

机器学习模拟投资者分歧

华泰研究

2024年6月15日 | 中国内地

深度研究

人工智能系列之 80: 机器学习用于模拟投资者分歧

投资者分歧是推动交易的原因之一,对金融市场的资产定价具有重要意义。本文运用机器学习来模拟投资者分歧。将不同的特征集输入给多个机器学习模型,模拟投资群体接收差异化信息源、并产生不同投资预测的过程。最后,根据模型预测差异构建的分歧度因子,能够有效刻画投资者分歧,对股票未来收益具有显著的预测作用。

投资者分歧度越高,股票预期收益越低

Miller (1977) 认为投资者分歧会造成股票价格过高,未来回报下降,原因是乐观投资者推高股价,而悲观投资者受到做空限制,对市场的影响力更弱。本文利用现有的多个深度学习因子,构建深度学习分歧度因子,验证了投资者分歧与股票未来收益的负相关性。样本空间为全 A 股,因子在 2017/1/4~2024/5/31 的回测期内周度 RankIC 均值为 6.06%,分 5 层 TOP 组合年化超额收益率为 9.75%,周频双边换手率为 117.70%。

树模型能够高效模拟投资者分歧

本文使用 50 个 LightGBM 模型来模拟投资者的预测观点。输入特征为股票截面特征,包含估值、成长、财务质量、量价、一致预期等类别,共 43 个。每个模型会随机选取 20 个特征进行训练,以模仿投资者接收差异化信息源。最后,计算所有模型对同一交易日同只股票的预测值的标准差,构建机器学习分歧度因子。因子在相同回测期内周度 RankIC 均值提升至 7.66%,分 5 层 TOP 组合年化超额收益率为 9.71%,周频双边换手率下降至 61.97%。

超参数选择对分歧度因子的影响基本不大

不同预测周期构建的分歧度因子回测表现接近,但预测周期越长, TOP 组合换手率越低,因为长线投资者的预测分歧在短期内变化不会太剧烈。不同投资者数量构建的分歧度因子回测表现差异不大,但投资者数量设置过少可能会造成 RankIC 偏低和换手率偏高。特征数量对分歧度因子的回测表现具有一定影响,随机选取的特征数量过多会导致投资者之间难以产生预测分歧。LightGBM、XGBoost 和 CatBoost 等不同机器学习模型构建的分歧度因子回测表现接近。

股票做空限制、动量和投资者认可度对分歧度因子具有显著影响

使用机构持股比例衡量个股的做空难易程度,机构持股比例低的股票池中分歧度因子多空表现更加显著,多空组合年化收益率从 32.89% 提升至 54.05%,说明做空限制是分歧度因子的坚实基础。近期涨幅最大的股票池中分歧度因子的多空组合年化收益率高达 63.04%,表明乐观投资者占据主导地位,观点分歧造成的买卖不对称性进一步增强。根据模型预测均值构建投资者认可度因子,发现一致看多的股票组合预期收益最高,“争议看空”的股票组合预期收益最低,分歧度与认可度因子等权重合成的复合因子 RankIC 和多头收益均有提升。

风险提示:借助机器学习构建的选股策略是历史经验的总结,存在失效的可能。机器学习存在过拟合的风险。本文回测暂未考虑交易费用。

研究员

SAC No. S0570516010001
SFC No. BPY421

林晓明

linxiaoming@htsc.com
+(86) 755 8208 0134

研究员

SAC No. S0570520080004
SFC No. BRB318

何康, PhD

hegang@htsc.com
+(86) 21 2897 0339

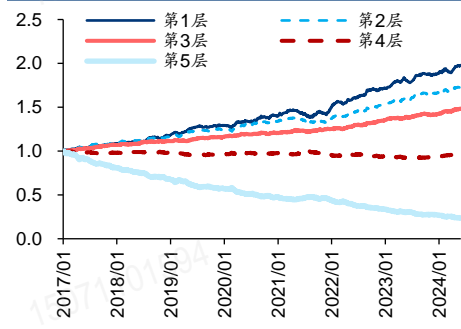
联系人

SAC No. S0570123070272

卢炯

lujiang@htsc.com
+(86) 755 8249 2388

机器学习分歧度因子分层相对净值



资料来源: Wind, 华泰研究

正文目录

投资者分歧与资产定价	4
机器学习模拟投资者分歧	4
基于机器学习的分歧度因子	6
初步探索	6
数据与方法	7
分歧度因子回测	8
参数敏感性分析	8
预测周期	8
投资者数量	9
特征数量	9
机器学习模型	10
相关性分析	10
影响因素分析	11
做空限制	11
动量	12
投资者认可度	13
总结	15
风险提示	15
参考文献	16

图表目录

图表 1: 机器学习模拟投资者分歧	5
图表 2: 机器学习分歧度因子 IC 值分析和分层回测结果汇总	5
图表 3: 机器学习分歧度因子分层组合相对净值	5
图表 4: 机器学习分歧度因子累计 RankIC	5
图表 5: 全频段量价模型体系	6
图表 6: 深度学习分歧度因子 IC 值分析和分层回测结果汇总	6
图表 7: 深度学习分歧度因子分层组合相对净值	7
图表 8: 深度学习分歧度因子累计 RankIC	7
图表 9: 机器学习模拟投资者分歧	7
图表 10: 机器学习分歧度因子 IC 值分析和分层回测结果汇总	8
图表 11: 机器学习分歧度因子分层组合相对净值	8
图表 12: 机器学习分歧度因子累计 RankIC	8
图表 13: 不同预测周期的分歧度因子 IC 值分析和分层回测结果汇总	8
图表 14: 不同预测周期的分歧度因子 TOP 组合相对净值	9
图表 15: 不同预测周期的分歧度因子累计 RankIC	9
图表 16: 不同投资者数量的分歧度因子 IC 值分析和分层回测结果汇总	9

图表 17: 不同投资者数量的分歧度因子 TOP 组合相对净值	9
图表 18: 不同投资者数量的分歧度因子累计 RankIC	9
图表 19: 不同投资者数量的分歧度因子 IC 值分析和分层回测结果汇总	9
图表 20: 不同特征数量的分歧度因子 TOP 组合相对净值	10
图表 21: 不同特征数量的分歧度因子累计 RankIC	10
图表 22: 不同机器学习模型的分歧度因子 IC 值分析和分层回测结果汇总	10
图表 23: 不同机器学习模型的分歧度因子 TOP 组合相对净值	10
图表 24: 不同机器学习模型的分歧度因子累计 RankIC	10
图表 25: 机器学习分歧度因子与常见量价因子的相关性	10
图表 26: 机器学习分歧度因子、换手率、特质波动率的多头超额净值对比	11
图表 27: 按照分歧度和机构持股比例分组的组合年化收益率	11
图表 28: 按照机构持股比例分组的分歧度因子 IC 值分析和分层回测结果汇总	11
图表 29: 按照机构持股比例分组的分歧度因子多空净值	12
图表 30: 按照机构持股比例分组的分歧度因子累计 RankIC	12
图表 31: 按照分歧度和近期收益分组的组合年化收益率	12
图表 32: 按照近期收益分组的分歧度因子 IC 值分析和分层回测结果汇总	12
图表 33: 按照近期收益分组的分歧度因子多空净值	13
图表 34: 按照近期收益分组的分歧度因子累计 RankIC	13
图表 35: 认可度因子 IC 值分析和分层回测结果汇总	13
图表 36: 认可度因子分层组合相对净值	13
图表 37: 认可度因子累计 RankIC	13
图表 38: 按照分歧度和认可度分组的组合年化收益率	14
图表 39: 按照认可度分组的分歧度因子 IC 值分析和分层回测结果汇总	14
图表 40: 按照认可度分组的分歧度因子多空净值	14
图表 41: 按照认可度分组的分歧度因子累计 RankIC	14
图表 42: 复合因子 IC 值分析和分层回测结果汇总	14
图表 43: 复合因子 TOP 组合相对净值	14
图表 44: 复合因子累计 RankIC	14

投资者分歧与资产定价

投资者分歧是推动交易的原因之一，对金融市场的资产定价具有重要意义，以往很多论文对投资者分歧开展了深入研究。Miller（1977）认为投资者分歧会造成股票价格过高，未来回报下降，原因是乐观投资者推高股价，而悲观投资者受到做空限制，对市场的影响力更弱。Diether 等（2002）发现分析师盈利预测的分歧也会导致更低的股票未来收益。

由于信息来源不同和行为偏差等原因，投资者会产生对股价预期的分歧。但投资者分歧并不容易直接观测，Berkman 等（2009）提出一些代理指标来间接反映这种分歧：

（1）历史盈利波动性。如果一家公司过去几年盈利的波动性很大，说明这家公司未来的盈利更难预测，投资者对公司的估值更容易产生分歧。这种方法的缺点是滞后性，无法反映近期市场分歧。

