估算个股收益率的协方差阵转化为估算因子收益率的协方差阵，从而有效降低数据分析的工作量，提高准确度。基于多因子模型的指数增强方法从统计的角度分析因子收益率的变化趋势，并从投资组合的层面对因子暴露进行管理，从而超越标的指数获取超额收益率。

我们基于华泰金工多因子系列报告以及因子跟踪周报系列的研究成果，选取了以下归属于 8 个风格的 18 个因子进行后续分析，详见图表 35。其中，最后一列因子方向的值为 1 代表因子暴露越大越好，值为 -1 代表因子暴露越小越好。

图表35： 本文使用的因子列表

因子类别	因子简称	因子计算方式	因子方向
估值	净利润（TTM）/总市值	Wind 因子 pe_ttm 的倒数	1
	净资产（TTM）/总市值	Wind 因子 pb_lf 的倒数	1
	营业收入（TTM）/总市值	Wind 因子 ps_ttm 的倒数	1
	股息率（TTM）	Wind 因子 dividendyield2	1
	经营现金流（TTM）/总市值	Wind 因子 pcf_ocf_ttm 的倒数	1
成长	净利润同比增长率	由 Wind 因子 net_profit_is 计算：（本季度值-去年同季度值）/去年同季度值的绝对值	1
	营业收入同比增长率	由 Wind 因子 oper_rev 计算：（本季度值-去年同季度值）/去年同季度值的绝对值	1
质量	净资产收益率（ROE）	Wind 因子 roe_avg	1
	资产收益率（ROA）	Wind 因子 roa	1
	毛利率	Wind 因子 grossprofitmargin	1
	净利率	Wind 因子 netprofitmargin	1
	权益乘数	wind 因子 tot_assets 除以 wind 因子 tot_equity	-1
规模	对数总市值	Wind 因子 mkt_cap_ard 的对数	-1
动量	近一个月收益率	个股最近 1 个月收益率	-1
	近三个月收益率	个股最近 3 个月收益率	-1
换手率	近一个月换手率	个股最近 1 个月内日均换手率	-1
波动率	近三个月波动率	个股最近 3 个月内日收益率序列的标准差	-1
beta	近 100 周相对上证综指的 beta	Wind 因子 beta_100w	-1

资料来源：Wind，华泰证券研究所

持仓风格的行业间一致性

行业的成长性、稳定性、周期性等情况不同，适用的选股因子也不尽相同。基金经理可能选择使用相同的因子在各行业间统一选股，也可能针对不同的行业制定不同的选股因子或选股策略。在本小节中，我们考察指数增强型基金持仓风格的行业间一致性，并探讨其与基金超额收益率的关系。

首先，在每个截面期上，对于全部 A 股，我们计算每只股票的每个因子经过中位数去极值、行业中性化、标准化之后的因子暴露值。计算方法依次如下：

- 1) 中位数去极值：计算全部 A 股在因子 k 上的因子暴露原始值 X_k^T ($N \times 1$ 维向量) 的中位数 $D_{M,k}$ 和 $|X_k^T - D_{M,k}|$ 的中位数 $D_{M1,k}$ ，接下来将向量 X^T 中所有大于 $D_{M,k} + 5D_{M1,k}$ 的元素重设为 $D_{M,k} + 5D_{M1,k}$ ，将向量 X^T 中所有小于 $D_{M,k} - 5D_{M1,k}$ 的元素重设为 $D_{M,k} - 5D_{M1,k}$ ，取处理后的因子暴露替代原始值，仍标记为 X_k^T ；
- 2) 行业中性化：用 OLS 线性回归拟合

$$X_k^T = \sum_{i=1}^I \text{Indus}_i^T \cdot \beta_i + \varepsilon$$

Indus_i^T ：当前截面期上行业 i 的哑变量（属于该行业为 1，否则为 0）， $N \times 1$ 维向量

β_i ， ε ：回归待拟合系数和残差， β_i 为数值， ε 为 $N \times 1$ 维向量

取上述回归方程的 ε 作为因子暴露的新值，仍记为 X_k^T ；

- 3) 标准化：将因子暴露 X_k^T 减去其均值，再除以其标准差，得到标准化之后的因子暴露新值，仍记为 X_k^T 。

之后，我们基于详细持仓数据计算基金在每个行业的因子偏好和风格偏好。在每个截面期上，我们将一级行业 i 内个股的因子值（标准化后）乘以个股持仓权重占行业 i 持仓总权重的比例，求和，得到基金在一级行业 i 上的因子暴露，并将因子暴露乘以因子方向，最终得到基金的行业-因子暴露矩阵 X^P ($K \times I \times T$ 维矩阵， K 为因子个数， I 为行业个数， T 为截面个数)。同理基于标的指数的成分股明细和成分股所属一级行业分类，计算得到标的指数的行业-因子暴露矩阵 X^B ($K \times I \times T$ 维矩阵)。二者相减，得到基金相对标的指数的行业-因子暴露差值矩阵 $X^{PB} = X^P - X^B$ 。如果基金或标的指数零配某个行业，则将该行业对应的行业-因子暴露数据设为无效值 (NaN)，不参与后续计算。

由于因子在不同基金、行业间差异较大，因此，我们将所有指数增强型基金的 X^P 和 X^{PB} 在因子维度上依次进行去极值、标准化、单位化的处理。对于每一个因子，去极值和标准化方法同上，单位化是将因子暴露减去其最小值，再除以其全距。其中，全距指最大值与最小值的差值。处理后， X^P 和 X^{PB} 矩阵中每一个元素（因子暴露）的取值范围均为 $[0, 1]$ 。

我们定义因子偏好和风格偏好，来表征基金的持仓风格特色。对于第 t 个截面期，行业 i 内因子 k 的因子偏好定义为因子 k 的因子暴露除以 18 个因子的因子暴露之和。考虑到因子数目较多，可能不便总结规律，于是我们将 18 个因子的因子偏好按照图表 33 第一列的风格大类等权重合成为 8 个风格的风格偏好。风格 j 的暴露值定义为风格 j 下所有因子的因子暴露均值，风格偏好定义为风格 j 的暴露值除以 8 个风格的暴露值之和。对于每个行业，因子偏好和风格偏好都是行业内和为 1 的向量，向量内每个元素（单个因子或风格的偏好值）的取值范围均为 $[0, 1]$ 。在每个截面期上，我们计算每个基金在一级行业上的自身因子偏好和风格偏好，以及相对指数的因子偏好和风格偏好。

为了衡量基金的因子偏好和风格偏好在行业间的一致性，我们定义了因子偏好行业间差异度和风格偏好行业间差异度，公式如下：

因子偏好行业间差异度

$$= \frac{1}{2 \times I^*} \sum_{k=1}^{18} \sum_{i=1}^{29} |\text{行业 } i \text{ 内因子 } k \text{ 的偏好值} \\ - \text{所有行业内因子 } k \text{ 的偏好均值}|$$

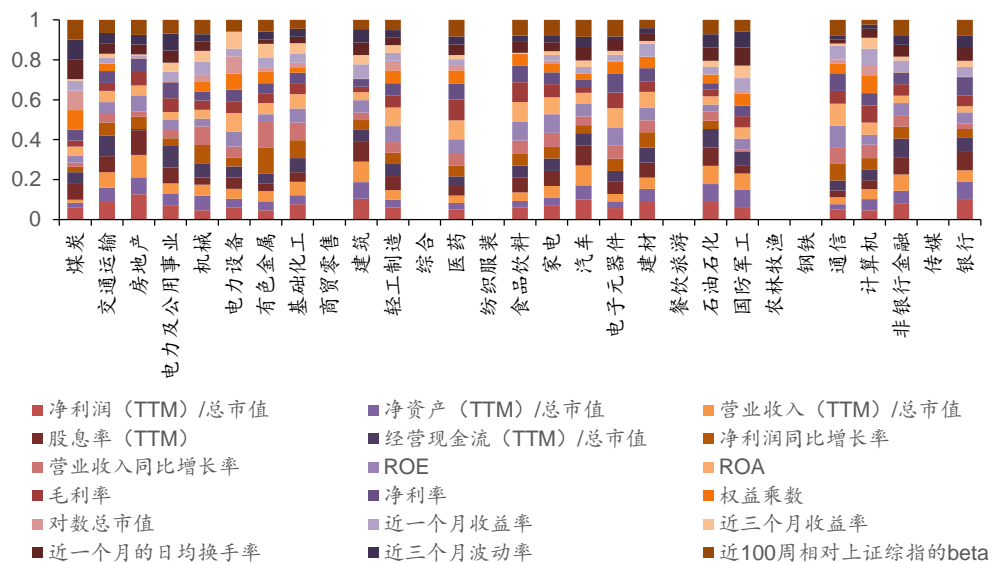
风格偏好行业间差异度

$$= \frac{1}{2 \times I^*} \sum_{j=1}^8 \sum_{i=1}^{29} |\text{行业 } i \text{ 内风格 } j \text{ 的偏好值} \\ - \text{所有行业内风格 } j \text{ 的偏好均值}|$$

