

因子风格择时策略（2021 年 11 月期）

本篇报告中我们通过 XGBoost 方法构建因子择时模型以及大小盘轮动模型。根据拟合得到 XGBoost 模型对于 barra 风格、大小盘风格未来收益进行预测，为投资者提供系统的风格配置参考。由于两套模型都用到宏观经济数据和市场情绪变量作为输入变量，且策略框架有一定共性，因此我们放在同一篇报告中做展示。

本报告包括两个相互独立的策略：

1) 策略 1：我们使用 XGboost 模型，对股票 barra 风格因子未来一个月因子收益正负做出预测。输入自变量包括因子拥挤度、宏观数据（多指标）、市场状态等。该策略对于做指数增强、smart beta 轮动策略的投资者有帮助。

模型回测显示，风格轮动择时效果明显。我们将择时结果与多因子选股组合优化框架进行结合，选取中证 500 指数为基准，选股范围中证 500 成分股。在控制行业偏离和风格偏离情况下，XGBoost 模型在测试集（2019 年至今）各风格因子上平均月度胜率为 61%，策略收益年化 32.64%，相对于中证 500 超额收益年化 27.77%。今年以来该模型超额收益稳健，除了 1 月其余各月基本上都跑赢基准，其中在 8 月份的收益尤为突出，主要是因为高配了动量、质量和估值因子。

模型在 11 月份建议高配 beta、质量、成长、估值、低换手因子，低配动量、小市值、低波动因子。从模型逻辑上看，当前经济周期还没有到全部复苏的阶段，市场流动性较为宽松，但市场主线尚不明确，没有明显的风格偏好，因此模型总体保持风格敞口的相对均衡配置。模型本期相对优选对于利率敏感性较强的小市值、成长因子，以及在整个投资环境趋势不明晰下控制质量和估值有一定敞口，另低换手率因子长期来看都值得高配。

2) 策略 2：我们使用嵌套 XGBoost 的打分加权模型，对沪深 300、中证 500 未来一个月收益高低做出预测。对于偏长期的宏观经济数据维度和偏短期的市场情绪、估值盈利差、风险偏好维度，我们分别构建两套不同的 XGBoost 模型来建立自变量到因变量的映射，然后对不同维度的信号进行等权打分，从而可判断当前市场环境是否更加有利于投资沪深 300 或者中证 500。策略选取业绩比较基准为“50%*沪深 300+50%*中证 500”。

由宏观经济数据为自变量构建的 XGBoost 模型的胜率分别为 76%、超额收益年化 6.53%，夏普比率 1.26。由市场情绪自变量构建的 XGBoost 模型胜率位 72%，夏普比率 1.5，超额年化收益 7.52%。我们将两个信号进行等权得到的综合信号。当两个信号方向不一致时，模型等权配置沪深 300 与中证 500。与单信号相比，综合信号虽然超额收益还是较之前有所降低，但是熨平了波动，策略夏普比率提高到 1.66。综合模型超额收益在 2017 年和 2020 年表现优异，18 年回撤较大，2021 年以来超额收益 8%。

当前 11 月份两个信号的方向有分歧，市场流动性偏宽松但是经济还未完全复苏。以宏观数据为基础的信号看多中证 500，而以偏短期视角为基础的信号看多沪深 300。

风险提示：模型测试基于历史数据，市场未来可能发生变化。

熊颖瑜 联系人
 (8621) 60753902
 xiongyingyu@gjzq.com.cn

张剑辉 分析师 SAC 执业编号: S1130519100003
 (8610)66211648
 zhangjh@gjzq.com.cn

内容目录

第一部分：A 股风格因子体系介绍	4
1.1 市场回顾:	4
1.2 最新建议:	5
1.3 多因子体系介绍:	6
第二部分：因子择时信号简介及有效性测试	7
2.1 离散度有效性测试	7
2.2 因子拥挤度有效性测试	7
2.3 宏观经济数据与市场情绪变量	9
第三部分：利用 XGBoost 模型构建风格择时策略	10
3.1 XGBoost 原理	10
3.2 构建流程	12
3.3 滚动回测	12
3.4 特征选择	13
第四部分：利用 XGBoost 模型构建指数轮动策略	16
4.1 指数历史走势情况	17
4.2 指数轮动策略回测	19

图表目录

图表 1：风格轮动策略框架	4
图表 2：近一年以来 A 股风格走势	5
图表 3：近四个月以来 A 股风格的拥挤度历史分位数	5
图表 4：近一年以来各风格因子收益情况	5
图表 5：中证 500 中近三个月因子配置建议	6
图表 6：近 3 个月风格因子月度收益	6
图表 7：多因子框架展示	6
图表 8：大类风格因子构成	6
图表 9：A 股纯因子收益：2016 年至 2021.10	7
图表 10：风格因子在离散度上的变化（2021.09 至 2021.10）	7
图表 11：风格因子在离散度上的变化	7
图表 12：风格因子在估值差上的变化（2021.09 至 2021.10）	8
图表 13：估值价差加权因子多空收益	8
图表 14：风格因子在配对相关性上的变化（2021.09 至 2021.10）	8
图表 15：配对相关性加权因子多空收益	8
图表 16：风格因子在多空波动率上的变化（2021.09 至 2021.10）	9
图表 17：多空波动率加权因子多空收益	9
图表 18：宏观指标：月度，数据区间：2016.01 至 2021.10	9
图表 19：宏观指标衍生：月度差分，数据区间：2016.01 至 2021.10	10
图表 20：市场情绪指标，数据区间：2016.01 至 2021.10	10

图表 21: 决策树原理.....	11
图表 22: 单颗决策树组成部分.....	11
图表 23: XGBoost 模型需要解决的问题.....	11
图表 24: XGB 模型构建流程图	12
图表 25: 滚动回测示意图.....	13
图表 26: beta 特征重要性.....	13
图表 27: 动量特征重要性.....	13
图表 28: 小市值特征重要性.....	14
图表 29: 质量特征重要性.....	14
图表 30: 低波动率特征重要性.....	14
图表 31: 成长性特征重要性.....	14
图表 32: 估值特征重要性.....	14
图表 33: 低换手率特征重要性.....	14
图表 34: 滚动回测测试集样本上的平均准确率.....	15
图表 35: 2016 年至今不同风格偏离约束下的组合优化超额收益.....	15
图表 36: 轮动模型 2021 年以来的月度超额收益.....	16
图表 37: XGBoost 模型多头收益	16
图表 38: XGBoost 模型超额收益	16
图表 39: 2010 年以来中证 500、沪深 300 指数走势.....	17
图表 40: 沪深 300 与中证 500 月度收益.....	17
图表 41: 沪深 300 指数成分股市值分布.....	18
图表 42: 中证 500 指数成分股市值分布.....	18
图表 43: 沪深 300 指数 PE 水平	18
图表 44: 中证 500 指数 PE 水平	18
图表 45: 沪深 300 指数行业权重分布.....	18
图表 46: 中证 500 指数行业权重分布.....	18
图表 47: 指数轮动模型构架搭建流程.....	19
图表 48: XGBoost 模型特征重要性（宏观经济数据）.....	19
图表 49: XGBoost 信号 1 有效性检验（宏观经济数据）.....	19
图表 50: XGBoost 模型特征重要性（市场情绪变量）.....	20
图表 51: XGBoost 信号 2 有效性检验（市场情绪变量）.....	20
图表 52: XGBoost 信号 1 滚动回测各期准确率.....	20
图表 53: XGBoost 信号 2 滚动回测各期准确率.....	20
图表 54: XGBoost 模型信号, 1 表示沪深 300, 0 表示中证 500.....	20
图表 55: 不同信号走势对比（2016.01~2021.10）.....	21
图表 56: 综合信号指数轮动收益（2016.01~2021.10）.....	21
图表 57: 综合信号走势对比（2016.01~2021.10）.....	21

第一部分：A 股风格因子体系介绍

因子择时必要性在不断提高。近年来随着因子投资和因子研究的深入，我们会看到因子的长期收益有时会偏离理论收益，出现因子失效的问题。如 2017 年，随着市场上量化策略的增多，许多之前有效的因子比如市值因子、动量因子、波动率因子都出现过阶段的失效，这就导致传统的多因子模型获取稳定的超额收益变得越发困难，因子择时也随之变得越发重要。本文认为造成因子失效现象的主要原因包括因子拥挤度提升以及不同宏观周期下因子投资性价比产生的变化等。本文将影响因子收益的这几个因素共同纳入到框架中讨论。

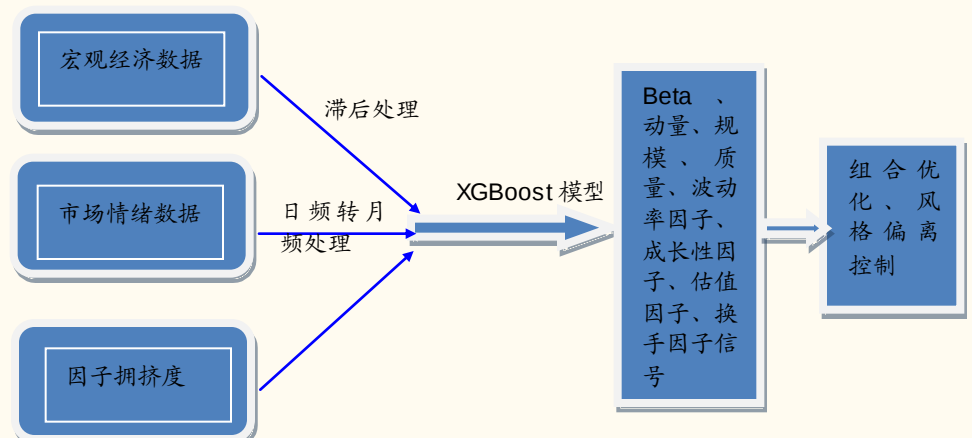
市场上进行因子风格轮动的方法主要是从两方面进行预测。一类是从因子收益本身的动量效应和因子拥挤度等内生指标出发，对未来收益正负进行预测，另一类是从市场宏观数据、市场状态、资金情绪等外部数据出发。

对于股票风格轮动，我们之前在《因子风格择时策略》系列报告中尝试用决策树方法对股票未来收益进行预测，所用到的自变量体系中的因素主要是利率趋势、A 股 roe 变化趋势，以及因子拥挤度等内生变量。决策树是一种描述对实例进行分类的树形结构，本身是一种弱分类器，其主要优点是模型具有可读性。构造决策树的核心问题是每一步对样本属性如何做拆分。

为了更加充分的利用宏观经济数据、市场情绪外部变量，我们准备在原有择时模型上扩充一些对 A 股市场有重要影响的宏观变量，形成“因子风格择时 2.0 版本”。我们将模型从弱学习器决策树改进为强学习器 XGBoost，增强模型的预测效果。在机器学习 XGBoost 的框架下对未来一个月的风格收益大小进行预测。新增的宏观变量包括经济增速、物价、货币政策、利率、进出口等数据，以及其一阶差分衍生指标。新增的市场情绪变量包括市场流动性、市场波动率、指数估值、指数涨幅表现等。这些变量对因子收益率影响的方式各不相同，这种复杂关系很难通过线性模型来刻画。

