《2023年度中国量化投资白皮书》编委会

QUANT

2023年度
中国量化投资白皮书

官方下载及活动链接—https://BigQuant.com/quantpaper/2023



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

傲慢与偏见

The Quantitative Parable of the Tower of Babel

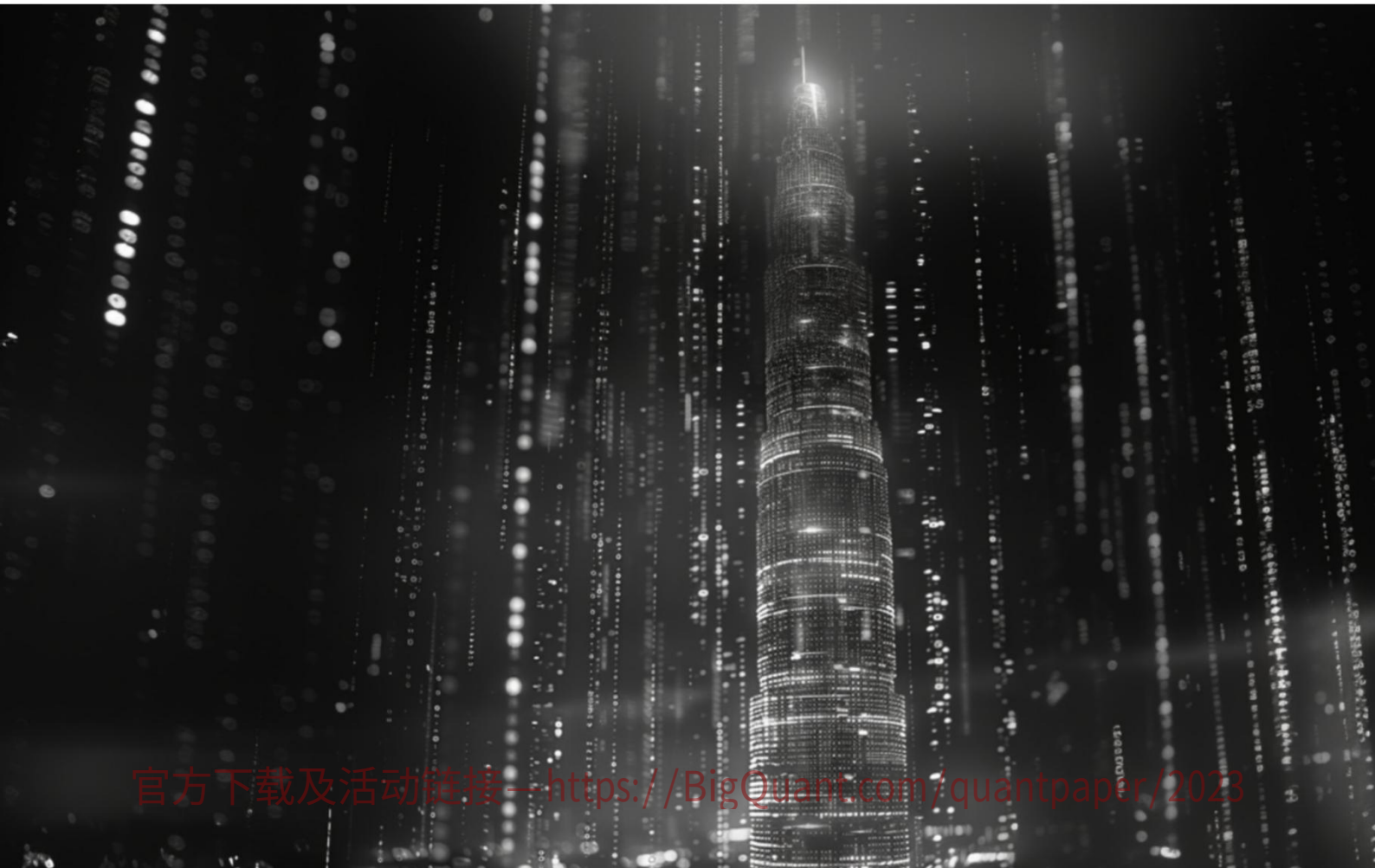
洪水之后，人们聚集在美索不达米亚平原，试图建造一座高耸入云的巴别塔，以表达对神的挑战和追求不朽的渴望。
然而，神看到来人们的傲慢和野心，为了制止人类，神让语言变得纷繁复杂，散乱、多样而神秘，不再能够被感官所通达。语言的封闭以及有意无意的模糊让人们互为聋哑，彼此为限，最终导致了混乱和困扰，巴别塔再也没有建成。

这像是量化与公众之间沟通的一个隐喻，过去十年，行业不断发展形成诸多词汇，当彼此之间沟通越加紧密之时，离其他语言也就越远。

过去两年，《中国量化投资白皮书》回收了663份问卷，描绘夯实行业共识，去形成国内行业、组织、群体、实践的共识语言，行至第三年，我们向市场发送两份问卷，一份面向量化机构，一份面向社会公众，在400份答案中寻找沟通的缝隙。
在问卷中，我们仍然可以看到量化机构百折不挠，于纷繁无序中，极致探寻算法价值，但我们同时也看到了行业在面向公众沟通的被动与无力。

认为量化对市场有害是一种偏见，但误认为公众对此看法单一、无法沟通何尝不是另外一种刻板化偏见，甚至傲慢，而我们知道，弱小和无知不是生存的障碍，傲慢才是。面对公众的质疑，量化机构选择了投资教育作为应对。但很多时候投资者教育成为量化机构了名为保持“低调”，实为并不沟通的安慰剂。