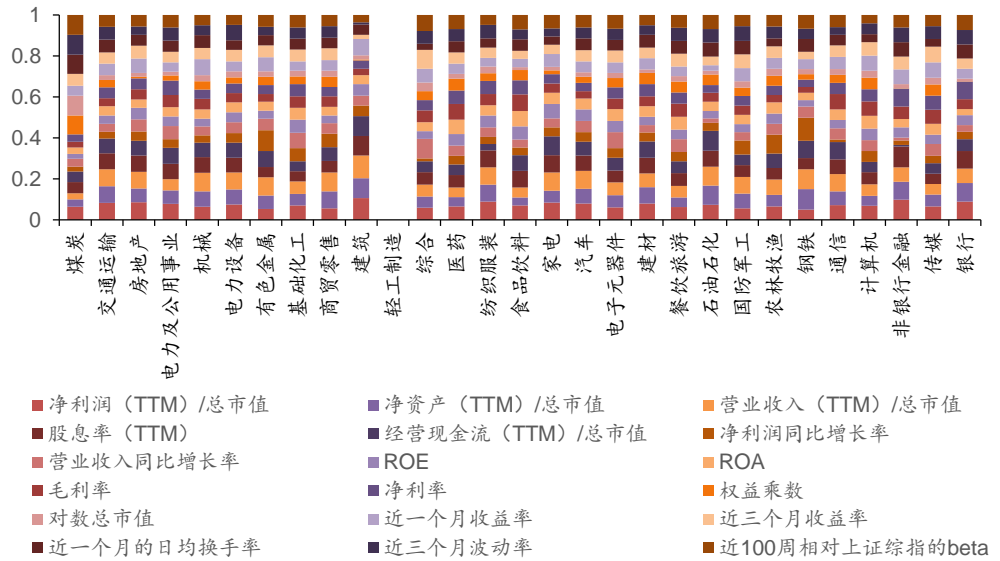
其中， I^* 为行业内偏好值非无效值的有效行业数目。因子偏好和风格偏好的行业间差异度定义类似单边换手率，取值范围为[0, 1]。值越靠近 0 代表行业间差异度越小，即因子偏好和风格偏好在行业间的一致性越高。

由于本指标的计算过程涉及多次处理转换，为了使读者对于因子、风格偏好的行业间差异度取值有一个直观的感受，我们以基金自身因子偏好行业间差异度为例，挑选了对应最大值（0.1624）的基金 A 的 2017 年报数据和对应最小值（0.0743）的基金 B 的 2017 半年报数据，并将二者的行业-自身因子偏好分布图绘制在图表 36-37 中。图中因子偏好为空的行业代表基金在当前截面期对该行业零配。其余持仓风格行业间一致性指标的举例见附录一。

图表36： 自身因子偏好行业间差异度最大值 0.1624：基金 A-2017 年报的自身因子偏好行业分布图

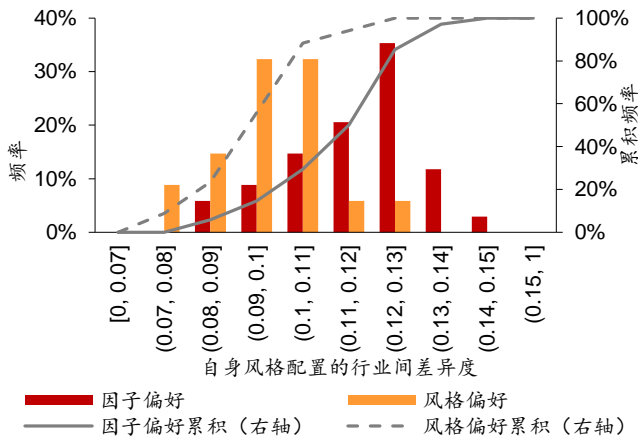


资料来源：Wind，华泰证券研究所

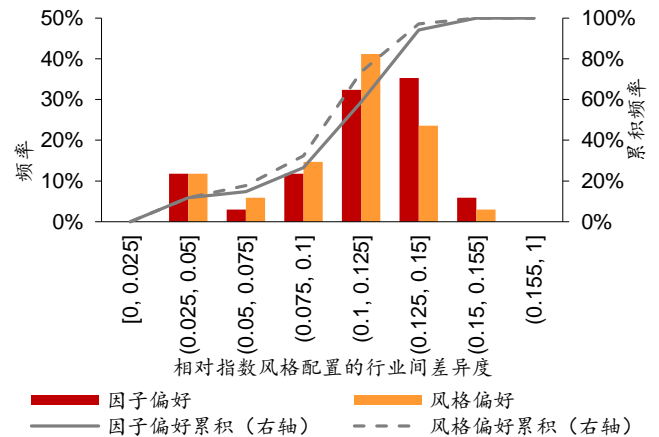
图表37： 自身因子偏好行业间差异度最小值 0.0743：基金 B-2017 半年报的自身因子偏好行业分布图

资料来源：Wind，华泰证券研究所

对于每个基金，我们将因子偏好和风格偏好的行业间差异度对截面期时间序列取均值。基金在自身因子偏好、风格偏好的行业间差异度，以及相对指数的因子偏好、风格偏好的行业间差异度上的分布直方图分别如图表 38-39 所示。在自身因子、风格偏好行业差异度上，基金分布较为集中且对称；在相对指数的因子、风格偏好行业差异度上，基金分布较为分散，约 12% 基金的因子、风格偏好行业差异度不高于 0.05，行业间一致性较高。

图表38： 自身持仓风格的行业间差异度的分布直方图

资料来源：Wind，华泰证券研究所

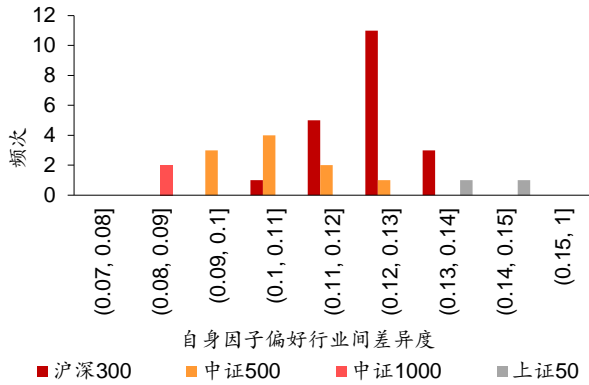
图表39： 相对指数持仓风格的行业间差异度的分布直方图

资料来源：Wind，华泰证券研究所

持仓风格的行业间一致性与超额收益率

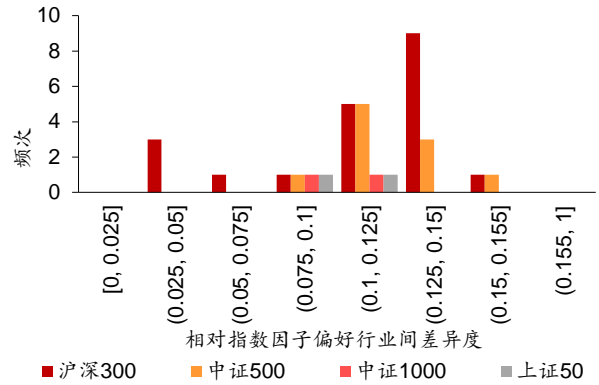
在本小节中，我们探讨持仓风格行业间一致性是否会影响基金的超额收益率。由于不同指数增强基金在自身因子偏好行业间差异度上的分布重叠度较低（见图表 40），不适合于对所有基金做整体的相关分析。而不同指数增强基金在相对基准的因子偏好的行业间差异度上的分布重叠度较高（见图表 41），并且投资人亦更加关注基金相对基准的因子暴露情况（而非关注基金本身的因子暴露情况），因此我们统一使用基金相对基准的因子、风格偏好行业间差异度来分析其与基金超额收益率的关系。

图表40：不同指数增强基金自身因子偏好行业间差异度的分布直方图



资料来源：Wind，华泰证券研究所

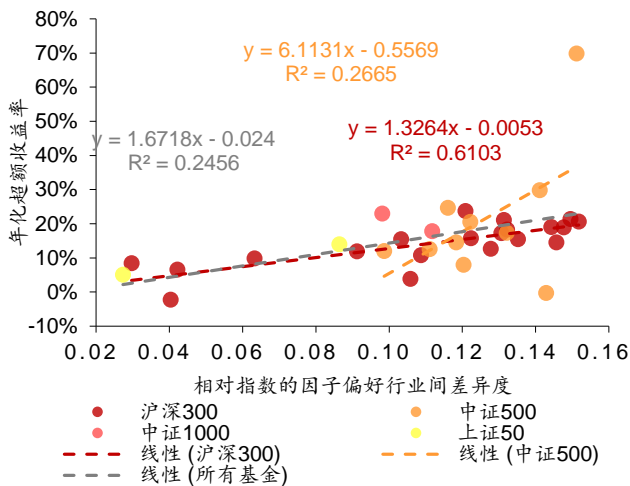
图表41：不同指数增强基金相对基准的因子偏好行业间差异度分布图



资料来源：Wind，华泰证券研究所

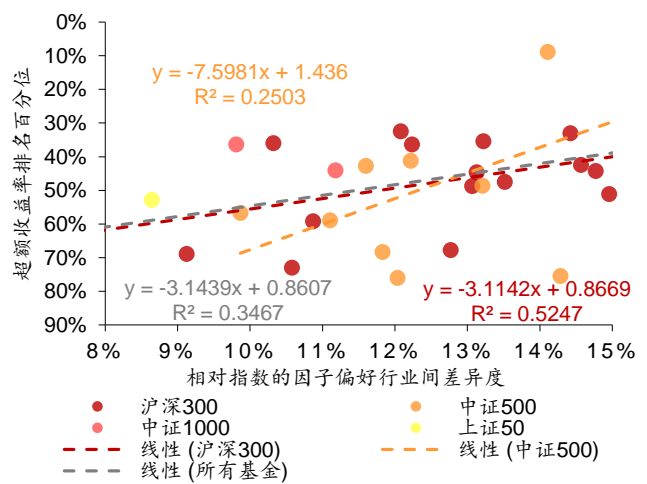
基金超额收益率与相对指数因子、风格偏好行业间差异度的关系图分别如图表 42-43、44-45 所示。由图可知，与风格偏好相比，基金相对基准的因子偏好行业间差异度与基金超额收益率表现的相关程度更高。整体来看，因子、风格偏好在行业间一致性程度越低，超额收益率表现大概率越好。因此，在不同行业间采取不同的选股因子，可能在概率意义上会获得更大超额收益。