XGBoost 是对通过反复地选择一个指向负梯度方向的函数，对一任意可导目标函数的优化过程。XGBoost 和 Adaboost 虽然都是由单颗决策树到多颗决策树进行优化，但是优化策略不同。通过对比两者效果，我们决定选用 XGBoost 模型。

图表 1：风格轮动策略框架



来源：国金证券研究所

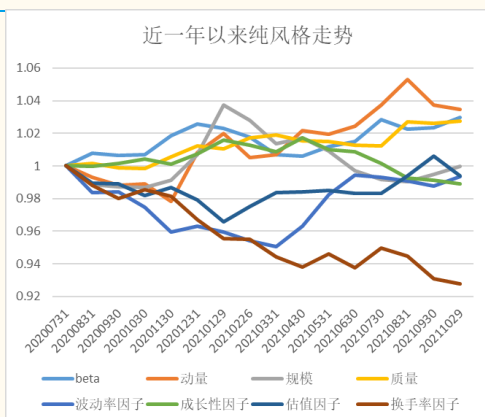
1.1 市场回顾：

自今年春节之后来到 9 月初，中证 500、中证 1000 表现优异，本轮中小盘行情是继 2013-2015 年之后再度持续时间较长的一轮行情。风格的持续性变化跟整个经济周期、市场流动性等外生变量相关性很大。风格的突然反向则跟风格拥挤度内生变量大幅上升有关。今年由于环保双控导致上游周期原材料供

给稀缺，而且新能源板块相对于其他蓝筹股估值更低，盈利预期更高，整个市场需要保持一定的流动性来对冲疫情后经济动能下滑的冲击。因此小市值因子表现比较持续优异。同时我们检测到小市值因子的拥挤度持续上升，在 9 月末模型已经提示市值因子的拥挤度超过 90%，且市值因子的估值价差对比其他因子已经比较高。在 9、10 月份小市值风格发生回撤，连续上涨 3 个月的动量因子也在 9、10 月份发生大幅回撤。

从 2020.07.31 至 2021.10.31 时段 A 股风格走势来看，除了长期持有低换手因子可以获取持续收益之外，其余因子切换整体还是比较频繁，能够持续 3 个月收益方向不变的因子不多。因此投资者在指数增强时如果对单向因子持续保留太多敞口，比如同时高配小市值、高波动、高换手、高成长、高 beta 等因子，组合收益的波动可能会比较剧烈。在指数增强的风格配置上我们更加推荐均衡配置，比如说高配小市值、高成长因子的同时高配一些质量、盈利因子，低配换手、波动因子。

图表 2：近一年以来 A 股风格走势



来源：国金证券研究所，wind

图表 3：近四个月以来 A 股风格的拥挤度历史分位数

	20210730	20210831	20210930	20211029
beta	10.00%	15.71%	18.57%	17.14%
动量	38.57%	51.43%	65.71%	54.29%
规模	77.14%	88.57%	94.29%	95.71%
质量	38.57%	47.14%	42.86%	35.71%
波动率因子	97.14%	95.71%	98.57%	100.00%
成长性因子	87.14%	82.86%	92.86%	97.14%
估值因子	98.57%	94.29%	97.14%	100.00%
换手率因子	91.43%	97.14%	81.43%	92.86%

来源：国金证券研究所，wind

图表 4：近一年以来各风格因子收益情况

	beta	动量	规模	质量	波动率因子	成长性因子	估值因子	换手率因子
20200930	-0.10	-0.52	-0.13	-0.30	0.03	0.18	-0.08	-0.81
20201030	0.04	0.12	-0.02	-0.02	-0.95	0.26	-0.70	0.53
20201130	1.16	-1.08	0.44	0.72	-1.58	-0.31	0.50	-0.42
20201231	0.69	3.06	1.58	0.65	0.38	0.63	-0.75	-1.43
20210129	-0.28	1.14	3.01	-0.16	-0.35	0.83	-1.38	-1.20
20210226	-0.51	-1.47	-0.91	0.65	-0.57	-0.28	0.98	-0.07
20210331	-1.05	0.19	-1.38	0.19	-0.38	-0.41	0.84	-1.13
20210430	-0.09	1.47	0.33	-0.38	1.33	0.85	0.05	-0.64
20210531	0.58	-0.22	-0.76	-0.04	1.97	-0.72	0.12	0.86
20210630	0.30	0.50	-1.19	-0.22	1.25	-0.11	-0.18	-0.93
20210730	1.30	1.26	-0.54	-0.02	-0.14	-0.73	0.00	1.30
20210831	-0.53	1.51	-0.13	1.45	-0.26	-0.90	1.06	-0.53
20210930	0.08	-1.49	0.44	-0.09	-0.32	-0.14	1.25	-1.46
20211029	0.58	-0.26	0.47	0.15	0.62	-0.22	-1.23	-0.34

来源：国金证券研究所，wind

1.2 最新建议：

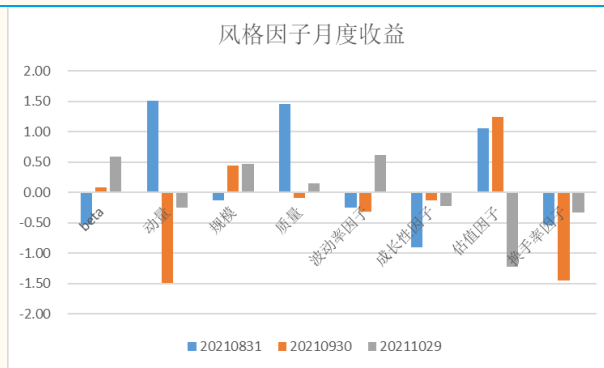
本文在 2021 年中推出的“因子风格择时策略”的基础上做了改进，我们采用维度更广的自变量，且将弱学习器升级为强学习器，构建了准确率更高的 A 股市场的因子择时指标体系 2.0。该策略主要以因子离散度、因子拥挤度、宏观变量、市场状态为输入变量，然后利用 XGboost 方法，预测每个风格因子未来的收益正负（排序）。相对于决策树模型，XGboost 可以相对提高因子择时胜率。最后我们将因子择时结果与多因子模型的组合风格偏离控制相结合。近几期因子风格择时参考如下：

图表 5：中证 500 中近三个月因子配置建议

	beta	动量	小市值	质量	低波动率因子	成长性因子	估值因子	低换手率因子
20210730	高配	高配	低配	低配	高配	高配	高配	高配
20210831	高配	高配	低配	高配	高配	高配	高配	高配
20210930	低配	高配	高配	高配	低配	高配	高配	高配

来源：国金证券研究所

图表 6：近 3 个月风格因子月度收益

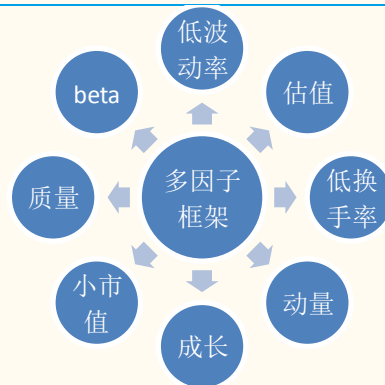


来源：国金证券研究所,wind

1.3 多因子体系介绍:

我们的因子体系选择如下：beta、动量、小市值因子、质量、低波动率因子、成长性因子、估值因子、低换手率因子。

图表 7：多因子框架展示



来源：Wind，国金证券研究所

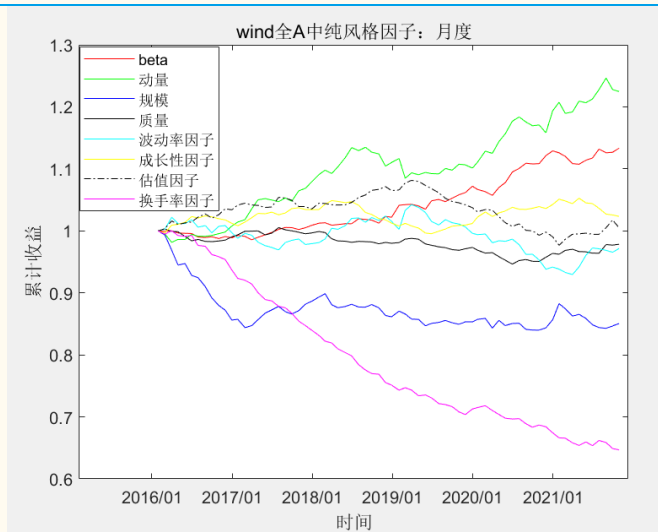
图表 8：大类风格因子构成

大类因子	风格因子构成
beta 因子	beta 因子
动量	RSTR 长期动量
市值因子	LNCAP 对数市值
质量因子	MLEV 市场杠杆、DTOA 资产负债率、BLEV 账面杠杆
波动率因子	CMRA 月超额波动率、DASTD 日超额收益波动率、HSIGMA
成长性因子	SGRO 营业收入增速、EGRO 利润同比增速
估值因子	PB
换手率因子	月流动性 STOM、季流动性 STOQ

来源：国金证券研究所

我们计算了上述 8 类因子 2012 年以来的纯因子收益。可以看出除了 beta 和成长性因子趋势性不太明显，其余因子趋势性都较强。

图表 9：A 股纯因子收益：2016 年至 2021.10



来源：国金证券研究所，wind

第二部分：因子择时信号简介及有效性测试

对于需要通过风格轮动来增加投资组合 beta 收益的投资者来说，选择中证 500 为基准做增强比较普遍。因此，我们的拥挤度计算和最终组合优化也在此范围进行构造。接下来我们使用更加直观的办法来观测各类信号有效性。

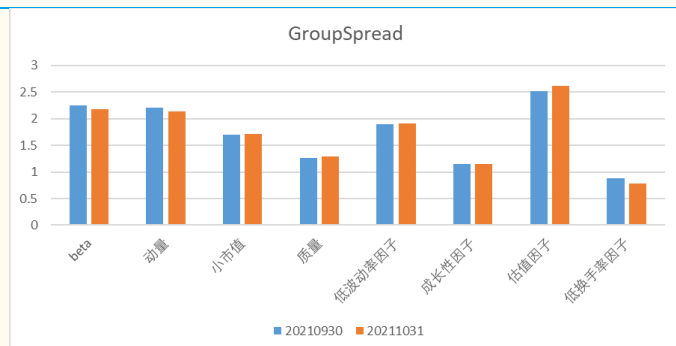
2.1 离散度有效性测试

因子离散度 spread 是在一个横截面上，根据某一个因子大小对股票池进行排序，前 10% 的股票因子均值与后 10% 的股票因子均值的差值即因子 spread。高离散度的因子在未来可能有更好的收益区分能力以及可以带来更高的信息比例。

基于 Spread 构建策略的方向和拥挤度指标相反。Spread 的投资逻辑主要是基于资产定价再极度无效之后会出现均值回复。当因子 spread 很大，代表这个因子已经被“过度忽视”、“欠定价较深”，此时投资该因子未来收益可能较大，所以性价比较高。比如低 ROE 股票的估值过低，高 ROE 股票的估值过高，PB 离散度高于 ROE 离散度，未来大概率 PB 有效性高于 ROE 因子。