对于技术与市场我们都可以更乐观一些，但生存本来就需要争取，从来不是理所当然。



官方下载及活动链接：https://BigQuant.com/quantpaper/2023



目录

CONTENTS

关键数据及发现 1
数据来源及结构 4
第一章 背离与博弈 6
第二章 阿尔法的生命力 27
第三章 风暴中的我们 77
第四章 市场与他者 98
第五章 量化的星辰大海 108
第六章 机构最佳实践 124
01 Man Group：我们的本土化创新 125
02 半鞅私募：“三个火枪手”的端到端量化之路 128
03 黑翼资产：控制回撤是第一位的 足够稳健才是长远目标 131
04 量派投资：万物皆存周期 行则将至 134
05 蒙玺投资：守正出奇 与卷同舒 137
06 涵德投资：慎思之 明辨之 笃行之 143
08 平方和：量化策略的匠心和价值 149
08 天算量化：人算不如天算 155
09 图灵基金：90 后掌门人的七十二般变化 159
10 稳博投资：模糊的正确与量化的精确 164
11 玄信资产：我们正在经历的风暴 169
12 致诚卓远：《破局》量化人的宏观书 173
13 中欧基金：曲径通幽 公募基金的基本面量化之路 177
14 上海卓胜：“老派”人工智能量化的反其道 181
第七章 附录 185

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

关键数据及发现

KEY FINDINGS

1.2023 年，A 股“冲高回落、震荡下跌”,量化进入存量博弈。私募规模扩张趋势放缓，新备案私募门槛增高，行业格局稳定性增强，已经形成了越来越快的“飞轮效应”：业绩好的机构形成口碑和规模，吸引了人才、资金、资源的大幅进入，同时算力、数据、投研等方面不断积累，资管规模不断扩大，而资源的增加又促成了良好业绩的产生。2023 年，规模 50 亿以上机构从 41 家增加至 50 家，占整个量化私募 5%。

2.量化机构的投研趋势之一体现在多资产、多信号、多频段融合。机构横跨多个资产类别，高中低频多个信号频段。

3.TB 级别（41.07%）和 PB 级别（27.38%）是当前主流存储规模，PB 级别（27.38%）则使得量化开始进入真正的“大数据”机构。低频量价数据的使用覆盖度最高（72.62%），其次是基本面财务数据（66.67%）和高频量价数据（58.33%）。高频量价数据维度大、噪声高，数据及因子一般不直接使用，需要进行大量数据清洗和加工，包括利用增量行情自动拼接全量行情，用快照数据重构订单簿，高频因子平滑处理等。

4.另类数据目前使用面最广的为分析师一致性预期数据及舆情数据，分析师一致性预期数据的分析主要有 Alpha-Capture 卖方预期评价、情绪打分、知识图谱构建等使用方式。事件、情绪维度的数据由于信息的滞后性则需要关注市场反身性避免形成反指。

5.量化机构通过研究经验、买方灵感、卖方研报、前沿信息获取研究灵感生成因子，按照因子种类、预测周期、产生背景，当前多因子已经形成非常庞杂的体系，量化机构一般会根据 300、500、1000 指增、中性等各类产品分配不同的因子比重，最常见的为 70-80%量价类因子+20-30%（基本面因子＋舆情和另类因子），以追求因子之间的低相关性。人工挖掘（覆盖度 77.38%）以逻辑性因子为主，依赖于研究人员的专业知识和直觉，机器挖掘（覆盖度 59.52%）以统计性因子为主，利用机器学习算法自动识别有预测能力的因子，发现人工方法难以识别的复杂模式和关系。

6.过往三年问卷调研数据显示，当前量化模型技术呈现多样性、复杂性，没有任何一种技术在所有环节中都占据绝对的优势，一个量化策略需要结合多种技术来处理不同问题。数据处理相关的环节模型应用较为集中（如数据清洗、特征提取、模型构建），而在直接的决策制定环节（如策略优化、高频交易）则较为分散。

7.组合优化环节，目前量化机构的优化器呈现更快、更复杂的趋势。更复杂表现在：除了回报（79.17%）和风险预测（76.79%）是优化的核心常见功能，交易成本（60.12%）、组合限制（48.21%）、业绩基准（42.86%）等、融券利率（19.64%）及基差成本（29.17%）等也有机构纳入了优化器功能之中。随着账户数量、投资标的和投资逻辑的不断增加，计算量更大，部分机构实现盘中分钟级别出交易清单。机构对市值、流动性、分散度、成分股等进行了风控，Barra 模型是目前国内的主流风险控制模型，但在实际使用中有许多局限性，例如预测周期偏长、没有考虑特异性股池，A 股的特点结合性有待提升等。所以很多机构除了根据其控制因子敞口，对标基准进行行业和风格约束之外，还会进行周期改动等优化调整。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

8.量化机构会采用股指期货、融券、期权等多种方式或结合的方式进行对冲，持续跟踪成本差异，以求得对冲成本、对冲精度的最优解。寻常认知中，公众对高频与做空最容易产生误解，但样本池内数据显示，当前公众对做空策略有相对较为成熟和均衡的看法，能够认识到做空（占比 83.84%）是市场正常机制的一部分，有助于发现价格泡沫和过高估值。只要不是恶意操纵市场，做空策略应该被允许（占比 65.5%）。