图表42：年化超额收益率与因子偏好行业间差异度的关系图



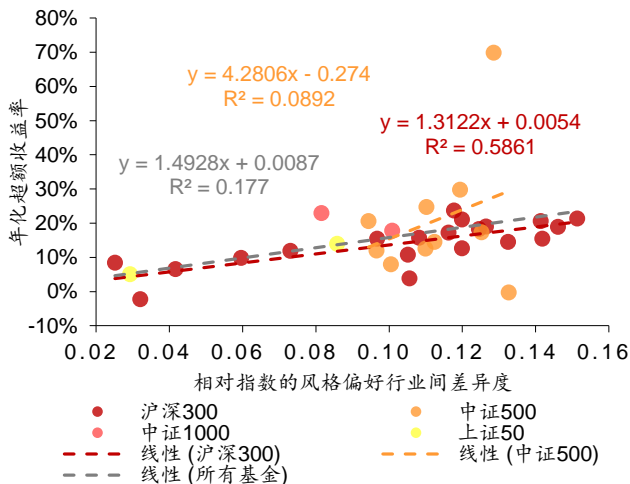
资料来源：Wind，华泰证券研究所

图表43：超额收益率排名百分位与因子偏好行业间差异度的关系图



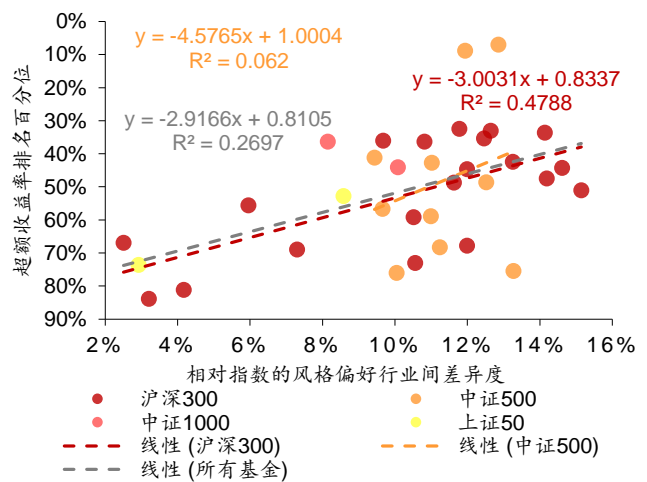
资料来源：Wind，华泰证券研究所

图表44：年化超额收益率与风格偏好行业间差异度的关系图



资料来源：Wind，华泰证券研究所

图表45：超额收益率排名百分位与风格偏好行业间差异度的关系图



资料来源：Wind，华泰证券研究所

持仓风格稳定性

在上一小节中，我们从单个截面期角度出发，探讨了基金持仓风格的行业间一致性。在本小节中，我们将从截面期时间序列角度出发，考察基金持仓风格在时间序列上的稳定性，并分析其与基金超额收益率的关系。

首先，在每个截面期上，对于全部 A 股，我们计算每只股票的每个因子经过中位数去极值、行业中性化、标准化之后的个股因子暴露值。之后，基于基金持仓明细，我们将基金持仓个股的因子暴露按个股持仓权重加权得到持仓个股因子暴露和，并将其乘以因子方向。最终得到基金的因子暴露矩阵 X^P ($K \times T$ 维矩阵)。接下来，我们将所有指数增强型基金的 X^P 在因子维度上依次进行去极值、标准化、单位化的处理，以消除因子间量纲不一致和基金间因子值差异较大的问题。处理后， X^P 矩阵中每一个元素（因子暴露）的取值范围均为 $[0, 1]$ 。

对于持仓风格指标，第 t 个截面期上因子 k 的因子偏好定义为因子 k 的因子暴露除以 18 个因子的因子暴露之和。风格 j 的风格偏好定义为风格 j 的暴露值除以 8 个风格的暴露值之和，其中风格 j 的暴露值定义为风格 j 内所有因子偏好的均值。在每个截面期上，因子偏好与风格偏好都是和为 1 的向量，向量内每个元素（单个因子或风格的偏好值）的取值范围均为 $[0, 1]$ 。

延续上一节中衡量基金行业配置稳定性的思路，我们用基金持仓风格前后期变动量和前后期相关系数两个指标来表征持仓风格在时间序列上的稳定性。这里，基金自身持仓风格指标不再存在不同指数增强基金分布重叠度过低的问题，因此在本节中，我们直接使用基金自身，而非相对指数的，因子偏好和风格偏好，进行后续分析。计算公式如下：令第 t 个截面期上的基金因子（或风格）偏好向量为

$$P_t = [P_{t,1}, P_{t,2}, \dots, P_{t,K}]^T, \quad \sum_{k=1}^K P_{t,k} = 1$$

则，第 t 期的因子（或风格）偏好前后期变动量为

$$\Delta_t^P = \frac{1}{2} \sum_{k=1}^K |P_{t,k} - P_{t-1,k}|, \quad t > 2$$

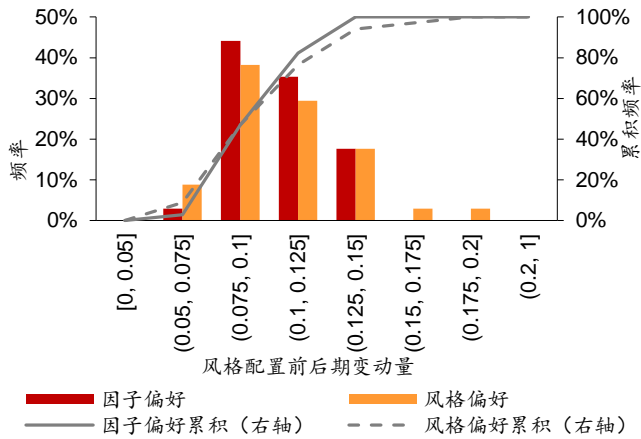
第 t 期的因子（或风格）偏好前后期相关系数为

$$Corr_t^P = \frac{Cov(P_t, P_{t-1})}{\sqrt{Var(P_t)} \cdot \sqrt{Var(P_{t-1})}}, \quad t > 2$$

因子（或风格）偏好前后期变动量类似单边换手率，取值范围为 $[0, 1]$ ，值越靠近 0，代表持仓风格稳定性越高。因子（或风格）偏好前后期相关系数的取值范围为 $[-1, 1]$ ，负值表示前、后期持仓风格呈负相关，正值表示前、后期持仓风格呈正相关，零值表示二者不相关。值越靠近 1，代表持仓风格稳定性越高。

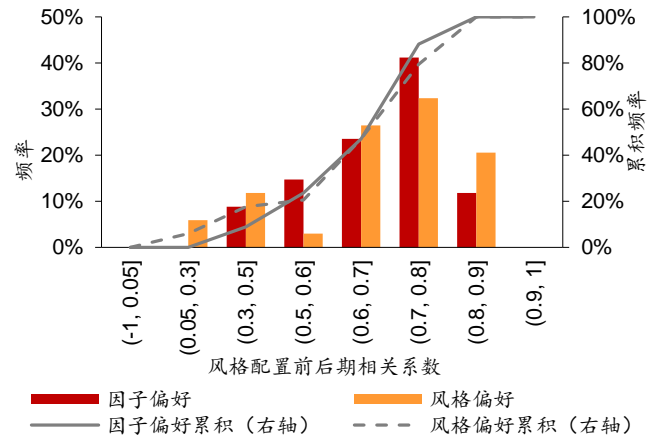
基金在持仓风格前后期变动量、前后期相关系数的分布直方图见图表 46-47。约 47% 的基金因子、风格偏好前后期变动量均不高于 0.1；约 53% 的基金因子、风格偏好前后期相关系数均大于 0.7，所有基金持仓风格前后期相关系数均不高于 0.9。