从长期来看，每个月选择离散度较高的因子进行超配还是会有趋势性收益的。近两个月各因子离散度变化不明显。

图表 10：风格因子在离散度上的变化（2021.09 至 2021.10）



来源：国金证券研究所，wind

图表 11：风格因子在离散度上的变化



来源：国金证券研究所，wind

2.2 因子拥挤度有效性测试

因子的拥挤度指标的研究最初源于国外研究机构。因子拥挤度指标越高，

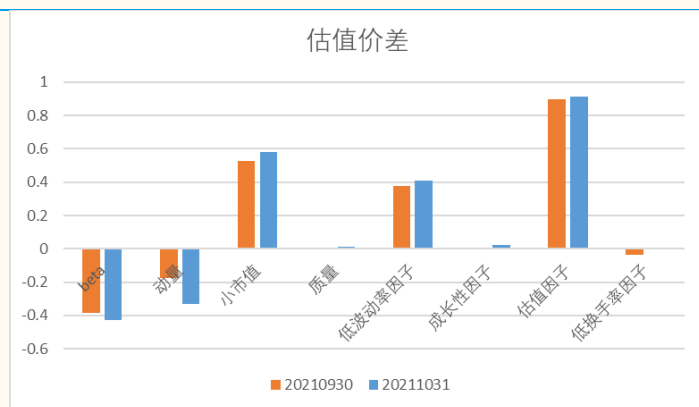
未来投资该因子的回报会越低。

因子拥挤是指当某个类型的因子被广泛认为有效时，会得到大量资金的追捧，这样的追捧尽管在一定程度上会加大因子的收益，但是另一方面也会加大因子收益的波动，使得因子的收益在一定时间段内出现收益降低或者收益大幅回撤的现象，从某种意义上来说，这个现象与量化投资策略的钝化现象有一定的相似性。因此，对因子拥挤现象的跟踪，可以在一定程度上规避或者控制因子失效带来的投资组合的亏损风险。从时间序列的维度上看，因子拥挤度越高，未来收益为负的可能性越大。从横截面的维度检验来看，拥挤度更高的因子下一期的收益也会比其他因子弱一些。

估值价差：估值价差实际上属于“因子的因子”概念范畴，即某个因子的多头组合的估值中位数减去空头组合的估值中位数就是该因子的估值价差。当资金追捧该因子时，多头因子的估值会上升，反之空头因子估值下降。估值价差就会变大。

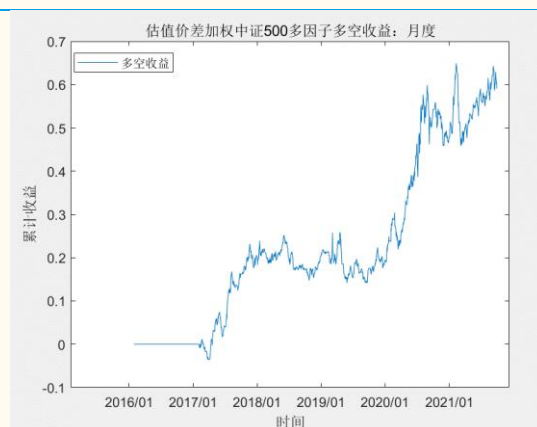
从长期来看，每个月选择估值价差较低的因子进行超配还是会有趋势性收益的。近两个月各因子估值价差变化不明显。市值风格、低波动、估值因子的估值价差处于高位。

图表 12：风格因子在估值差上的变化（2021.09 至 2021.10）



来源：国金证券研究所,wind

图表 13：估值价差加权因子多空收益

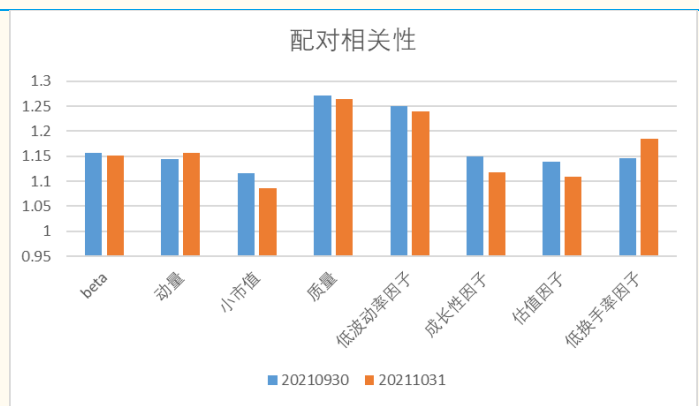


来源：国金证券研究所,wind

配对相关性：当某个因子过热时，投资者可能会同时买入多头组合的股票，或者同时卖出空头组合的股票，因此多头、空头组合内部的股票可能会同涨同跌的态势。

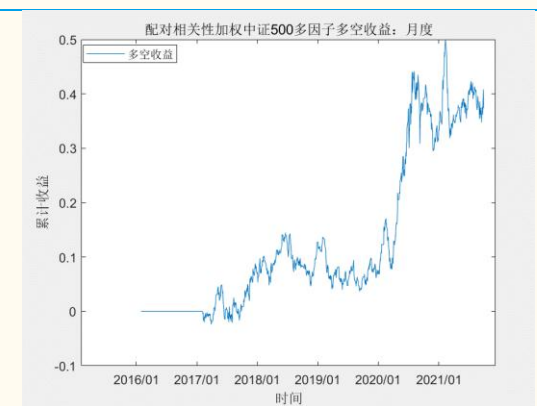
从长期来看，每个月选择配对相关性较低的因子进行超配，收益波动比较大，趋势不明显。近两个月小市值、成长性、估值因子的配对相关性有所下降，低换手因子配对相关性提升。

图表 14：风格因子在配对相关性上的变化（2021.09 至 2021.10）



来源：国金证券研究所,wind

图表 15：配对相关性加权因子多空收益

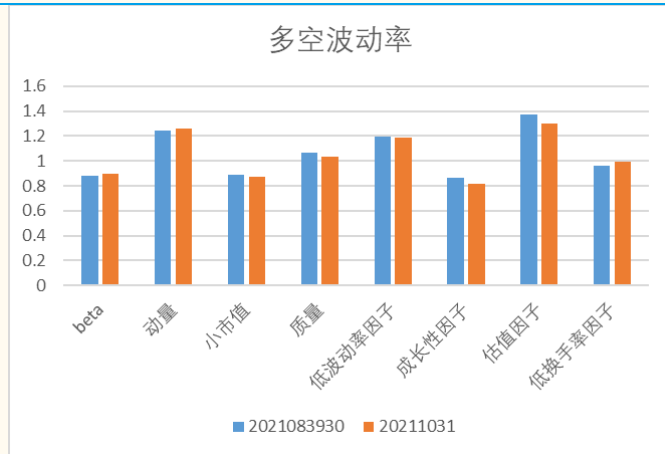


来源：国金证券研究所,wind

多空收益波动率：主要描述当资金过度集中在某一个因子上时，该因子的日收益率的波动会加大；

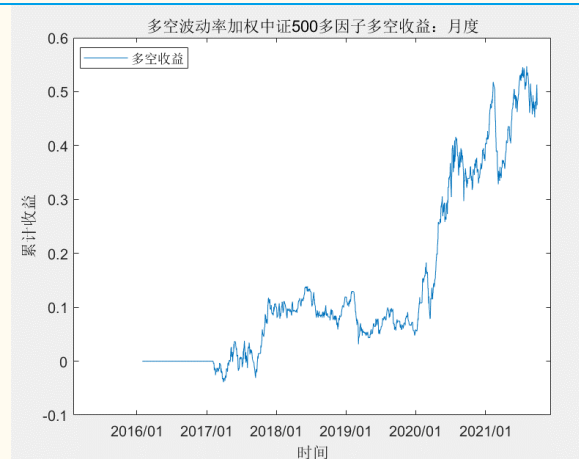
从长期来看，每个月选择多空波动率较低的因子进行超配，收益虽然有一定的波动，但也有明显的趋势。近两个月估值因子和成长性因子的多空波动率有所下滑，其他风格因子多空波动率变化不大。

图表 16：风格因子在多空波动率上的变化（2021.09 至 2021.10）



来源：国金证券研究所,wind

图表 17：多空波动率加权因子多空收益



来源：国金证券研究所,wind

2.3 宏观经济数据与市场情绪变量

宏观经济数据是一类不可忽视的变量。无论是风格轮动、指数轮动，一个重要的因素就是由于经济逻辑驱动的。宏观经济对于市值风格的影响是显然的，当宏观经济向好，无风险利率较低时，小盘股通常能够获得较好的收益。

宏观变量主要是从经济增长、价格水平、货币环境、国际贸易、利率水平五大类选取指标作为代表变量。每个大类下我们选取了一些具有代表性的细分指标。表征经济增长类的指标包括工业增加值同比、工业企业产销量。表征价格水平的包括 CPI、PPI 等指标。表征货币环境的价包括 M0、M1、M2 增速以及其剪刀差，人民币贷款余额增速，社会融资规模增速。表征货币环境的量即利率水平，我们纳入 3 个月、6 个月、1 年、10 年国债利率。还有国际贸易方面的进出口金额等。

针对投资者关注的某些指标（如图表 18）我们做了一阶差分处理。由于统计局发布数据往往滞后一个月，我们对该部分指标进行滞后处理，比如在 9 月末我们用 8 月公布的统计局数据作为最新自变量。另外我们将日频数据取月末截面值转换为月频数据。

图表 18：宏观指标：月度，数据区间：2016.01 至 2021.10

M0000545	工业增加值：当月同比	M0001380	M0
M0000011	工业增加值：累计同比	M0001382	M1
M0007452	工业企业：产销量：当月值	M0001384	M2
M0000706	CPI:食品:环比	M0009978	存贷差
M0000612	CPI同比	M0009969	金融机构：各项贷款余额
M0001228	PPI:生产资料：同比	M0024054	公共财政收入
M0001232	PPI：生活资料：同比	M0024057	税收收入
M0001227	PPI：全部工业品：同比	S0059742	国债利率：6个月
M0000608	进口金额:当月值	S0059744	国债利率：1年
M0000606	出口金额:当月值	M0000185	美元兑人民币汇率
M0000273	固定资产投资累计同比	S0059741	国债利率：3个月
M0001428	社会消费品零售总额:当月同比	S0059749	国债利率：10年
M0001427	社会消费品零售总额:当月值	Realized_Vol_20D	中证全指收益率过去20天波动率

来源：国金证券研究所,wind

图表 19：宏观指标衍生：月度差分，数据区间：2016.01 至 2021.10

	原指标	衍生指标
工业企业：产销量：当月值	M0007452	M0007452diff
进口金额：当月值	M0000608	M0000608diff
出口金额：当月值	M0000606	M0000606diff
社会消费品零售总额：当月值	M0001427	M0001427diff
M0	M0001380	M0001380diff
M1	M0001382	M0001382diff
M2	M0001384	M0001384diff
金融机构：各项贷款余额	M0009969	M0009969diff
公共财政收入	M0024054	M0024054diff
税收收入	M0024057	M0024057diff
国债利率：6个月	S0059742	S0059742diff
国债利率：1年	S0059744	S0059744diff
美元兑人民币汇率	M0000185	M0000185diff
国债利率：3个月	S0059741	S0059741diff
国债利率：10年	S0059749	S0059749diff
中证全指收益率过去20天波动率	Realized_Vol_20D	Realized_Vol_20Ddi

来源：国金证券研究所,wind

除宏观基本面指标外，微观市场结构在风格轮动上也有一定逻辑与价值。具体包括市场涨跌状况、市场波动率、市场交易量，市场换手率等，不同市场结构下可能会呈现完全不同的轮动规律。