9.不同的交易频率有其适应的市场行情，如果仅仅是擅长某一类因子或者某一个频段的预测，在多变的市场环境下可能会遭遇逆周期时出现回撤，所以机构需要捕捉市场不同周期收益来源，逐步丰富覆盖低、中、高多个频段，全频段覆盖及融合的挑战在于：1.无明显频段短板；2.平衡、分配不同预测周期的权重。有的从特征层面融合，有的从优化器层面融合，还有从组合层面融合。

10.数据显示，目前机构认为业绩归因 Smart Beta 超越 Pure Alpha，持仓收益超过交易收益.值得注意的是，1 个亿以下规模量化机构（平均分 5.29）和 50-100 亿规模机构（平均分 4.55）对自己创造 Pure Alpha 的能力给出了相对较高的评分。

11.面向未来一年，量化投研人员对未来市场的预期较为中性，部分因素略偏向乐观（如市场行情 Beta 走势、成交量和流动性、科技创新），而对于监管政策及舆论导向和交易、策略集中度则较为担忧。

12.极端行情期间，量化机构调整策略与未调整的人各占一半，主要调整内容包括因子权重（51.18%）、杠杆比率（49.41%）、交易执行（48.82%）、指数成分（45.88%）。调研认为本轮危机中，一些软性的实力放在了更显著的位置，排名前 5 的为心理韧性、危机沟通能力、团队协作、快速响应、合规风控，超越策略本身。

13.量化交易对 A 股市场影响方面，大多数公众认为量化交易对 A 股市场有积极的影响，特别是在提升市场效率（71.18%）、促进市场创新（64.63%）和增强流动性方面（79.91%）。同时，也有一定比例的公众担心量化交易可能会增加市场波动，反映了对量化交易影响的复杂性和双刃剑特性的认识。

14.四成的受访者对量化机构没有负面情绪，剩下的有情绪的公众中归因主要方向是：信息不对称（31.88%）、量化复杂性（27.51%）、媒体报道（28.82%）。

15.面对公众的质疑，7 成机构选择了投资教育作为应对公众误解的主要策略。但传播与教育效果并未如预期，公众教育甚至成为机构名为保持“低调”，实为并不沟通的安慰剂，本质上是一个静默策略，实际执行过程中，自行组织视频投教仅为 22.35%，部分机构甚至未与任何渠道做过任何沟通，投资者教育活动可能未能达到预期的覆盖面或深度。

16.公众对于将量化投资策略纳入个人投资组合持积极态度。多数受访者（79.04%）表达了对量化投资的信任，愿意考虑将其作为投资组合的一部分。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

17.若因子与模型的经济学意义与人工智能代表两个相反方向，可以大致对量化机构派别进行划分为：传统经济学派（占比 44.12%）、平衡派（占比 22.94%）及人工智能派（占比 33.53%），预计未来三年，这个比例会调整为 20.59%：18.24%：60.59%。

18.大模型当前应用分为创造 Alpha、提升效率以及投资者关系三个方面：其中最明显的作用在于思维启发、投研思路拓展（64.71%）。代码生产、编程辅助或 Debug 占比 34.71%。

19.数据质量是限制另类数据使用最核心的因素，包括数据不完整（61.76%）缺少关键信息，影响分析的完整性。清洗难度大（48.82%）意味着需要投入更多的时间和资源。供应商杂乱（47.65%）多且标准不一，会给数据采购和质量控制带来挑战。

20.在应用另类数据的时候，量化机构最核心的还是是否具有强逻辑。量化机构需要寻找逻辑链条较短、逻辑清晰、历史更长的产业链数据或行业特有数据，目前另类数据中，结构化数据仍然是最主要的数据来源（权重占比 55.29%），非结构化数据如文字、语音、视频和图像还需机构自行清理，所以应用范围较少，权重占比分别为 18.88%、7.03%、11.78%。

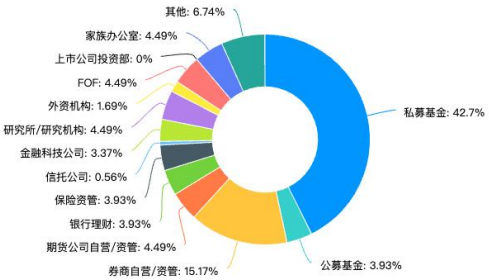
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