图表46: 持仓风格前后期变动量的分布直方图



资料来源: Wind, 华泰证券研究所

图表47: 持仓风格前后期相关系数的分布直方图

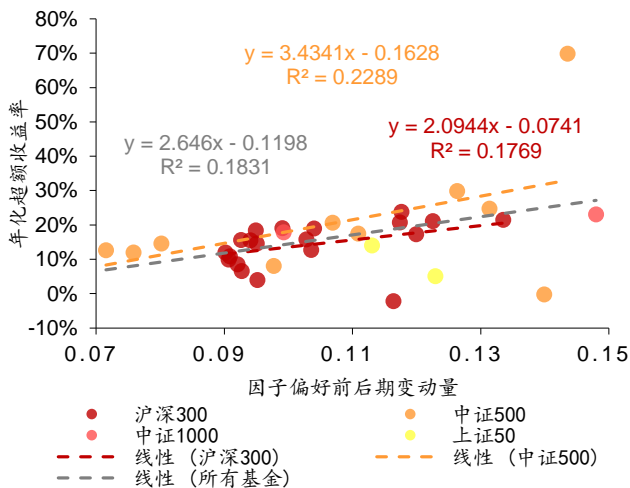


资料来源: Wind, 华泰证券研究所

持仓风格稳定性与超额收益率

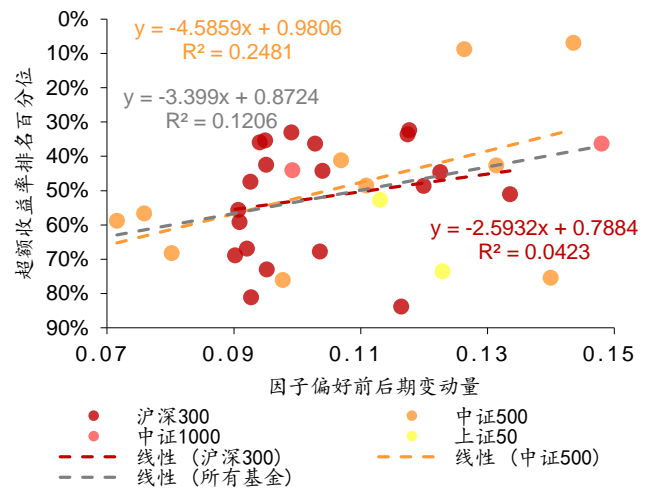
基金超额收益率与因子偏好前后期变动量、前后期相关关系的关系图分别如图表 48-49、50-51 所示。从图中可以看出，一定程度上，因子偏好前后期变动量越大，基金超额收益率表现越好。但是，超额收益率与因子偏好前后期相关关系的相关性较差。基金超额收益率与风格偏好稳定性的相关程度较低，这里就不做赘述，详见附录二。综上，基金的持仓风格稳定性指标中，只有因子偏好前后期变动量对基金超额收益率表现具有一定的解释效力。当用因子偏好前后期变动量表征基金持仓风格的稳定性时，持仓风格稳定性较低的基金，其超额收益率大概率较好。

图表48: 年化超额收益率与因子偏好前后期变动量的关系图



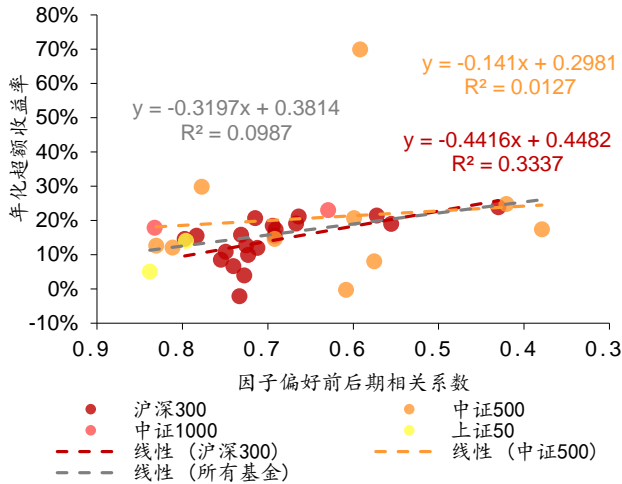
资料来源: Wind, 华泰证券研究所

图表49: 超额收益率排名百分位与因子偏好前后期变动量的关系图



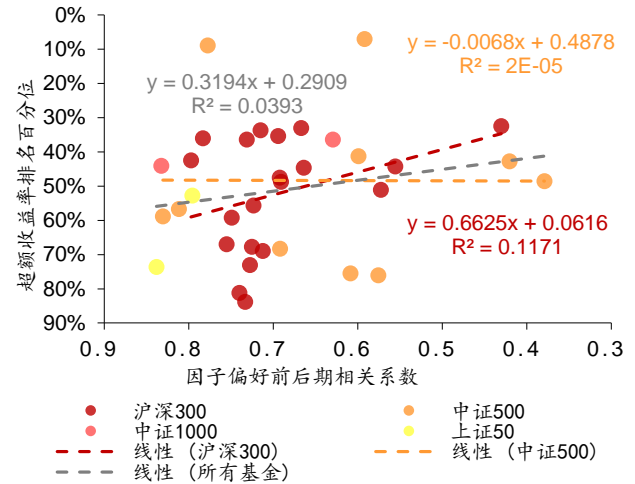
资料来源: Wind, 华泰证券研究所

图表50： 年化超额收益率与因子偏好后期相关系数的关系图



资料来源：Wind，华泰证券研究所

图表51： 超额收益率排名百分位与因子偏好后期相关系数的关系图



资料来源：Wind，华泰证券研究所

本节小结

持仓风格的行业间一致性较低的基金，其超额收益率大概率较高、超额收益率排名较靠前。

持仓风格稳定性上，因子偏好后期变动量较大的基金，其超额收益率大概率较高、超额收益率排名较靠前。其余稳定性指标，如因子偏好后期相关系数，则与超额收益表现相关性较弱。

配置特色：持仓集中度

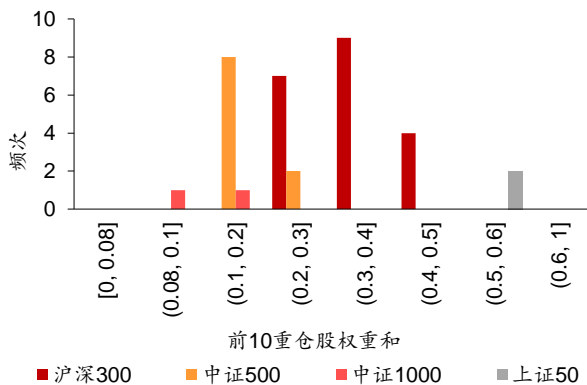
投资圈里流传着这样一句警示名言——“不能把鸡蛋放在一个篮子里”。目的在于建议投资者们应当分散投资，而不要孤注一掷，从而分散风险，尽可能避免一次决策失败可能带来的巨额损失。但是，投资者们也都深谙“机遇与风险并存”的道理，面对看好的盈利机会，大比例投注虽然可能承担较大的风险，但也可能带来傲人的收益。

在指数增强型基金层面上，投资比例带来的收益与风险的取舍就反映在基金的持仓集中度指标上。持仓集中度指基金投资于前 N 大重仓股的金额占其股票投资总额的比例，反映了基金倾向于“将鸡蛋放在一个篮子里”的意愿程度。高持仓集中度的基金在选股良好的情况下，可能获得更为丰厚的收益，但也需要承担选股失误或其他意外因素带来的更大的损失风险。低持仓集中度的基金配股分散程度更高，风险暴露程度更低，但基金的收益也可能相对折损。那么集中投资策略和分散投资策略用于指数增强究竟孰优孰劣呢？

在本节中，我们考察指数增强型基金在前 10 和前 20 重仓股的持仓集中度上的整体表现，并探讨其与超额收益率和跟踪误差的关系。

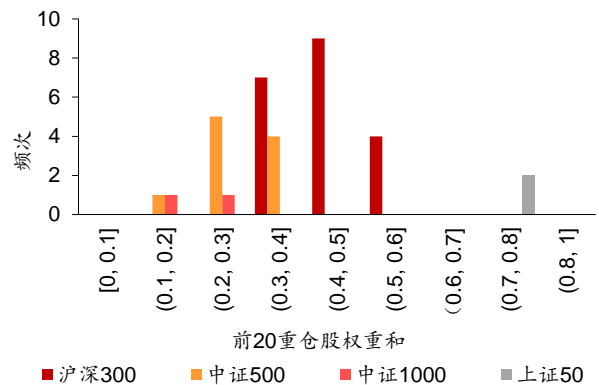
基金在前 10 和前 20 重仓股权重和上分布直方图分别如图表 52、53 所示。从图中可以看出，跟踪同一指数的基金的持仓集中度分布较为集中，而不同指数的基金的持仓集中度分布较为分散，且其大小关系与其跟踪的标的指数的成分股数目有关。整体看来，标的指数成分股数目越多，其基金的持仓集中度均值越小。