图表 20：市场情绪指标，数据区间：2016.01 至 2021.10

市场情绪指标		
市场流动性	VOLUME	交易量
	vol20	20日均交易量
	vol60	60日均交易量
	vol120	120日均交易量
市场估值	PE_TTM	估值
	PB_LF	估值
	DIVIDENDYIELD2	股息率
市场波动率	VOLATILITYRATIO	波动率
市场表现	CLOSE	收盘价
	tech_MA20	ma20
	chgpct20	20日涨跌幅
	chgpct60	60日涨跌幅
	chgpct120	120日涨跌幅

来源：国金证券研究所，wind

第三部分：利用 XGBoost 模型构建风格择时策略

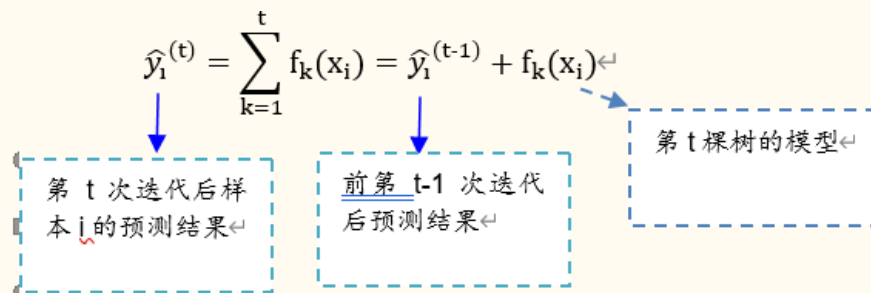
3.1 XGBoost 原理

XGBoost 的全称是 eXtreme Gradient Boosting，它是经过优化的分布式梯度提升库，旨在高效、灵活且可移植。XGBoost 是大规模并行 boosting tree 的工具，它是目前最快最好的开源 boosting tree 工具包，比常见的工具包快 10 倍以上。

原理：XGBoost 算法可以说是一种集成式提升算法，是将许多基础模型集成在一起，形成一个很强的模型。这里的基模型可以是回归决策树 CART。该算法会不断的添加树，并且不断的进行特征分裂来生长一棵树，每次添加一个

树，其实就是学习一个新函数，去拟合上次预测的残差，当我们训练完成得到 t 棵树，假设我们第 t 次迭代要训练的树模型是 $f_k(x)$ ，则有：

图表 21：决策树原理

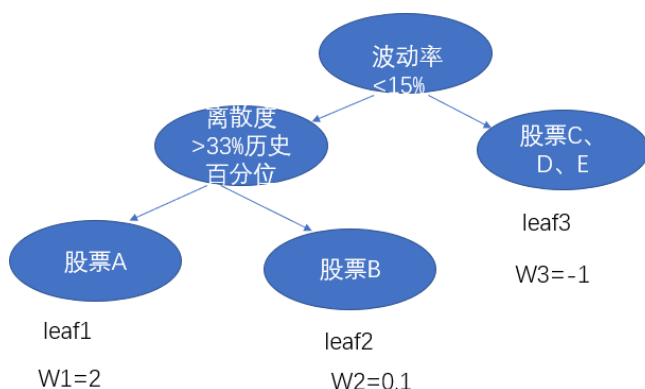


来源：国金证券研究所

定义一棵决策树，主要包括两个部分：叶子节点的权重向量 w 以及实例样本到叶子节点的映射关系 q (本质是树的分支结构)

图表 22：单颗决策树组成部分

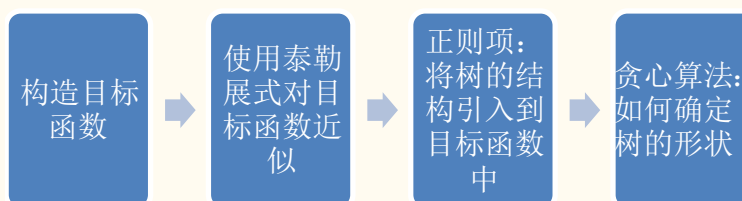
- $f_t(x) = w_q(x)$ ，其中 $w \in R^T$ ， w 是长度为 T 的一维向量，代表树 q 各叶子节点的权重。 $q: R^d \rightarrow \{1, 2, 3, \dots, T\}$ q 代表一棵树的结构，作用是将输入 x_i 映射到某个叶子节点，假设这个树有 T 个叶子节点。



$q(\text{股票A})=1$
 $q(\text{股票B})=2$
 $q(\text{股票C})=3$
 $q(\text{股票D})=3$
 $q(\text{股票E})=3$

来源：国金证券研究所

图表 23：XGBoost 模型需要解决的问题



来源：国金证券研究所

XGBoost 的目标函数：模型的预测精度由模型的偏差和方差共同决定，损失函数代表了模型的偏差，想要方差小则需要在目标函数中添加正则项，用于防止过拟合。所以目标函数由模型的损失函数 L 与抑制模型复杂度的正则项 Ω

组成，目标函数的定义如下：

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{i=1}^t \Omega(f_i)$$

其中，前一项 $L = \sum_{i=1}^n l(y_i, \hat{y}_i)$ 表示损失函数，是由预测值与真实值之间的偏差决定的。后一项 $\sum_{i=1}^t \Omega(f_i)$ 表示抑制模型复杂度的正则项，将全部 t 棵树的复杂度进行求和，用于防止模型过拟合。

对于如何将上述 Obj 用泰勒公式展开的过程我们暂且省略，得到

$$Obj^t \sim \sum_{i=1}^n [g_i f_t(x_i) + 1/2 h_i f_t^2(x_i)] + \Omega(f_t)$$

其中 g_i 表示损失函数的一阶导（可求）， h_i 表示损失函数的二阶导（可求）。我们只需要求出每一步损失函数的一阶导和二阶导的值然后最优化目标函数就可以得到每一步的 $f(x)$ ，最后根据加法模型得到一个整体模型。

决策树的复杂程度 Ω 可由叶子数 T 组成，叶子节点越少模型越简单。此外叶子节点也不应该含有过高的权重 w ，所以目标函数的正则项由生成的所有决策树的叶子节点数量，和所有节点权重所组成的向量的 L_2 范式共同决定，其中 T 表示叶子总数， w 表示叶子节点权重向量范式

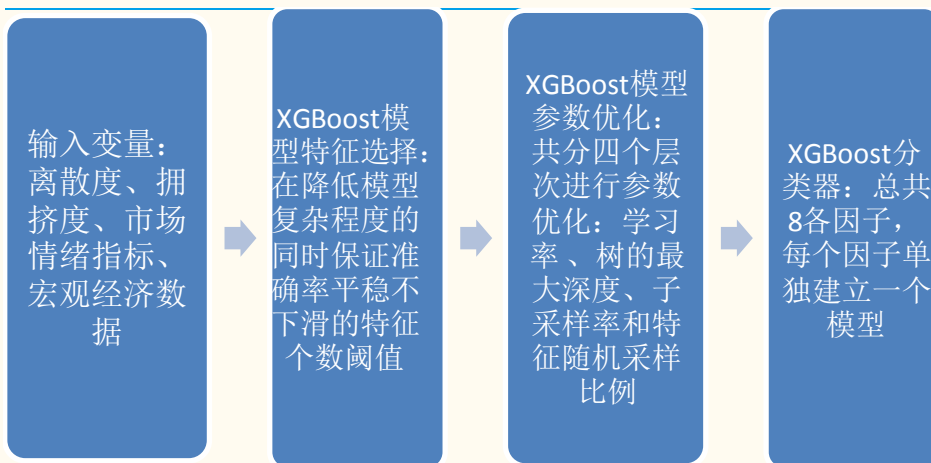
$$\Omega(f_t) = \gamma T + 1/2\lambda \sum_{j=1}^T w_j^2$$

关于树的特征分裂方法，本质上是找到使得目标函数降低的越多越好的分裂点。我们首先找到基于每个特征找到收益（收益是指基于当前特征分裂点前后，目标函数的差值）最大的分割点，然后基于所有特征找到收益最大的特征。然后在这所有的循环中，挑出增益最大的那个分割点。对于不同的特征的特征划分点，XGBoost 分别在不同的线程中并行选择分裂的最大增益。

3.2 构建流程

接下来，涉及到决策树模型的构建。假设我们需要预测因子下一期收益的正负（-1, 1 离散因变量），现有自变量包括离散度、估值价差、配对相关性、多空波动率以及宏观趋势指标。自变量和因变量之间的关系并非简单的线性关系。我们为模型提供特征（自变量）和标签（因变量），以便模型学习根据特征对点进行分类。

图表 24：XGB 模型构建流程图



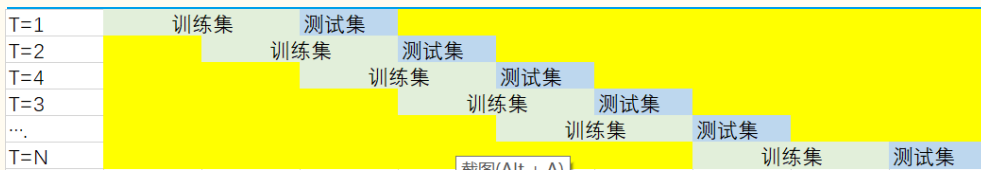
来源：国金证券研究所

3.3 滚动回测

训练集：使用前 40 期~前 10 期的数据

测试集：使用最近 10 期的数据

图表 25：滚动回测示意图



来源：国金证券研究所，wind

滚动回测是一种非常常见的回测方式。我们使用滚动回测窗口为 40 期，在第 T 期实际训练时，训练集为第 T-40~T-10 期，测试集为 T-10~T。通过模型在验证集上的表现来确定模型的 XGBoost 模型参数。XGBoost 模型需要调节的参数主要包括学习率、树的最大深度、子采样率和特征随机采样比例。

3.4 特征选择

梯度增强算法可以为我们提供检索每个属性的重要性得分。单个决策树的重要性是通过每个属性拆分点提高性能指标的数量来计算的，并由节点负责的观察次数来加权。性能度量可以是用于选择拆分点的纯度（基尼系数），也可以是其他更特定的误差函数。然后，将特征重要性在模型中所有决策树之间平均。

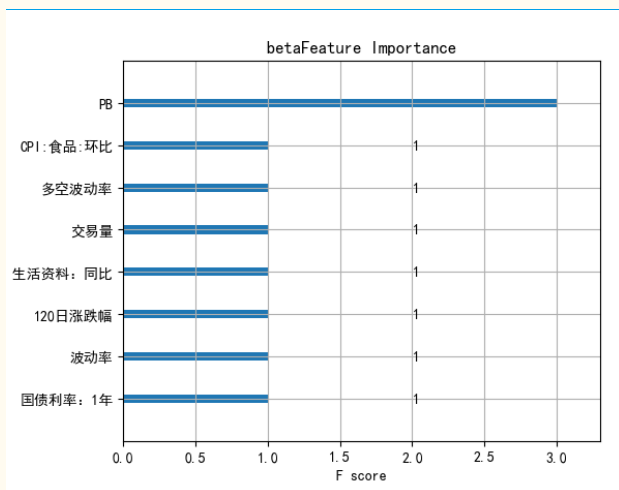
XGBoost 重要性评估方法：

- weight：该特征在所有树中被用作分割样本的特征的总次数。
- gain：该特征在其出现过的所有树中产生的平均增益。
- cover：该特征在其出现过的所有树中的平均覆盖范围。