数据来源及结构

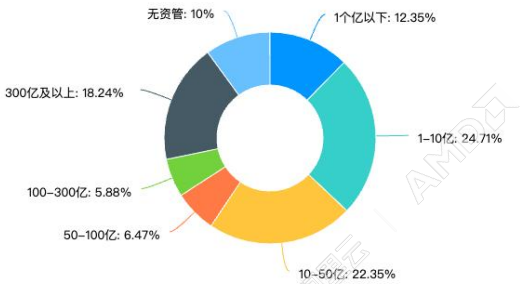
DATA SOURCE AND STRUCTURE

量化机构版——

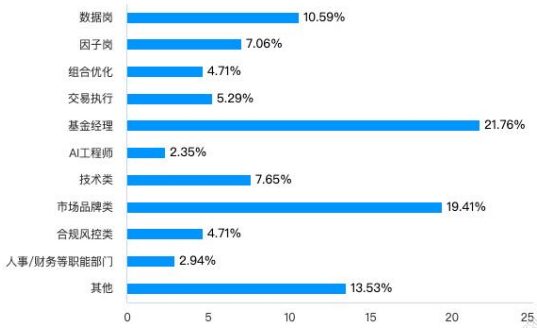
■图表 机构类型



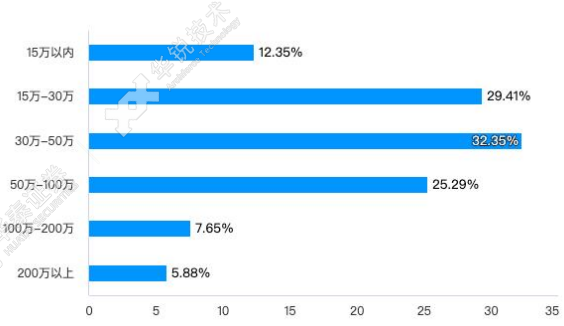
■图表 机构当前管理规模



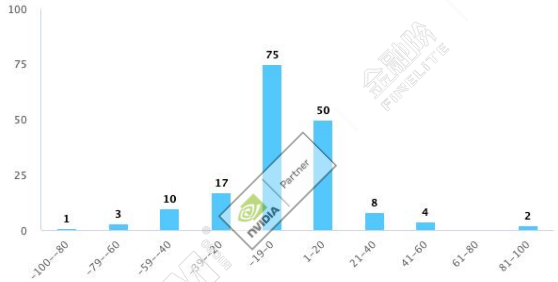
■图表 岗位类别



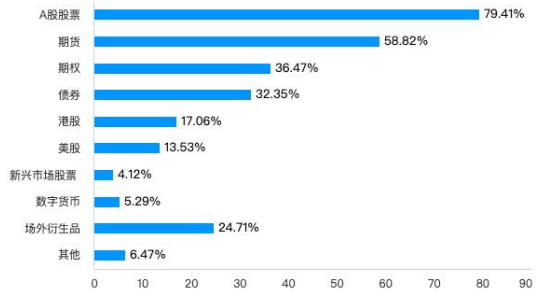
■图表 薪酬区间【总包】



■图表 薪酬涨跌幅



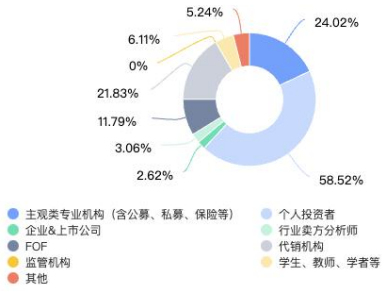
■图表 投资标的



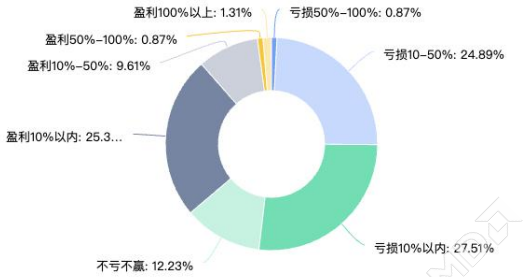
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

社会公众版——

■图表 身份类型



■图表 过去一年公众投资总体盈亏情况



数据说明——

风险提示：本次问卷受限于问卷设计、样本获取、数据整理方式，可能与实际市场情况存在偏差与测算主观性问题。

白皮书涉及问卷数据只能反映受访者对自己理解程度的主观评价，并不一定准确反映其实际能力水平。

不同受访者对评分的标准可能存在差异，因此这些数据更多地提供了一个关于受访者自我认知的总体趋势，而不是一个严格的能力测试结果。

本问卷投放暴露于量化行业或者关注量化的社会公众，其认知与态度可能存在幸存者偏差。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

CHAPTER

第一章

背离与博弈

```
divergence = lambda price, indicator: abs(price - indicator)
market_play = divergence(100, 105) > divergence(100, 102)
print("Strategic divergence" if market_play else "Consensus alignment")
```



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

01 量化与主流机构的背离

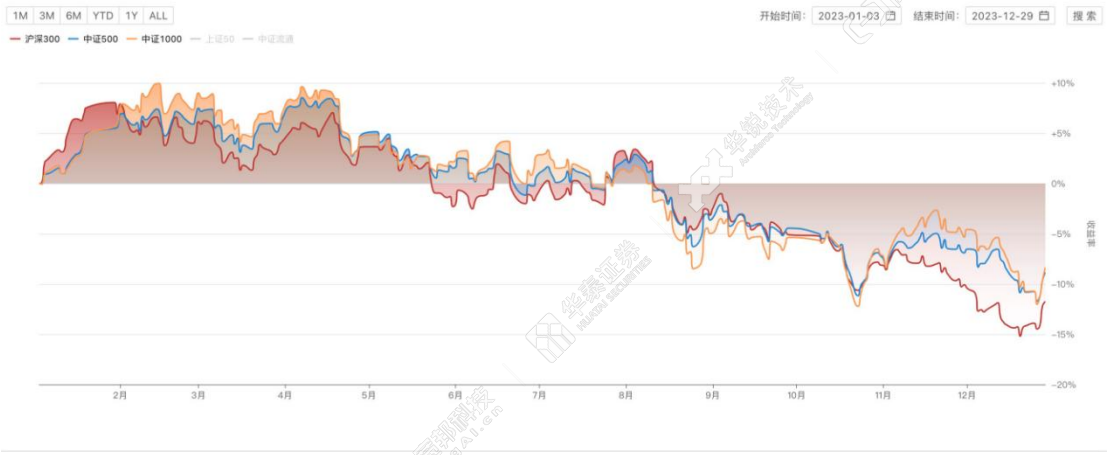
1.冲高回落 几家欢喜

“冲高回落、震荡下跌”，几乎所有机构面对 2023 年 A 股市场时，都会如此形容。

2023 年开年伊始，ChatGPT 横空出世，引爆相关行业。下半年，俄乌冲突一波未平，巴以冲突又起，全球地缘政治割裂、动荡，叠加美联储持续加息，北向资金出现持续净卖出。