（2）公司年龄。年龄越大的公司具有更长的经营数据，而且一般已经进入成熟期，发展前景的不确定性更强，因此投资者分歧更低。这种方法同样难以刻画近期市场观点。

（3）换手率。股票换手率越大，说明股票交易活跃，买卖双方的预期具有更大分歧，尽管股票交易并不完全是由分歧造成的。

（4）特质波动率。特质波动率衡量股票的非系统性风险和不确定性，不确定性会带来投资者分歧。

（5）分析师预测标准差。卖方分析师盈利预测的标准差能够反映不同分析师预期的分歧，不过分析师观点不一定会形成交易行为，而且股票覆盖度有限。

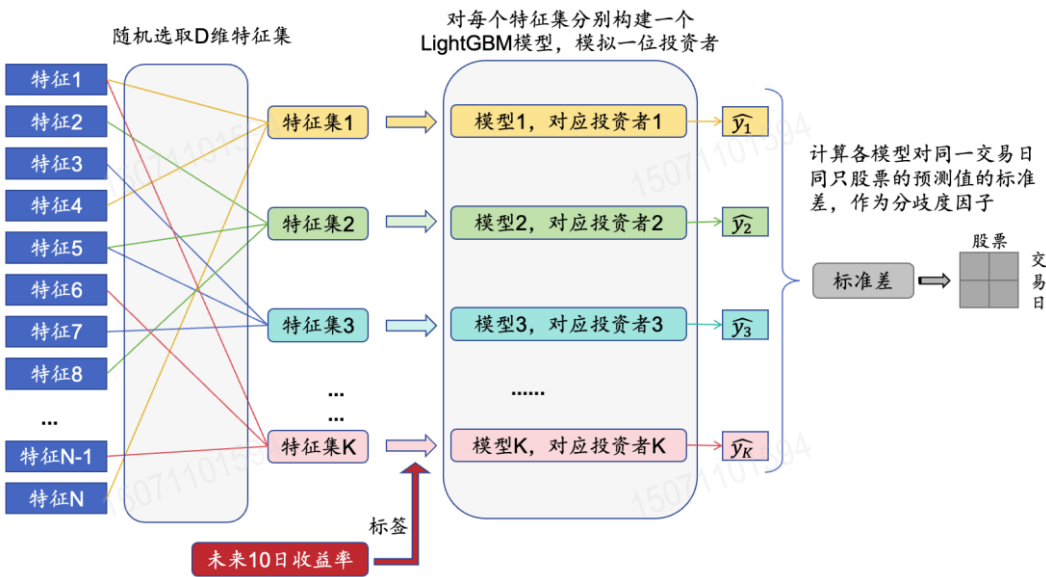
机器学习模拟投资者分歧

Bali 等人于 2023 年在 NBER 上发表论文《Machine Forecast Disagreement》，旨在研究金融市场中的投资者信念分歧及其对股票未来收益的影响。论文提出一种新的投资者信念分歧的测量方法，利用不同机器学习模型来模拟投资者的异质性预测，称之为“机器预测分歧”（Machine Forecast Disagreement, MFD）。论文的主要贡献包括：提出了一种新的资产层面信念分歧测量方法；证明了 MFD 在解释个股截面收益方面的强大预测能力；探讨了 MFD alpha 的经济学基础，特别是卖空限制条件的影响。

论文使用来自 Jensen 等（2022）提供的数据集，包括股票回报和 153 个特征信息。样本涵盖 1966 年 7 月至 2022 年 12 月期间在 NYSE、AMEX 和 NASDAQ 交易的普通股。为了避免小盘和低流动股票的影响，剔除价格低于 5 美元的股票。研究采用 10 年滚动窗口估计随机森林回归模型，每个投资者被赋予一个特定的模型和特征子集。最后通过计算每只股票在每个月的投资者预测分歧的标准差来构建 MFD。超参数设定包括：投资者数量为 100，每个投资者的特征数量为 76，决策树的最大深度为 3，决策树的数量为 2000，样本采样率为 0.05，列采样率为 0.05。

论文发现，MFD 能够显著预测个股的未来收益，高 MFD 的股票未来收益显著低于低 MFD 的股票。按 MFD 分组的最高和最低十分位数股票的市值加权（等权）组合，其月度回报差异为 1.14%（1.32%），具有很高的统计显著性。此外，MFD 与分析师预测分歧（AFD）相比，表现出更强的预测能力和更高的数据覆盖率。进一步的分析表明，MFD alpha 的产生与股票卖空成本和套利限制有关，在高卖空成本和高散户持有率的股票中，MFD alpha 更为显著。论文还发现，MFD 在股票盈利公告期间的预测能力更为明显，表明 MFD 捕捉到了由投资者信念分歧引起的股票价格错配。综上所述，该论文不仅提出了一种新的信念分歧测量方法，还揭示了其在股票定价中的重要作用。

图表1：机器学习模拟投资者分歧



资料来源：华泰研究

本篇报告借鉴 Bali 等（2023）的研究，运用机器学习来模拟投资者分歧。将不同的特征集输入给多个机器学习模型，模拟投资群体接收差异化信息源、并产生不同投资预测的过程。最后，根据模型预测差异构建的分歧度因子，能够有效刻画投资者分歧，对股票未来收益具有显著的预测作用。

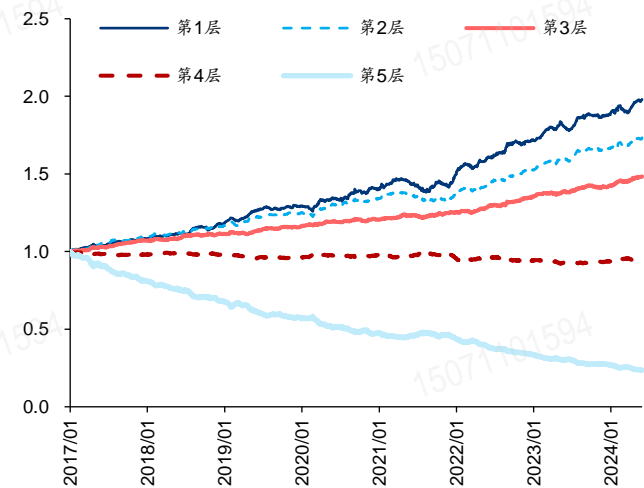
图表2：机器学习分歧度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	TOP 组合年化超额收益率	TOP 组合信息比率	TOP 组合胜率	TOP 组合换手率
机器学习分歧度因子	7.66%	10.18%	0.75	77.78%	9.71%	2.46	68.54%	61.97%

注：回测期为 2017/1/4~2024/5/31，换手率为周频单次调仓双边换手率

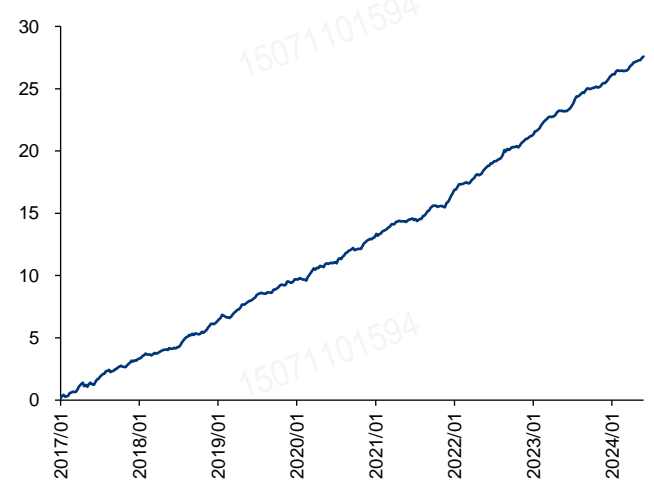
资料来源：Wind，华泰研究

图表3：机器学习分歧度因子分层组合相对净值



资料来源：Wind，华泰研究

图表4：机器学习分歧度因子累计 RankIC



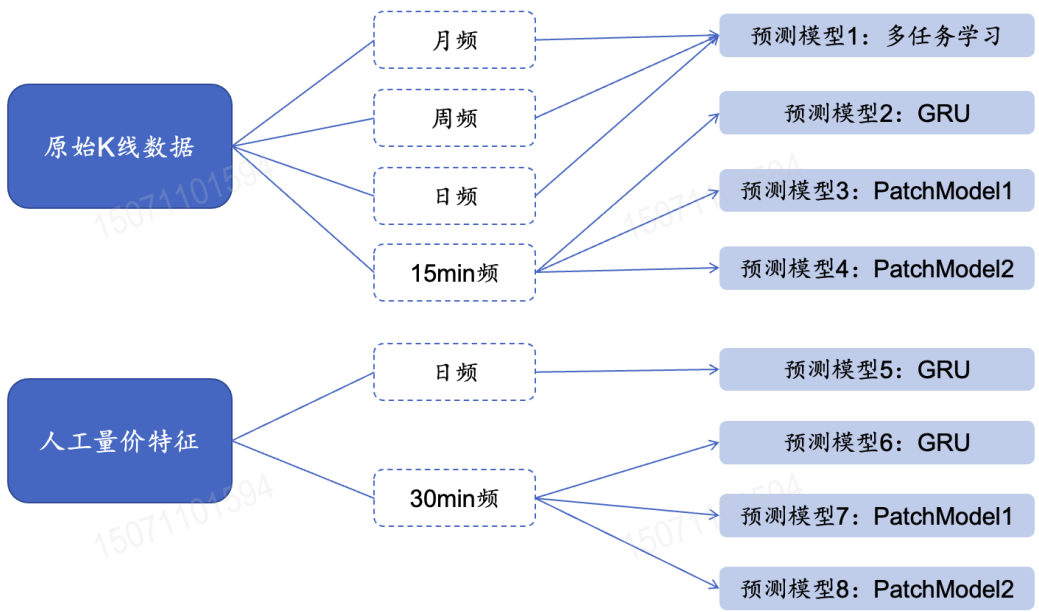
资料来源：Wind，华泰研究

基于机器学习的分歧度因子

初步探索

华泰金工前期报告《基于全频段量价特征的选股模型》(2023.12.8)和《如何捕捉长时间序列量价数据的规律》(2024.3.14),利用日k线、周k线、月k线、分钟k线、逐笔成交和逐笔委托等数据,训练多个深度学习模型。这些模型对股票收益具有较强的预测能力,且相关性不高,能够初步模拟投资者分歧。