模型的性能通常随着所选特征的数量而下降，一般为了降低模型的复杂程度我们会针对筛选之后的特征来做预测。但是由于该策略所用的特征总数数量不多（40 个），因此我们不对特征做进一步选择。特征重要性仅供参考。

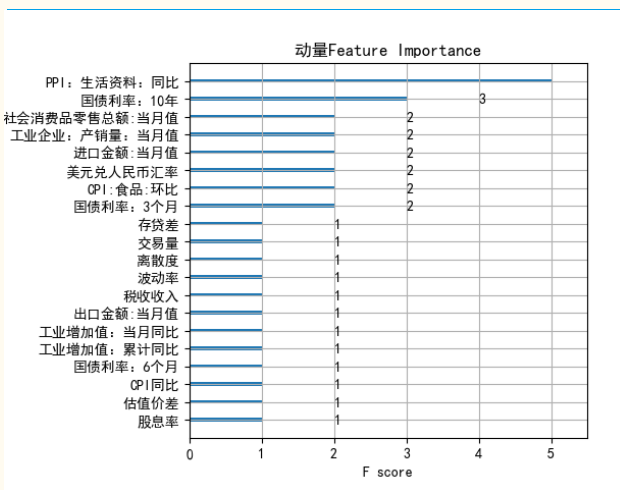
由下图看到不同风格筛选出来的重要性特征不同。动量、低波动、估值因子与宏观经济因素的相关性更大，市场情绪对其影响较小。其余因子的重要性特征中市场情绪、宏观经济占比较为均衡。由于宏观经济属于偏长期的视角，因此模型对于动量、低波动、估值因子的预测方向变化不会很频繁。

图表 26：beta 特征重要性



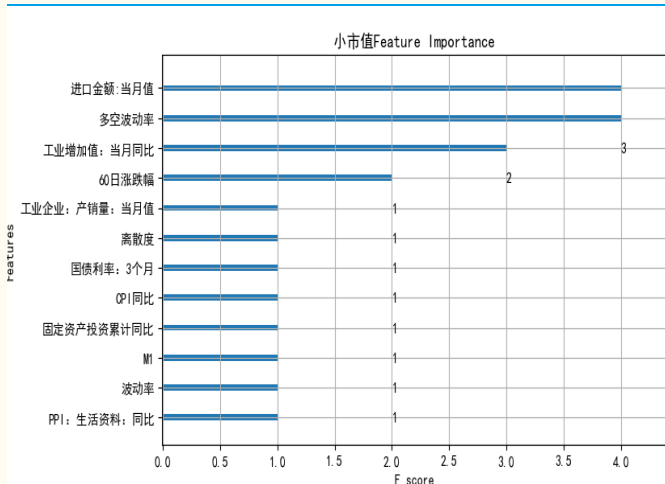
来源：国金证券研究所，wind

图表 27：动量特征重要性



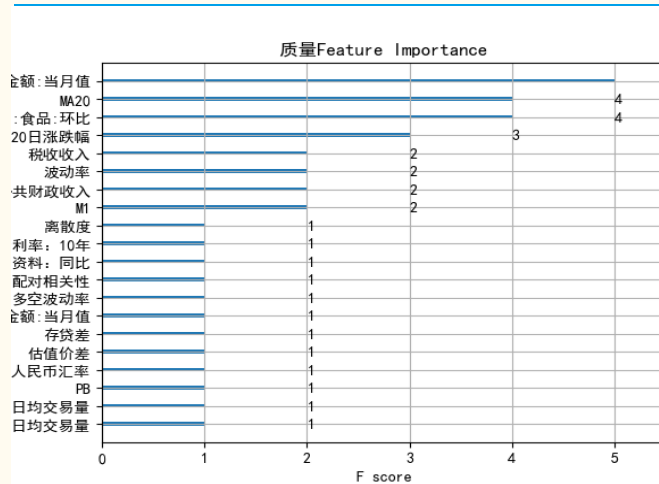
来源：国金证券研究所，wind

图表 28：小市值特征重要性



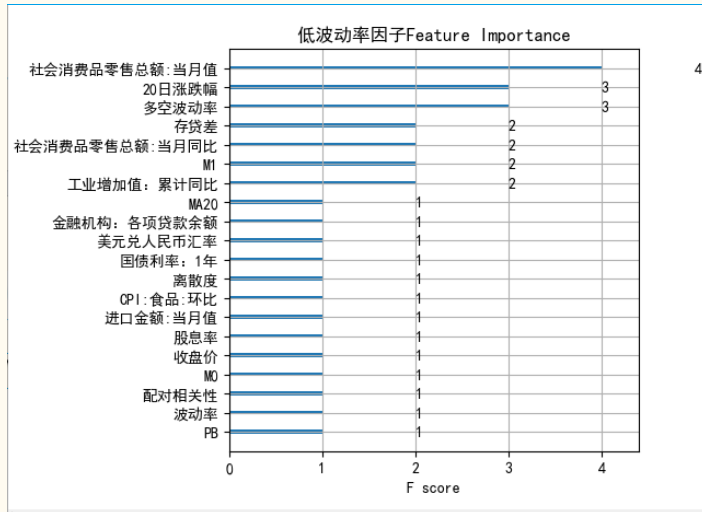
来源：国金证券研究所，wind

图表 29：质量特征重要性



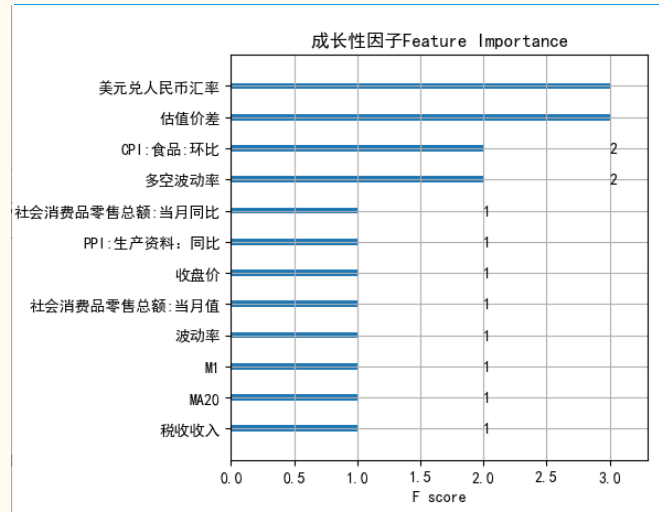
来源：国金证券研究所，wind

图表 30：低波动率特征重要性



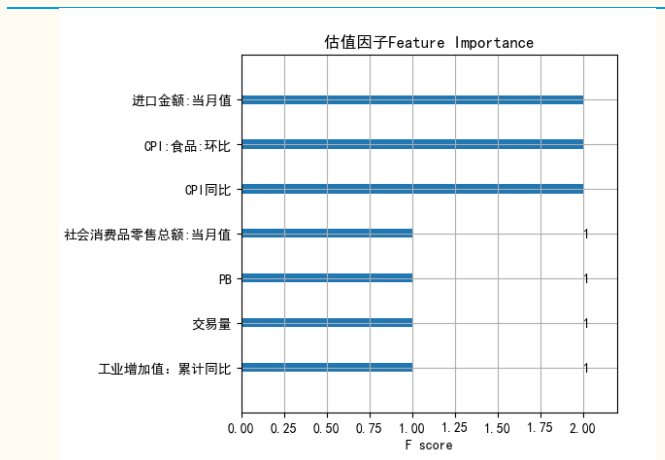
来源：国金证券研究所，wind

图表 31：成长性特征重要性



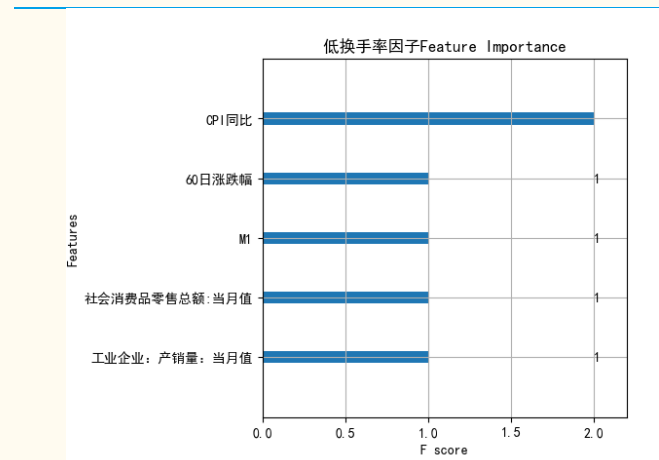
来源：国金证券研究所，wind

图表 32：估值特征重要性



来源：国金证券研究所，wind

图表 33：低换手率特征重要性

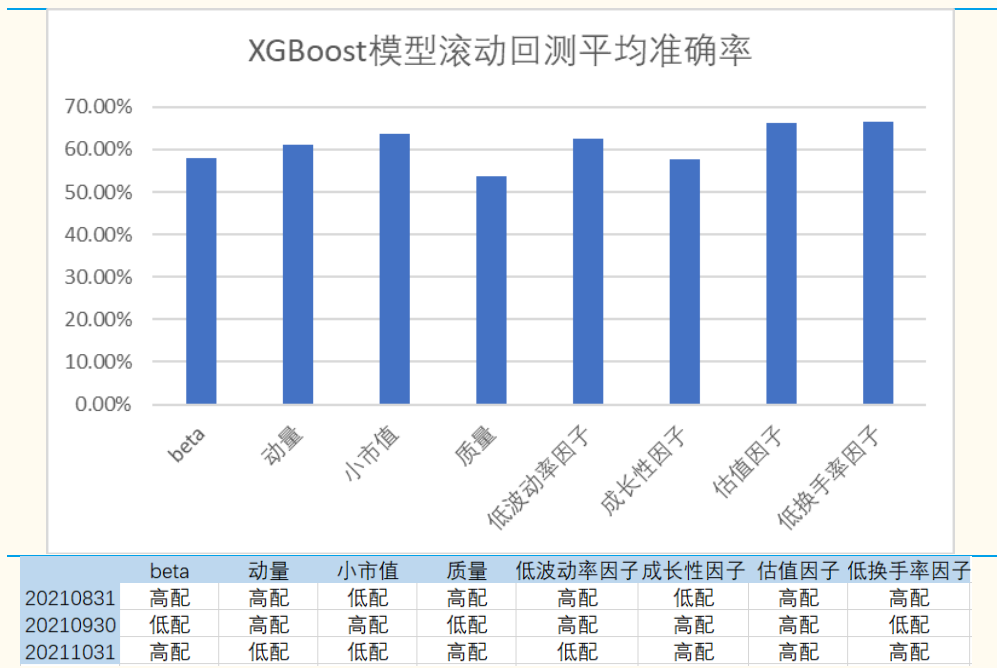


来源：国金证券研究所，wind

当我们确定下最终的模型之后，统计不同回测窗口 XGBoost 模型的准确率，各风格在测试集上的平均准确率有 61%，还是比较稳定的。

模型在 11 月份建议高配 beta、质量、成长、估值、低换手因子，低配动量、小市值、低波动因子。从模型逻辑上看，当前经济周期还没有到全部复苏的阶段，市场流动性较为宽松，但市场主线尚不明确，没有明显的风格偏好，因此模型总体保持风格敞口的相对均衡配置。模型本期相对优选对于利率敏感性较强的成长因子，以及在整个投资环境趋势不明晰下控制质量和估值有一定敞口，另低换手率因子长期来看都值得高配。