上证指数从年初 3000 点附近一路震荡反弹，一度攀升至 3418.95 点，随后二季度开始震荡起伏，三季度末收盘于 3110.48 点。四季度市场一度震荡下跌，最低跌破 2900 点，达到 2882.0 点，随后震荡反弹，12 月底回到 2974.93 点上方，年内小幅下跌 3.7%。

■图表 2023 年 A 股市场表现



资料数据来源：火富牛

2023 年市场先扬后抑，板块轮动剧烈，中特估、ChatGPT、华为产业链、室温超导、北交所，各类主题昙花一现。仅剩微盘股全年表现一枝独秀。使得不管是持有型机构还是交易型机构/个人都在市场里备受煎熬。

2023 年底，招商证券统计了全 A 个股两两相关系数中位数。发现全 A 个股相关性有所降低，这也为量化基金在 2023 年获得超额收益提供了市场条件。同时可以采用方差比检验来考察市场的定价效率，显示定价效率在 2023 年有所提升，有利于基于历史数据构建策略以获得超额收益。——《公募量化基金 2023 年度盘点：激流勇进，再接再厉》2024/01/30 招商证券

主题突出、切换频繁的行情使得市场分化较明显。主观基金与量化基金风险承担和盈亏情况上存在显著差异。尤其是“大白马”、核心资产、大盘成长的基金整体表现不佳。从基金产品收益率平均值和中位数上看，主观基金与量化基金虽然各有盈亏，但量化基金的回撤、集中性表现更好。

■图表 2023 年主观多头 vs 量化多头 vs 股票市场中性策略业绩分布情况

		主观多头策略			量化多头策略			股票市场中性策略		
	公司规模	平均值	中位数	标准差	平均值	中位数	标准差	平均值	中位数	标准差
收	0–5 亿	5.35%	6.88%	23.43%	2.41%	0.59%	19.79%	6.12%	4.83%	11.92%
益	5–10 亿	2.48%	5.69%	23.31%	4.74%	3.75%	15.71%	10.44%	8.06%	18.39%
率	10–20 亿	4.84%	6.18%	18.64%	3.69%	2.49%	10.75%	6.96%	6.84%	9.37%

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

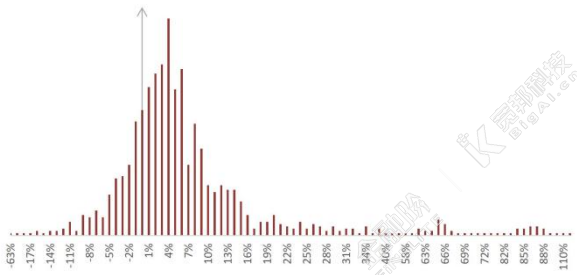
	20–50 亿	2.77%	3.57%	15.92%	4.07%	3.76%	10.38%	6.82%	7.55%	8.63%
	50–100 亿	5.20%	–7.94%	12.90%	5.43%	4.69%	9.31%	8.64%	7.66%	11.39%
	100 亿+	–1.74%	6.21%	15.97%	5.50%	4.45%	6.55%	6.67%	5.63%	4.83%
最大回撤	0–5 亿	–23.65%	–21.52%	13.89%	–15.28%	–12.52%	10.87%	–6.36%	3.65%	6.99%
	5–10 亿	–20.29%	–18.95%	10.84%	–12.64%	–10.95%	7.05%	–3.15%	–1.86%	5.22%
	10–20 亿	–20.11%	–18.06%	11.57%	–12.33%	–11.31%	6.19%	–3.90%	–2.11%	6.31%
	20–50 亿	–19.10%	–17.62%	10.79%	–12.20%	–11.74%	6.07%	–4.73%	–1.91%	9.63%
	50–100 亿	–18.58%	–18.32%	8.50%	–11.09%	–10.36%	3.49%	4.27%	–2.13%	5.36%
	100 亿+	–16.89%	–16.62%	6.01%	–10.32%	–9.83%	2.85%	–2.10%	–1.39%	2.96%

数据来源：白小白的学习笔记

在收益方面：量化代表机构（规模 50 亿以上）平均值和中位数收益表现更好，在百亿私募 TOP20 的榜单上，量化策略 16 家上榜，占比 80%。

从指数增强超额收益中位数来看，300 指增约 8%，500 指增约 10%，1000 指增超过 15%，量化多头 2023 年收益分布如下图，80%收益集中在[-5%，14%]，收益为正的占比 78%。收益分布的集中度显著高于传统多头，这意味着投资量化多头产品有更稳定的预期表现。

■图表 2023 年量化多头收益分布



■图表 2023 年量化多头仓位变化图



数据来源：招商私募指数数据库

■图表 公、私募量化基金超额收益业绩统计

	300 指增		500 指增		1000 指增		主动增强	
	公募	私募	公募	私募	公募	私募	公募	私募
2017	–0.21%	3.00%	11.72%	18.12%	13.58%	14.17%	0.36%	1.32%
2018	4.82%	15.50%	11.47%	27.40%	26.97%	43.63%	11.78%	27.17%
2019	1.23%	6.42%	4.480%	11.96%	8.21%	20.78%	1.79%	–1.03%
2020	11.15%	10.81%	11.17%	15.71%	14.17%	14.57%	13.81%	8.45%
2021	5.96%	13.76%	2.65%	10.63%	4.74%	12.12%	2.02%	11.56%
2022	2.87%	10.19%	4.66%	13.00%	8.41%	18.77%	0.34%	13.23%
2023	1.94%	5.66%	2.18%	11.39%	7.28%	15.06%	0.27%	7.44%
年化收益率	4.13%	9.11%	7.07%	15.92%	12.81%	19.78%	4.34%	9.73%
年化波动率	2.49%	3.58%	3.92%	4.26%	7.40%	4.94%	4.55%	6.52%