图表5: 全频段量价模型体系



资料来源: 华泰研究

通过计算深度学习模型间的标准差,构建深度学习分歧度因子,并进行单因子测试。投资者分歧度越高,未来预期收益越低,因此在回测前先进行方向调整(下同)。单因子测试方法如下:

1. 股票池: 全A股,剔除ST股票,剔除每个截面期下一交易日停牌、涨停的股票。
2. 回测区间: 2017/1/4~2024/5/31。
3. 调仓周期: 周频,不计交易费用。
4. 因子预处理: 去极值、行业市值中性化、标准化。
5. 测试方法: IC值分析,因子分5层测试。

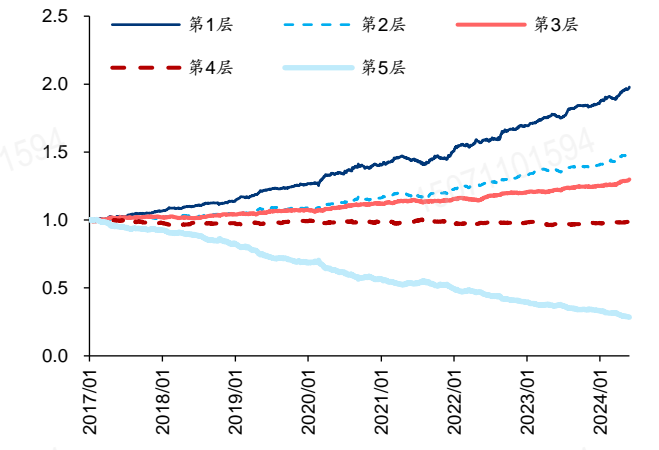
测试结果如下所示。方向调整后的深度学习分歧度因子在分层回测中呈现了较好的单调性,第一组(分歧度最低的20%)收益明显高于其他组,第五组(分歧度最高的20%)收益明显低于其他组。

图表6: 深度学习分歧度因子IC值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	TOP 组合年化超额收益率	TOP 组合信息比率	TOP 组合胜率	TOP 组合换手率
深度学习分歧度因子	6.06%	7.13%	0.85	81.94%	9.75%	3.03	83.15%	117.70%

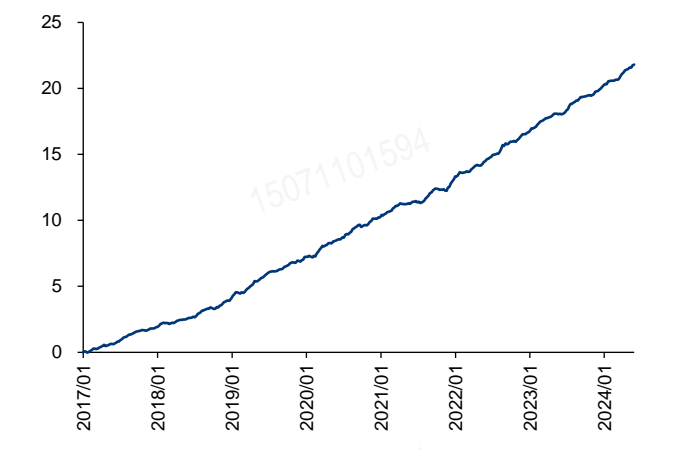
资料来源: Wind, 华泰研究

图表7：深度学习分歧度因子分层组合相对净值



资料来源：Wind，华泰研究

图表8：深度学习分歧度因子累计 RankIC



资料来源：Wind，华泰研究

数据与方法

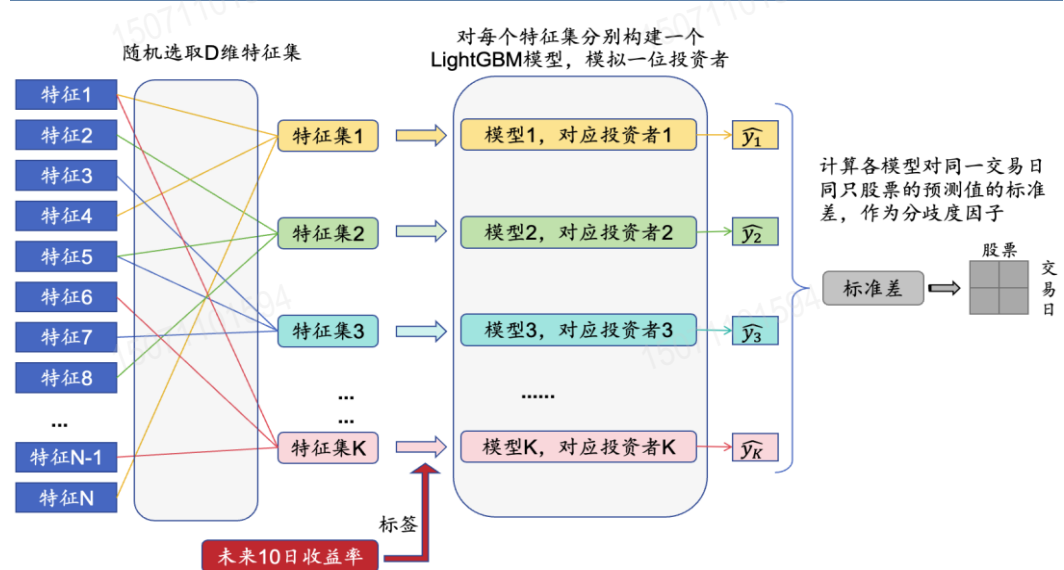
由多个深度学习模型构建的分歧度因子具有显著的选股效果，但模型训练成本较高，只能模仿少数投资者观点，难以大规模应用。参考 Bali 等（2023）的研究，我们采用树模型来提升预测效率。

输入特征为股票截面特征，包含估值、成长、财务质量、量价、一致预期等类别，共 43 个。为了模拟投资者接收差异化信息源，每个模型会随机选取 20 个特征进行训练，而非采用全部特征。预测目标是股票未来 10 个交易日的收益率。

模型采用 LightGBM，数量为 50 个，即模拟 50 位投资者的预测观点。模型采用默认超参数，没有进行超参数寻优，以节省运算时间。模型每年重新训练一次，每次训练使用过去 6 年数据，按照 7:3 比例划分训练集和验证集。

最后，计算所有模型对同一交易日同只股票的预测值的标准差，构建机器学习分歧度因子。

图表9：机器学习模拟投资者分歧



资料来源：华泰研究

分歧度因子回测

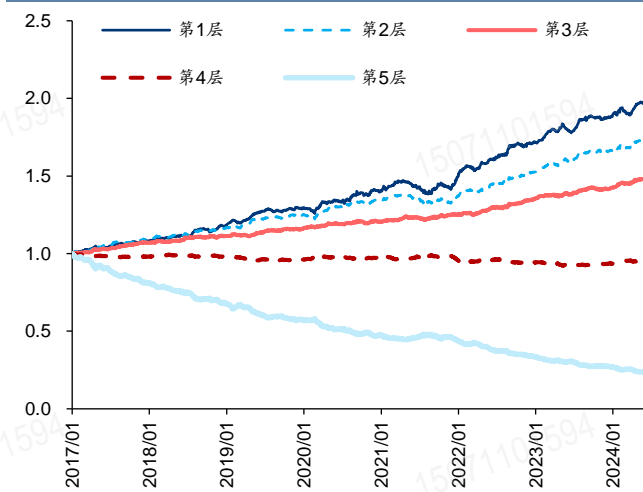
对机器学习分歧度因子进行单因子测试,测试方法与前文保持一致。机器学习分歧度因子同样具有较好的分层单调性。相比于深度学习分歧度,该因子的第一层超额收益接近,而 RankIC 明显提升,换手率大幅下降。

图表10: 机器学习分歧度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	TOP 组合年化超额收益率	TOP 组合信息比率	TOP 组合胜率	TOP 组合换手率
机器学习分歧度因子	7.66%	10.18%	0.75	77.78%	9.71%	2.46	68.54%	61.97%