图表 52：不同指数增强基金前 10 重仓股权重和的分布直方图



资料来源：Wind，华泰证券研究所

图表 53：不同指数增强基金前 20 重仓股权重和的分布直方图



资料来源：Wind，华泰证券研究所

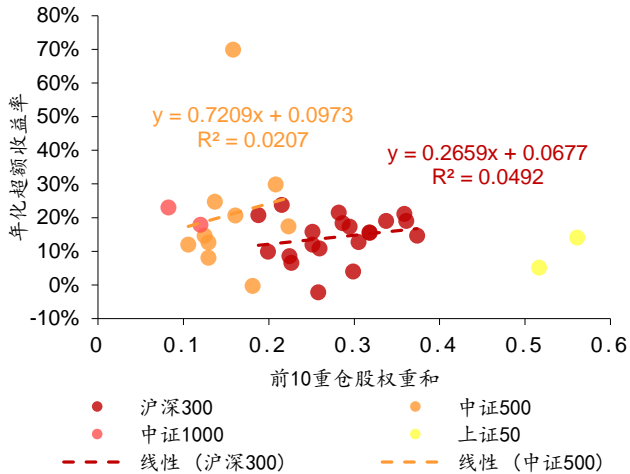
持仓集中度与超额收益率

由于不同指数的基金在持仓集中度上分布较为分散，不适合于对所有基金做整体的相关分析，因此在本小节和下一小节中，我们只针对沪深 300 和中证 500 这两个基金池进行线性拟合。基金的年化超额收益率与前 10、前 20 重仓股权重和的关系分别如图表 54、55 所示。从图中可以看出，基金超额收益率与基金持仓集中度的相关关系较弱。因此，对于指数增强型基金而言，集中投资策略与分散投资策略相比，可能并不具备获益上的明显优势。

持仓集中度与跟踪误差

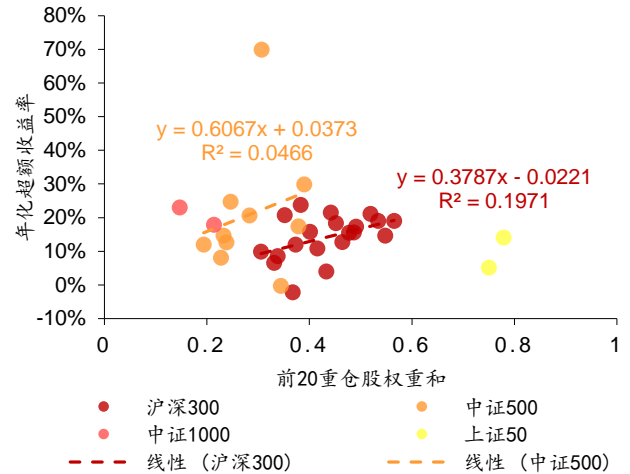
既然集中投资策略并不能占领超额收益上的高地，那它是否仍处于风险控制上的低谷呢？我们考察基金的年化跟踪误差与其前 10、前 20 重仓股的持仓集中度的关系，见图表 56、57。由图可知，越高的持仓集中度，可能对应越大的年化跟踪误差。综上，与分散投资策略相比，集中投资策略不但不能具备收益上的明显优势，还大概率需要承担更大的风险。

图表54： 年化超额收益率与前 10 重仓股权重和的关系图



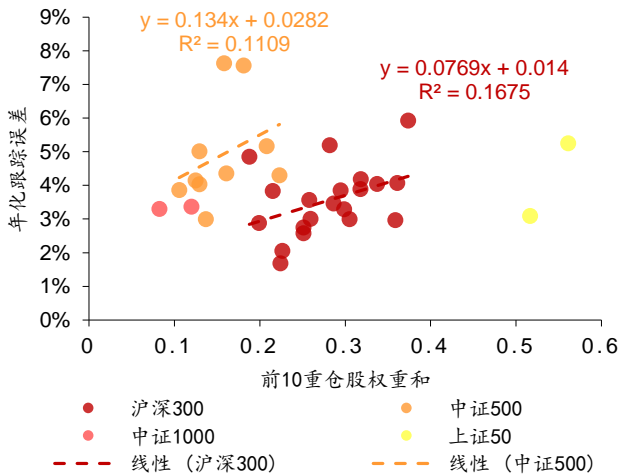
资料来源：Wind，华泰证券研究所

图表55： 年化超额收益率与前 20 重仓股权重和的关系图



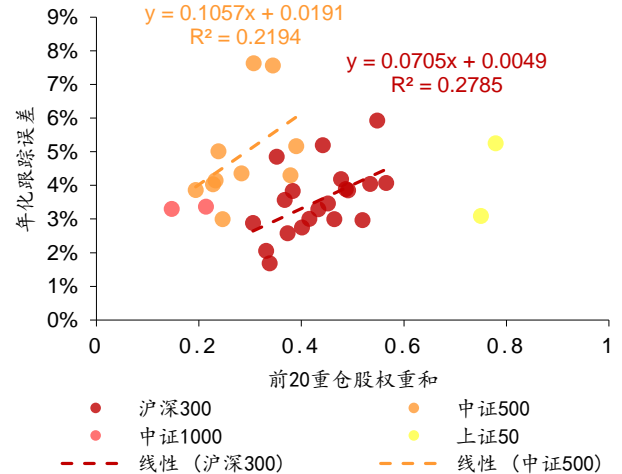
资料来源：Wind，华泰证券研究所

图表56： 年化跟踪误差与前 10 重仓股权重和的关系图



资料来源：Wind，华泰证券研究所

图表57： 年化跟踪误差与前 20 重仓股权重和的关系图



资料来源：Wind，华泰证券研究所

本节小结

持仓集中度与标的指数有关：标的指数成分股数目越多，其基金的持仓集中度均值越小。

持仓集中度较高的基金，其超额收益表现不一定较好，但是其跟踪误差大概率较大。

基于配置特色的基金聚类

在前面三节中，我们依次探讨了行业、风格、持股集中度配置特色中的单个配置指标与基金超额收益率的相关关系。在本节中，我们首先汇总单指标的分析结果，根据单指标的相关性和对超额收益的拟合优度筛选有效指标，并构建配置指标集合。之后基于 k-means 方法将基金聚类，考察基金类别配置特色与超额收益的关系。

指标集构建

在前文行业、风格、持股集中度三节中分析的所有配置指标中，行业择时准确度定义与收益率直接挂钩，可操作性较差，难以控制；持仓集中度在不同指数增强基金池间的取值较为分散，不适合对所有基金进行整体相关分析。因此，我们忽略这两个指标，将剩余配置指标作为初始指标集。

初始指标集如图表 58 所示。其中，拟合优度定义为在所有基金池内，单个配置指标对年化超额收益率和超额收益排名百分位的线性拟合优度的均值。指标方向表示配置指标对超额收益的预测方向：值为-1 表示指标越小，超额收益表现越好；值为 1 表示指标越大，超额收益表现越好；值为 0 表示指标的拟合优度较低，方向不明确。我们舍去拟合优度过低的指标（拟合优度小于 0.1）。

图表58： 初始指标集内单个配置指标的拟合优度及方向

分析维度		配置指标	拟合优度	指标方向
行业	相对指数的行业中性程度	整体行业偏离度	0.2361	1
	行业配置稳定性	行业权重前后期变动量	0.1920	1
		行业权重前后期相关系数	0.1522	-1
风格	持仓风格的行业间一致性	因子偏好行业间差异度	0.2900	1
		风格偏好行业间差异度	0.2144	1
	持仓风格稳定性	因子偏好前后期变动量	0.1524	1
		因子偏好前后期相关系数	0.0709	0
		风格偏好前后期变动量	0.0291	0
		风格偏好前后期相关系数	0.0086	0

资料来源：Wind，华泰证券研究所

图表59： 初始指标集的指标间相关关系矩阵

	整体行业偏离度	行业权重前后期变动量	行业权重前后期相关系数	因子偏好行业间差异度	风格偏好行业间差异度	因子偏好前后期变动量	因子偏好前后期相关系数	风格偏好前后期变动量	风格偏好前后期相关系数
整体行业偏离度	1.00	0.85	-0.73	0.52	0.52	0.46	-0.39	0.32	-0.22
行业权重前后期变动量	0.85	1.00	-0.89	0.59	0.59	0.57	-0.54	0.42	-0.37
行业权重前后期相关系数	-0.73	-0.89	1.00	-0.42	-0.41	-0.53	0.60	-0.41	0.49
因子偏好行业间差异度	0.52	0.59	-0.42	1.00	0.97	0.22	-0.38	0.20	-0.28
风格偏好行业间差异度	0.52	0.59	-0.41	0.97	1.00	0.16	-0.37	0.10	-0.25
因子偏好前后期变动量	0.46	0.57	-0.53	0.22	0.16	1.00	-0.45	0.72	-0.30
因子偏好前后期相关系数	-0.39	-0.54	0.60	-0.38	-0.37	-0.45	1.00	-0.38	0.84
风格偏好前后期变动量	0.32	0.42	-0.41	0.20	0.10	0.72	-0.38	1.00	-0.37
风格偏好前后期相关系数	-0.22	-0.37	0.49	-0.28	-0.25	-0.30	0.84	-0.37	1.00

资料来源：Wind，华泰证券研究所

图表 59 为初始指标集的指标间相关关系矩阵。指标间相关性较高会影响聚类结果，因此我们在相关性过高（相关系数的绝对值大于 0.7）的指标中舍去拟合优度较低的指标。