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

夏普比率	0.85	1.99	1.29	3.27	1.46	3.60	0.51	1.19
最大回撤	-5.01%	-3.95%	-6.30%	-3.82%	-11.22%	-4.12%	-8.98%	-7.81%

数据统计区间：2017.01.01-2023.12.29

数据来源：wind、朝阳永续、国泰君安证券研究

在回撤方面：量化多头策略在所有规模区间的公司最大回撤都比主观多头策略小，在不利市场情况下的表现更为稳健。在集中性方面：量化多头策略的收益标准差普遍比主观多头策略低，收益分布更为集中，主观前后收益分化则更大。这就使得量化基金的投资人在 2023 年能享有更好的投资体验，误踩概率更小。同时也使得量化机构遭受资金热捧，投资占比提升。

■图表 私募基金策略占比变化情况

时间	主观多头	量化多头	股票中性	债券	CTA	FOF	混合类	期权
2023-1	38.1%	7.2%	2.9%	7.0%	11.7%	18.0%	12.0%	3.00%
2023-2	38.9%	7.1%	2.9%	7.1%	11.7%	17.6%	11.6%	3.0%
2023-3	39.2%	7.1%	3.0%	7.2%	11.7%	17.30%	11.3%	3.20%
2023-4	39.3%	6.9%	2.8%	7.3%	12.1%	17.4%	12.0%	2.3%
2023-5	38.4%	6.8%	3.1%	7.3%	11.4%	17.6%	11.8%	3.8%
2023-6	38.1%	6.9%	3.1%	7.4%	11.0%	17.8%	11.8%	3.9%
2023-7	38.0%	6.8%	3.1%	7.5%	10.9%	17.9%	11.8%	4.0%
2023-8	36.9%	7.0%	3.1%	7.5%	10.8%	18.1%	12.4%	4.3%
2023-9	36.6%	7.1%	3.1%	7.6%	10.7%	18.4%	12.3%	4.2%
2023-10	36.3%	7.2%	3.1%	7.7%	10.5%	18.4%	12.5%	4.3%
2023-11	36.4%	7.3%	3.2%	7.7%	10.5%	18.5%	12.0%	4.4%
2023-12	36.1%	7.4%	3.3%	7.7%	10.1%	18.7%	12.1%	4.5%

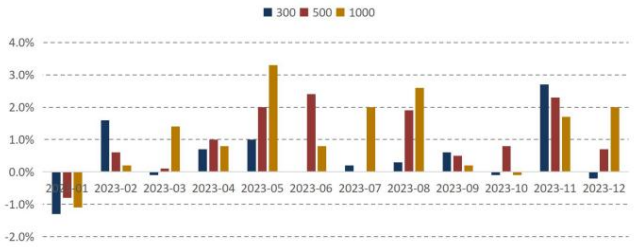
数据资料来源：招商私募指数数据库

指增方面：对比往年月度超额，指增超额存在进一步收缩下滑，单月超额超 2%在 2023 年仅为个别月份，往年则相对普遍。年内 1000 指增表现优于 500 优于 300。同时由于近两年 Beta 较弱，大部分产品净值没有突破 2021 年 9 月高点，管理人难以计提业绩报酬。

中性方面：超额 Alpha 环境友好，叠加对冲成本降低，2023 年中性策略业绩平均收益 8.2%，80%收益集中收益区间在 (0%, 14%) ，收益为正的占比达到 91.6%，波动率低于 3%。2023 年，IC 跨期基差年化水平在 3%+，相比 2021 年 10%+、2022 年 5%+的对冲成本更低。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 指数增强平均超额变化趋势



■图表 2023 年股票市场中性收益表现



数据资料来源：招商私募指数数据库

公募量化基金方面，2023 年快速扩容，截至 2023 年底，公募量化产品总规模达 3033.49 亿元，其中主动量化产品 1015.85 亿元，指数增强产品 1935.01 亿元，对冲型产品 82.64 亿元，比 2022 年底增加 527.62 亿元。对比主动权益基金而言，收益来源存异，Alpha 呈现较低相关性；对比量化私募，超额差异缩窄，整体相关性呈抬升趋势。

Beta 层面，相较同类产品，绩优基金呈现更为明显的小市值、低估值与低波动特征；策略层面，相较同类产品，绩优基金的基本面因子暴露更低，价量因子暴露略高。

从产品类型来看，指数增强占据绝对优势，500 量化产品规模扩张明显，以中证 1000 和国证 2000 为代表的小微盘宽基指数量化产品也不容小觑，同时主题与行业指增产品正呈现多点开花，逐步抬升趋势。《方兴未艾，砥砺前行—公募量化市场概况与策略解析》中金量化及 ESG，2023-11-28

从交易状态和体感来看，《2023 中国量化投资白皮书问卷调研》调研了从业者交易心情图和记忆图，大家对于 A 股的记忆贴近 A 股交易现实，大部分皆为“冲高回落、震荡下跌”，但是交易心情却是震荡模糊，很难对其归类。