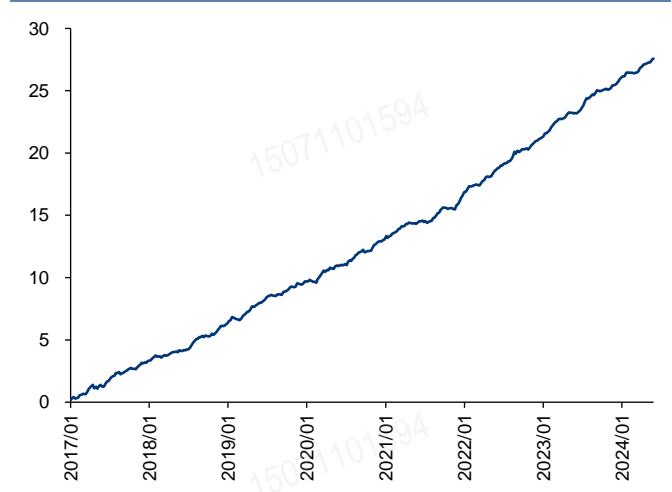
资料来源: Wind, 华泰研究

图表11: 机器学习分歧度因子分层组合相对净值



资料来源: Wind, 华泰研究

图表12: 机器学习分歧度因子累计 RankIC



资料来源: Wind, 华泰研究

参数敏感性分析

机器学习的超参数可能对分歧度因子的计算产生影响,以下对几个参数进行敏感性分析,包括预测周期、投资者数量、特征数量和机器学习模型。

预测周期

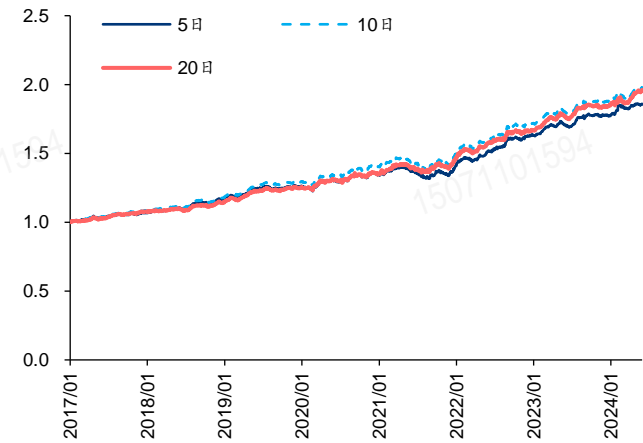
不同预测周期构建的分歧度因子回测表现接近,其中预测 10 日收益率的因子 RankIC 和多头收益表现略好。一个明显的规律是,预测周期越长, TOP 组合换手率越低,这是因为长线投资者的预测分歧在短期内变化不会太剧烈。

图表13: 不同预测周期的分歧度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	TOP 组合年化超额收益率	TOP 组合信息比率	TOP 组合胜率	TOP 组合换手率
分歧度因子 (5 日)	7.65%	10.21%	0.75	78.33%	8.78%	2.26	67.42%	71.47%
分歧度因子 (10 日)	7.66%	10.18%	0.75	77.78%	9.71%	2.46	68.54%	61.97%
分歧度因子 (20 日)	7.30%	9.87%	0.74	75.83%	9.58%	2.55	68.54%	56.73%

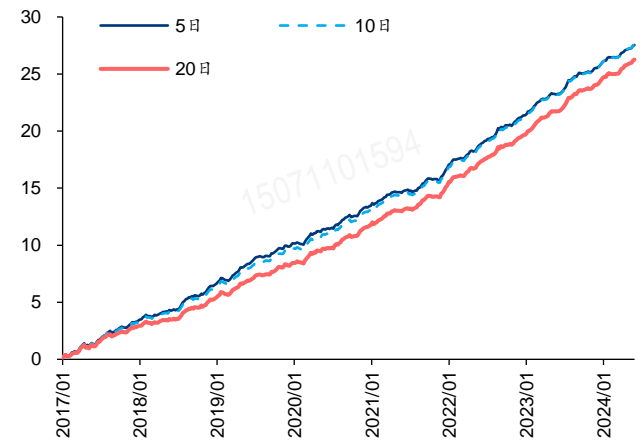
资料来源: Wind, 华泰研究

图表14: 不同预测周期的分歧度因子 TOP 组合相对净值



资料来源: Wind, 华泰研究

图表15: 不同预测周期的分歧度因子累计 RankIC



资料来源: Wind, 华泰研究

投资者数量

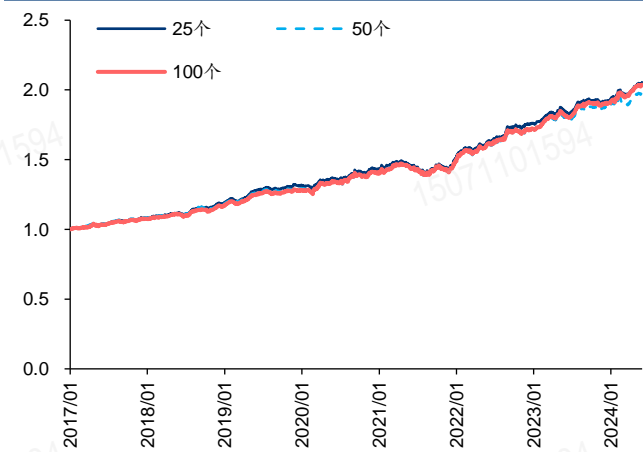
不同投资者数量构建的分歧度因子回测表现差异不大。随着投资者数量增加, RankIC 有所提升, 换手率有所下降。深度学习分歧度因子也验证了这一发现, 由于投资者数量过少 (8 个深度学习因子), 其 RankIC 偏低, 换手率过高。

图表16: 不同投资者数量的分歧度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	TOP 组合年化超额收益率	TOP 组合信息比率	TOP 组合胜率	TOP 组合换手率
分歧度因子 (25 个投资者)	7.38%	9.86%	0.75	77.22%	10.26%	2.70	71.91%	68.61%
分歧度因子 (50 个投资者)	7.66%	10.18%	0.75	77.78%	9.71%	2.46	68.54%	61.97%
分歧度因子 (100 个投资者)	7.70%	10.11%	0.76	78.06%	10.18%	2.57	70.79%	60.71%

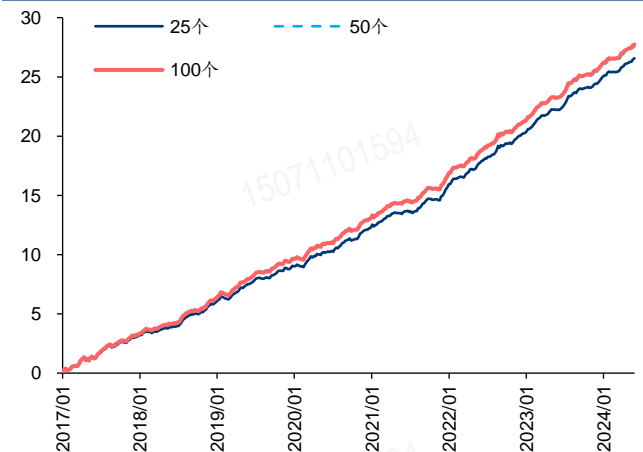
资料来源: Wind, 华泰研究

图表17: 不同投资者数量的分歧度因子 TOP 组合相对净值



资料来源: Wind, 华泰研究

图表18: 不同投资者数量的分歧度因子累计 RankIC



资料来源: Wind, 华泰研究

特征数量

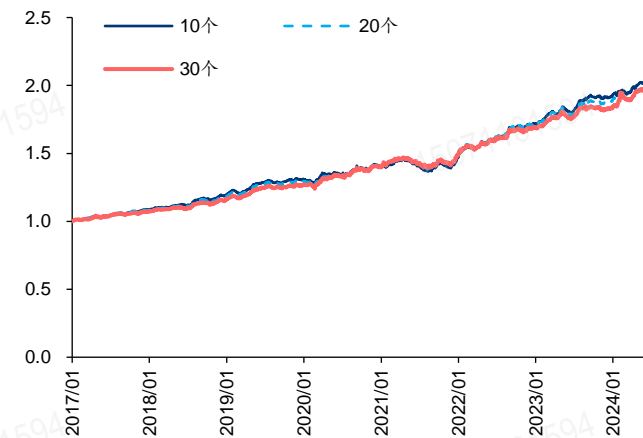
特征数量对分歧度因子的回测表现具有一定影响。随机选取的特征数量越多, 因子的 RankIC 和多头收益越低, 而换手率越高。利用机器学习构建分歧度因子的逻辑是, 通过随机选取特征集, 模拟投资者接收差异化信息源。如果特征数量过多, 投资者之间就难以产生预测分歧。