经过拟合优度和指标间相关性筛选后，我们得到了有效指标集，共包含以下 3 个有效指标：整体行业偏离度、因子偏好行业间差异度和因子偏好前后期变动量。为了消除指标间量纲不一致的情况，我们将指标值归一化——指标值减去其最小值，再除以其最大值与最小值的差值。归一化后，所有指标的取值范围均为 [0, 1]。

k-means 聚类

k-means 聚类方法是典型、常用的无监督学习方法。利用 k-means 方法，我们可以在不需要事先定义类别标签的情况下，按照基金在配置指标空间上分布的亲疏远近，将基金聚类为若干类别，使得同一类别基金在配置特色上更为相似，而不同类别基金在配置特色上相差较大，从而在多个配置指标的维度上，探寻配置指标与基金超额表现的关系。

k-means 聚类方法的一般过程如下：

- (1) 指定 k 个类别中心的初始值。
- (2) 对于每个数据点，计算其到 k 个类别中心的距离，并将其归类到距离值最小的类别中。得到所有数据点的聚类结果后，更新聚类误差，即所有数据点到对应类别中心的距离和。
- (3) 对于每个类别，将其类别中心值更新为该类别内所有数据点的均值。
- (4) 迭代执行第 2-3 步，直到满足以下任意一个终止条件：(a) 聚类误差不再减小；(b) 达到最大迭代次数。

一般地，k-means 分类方法需要指定以下参数：类别数目 k，类别中心的初始化方法，距离定义方法，最大迭代次数。其中，类别数目 k 的选择可以通过碎石图辅助判断。碎石图的 x 轴为类别数目 k，y 轴为 k 类别下最终输出的聚类误差。随着 k 值增大，聚类误差逐渐减小。如果碎石图中存在拐点，即聚类误差在该点之后的下降速度明显慢于在该点之前的下降速度，则拐点对应 k 值为类别数目的建议取值。但是，很多情况下，碎石图中没有明显拐点，即随着 k 值增大，聚类误差匀速减小。此时，就需要根据实际需求取值。常用的类别中心的初始值选择方法有数据样本随机抽样、指标区间均匀抽样、10% 样本数据预分类以及 k-means++ 等优化算法。类别中心初始值的选择不当很可能导致聚类结果陷入局部最优解，因此一般需要尝试多种初始化方法并多次聚类，选择聚类误差最小的结果作为聚类结果。距离可以定义为欧式距离的平方、绝对值距离、余弦距离、相关系数等，其中最常用的是欧式距离的平方。最大迭代次数视具体需求选择即可。

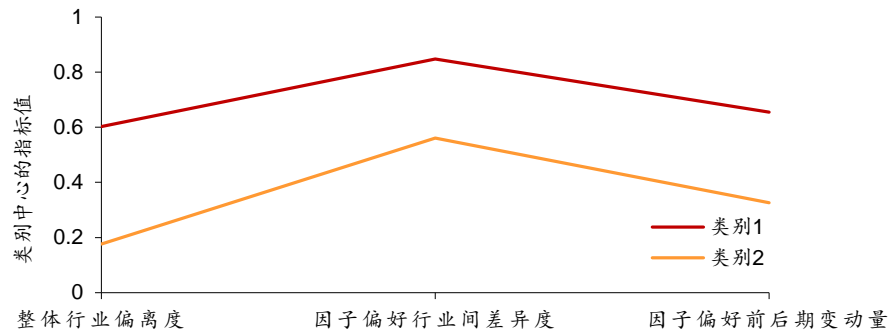
我们利用 k-means 基于有效指标集进行基金聚类。由于碎石图中无明显拐点，因此结合实际意义，我们将 k 值设为 2。距离设为欧式距离的平方，最大迭代次数设为 200。类别中心初始化时遍历了以上提到的所有初始化方法，并逐一执行 20 次，最终聚类误差最小值对应的方法为 k-means++ 算法。

基金基于整体行业偏离度、因子偏好行业间差异度和因子偏好前后期变动量这 3 个有效指标被分为 2 类。其中类别 1 包含 14 只基金，类别 2 包含 20 只基金。图表 60 为类别中心在各个有效指标上的值。由图可知，类别 1 在三个配置指标上都大于类别 2，说明与类别 2 相比，类别 1 基金相对指数的行业中性程度更低，持仓风格的行业间一致性更低，持仓风格的稳定性更低。

图表 61、62 分别为基金类别在年化超额收益率和超额收益率排名百分位上的表现。图中红色和橙色的散点分别表示类别 1、2 内的基金数据，黑色实线表示每个类别内基金的均值，黑色误差线表示每个类别内基金的标准误。整体来看，与类别 2 相比，类别 1 内基金的年化超额收益率上大概率更大，超额收益率排名百分位更靠前。

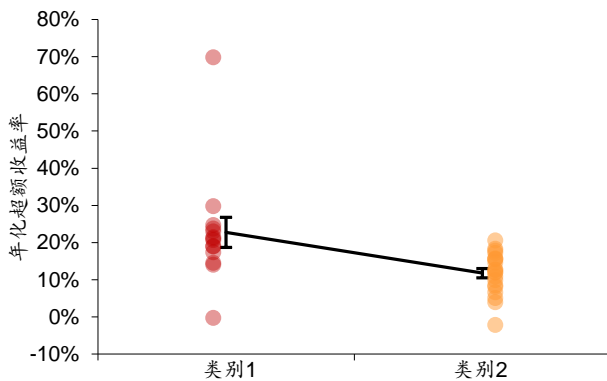
综上，相对指数行业中性程度较低、持仓风格行业间一致性较低、持仓风格稳定性较低的基金，大概率能获得较好的超额收益率表现。我们在多维配置空间上发现的这个趋势，与前文中的单个配置指标上的趋势一致。

图表60：基金类别中心的指标值



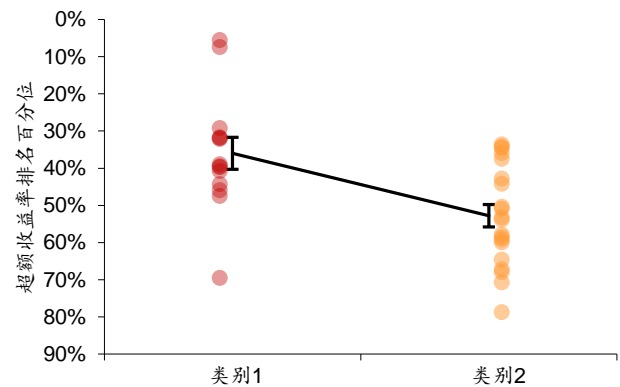
资料来源：Wind，华泰证券研究所

图表61：基金类别的年化超额收益率



资料来源：Wind，华泰证券研究所

图表62：基金类别的超额收益率排名百分位



资料来源：Wind，华泰证券研究所

本节小结

基于前文的单个配置指标分析结果，经拟合优度和指标间相关性筛选后，我们选择了三个指标构成有效指标集，这三个指标分别表征相对指数的行业中性程度、持仓风格的行业间一致性、持仓风格稳定性。

我们利用 k-means 方法，将基金在多维配置指标空间上分为 2 类，并考察基金类别的配置特色与基金类别超额收益表现的关系。我们发现，相对指数行业中性程度较低、持仓风格行业间一致性较低、持仓风格稳定性较低的基金，大概率能获得较好的超额收益率表现。此趋势与前文中在单个配置指标上发现的趋势一致。

全文总结

本文基于持仓数据将现有指数增强型基金进行收益分析，首先从收益分解角度出发，对比不同分解维度的相对获益难易；之后从收益相关角度出发，按照从单维到多维的顺序，探究配置特色与超额收益率的相关关系。主要结论列举如下：

1. 收益分解：

- 1) 与行业选择相比，行业内选股更可能获取高期望值的超额收益率。但是二者在获胜率和获益稳定性上没有明显区别；在 29 个一级行业内，相对容易获取高期望、高胜率、稳定的超额收益率的行业为轻工制造和机械。
- 2) 与在指数成分股内选股相比，在指数成分股外选股更容易获得高期望、高胜率、稳定的超额收益率。

2. 配置特色与收益的相关：

1) 单个指标：

- (a) 行业：相对指数的行业中性程度较低、行业配置稳定性较低、行业择时准确度较高的基金，其超额收益大概率较好。
- (b) 风格：持仓风格的行业间一致性较低、持仓风格稳定性较低的基金，其超额收益大概率较好。
- (c) 持仓集中度：持仓集中度较高的基金，其超额收益不一定较好，但是其跟踪误差大概率较大。