■图表 问卷 A 股交易心情图



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 问卷 A 股交易记忆图



数据资料来源：《2023 中国量化投资白皮书》问卷调研

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

2.方法派系 迥然不同

招商证券策略团队把现有的 A 股投资人分为价值派和市场派，根据持股时间、预期波动以及预期回报分为 9 大类投资人，包括长期价值、逆向投资等。

不同类型投资人在强势期或增量期主导不同市场风格，本问卷也同时邀请各类投资人对自身策略进行标签。最受欢迎的策略是“自上而下”（36.77%），其次是“资产配置”（32.29%）以及“行业景气度”（24.66%）。这些策略通常涉及较为宏观、长期的分析和投资视角。例如自上而下需要识别宏观环境和产业趋势→确定适用的定价模型→确定选择属性→确认风格指数→确认选择行业→用核心指标筛选更有优势的具体标的→定性筛查，最终筛选出标的。这与量化投资的语境截然不同，或许也正因此，在社会舆论中难以在一开始获得理解。

■图表 A 股机构投资者的分类与其主要特点

属性	持股时间	预期波动	预期回报	流派
价值派	长	高	高	长期价值
	长	低	高	深度价值
	长	低	低	类固收/高分红
	中	低	高	逆向投资
	中	高	高	价值趋势
市场派	中	低	低	严格交易策略
	短	高	高	热点追击/事件驱动
	短	低	低	量化套利
	短	低	高	高频交易

资料来源：《A 股投资启示录》，招商证券

■图表 A 股投资者的标签



数据资料来源：《2023 中国量化投资白皮书》问卷调研

■图表 增量资金投资人主导的风格

个人投资者	一般情况下，个人投资者会更希望追求更大的价格弹性，个人投资者普遍对于上市公司基本面信息获取的渠道有限，股票选择时会更加倾向于叙事宏大、空间大、弹性大的股票，对于容易理解的大产业趋势相对更加偏好。也就是前面所述，更偏好的 FMvD 模型。因此，当个人投资者加速入市，体现为融资融券规模大幅增加时，市场可能会体现为小盘成长风格。
公募基金	主动偏股公募基金经理普遍对于基本面要求更高，对于 PEG、DCF 模型更加熟悉，基金经理可以通过深入研究调研，分析上市公司的业绩趋势，寻找业绩增速和估值匹配的标的；或者通过寻找股价低于内在价值的低估标的，那么公募基金成为主力增量资金时，市场通常会呈现成长风格占优。
外资	外资尤其是中长期外资，在选择 A 股标的时，更加看重上市公司经营和长期增长空间，更加愿意用 DCF 模型来看待 A 股的估值，因此，外资占据主导时，通常为呈现大盘风格，大盘价值或者大盘成长。
私募基金	私募基金有很多种类型，主动股票策略私募与主动公募基金类似，不过近些年规模扩张较快的是量化私募，量化私募更愿意在中小市值公司中寻找价格和基本面信号，量化私募占据主导时，市场呈现小盘风格。
保险社保	这两类机构更加追求稳健收益，对安全边际要求更高，更加偏好低估值高分红，因此，保险社融占据主导时，市场通常呈现大盘价值风格。从 2013 年开始 A 股的主力增量资金经历过多轮变迁。

资料来源：机构投资者业绩突围的八个思维模式和陷阱——A 股投资启示录（二十） 招商策略研究 2022-09-08

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

调研数据显示，有超过一半（55.61%）的受访者对量化投资持有负面情绪，主要原因包括信息不对称、量化复杂性、媒体报道、负面新闻和利益冲突。其中最主要的原因信息不对称（32.29%），不少受访者认为量化投资缺少透明度和沟通，导致其不明白量化团队的操作和决策过程。而量化本身的复杂性（26.01%），涉及高级数学和统计方法，对非专业人士来说较难掌握。仅有一小部分受访者（11.21%）是因为利益冲突，担心量化投资会影响到他们的盈利，尤其是在短线交易和高频交易领域。

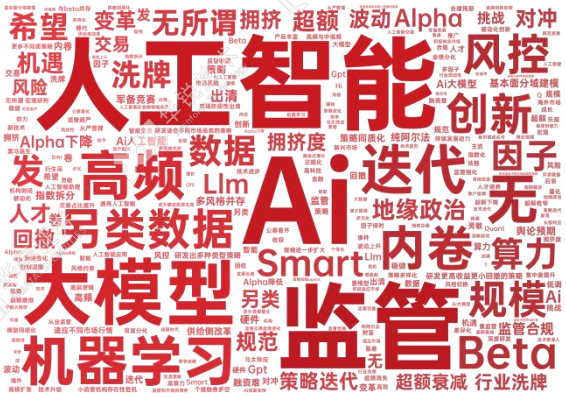
具体来看，公众自认为对“策略”这个词理解程度最高，可能源自策略是量化投资中最直观和最容易与投资成果联系起来的部分，也有可能是策略相对于其他更技术性的环节更易于理解和沟通。其次为所有投资人都会观察和分析的数据。

但公众对于因子、模型、组合优化、交易执行等的理解程度逐步降低，因为相对更为复杂和技术性的环节，通常需要较强的数学、统计学和计算机科学背景，尤其是组合优化，涉及风险管理、资产配置等概念，而交易执行则涉及市场微观结构、交易成本估计等。【详见本书第三章】