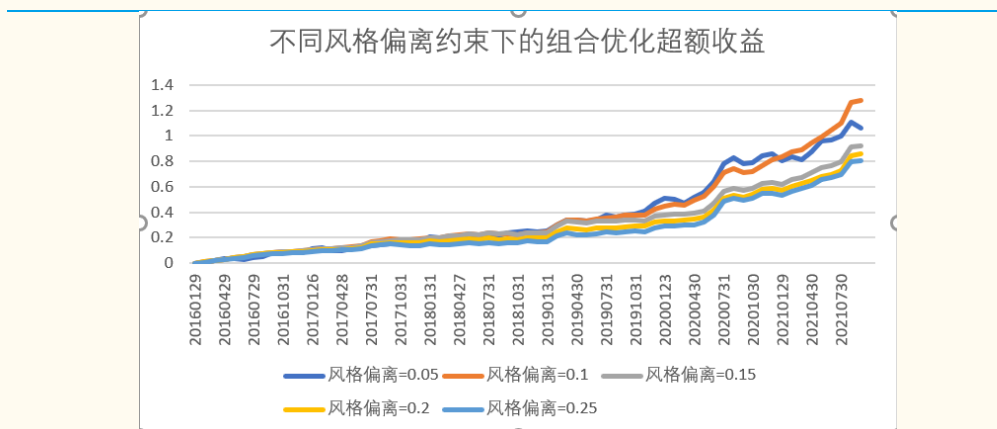
图表 34：滚动回测测试集样本上的平均准确率



来源：国金证券研究所，wind

模型给出每个月确定的风格暴露方向之后，我们通过多因子模型组合优化，以中证 500 为基准进行选股择时。下一期股票的预期收益设定为“因子暴露*模型输出的因子方向（1 或者 -1）”。在风险控制步骤中，我们对风格暴露相对于基准的偏离幅度进行参数敏感性分析，依次设置为[5%,10%,15%,20%,25%]，发现当“风格偏离=10%”的时候超额收益最高。例如当模型输出下期应该看多动量因子时，我们约束组合在动量因子上的暴露为基准在动量因子上的暴露的 1.1 倍，反之为 0.9 倍。

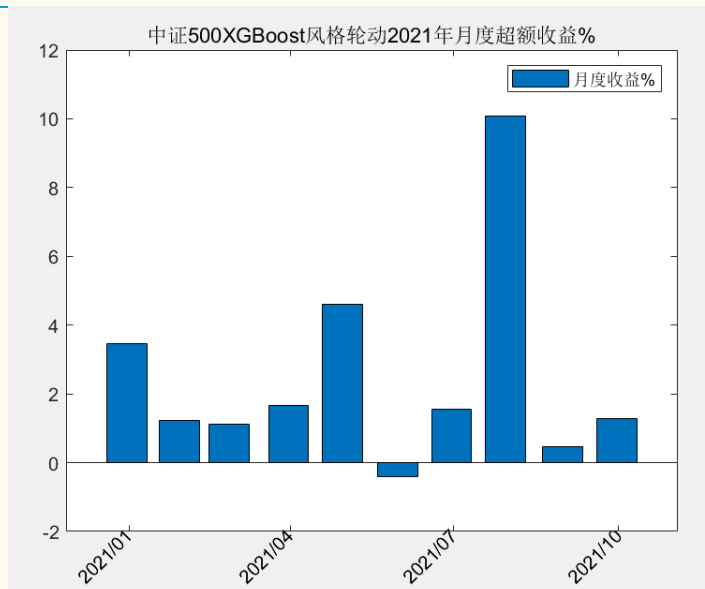
图表 35：2016 年至今不同风格偏离约束下的组合优化超额收益



来源：国金证券研究所，wind

今年以来该模型超额收益稳健，除了 1 月其余各月基本上都跑赢基准，其中在 8 月份的收益尤为突出，主要是因为高配了动量、质量和估值因子。

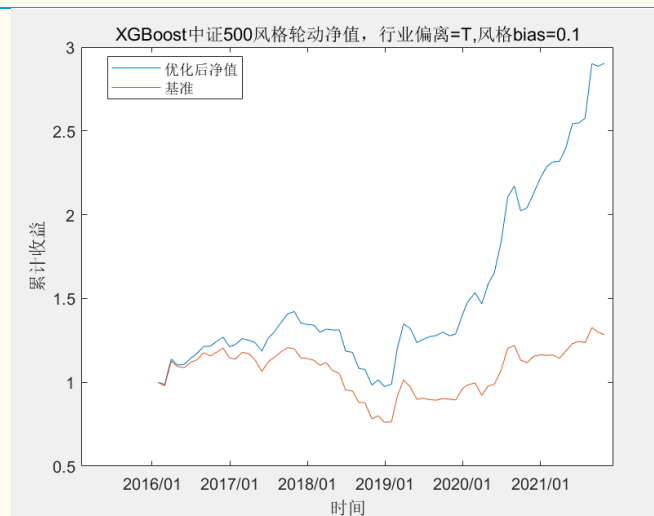
图表 36：轮动模型 2021 年以来的月度超额收益



来源：国金证券研究所，wind

在控制行业偏离和风格偏离的情况下，XGBoost 策略收益年化 32.64%，相对于中证 500 超额收益年化 27.77%。

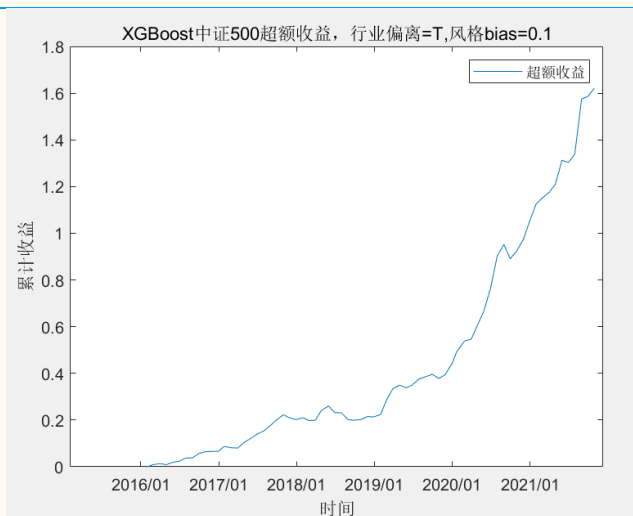
图表 37：XGBoost 模型多头收益



1	2	3	4	5
AnnRet	AnnVol	Sharpe	MDD	Calmar_R
0.3264	0.1845	1.7693	0.3142	1.0389

来源：国金证券研究所，wind

图表 38：XGBoost 模型超额收益



1	2	3	4	5
AnnRet	AnnVol	Sharpe	MDD	Calmar_R
0.2777	0.0734	3.7845	0.0485	5.7245

来源：国金证券研究所，wind

第四部分：利用 XGBoost 模型构建指数轮动策略

对于指数轮动，我们主要讨论沪深 300 以及中证 500 的收益差。我们认为市场风险偏好、宏观经济数据、市场情绪指标，还有指数本身的估值、盈利水平的差异都会影响不同指数的潜在风险收益。宏观数据偏长周期视角，而市场情绪、估值盈利差、风险偏好为短期视角。对于偏长期的宏观经济数据维度和

偏短期的自变量，我们分别构建两套不同的 XGBoost 模型来建立自变量到因变量的映射，并对不同维度的信号进行等权打分，从而可判断当前宏观、市场、资金环境是否更加有利于投资沪深 300 或者中证 500。策略选取业绩比较基准为“50%*沪深 300+50%*中证 500”。

4.1 指数历史走势情况

下图为 2010 年以来两只指数的走势。沪深 300 与中证 500 虽然两者走势相关性较大，但是指数本身的波动有差异，投资者很难依据简单的趋势动量来判断短期持有哪个指数获利更多。

图表 39：2010 年以来中证 500、沪深 300 指数走势



来源：国金证券研究所，wind

在 2017 年前后由于注册制推出导致小盘股溢价消失，市值风格发生了转向。我们利用 2010 年 1 月至 2021 年 10 月的数据进行回测，统计总共 142 个月。在 2017 年以前，中证 500 风格占优比例较高为 52%，沪深 300 占有比例为 47%。2017 年以后，沪深 300 占优的比例达到 56%，中证 500 占优的比例为 44%。下图为 2016 年以来的沪深 300、中证 500 月度收益图。自 2020 年四季度以来的抱团行情导致沪深 300 持续占优，今年春节之后的疫情影响带来新能源、上游周期股的机会导致中证 500 持续占优。投资者对于大小盘的偏好呈现出周期性的转变，这个转变的拐点与宏观经济、市场情绪以及指数本身估值、收益预期关系比较紧密。

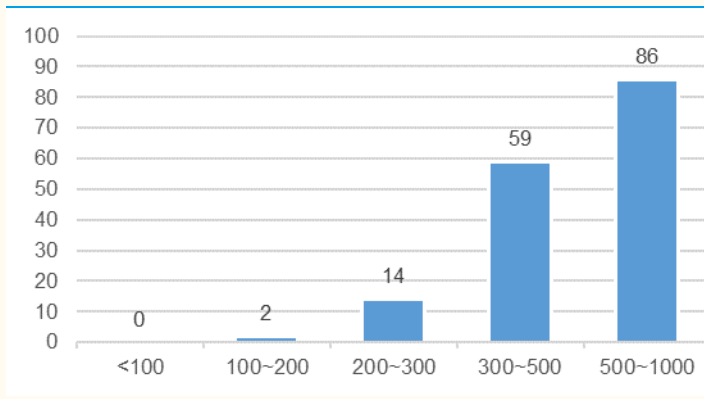
图表 40：沪深 300 与中证 500 月度收益

日期	中证500	沪深300	中证500	沪深300	中证500	沪深300
20160129	-28.20	-21.04	20180131	-0.98	20200123	2.09
20160229	-2.17	-2.33	20180228	-2.68	20200228	1.37
20160331	15.05	11.84	20180330	1.51	20200331	-7.52
20160429	-2.77	-1.91	20180427	-4.15	20200430	6.20
20160531	-0.64	0.41	20180531	-1.82	20200529	0.98
20160630	2.96	-0.49	20180629	-9.33	20200630	8.47
20160729	1.30	1.59	20180731	-0.56	20200731	12.20
20160831	3.73	3.87	20180831	-7.21	20200831	1.35
20160930	-1.65	-2.24	20180928	-0.29	20200930	-7.14
20161031	1.96	2.55	20181031	-11.00	20201030	-1.31
20161130	2.07	6.05	20181130	2.44	20201130	3.43
20161230	-4.89	-6.44	20181228	-4.77	20201231	0.73
20170126	-0.64	2.35	20190131	0.20	20210129	-0.33
20170228	3.65	1.91	20190228	20.32	20210226	0.29
20170331	-0.76	0.09	20190329	10.39	20210331	-1.73
20170428	-2.96	-0.47	20190430	-4.33	20210430	3.70
20170531	-6.24	1.54	20190531	-7.45	20210531	3.75
20170630	5.39	4.98	20190628	0.78	20210630	1.18
20170731	2.60	1.94	20190731	-0.95	20210730	-0.60
20170831	2.74	2.25	20190830	-0.34	20210831	7.21
20170929	2.05	0.38	20190930	1.11	20210930	1.26
20171031	-0.65	4.44	20191031	-0.48	20211029	-1.14
20171130	-4.52	-0.02	20191129	-0.46		
20171229	-0.20	0.62	20191231	7.61		