图表19: 不同投资者数量的分歧度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	TOP 组合年化超额收益率	TOP 组合信息比率	TOP 组合胜率	TOP 组合换手率
分歧度因子 (10 个特征)	8.15%	10.45%	0.78	78.33%	10.06%	2.35	70.79%	59.67%
分歧度因子 (20 个特征)	7.66%	10.18%	0.75	77.78%	9.71%	2.46	68.54%	61.97%
分歧度因子 (30 个特征)	7.21%	9.81%	0.73	77.22%	9.65%	2.52	69.66%	62.74%

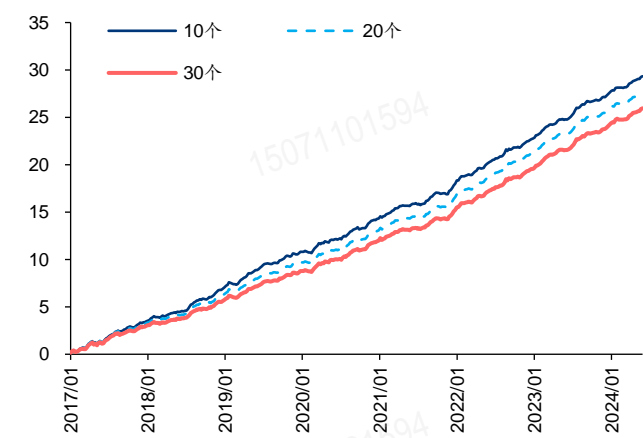
资料来源: Wind, 华泰研究

图表20: 不同特征数量的分歧度因子 TOP 组合相对净值



资料来源: Wind, 华泰研究

图表21: 不同特征数量的分歧度因子累计 RankIC



资料来源: Wind, 华泰研究

机器学习模型

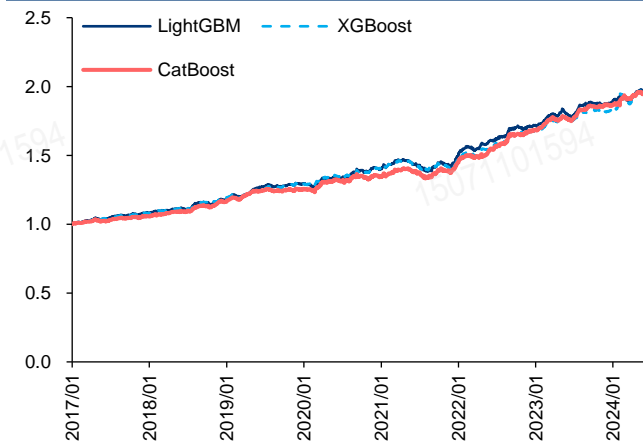
不同机器学习模型构建的分歧度因子回测表现接近, 其中 LightGBM 模型的 RankIC 和多头收益表现较好, XGBoost 模型的换手率较低。

图表22: 不同机器学习模型的分歧度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	TOP 组合年化超额收益率	TOP 组合信息比率	TOP 组合胜率	TOP 组合换手率
分歧度因子 (LightGBM)	7.66%	10.18%	0.75	77.78%	9.71%	2.46	68.54%	61.97%
分歧度因子 (XGBoost)	6.99%	8.70%	0.80	80.56%	9.62%	2.61	75.28%	56.30%
分歧度因子 (CatBoost)	7.29%	9.56%	0.76	77.78%	9.63%	2.56	70.79%	61.84%

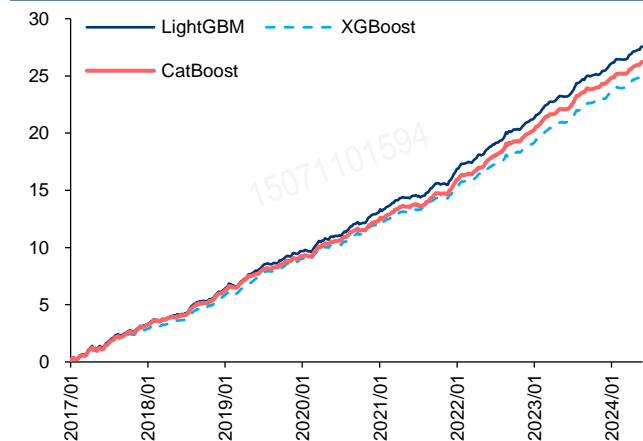
资料来源: Wind, 华泰研究

图表23: 不同机器学习模型的分歧度因子 TOP 组合相对净值



资料来源: Wind, 华泰研究

图表24: 不同机器学习模型的分歧度因子累计 RankIC



资料来源: Wind, 华泰研究

相关性分析

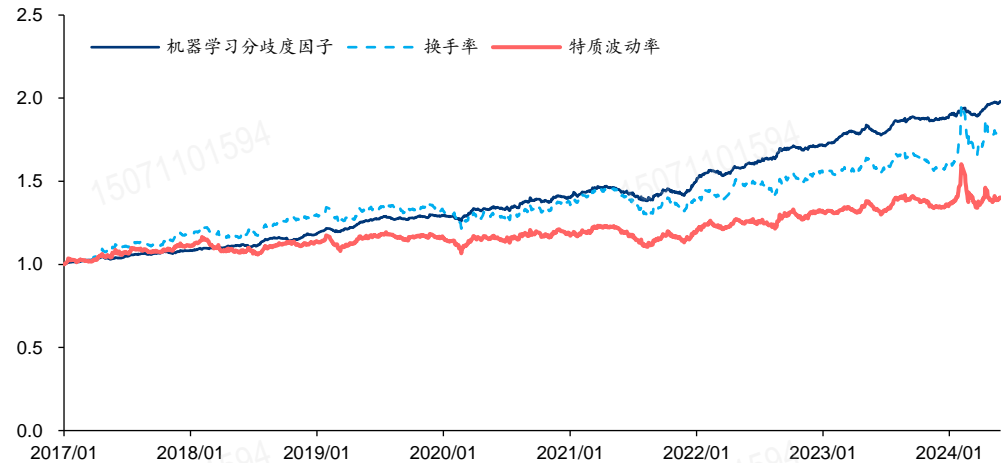
机器学习分歧度因子与常见量价因子相关性如下, 其中与换手率和特质波动率的相关性较高, 这两个因子也是文献中常用来刻画投资者分歧的代理指标。对比因子的多头超额收益, 机器学习分歧度因子表现比换手率、特质波动率更加稳定, 近期也没有出现大幅回撤。

图表25: 机器学习分歧度因子与常见量价因子的相关性

	反转	市值	成交金额	换手率	beta	特质波动率
机器学习分歧度因子	0.20	-0.09	0.35	0.60	0.30	0.61

资料来源: Wind, 华泰研究

图表26：机器学习分歧度因子、换手率、特质波动率的多头超额净值对比



资料来源：Wind，华泰研究

影响因素分析

除了模型超参数外，其他因素也会对分歧度因子的表现产生影响，比如做空限制、动量和投资者认可度。

做空限制

如果股票容易做空，即使投资者分歧度高，悲观投资者同样能够充分表达观点，与乐观投资者分庭抗礼，股价虚高的可能性降低。机构持股比例可用来衡量个股的做空难易程度，机构持股比例高的股票做空相对容易，机构持股比例低的股票做空相对困难。

通过 Wind 获取个股每季度的机构持股比例，由于报告披露有一定的滞后性，为避免使用未来信息，1~4 月份用去年三季报数据，5~8 月份用一季报数据，9~10 月份用半年报数据，11~12 月份用三季报数据。

参考 Fama 和 French（2015）提出的方法，沿着分歧度和机构持股比例两个维度将股票分成 5x5 的 25 宫格组合，计算每个组合的年化收益率。从左往右观察，分歧度越高，股票组合预期收益越低；从上往下观察，机构持股比例越高，股票组合预期收益越高。在机构持股比例低的股票池中分歧度因子的多空表现更好，特别是空头端，说明做空限制是分歧度因子的坚实基础。

图表27：按照分歧度和机构持股比例分组的组合年化收益率

	分歧度低	2	3	4	分歧度高
机构持股比例低	8.68%	5.30%	0.40%	-8.94%	-30.44%
2	10.06%	6.17%	5.79%	-0.99%	-19.45%
3	10.83%	6.59%	3.49%	-1.96%	-15.55%
4	8.86%	7.21%	4.41%	2.52%	-13.87%
机构持股比例高	10.97%	11.21%	9.75%	3.37%	-9.20%

资料来源：Wind，华泰研究

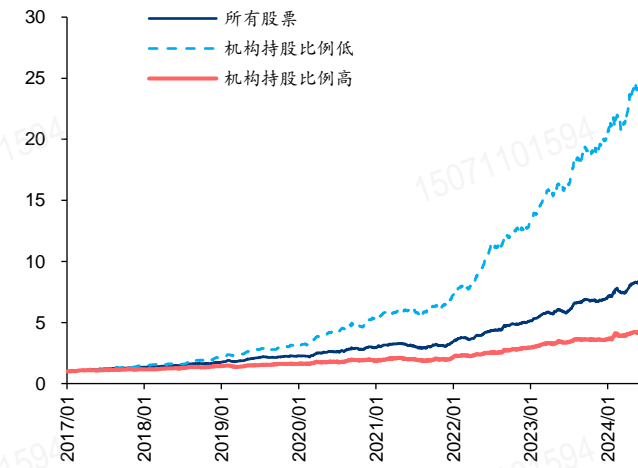
图表28：按照机构持股比例分组的分歧度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	多空组合年化收益率	多空组合夏普比率
分歧度因子（所有股票）	7.66%	10.18%	0.75	77.78%	32.89%	3.48
分歧度因子（机构持股比例低）	9.66%	13.31%	0.73	75.56%	54.05%	4.89
分歧度因子（机构持股比例高）	5.63%	9.11%	0.62	73.06%	21.14%	2.19