- 2) 多个指标的基金聚类：相对指数行业中性程度较低、持仓风格行业间一致性较低、持仓风格稳定性较低的基金，其超额收益大概率较好。

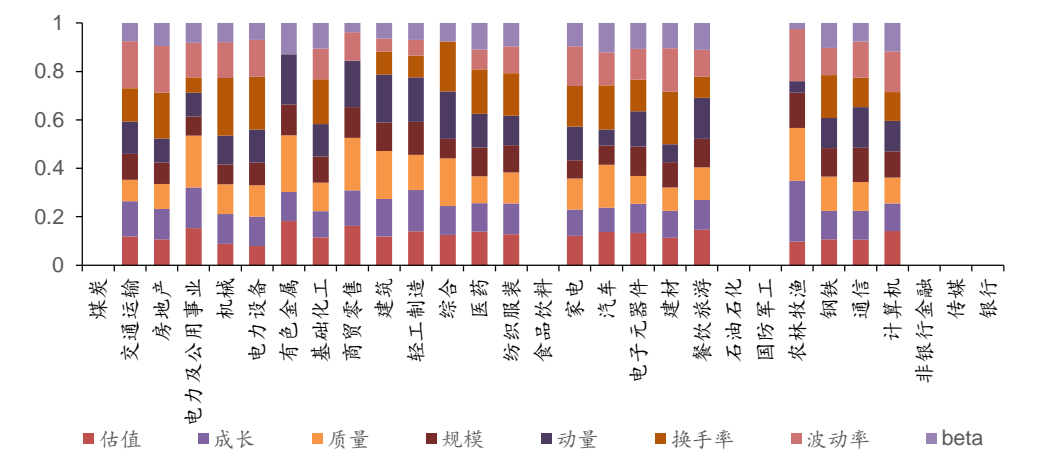
附录一 持仓风格行业间差异度举例

对于自身风格偏好行业间差异度，我们挑选了对应最大值（0.1343）的基金 C 的 2016 半年报数据和对应最小值（0.061）的基金 D 的 2018 半年报数据，并将二者的行业-自身风格偏好分布图分别绘制在图表 63、64 中。

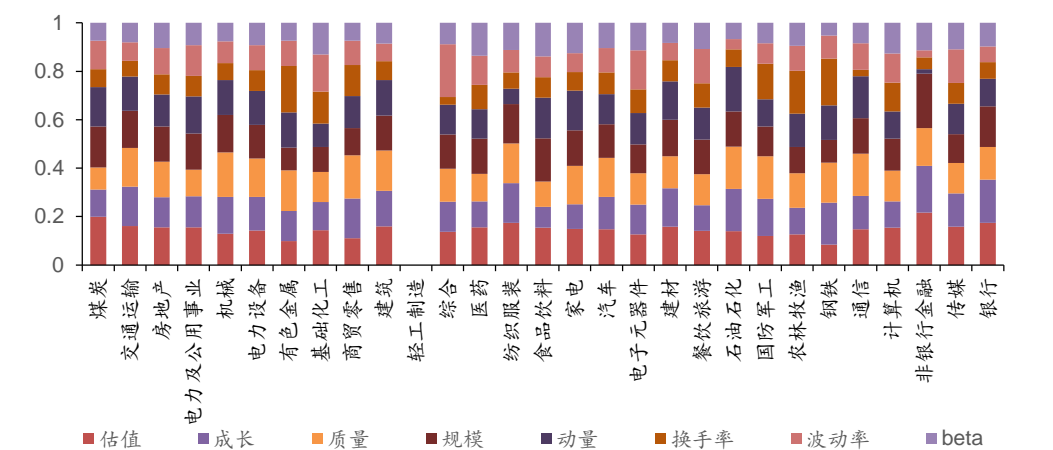
对于相对指数的因子偏好行业间差异度，我们挑选了对应最大值（0.2522）的基金 E 的 2015 年报数据和对应最小值（0.0025）的基金 F 的 2016 年报数据，并将二者的行业-相对指数因子偏好分布图分别绘制在图表 65、66 中。

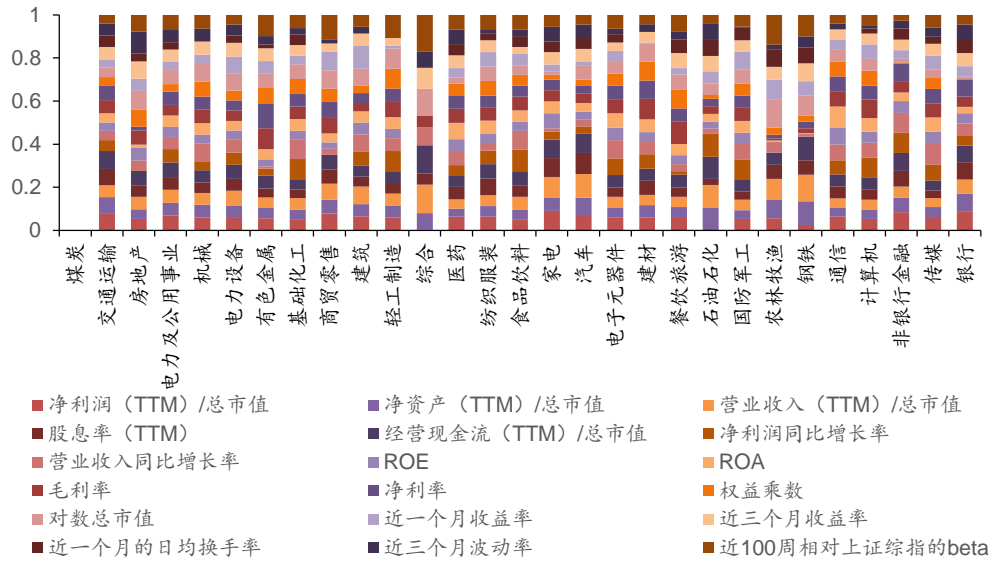
对于相对指数的风格偏好行业间差异度，我们挑选了对应最大值（0.2695）的基金 G 的 2016 半年报数据和对应最小值（0.0025）的基金 F 的 2016 年报数据，并将二者的行业-相对指数风格偏好分布图分别绘制在图表 67、68 中。