来源：国金证券研究所，wind

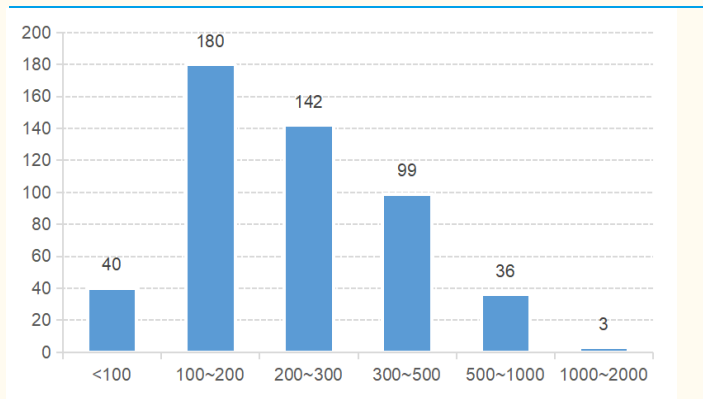
沪深 300 和中证 500 在市值、估值、行业分布、盈利预期上都有差异。从行业分布上看, 沪深 300 以消费金融医药为主, 而中证 500 以新能源电子周期为主, 受到宏观因素的影响不同。由于指数的盈利预期和估值分布的差异, 投资者不同的风险偏好也会导致对两者投资性价比判断分化。

图表 41: 沪深 300 指数成分股市值分布



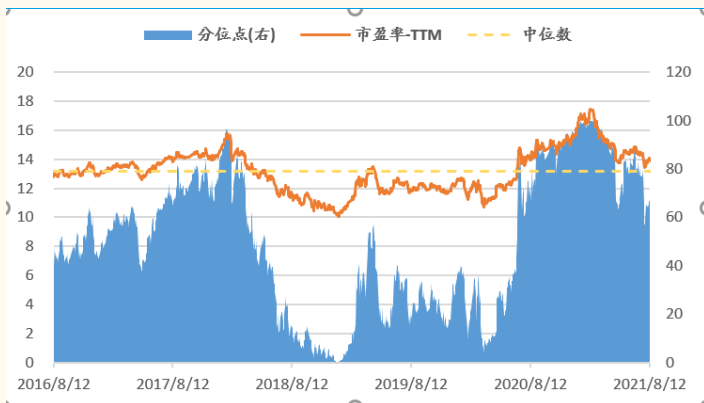
来源: 国金证券研究所, wind

图表 42: 中证 500 指数成分股市值分布



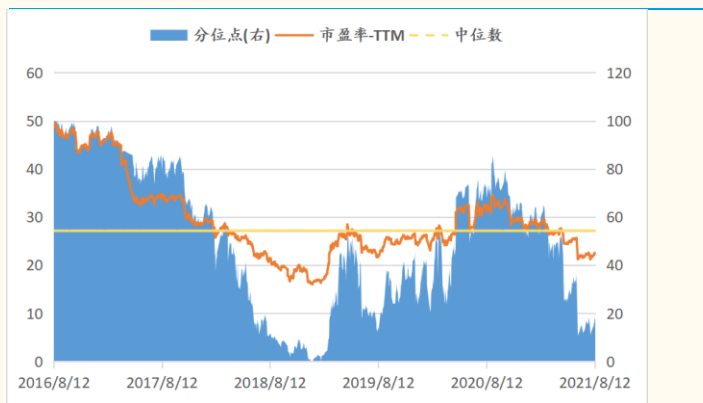
来源: 国金证券研究所, wind

图表 43: 沪深 300 指数 PE 水平



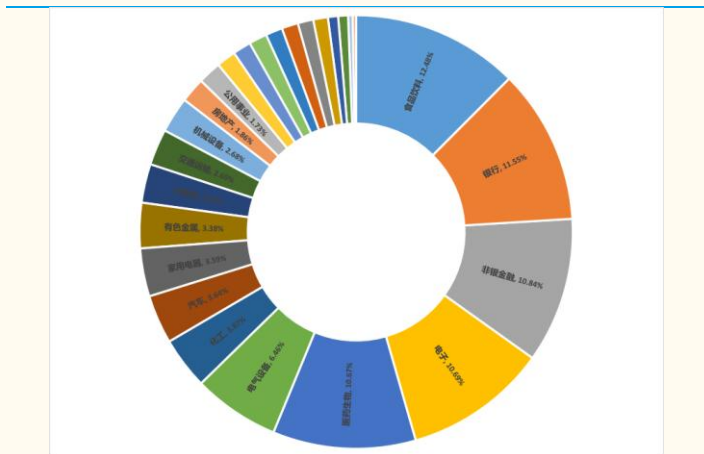
来源: 国金证券研究所, wind

图表 44: 中证 500 指数 PE 水平



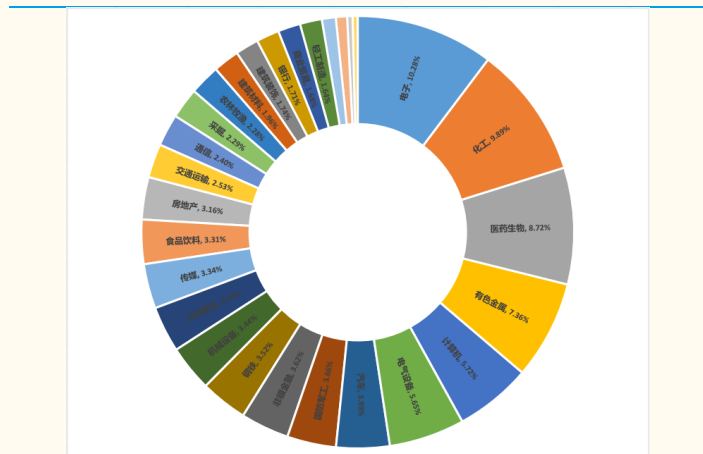
来源: 国金证券研究所, wind

图表 45: 沪深 300 指数行业权重分布



来源: 国金证券研究所, wind

图表 46: 中证 500 指数行业权重分布

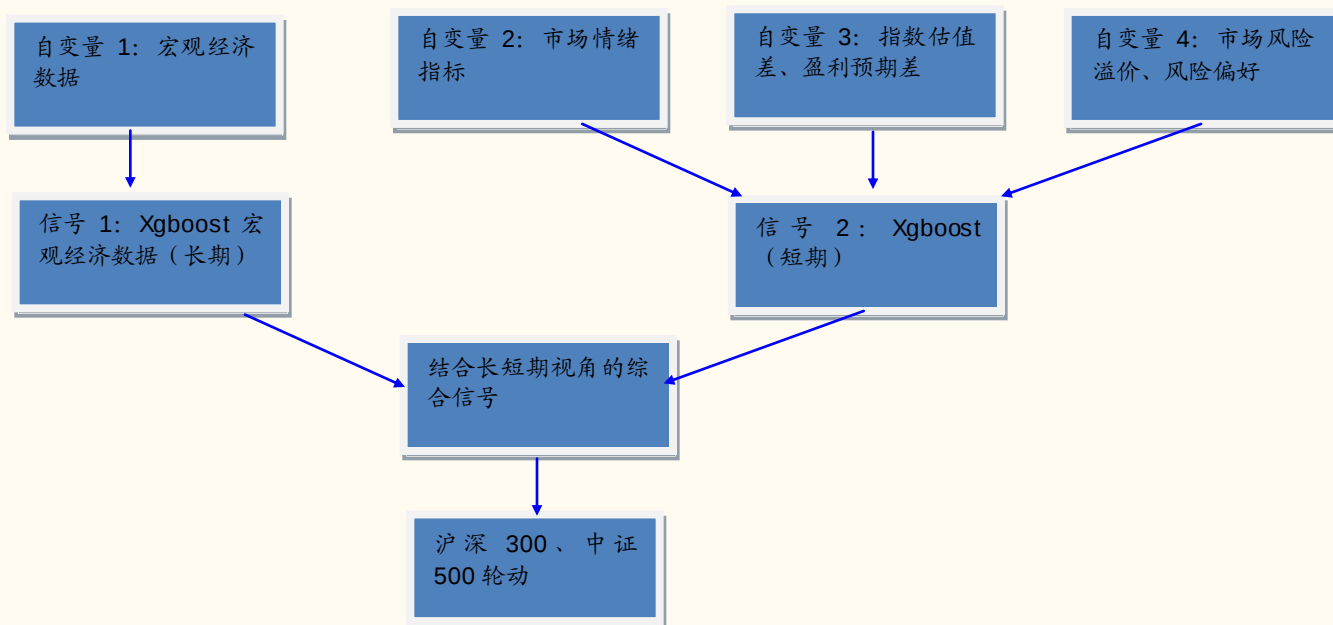


来源: 国金证券研究所, wind

4.2 指数轮动策略回测

在 A 股中构建大小盘轮动风格具有重要价值。我们将不同维度的自变量区分开来，偏长期维度的信号以宏观经济数据为自变量，偏短期维度的信号以市场情绪、指数估值盈利差、风险溢价等因素为自变量，分别构建 XGBoost 模型，最终等权得到综合信号，帮助投资者进行多维度的判断。

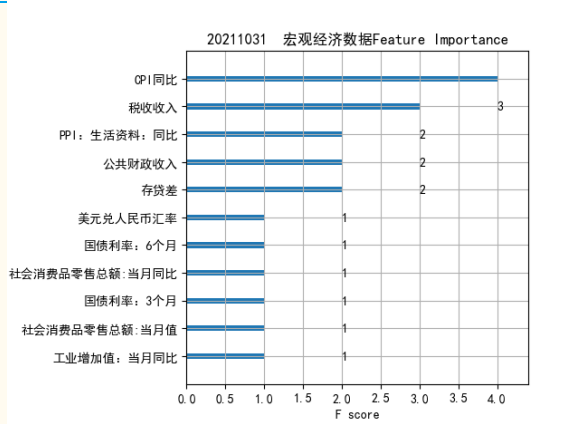
图表 47：指数轮动模型构架搭建流程



来源：国金证券研究所，wind

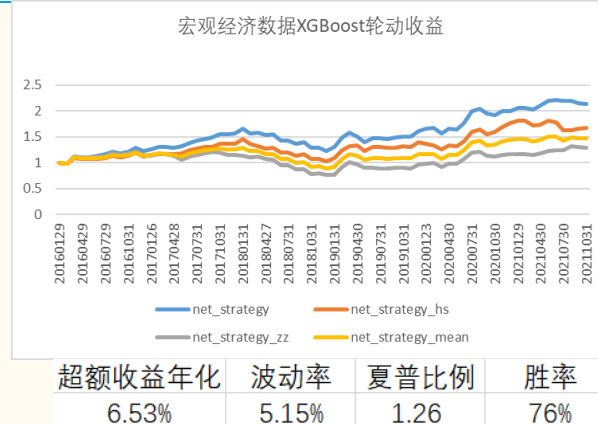
由宏观经济数据为自变量构建的 XGBoost 模型的胜率分别为 76%、超额收益年化 6.53%，夏普比率 1.26，其中 CPI 同比、税收收入、PPI、财政收入同比等因素为重要性靠前的特征。由市场情绪自变量构建的 XGBoost 模型胜率位 72%，夏普比率 1.5，超额年化收益 7.52%。其中指数估值差 diff、中证全指 120 日涨跌幅，沪深 300 与中证 500 盈利预期差、PB 等因素为重要性靠前的特征。