注：多空组合收益取分层回测第 1 层和第 5 层净值，计算每日收益之差，再求累计收益

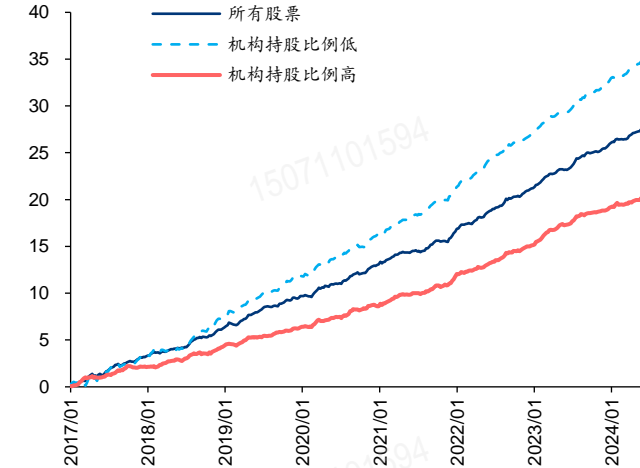
资料来源：Wind，华泰研究

图表29：按照机构持股比例分组的分歧度因子多空净值



资料来源：Wind，华泰研究

图表30：按照机构持股比例分组的分歧度因子累计 RankIC



资料来源：Wind，华泰研究

动量

计算个股最近 20 个交易日涨跌幅，沿着分歧度和近期收益两个维度将股票分成 5x5 的 25 宫格组合，计算每个组合的年化收益率。从左往右观察，分歧度越高，股票组合预期收益越低；从上往下观察，当近期收益过高时，股票组合预期收益非常低，体现出反转效应。近期上涨的股票池中分歧度因子的多空表现更加显著，空头效应很强，而近期下跌的股票池中分歧度因子的分组收益不严格单调。如果股票近期下跌，说明悲观投资者的观点已经有所释放，观点分歧造成的买卖不对称性减弱；如果股票近期上涨，说明乐观投资者占据主导地位，观点分歧造成的买卖不对称性进一步增强。

图表31：按照分歧度和近期收益分组的组合年化收益率

	分歧度低	2	3	4	分歧度高
近期收益低	10.17%	4.78%	5.78%	0.51%	-4.80%
2	8.77%	8.12%	8.84%	6.54%	-4.70%
3	11.41%	8.07%	7.72%	4.25%	-6.79%
4	12.64%	10.26%	8.34%	1.31%	-11.51%
近期收益高	2.99%	-0.98%	-9.59%	-25.06%	-38.20%

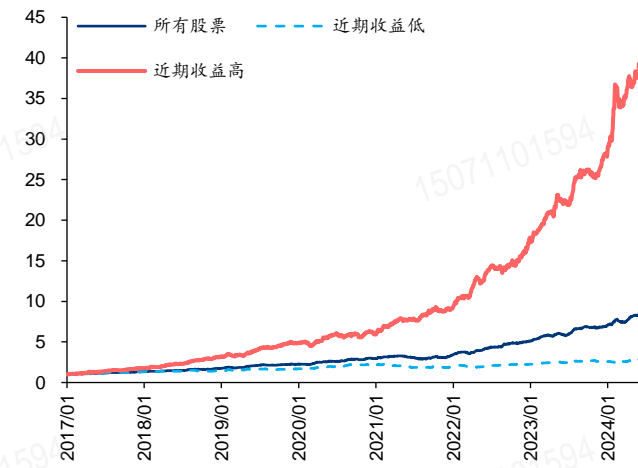
资料来源：Wind，华泰研究

图表32：按照近期收益分组的分歧度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	多空组合年化收益率	多空组合夏普比率
分歧度因子（所有股票）	7.66%	10.18%	0.75	77.78%	32.89%	3.48
分歧度因子（近期收益低）	5.33%	11.44%	0.47	69.72%	14.89%	1.65
分歧度因子（近期收益高）	9.80%	12.03%	0.81	80.56%	63.04%	4.47

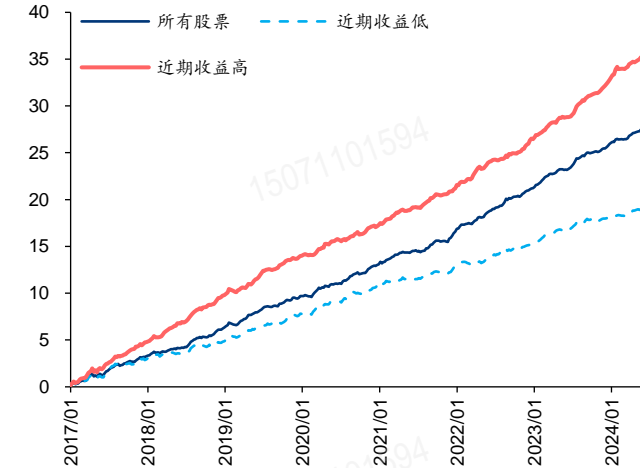
资料来源：Wind，华泰研究

图表33: 按照近期收益分组的分歧度因子多空净值



资料来源: Wind, 华泰研究

图表34: 按照近期收益分组的分歧度因子累计 RankIC



资料来源: Wind, 华泰研究

投资者认可度

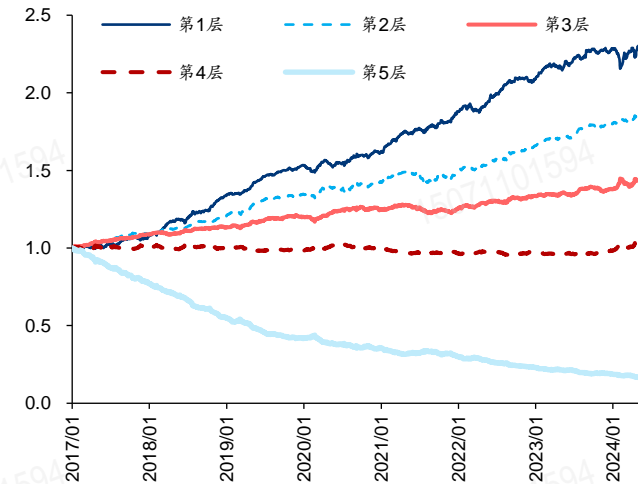
如果说标准差能够模拟投资者观点分歧,那么均值则可以刻画投资者认可度。计算所有模型对同一交易日同只股票的预测值的均值,构建认可度因子。该因子对未来收益具有正向的预测能力,机器学习中的集成学习也体现了这一思想。

图表35: 认可度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	TOP 组合年化超额收益率	TOP 组合信息比率	TOP 组合胜率	TOP 组合换手率
认可度因子	8.34%	9.01%	0.93	81.94%	12.57%	3.06	76.40%	65.45%

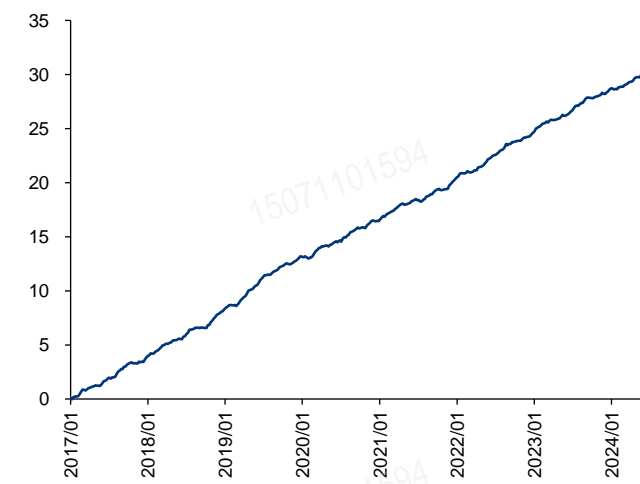
资料来源: Wind, 华泰研究

图表36: 认可度因子分层组合相对净值



资料来源: Wind, 华泰研究

图表37: 认可度因子累计 RankIC



资料来源: Wind, 华泰研究

沿着分歧度和认可度两个因子维度将股票分成 5x5 的 25 宫格组合,计算每个组合的年化收益率。从左往右观察,分歧度越高,股票组合预期收益越低;从上往下观察,认可度越高,股票组合预期收益越高,均与前文研究结论一致。投资者一致看多(认可度高,分歧度低)的股票组合收益最高,与常识相符;然而收益最低的却不是投资者一致看空(认可度低,分歧度低)的股票组合,而是“争议看空”(认可度低,分歧度高),并且认可度低的股票池中分歧度因子的多空收益更为显著。