图表63： 自身风格偏好行业间差异度最大值0.1343：基金C-2016半年报的自身风格偏好行业分布图

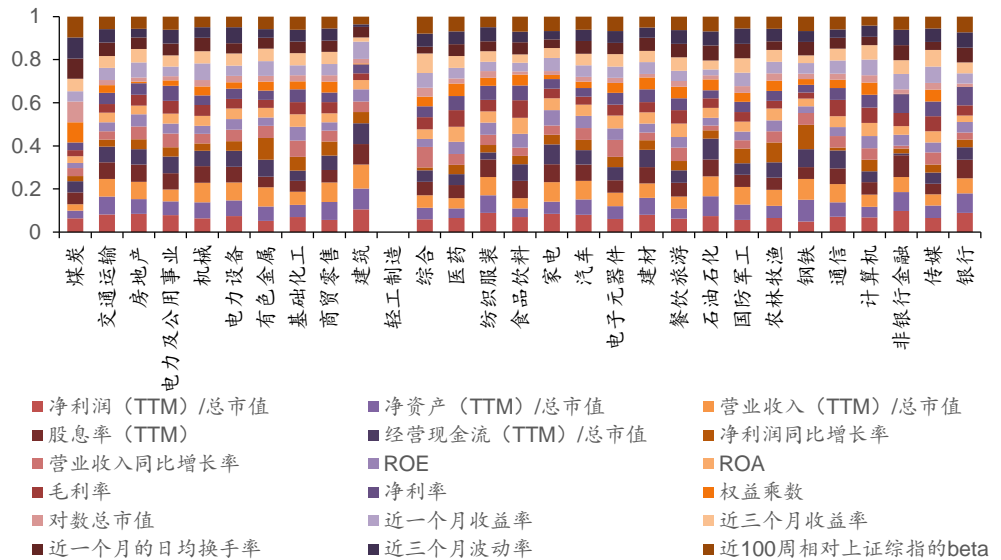


图表64： 自身风格偏好行业间差异度最小值0.061：基金D-2018半年报的自身因子偏好行业分布图

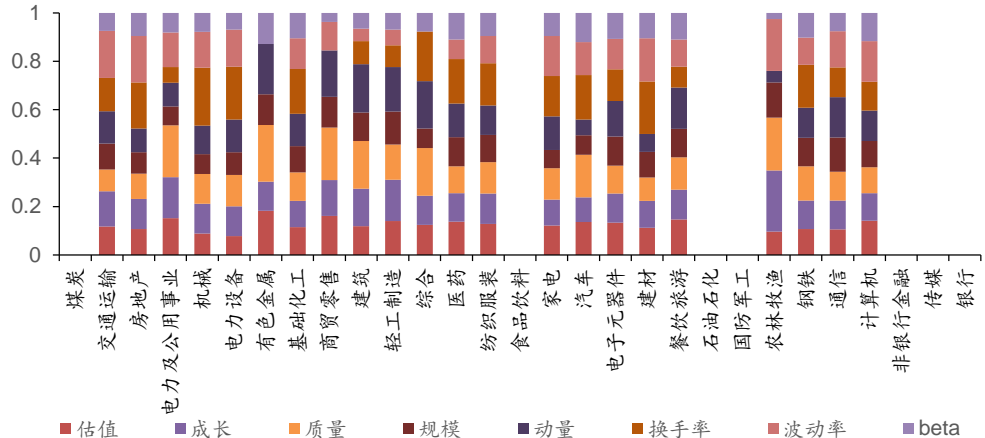


图表65： 相对指数因子偏好行业间差异度最大值 0.2522：基金 E-2015 年报的相对指数因子偏好行业分布图

资料来源：Wind，华泰证券研究所

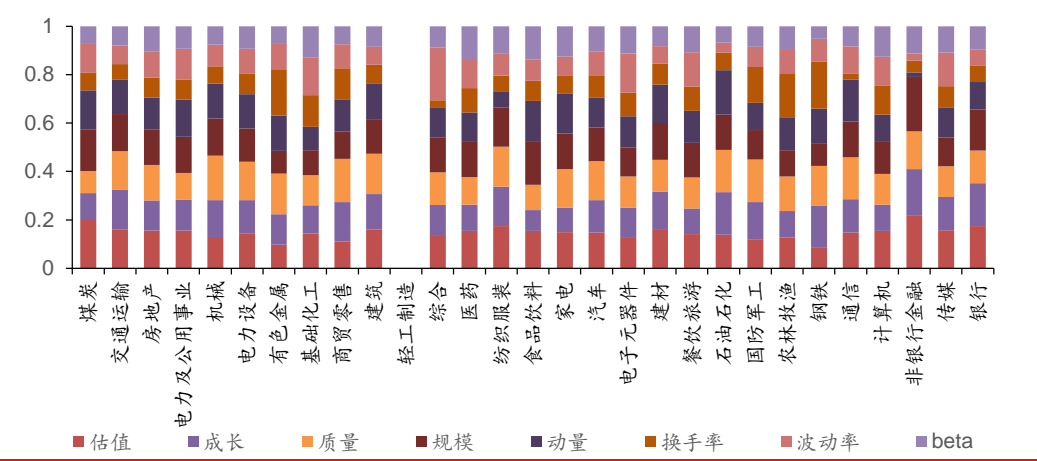
图表66： 相对指数因子偏好行业间差异度最小值 0.0025：基金 F-2016 年报的相对指数因子偏好行业分布图

资料来源：Wind，华泰证券研究所

图表67： 相对指数风格偏好行业间差异度最大值 0.2695：基金 G-2016 半年报的相对指数风格偏好行业分布图

资料来源：Wind，华泰证券研究所

图表68： 相对指数风格偏好行业间差异度最小值 0.0025：基金 F-2016 年报的相对指数风格偏好行业分布图

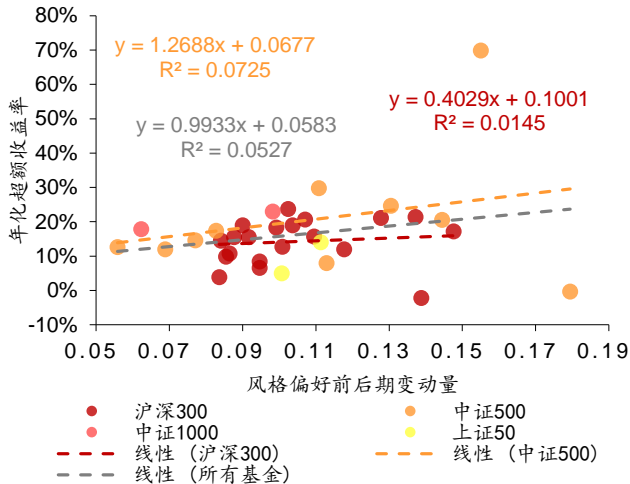


资料来源：Wind，华泰证券研究所

附录二 风格偏好稳定性与超额收益率的关系

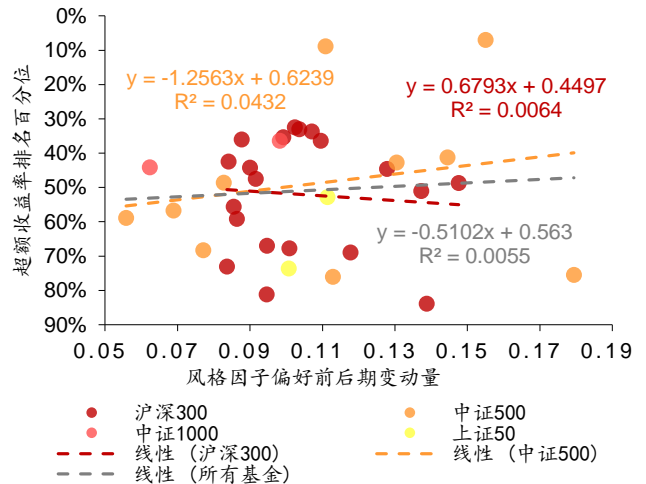
基金超额收益率与风格偏好前后期变动量、前后期相关关系的关系图分别如图表 69-70、71-72 所示。从图中可以看出，基金超额收益率与风格偏好稳定性的相关程度较低。

图表69： 年化超额收益率与风格偏好前后期变动量的关系图



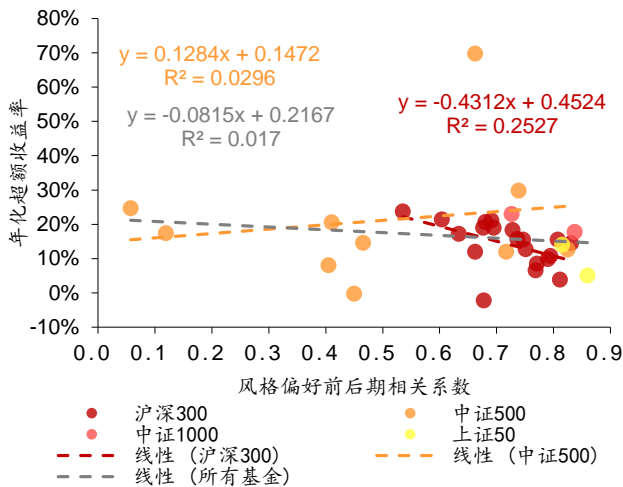
资料来源：Wind，华泰证券研究所

图表70： 超额收益率排名百分位与风格偏好前后期变动量的关系图



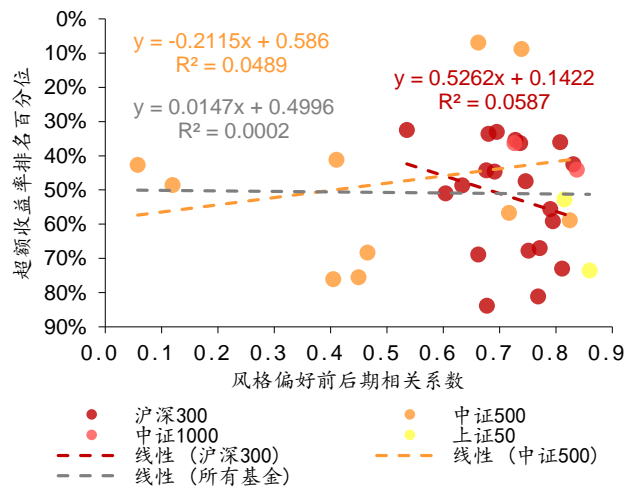
资料来源：Wind，华泰证券研究所

图表71： 年化超额收益率与风格偏好前后期相关系数的关系图



资料来源：Wind，华泰证券研究所

图表72： 超额收益率排名百分位与风格偏好前后期相关系数的关系图



资料来源：Wind，华泰证券研究所

风险提示

本报告中的收益分析是基于历史数据，由此总结的规律在未来存在失效的可能。本报告仅使用基金的半年报和年报公布的持仓数据，时间跨越尺度较大，存在一定局限性。由于现有宽基指数的公募指数增强型基金数目有限，本报告只考察了基金配置特色与超额收益的线性相关关系，未讨论可能的非线性关系，敬请注意。

免责声明

本报告仅供华泰证券股份有限公司（以下简称“本公司”）客户使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司对该等信息的准确性及完整性不作任何保证。本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，本公司可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。本公司不保证本报告所含信息保持在最新状态。本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成所述证券的买卖出价或征价。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，本公司及作者均不承担任何法律责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

本公司及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，也可能为之提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权力。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

本公司具有中国证监会核准的“证券投资咨询”业务资格，经营许可证编号为：91320000704041011J。

全资子公司华泰金融控股（香港）有限公司具有香港证监会核准的“就证券提供意见”业务资格，经营许可证编号为：A0K809

©版权所有 2018 年华泰证券股份有限公司

评级说明

行业评级体系

一报告发布日后的 6 个月内的行业涨跌幅相对同期的沪深 300 指数的涨跌幅为基准；

一投资建议的评级标准

增持行业股票指数超越基准

中性行业股票指数基本与基准持平

减持行业股票指数明显弱于基准

公司评级体系

一报告发布日后的 6 个月内的公司涨跌幅相对同期的沪深 300 指数的涨跌幅为基准；

一投资建议的评级标准

买入股价超越基准 20%以上

增持股价超越基准 5%-20%

中性股价相对基准波动在-5%~5%之间

减持股价弱于基准 5%-20%

卖出股价弱于基准 20%以上

华泰证券研究

南京

南京市建邺区江东中路 228 号华泰证券广场 1 号楼/邮政编码：210019

电话：86 25 83389999/传真：86 25 83387521

电子邮件：ht-rd@htsc.com

深圳

深圳市福田区深南大道 4011 号香港中旅大厦 24 层/邮政编码：518048

电话：86 755 82493932/传真：86 755 82492062

电子邮件：ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同 28 号太平洋保险大厦 A 座 18 层

邮政编码：100032

电话：86 10 63211166/传真：86 10 63211275

电子邮件：ht-rd@htsc.com

上海

上海市浦东新区东方路 18 号保利广场 E 栋 23 楼/邮政编码：200120

电话：86 21 28972098/传真：86 21 28972068

电子邮件：ht-rd@htsc.com