图表 48：XGBoost 模型特征重要性（宏观经济数据）



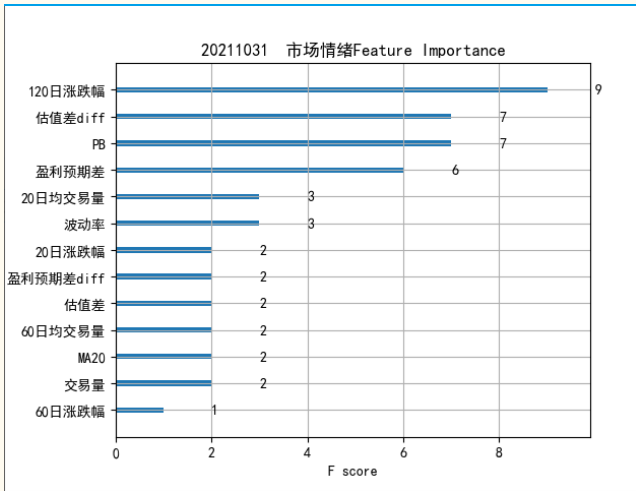
来源：国金证券研究所，wind

图表 49：XGBoost 信号 1 有效性检验（宏观经济数据）



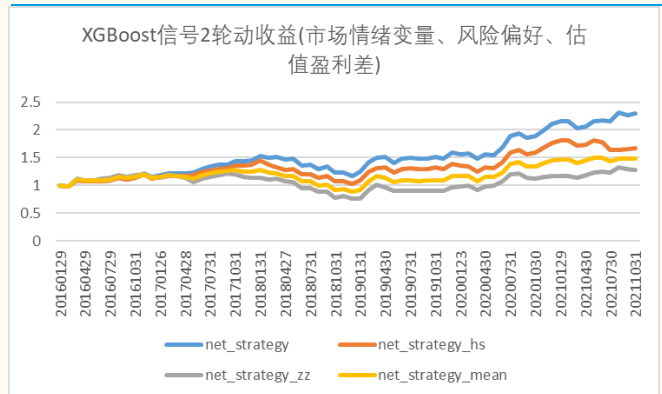
来源：国金证券研究所，wind

图表 50: XGBoost 模型特征重要性 (市场情绪变量)



来源: 国金证券研究所, wind

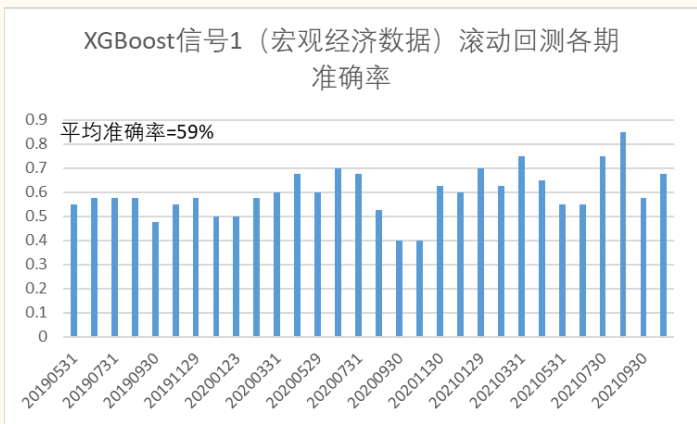
图表 51: XGBoost 信号 2 有效性检验 (市场情绪变量)



超额收益年化	波动率	夏普比例	胜率	滚动平均准确率
7.52%	5.01%	1.50	72.86%	54.33%

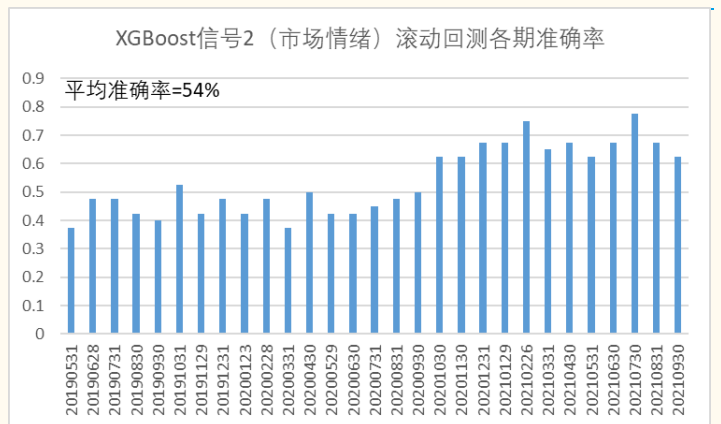
来源: 国金证券研究所, wind

图表 52: XGBoost 信号 1 滚动回测各期准确率



来源: 国金证券研究所, wind

图表 53: XGBoost 信号 2 滚动回测各期准确率



来源: 国金证券研究所, wind

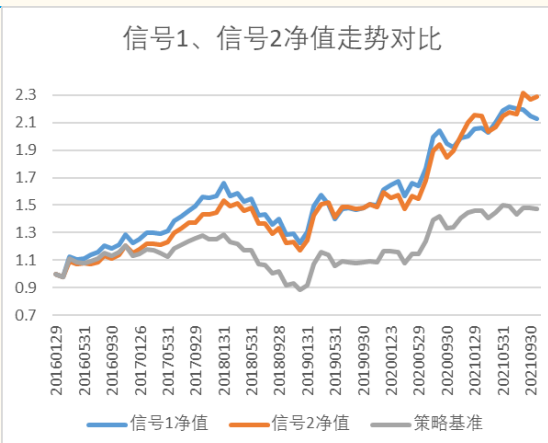
当前 11 月份两个信号的方向有分歧, 市场流动性偏宽松但是经济还未完全复苏。以宏观数据为基础的信号看多中证 500, 而以偏短期视角为基础的信号看多沪深 300。

图表 54: XGBoost 模型信号, 1 表示沪深 300, 0 表示中证 500

	XGBoost 信号 1: 宏观经济数 据	XGBoost 信号 2: 市场情绪、 盈利估值差
20201030	0	1
20201130	0	1
20201231	1	1
20210129	0	1
20210226	0	1
20210331	0	1
20210430	1	1
20210531	0	0
20210630	0	0
20210730	1	0
20210831	0	0
20210930	0	1
20211031	0	1

来源: 国金证券研究所, wind

图表 55：不同信号走势对比（2016.01~2021.10）

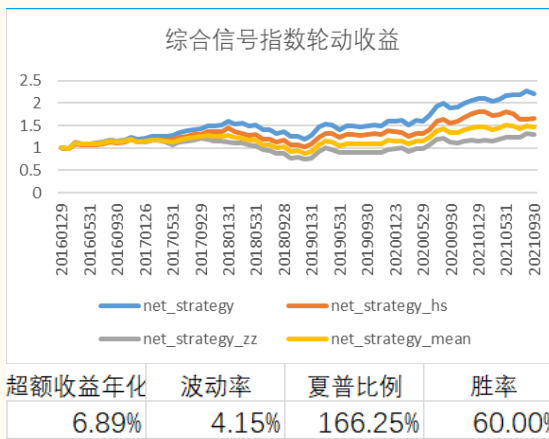


来源：国金证券研究所，wind

从历史回测来看，我们的训练期数总共 70 期，其中两个模型给出一致观点有 49 期数，一致性比例有 71%。当两个模型信号一致的时候，其信号准确率有 85%。从今年行情来看，以市场情绪、指数估值差、盈利差为信号的 XGBoost 模型在近一年（滚动样本外）给出的信号胜率都很高，而以宏观数据为信号的 XGBoost 模型在近一年给出的信号准确率降低。

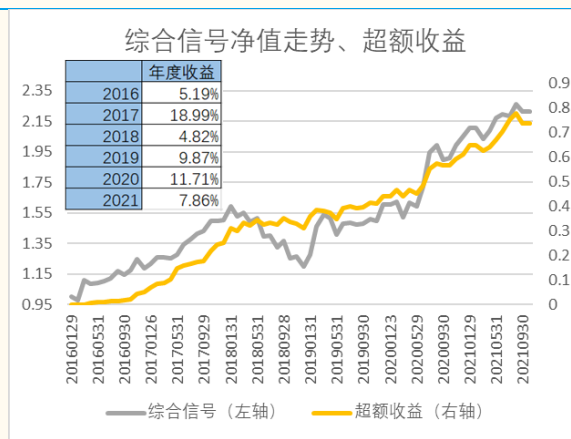
我们将两个信号进行等权得到综合信号。当两个信号方向不一致时，模型等权配置沪深 300 与中证 500。与单信号相比，综合信号虽然超额收益还是较之前有所降低，但是熨了波动，策略夏普比率提高到 1.66。模型在大市值环境下收益都不错，例如 2017 年和 2020 年。在 2018 年权益指数下跌的过程中超额收益不明显，波动较大，超额收益的最大回撤发生在 2018 年底。

图表 56：综合信号指数轮动收益（2016.01~2021.10）



来源：国金证券研究所，wind

图表 57：综合信号走势对比（2016.01~2021.10）



来源：国金证券研究所，wind

结论： 本文通过机器学习的算法分别构建了因子择时模型以及指数轮动模型，不同模型所用到的自变量大致都结合宏观经济指标、市场情绪指标、因子拥挤度、指数估值盈利预期差等因素。风格轮动模型通过组合优化的方法在控制行业中性风格和风格偏离的情况下超额年化收益 27.77%，夏普比率 3.78。策略今年以来取得稳定收益。指数轮动模型中的两个信号都能获取 70%以上的胜率，以市场情绪为信号的策略在今年以来基本预测到了沪深 300、中证 500 切换的拐点。信号融合之后模型的波动进一步减小，综合信号在 2018 年和 2020 年表现优异。后续我们会继续跟踪以上两个策略并将其推广至行业轮动模型中。

风险提示： 模型测试基于历史数据，市场未来可能发生变化。

特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证，对由于该等问题产生的一切责任，国金证券不作出任何担保。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整。

本报告中的信息、意见等均仅供参考，不作为或被视为出售及购买证券或其他投资标的邀请或要约。客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告反映编写分析员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，且收件人亦不会因为收到本报告而成为国金证券的客户。

根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；非国金证券 C3 级以上（含 C3 级）的投资者擅自使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

此报告仅限于中国大陆使用。

上海

电话：021-60753903

传真：021-61038200

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路 1088 号

紫竹国际大厦 7 楼

北京

电话：010-66216979

传真：010-66216793

邮箱：researchbj@gjzq.com.cn

邮编：100053

地址：中国北京西城区长椿街 3 号 4 层

深圳

电话：0755-83831378

传真：0755-83830558

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：中国深圳福田区深南大道 4001 号

时代金融中心 7GH