图表38: 按照分歧度和认可度分组的组合年化收益率

	分歧度低	2	3	4	分歧度高
认可度低	-6.49%	-12.02%	-21.19%	-31.51%	-37.24%
2	4.51%	3.49%	1.52%	-2.60%	-6.47%
3	6.59%	6.68%	5.70%	4.55%	1.68%
4	12.99%	11.36%	10.49%	8.72%	2.33%
认可度高	16.68%	14.75%	13.36%	11.02%	6.52%

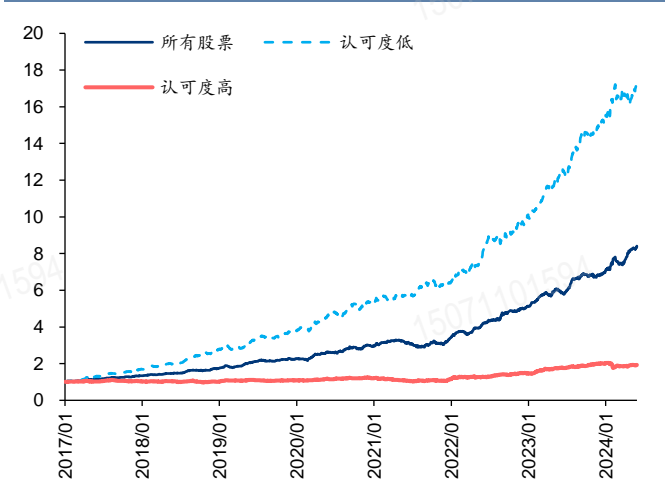
资料来源: Wind, 华泰研究

图表39: 按照认可度分组的分歧度因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	多空组合年化收益率	多空组合夏普比率
分歧度因子 (所有股票)	7.66%	10.18%	0.75	77.78%	32.89%	3.48
分歧度因子 (认可度低)	7.84%	9.87%	0.79	80.28%	47.17%	4.17
分歧度因子 (认可度高)	3.33%	10.50%	0.32	62.78%	9.42%	1.11

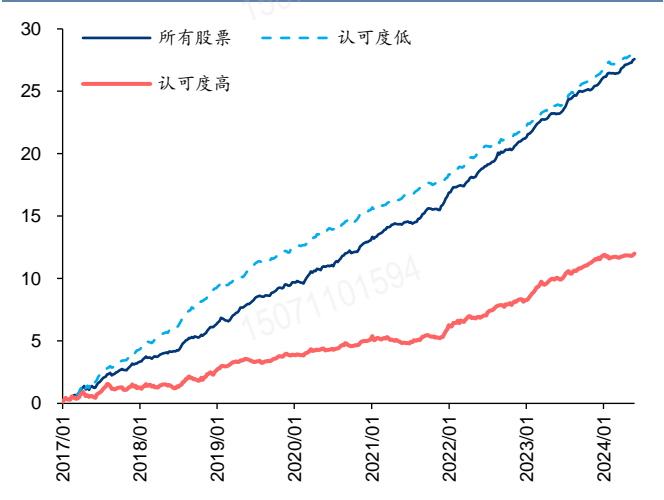
资料来源: Wind, 华泰研究

图表40: 按照认可度分组的分歧度因子多空净值



资料来源: Wind, 华泰研究

图表41: 按照认可度分组的分歧度因子累计 RankIC



资料来源: Wind, 华泰研究

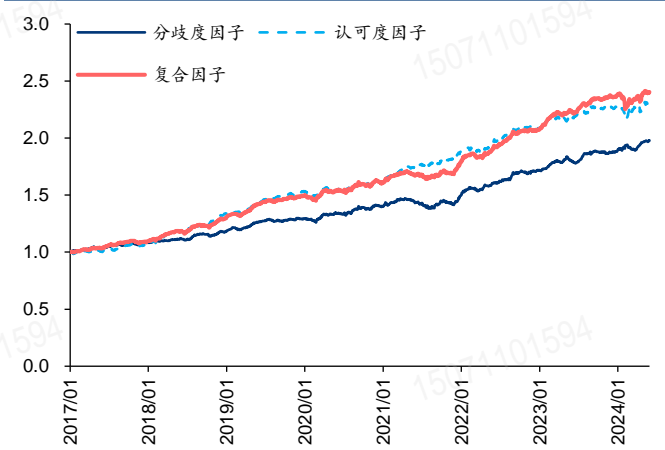
将分歧度因子与认可度因子等权合成, 得到复合因子的 RankIC 和多头收益均有提升。

图表42: 复合因子 IC 值分析和分层回测结果汇总

	RankIC 均值	RankIC 标准差	IC_IR	IC>0 占比	TOP 组合年化超额收益率	TOP 组合信息比率	TOP 组合胜率	TOP 组合换手率
分歧度因子	7.66%	10.18%	0.75	77.78%	9.71%	2.46	68.54%	61.97%
认可度因子	8.34%	9.01%	0.93	81.94%	12.57%	3.06	76.40%	65.45%
复合因子	9.02%	9.47%	0.95	81.94%	13.07%	3.25	74.16%	64.41%

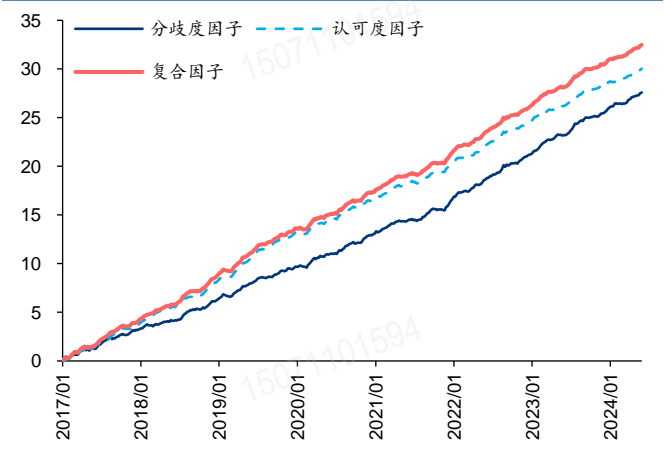
资料来源: Wind, 华泰研究

图表43: 复合因子 TOP 组合相对净值



资料来源: Wind, 华泰研究

图表44: 复合因子累计 RankIC



资料来源: Wind, 华泰研究

总结

投资者分歧是推动交易的原因之一,对金融市场的资产定价具有重要意义。本文运用机器学习来模拟投资者分歧。将不同的特征集输入给多个机器学习模型,模拟投资群体接收差异化信息源、并产生不同投资预测的过程。最后,根据模型预测差异构建的分歧度因子,能够有效刻画投资者分歧,对股票未来收益具有显著的预测作用。

投资者分歧度越高,股票预期收益越低。Miller (1977) 认为投资者分歧会造成股票价格过高,未来回报下降,原因是乐观投资者推高股价,而悲观投资者受到做空限制,对市场的影响力更弱。本文利用现有的多个深度学习因子,构建深度学习分歧度因子,验证了投资者分歧与股票未来收益的负相关性。样本空间为全 A 股,因子在 2017/1/4~2024/5/31 的回测期内周度 RankIC 均值为 6.06%,分 5 层 TOP 组合年化超额收益率为 9.75%,周频双边换手率为 117.70%。

树模型能够高效模拟投资者分歧。本文使用 50 个 LightGBM 模型来模拟投资者的预测观点。输入特征为股票截面特征,包含估值、成长、财务质量、量价、一致预期等类别,共 43 个。每个模型会随机选取 20 个特征进行训练,以模仿投资者接收差异化信息源。最后,计算所有模型对同一交易日同只股票的预测值的标准差,构建机器学习分歧度因子。因子在相同回测期内周度 RankIC 均值提升至 7.66%,分 5 层 TOP 组合年化超额收益率为 9.71%,周频双边换手率下降至 61.97%。

超参数选择对分歧度因子的影响基本不大。不同预测周期构建的分歧度因子回测表现接近,但预测周期越长, TOP 组合换手率越低,因为长线投资者的预测分歧在短期内变化不会太剧烈。不同投资者数量构建的分歧度因子回测表现差异不大,但投资者数量设置过少可能会造成 RankIC 偏低和换手率偏高。特征数量对分歧度因子的回测表现具有一定影响,随机选取的特征数量过多会导致投资者之间难以产生预测分歧。LightGBM、XGBoost 和 CatBoost 等不同机器学习模型构建的分歧度因子回测表现接近。

股票做空限制、动量和投资者认可度对分歧度因子具有显著影响。使用机构持股比例衡量个股的做空难易程度,机构持股比例低的股票池中分歧度因子多空表现更加显著,多空组合年化收益率从 32.89%提升至 54.05%,说明做空限制是分歧度因子的坚实基础。近期涨幅最大的股票池中分歧度因子的多空组合年化收益率高达 63.04%,表明乐观投资者占据主导地位,观点分歧造成的买卖不对称性进一步增强。根据模型预测均值构建投资者认可度因子,发现一致看多的股票组合预期收益最高,“争议看空”的股票组合预期收益最低,分歧度与认可度因子等权重合成的复合因子 RankIC 和多头收益均有提升。

风险提示

借助机器学习构建的选股策略是历史经验的总结,存在失效的可能。机器学习存在过拟合的风险。本文回测暂未考虑交易费用。

参考文献

- [1] Miller E M. Risk, uncertainty, and divergence of opinion[J]. The Journal of Finance, 1977, 32(4): 1151-1168.
- [2] Diether K B, Malloy C J, Scherbina A. Differences of opinion and the cross section of stock returns[J]. The Journal of Finance, 2002, 57(5): 2113-2141.
- [3] Berkman H, Dimitrov V, Jain P C, et al. Sell on the news: Differences of opinion, short-sales constraints, and returns around earnings announcements[J]. Journal of Financial Economics, 2009, 92(3): 376-399.
- [4] Bali T G, Kelly B T, Mörke M, et al. Machine Forecast Disagreement[R]. National Bureau of Economic Research, 2023.
- [5] Jensen T I, Kelly B, Pedersen L H. Is there a replication crisis in finance?[J]. The Journal of Finance, 2023, 78(5): 2465-2518.
- [6] Fama E F, French K R. A five-factor asset pricing model[J]. Journal of Financial Economics, 2015, 116(1): 1-22.

免责声明

分析师声明

本人,林晓明、何康,兹证明本报告所表达的观点准确地反映了分析师对标的证券或发行人的个人意见;彼以往、现在或未来并无就其研究报告所提供的具体建议或所表达的意见直接或间接收取任何报酬。

一般声明及披露

本报告由华泰证券股份有限公司(已具备中国证监会批准的证券投资咨询业务资格,以下简称“本公司”)制作。本报告所载资料是仅供接收人的严格保密资料。本报告仅供本公司及其客户和其关联机构使用。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,但本公司及其关联机构(以下统称为“华泰”)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,华泰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。以往表现并不能指引未来,未来回报并不能得到保证,并存在损失本金的可能。华泰不保证本报告所含信息保持在最新状态。华泰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。

本公司不是 FINRA 的注册会员,其研究分析师亦没有注册为 FINRA 的研究分析师/不具有 FINRA 分析师的注册资格。

华泰力求报告内容客观、公正,但本报告所载的观点、结论和建议仅供参考,不构成购买或出售所述证券的要约或招揽。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,华泰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。

除非另行说明,本报告中所引用的关于业绩的数据代表过往表现,过往的业绩表现不应作为日后回报的预示。华泰不承诺也不保证任何预示的回报会得以实现,分析中所做的预测可能是基于相应的假设,任何假设的变化可能会显著影响所预测的回报。

华泰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,华泰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,为该公司提供投资银行、财务顾问或者金融产品等相关服务或向该公司招揽业务。

华泰的销售人员、交易人员或其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。华泰没有将此意见及建议向报告所有接收者进行更新的义务。华泰的资产管理部门、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。投资者应当考虑到华泰及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。有关该方面的具体披露请参照本报告尾部。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布的机构或人员,也并非意图发送、发布给因可得到、使用本报告的行为而使华泰违反或受制于当地法律或监管规则的机构或人员。

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(无论整份或部分)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,并需在使用前获取独立的法律意见,以确定该引用、刊发符合当地适用法规的要求,同时注明出处为“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利。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。

中国香港

本报告由华泰证券股份有限公司制作,在香港由华泰金融控股(香港)有限公司向符合《证券及期货条例》及其附属法律规定的机构投资者和专业投资者的客户进行分发。华泰金融控股(香港)有限公司受香港证券及期货事务监察委员会监管,是华泰国际金融控股有限公司的全资子公司,后者为华泰证券股份有限公司的全资子公司。在香港获得本报告的人员若有任何有关本报告的问题,请与华泰金融控股(香港)有限公司联系。

香港-重要监管披露

- 华泰金融控股（香港）有限公司的雇员或其关联人士没有担任本报告中提及的公司或发行人的高级人员。
- 有关重要的披露信息，请参华泰金融控股（香港）有限公司的网页 https://www.htsc.com.hk/stock_disclosure 其他信息请参见下方 “美国-重要监管披露”。

美国

在美国本报告由华泰证券（美国）有限公司向符合美国监管规定的机构投资者进行发表与分发。华泰证券（美国）有限公司是美国注册经纪商和美国金融业监管局（FINRA）的注册会员。对于其在美国分发的研究报告，华泰证券（美国）有限公司根据《1934年证券交易法》（修订版）第15a-6条规定以及美国证券交易委员会人员解释，对本研究报告内容负责。华泰证券（美国）有限公司联营公司的分析师不具有美国金融监管（FINRA）分析师的注册资格，可能不属于华泰证券（美国）有限公司的关联人员，因此可能不受 FINRA 关于分析师与标的公司沟通、公开露面和所持交易证券的限制。华泰证券（美国）有限公司是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。任何直接从华泰证券（美国）有限公司收到此报告并希望就本报告所述任何证券进行交易的人士，应通过华泰证券（美国）有限公司进行交易。

美国-重要监管披露

- 分析师林晓明、何康本人及相关人士并不担任本报告所提及的标的证券或发行人的高级人员、董事或顾问。分析师及相关人士与本报告所提及的标的证券或发行人并无任何相关财务利益。本披露中所提及的“相关人士”包括 FINRA 定义下分析师的家庭成员。分析师根据华泰证券的整体收入和盈利能力获得薪酬，包括源自公司投资银行业务的收入。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或不时会以自身或代理形式向客户出售及购买华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或其高级管理层、董事和雇员可能会持有本报告中所提到的任何证券（或任何相关投资）头寸，并可能不时进行增持或减持该证券（或投资）。因此，投资者应该意识到可能存在利益冲突。

新加坡

华泰证券（新加坡）有限公司持有新加坡金融管理局颁发的资本市场服务许可证，可从事资本市场产品交易，包括证券、集体投资计划中的单位、交易所交易的衍生品合约和场外衍生品合约，并且是《财务顾问法》规定的豁免财务顾问，就投资产品向他人提供建议，包括发布或公布研究分析或研究报告。华泰证券（新加坡）有限公司可能会根据《财务顾问条例》第32C条的规定分发其在华泰内的外国附属公司各自制作的信息/研究。本报告仅供认可投资者、专家投资者或机构投资者使用，华泰证券（新加坡）有限公司不对本报告内容承担法律责任。如果您是非预期接收者，请您立即通知并直接将本报告返回给华泰证券（新加坡）有限公司。本报告的新加坡接收者应联系您的华泰证券（新加坡）有限公司关系经理或客户主管，了解来自或所分发的信息相关的事宜。

评级说明

投资评级基于分析师对报告发布日后6至12个月内行业或公司回报潜力（含此期间的股息回报）相对基准表现的预期（A股市场基准为沪深300指数，香港市场基准为恒生指数，美国市场基准为标普500指数，台湾市场基准为台湾加权指数，日本市场基准为日经225指数，新加坡市场基准为海峡时报指数，韩国市场基准为韩国有价证券指数），具体如下：

行业评级

增持：预计行业股票指数超越基准

中性：预计行业股票指数基本与基准持平

减持：预计行业股票指数明显弱于基准

公司评级

买入：预计股价超越基准15%以上

增持：预计股价超越基准5%~15%

持有：预计股价相对基准波动在-15%~5%之间

卖出：预计股价弱于基准15%以上

暂停评级：已暂停评级、目标价及预测，以遵守适用法规及/或公司政策

无评级：股票不在常规研究覆盖范围内。投资者不应期待华泰提供该等证券及/或公司相关的持续或补充信息

法律实体披露

中国: 华泰证券股份有限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

香港: 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:AOK809

美国: 华泰证券(美国)有限公司为美国金融业监管局(FINRA)成员,具有在美国开展经纪交易商业业务的资格,经营业务许可编号为:CRD#:298809/SEC#:8-70231

新加坡: 华泰证券(新加坡)有限公司具有新加坡金融管理局颁发的资本市场服务许可证,并且是豁免财务顾问。公司注册号:202233398E

华泰证券股份有限公司

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

深圳市福田区益田路5999号基金大厦10楼/邮政编码:518017

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

华泰金融控股(香港)有限公司

香港中环皇后大道中99号中环中心53楼

电话:+852-3658-6000/传真:+852-2567-6123

电子邮件:research@htsc.com

http://www.htsc.com.hk

华泰证券(美国)有限公司

美国纽约公园大道280号21楼东(纽约10017)

电话:+212-763-8160/传真:+917-725-9702

电子邮件:Huatai@htsc-us.com

http://www.htsc-us.com

华泰证券(新加坡)有限公司

滨海湾金融中心1号大厦, #08-02, 新加坡 018981

电话:+65 68603600

传真:+65 65091183

©版权所有2024年华泰证券股份有限公司