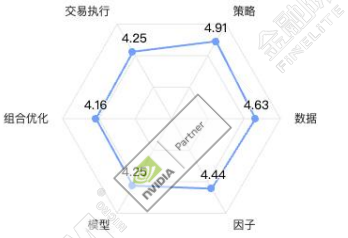
■图表 回顾 2023 量化三大关键词词云



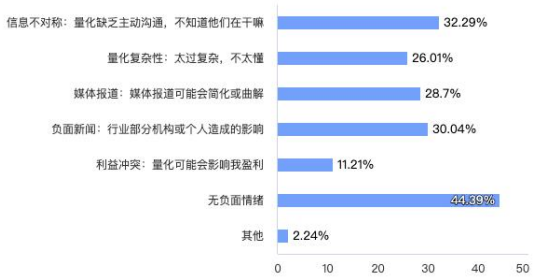
■图表 展望未来 3 年，量化行业三大关键词词云



■图表 公众对于量化投资的各个流程理解程度



■图表 公众认为对量化负面印象的成因



数据资料来源：《2023 中国量化投资白皮书》问卷调研

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

3.舆论场域 深度背离

正如投资人所说：当下市场很多东西都摇摇欲坠的时刻，每一个“瓜”的出现都会动摇行业可信度的根基，提升未来的沟通成本，让整个行业多年才逐渐建立起来的认同和美誉度以比建立时快得多的速度崩塌。

从 2013 年底，量化关键词搜索指数逐步上升，到 2021 年量化私募进入 1 万亿规模，行业一举一动都处于媒体探照灯中，百度指数量化关键词咨询指数呈现脉冲状，且中性偏负面新闻居多。

从 2021 年开始就已经有“全国限电影响量化机构的硬件算力”“百亿私募被管理层强平”等谣言。但更屡见不鲜的是“杀一个量化祭天”“量化滚出 A 股”“天天过山车，月月轮动，肯定是量化干的”等发言。

过去三年量化在非议中一路发展，时至 2023 年情况越演越烈，量化私募在不断辟谣、道歉、反思，有的机构认为行业是“背着锅把钱挣了”，其中辟谣内容方向主要有以下几种类型：

•信息谣言型：主要为量化私募“面临破产”、“老板跑路”、“大幅清盘”等新闻。典型如 2024 年 2 月，盛传幻方“大几千亿产品清盘、自营盘巨亏 15 亿、DMA 欠款 60 亿、公司破产”。

•花边新闻型：媒体对于量化机构的私人生活、高额薪酬保持较高关注。

•情绪爽文型：许多自媒体以迎合情绪为主要目标，在不了解量化策略具体实施的情况下，炮轰量化，将其与其他投资人的亏损或大盘的下跌紧密挂靠起来。

如 2023 年 8 月的热文《量化基金，今天砸盘的主力军》，掀起了主观投资者与量化投资者的论战。当晚，九坤投资创始合伙人、CEO 王琛在朋友圈转发一篇名为《砸你妹的盘，量化才是大 A 的脊梁！》文章，并发表点评称：“中国量化已经承受了太多莫须有的恶意，说量化砸盘纯属无知，永远满仓的中国量化基金，才是大 A 的脊梁”。次日上午，靖奇投资在其官方公众号发布文章《对不起，是量化的错！》称：“我们是量化，我们是最坚定的做多者，永远满仓热泪盈眶，不背锅。”

但从《2023 中国量化投资白皮书》调研的结果来看：《满仓交易，大 A 脊梁》这个观点所受到的公众支持度仅为 1/3。

量化机构在反驳量化有害时，公众最为认同的观点分别是：1.增强市场流动性；2.提升市场效率；3.促进市场创新；

最不认同的观点分别是：1.没有影响个人投资者；2.满仓交易，大 A 脊梁；3.全市场统一规则，交易公平。

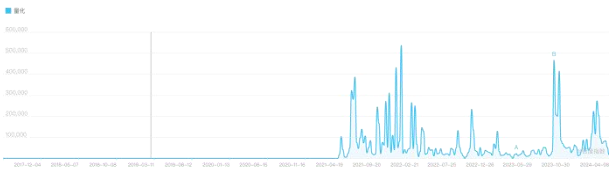
■图表 百度指数量化关键词搜索指数（2011-2024）



该数据为：互联网用户对关键词搜索关注程度及持续变化情况。

算法说明：以网民在百度的搜索量为数据基础，以关键词为统计对象，科学分析并计算出各个关键词在百度网页搜索中搜索频次的加权。根据数据来源的不同，搜索指数分为 PC 搜索指数和移动搜索指数。

■图表 百度指数量化关键词资讯指数（2017-2024）



该数据为：新闻资讯在互联网上对特定关键词的关注及报道程度及持续变化。

资讯指数：以百度智能分发和推荐内容数据为基础，将网民的阅读、评论、转发、点赞、不喜欢等行为的数量加权求和得出资讯指数。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

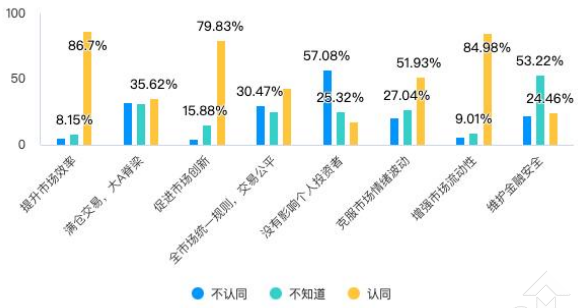
■图表 百度量化搜索指数-需求图谱



该数据：通过用户在搜索该词的前后的搜索行为变化中表现出来的相关检索词需求

算法说明：综合计算关键词与相关词的相关程度，以及相关词自身的搜索需求大小得出。相关词距圆心的距离表示相关词与中心检索词的相关性强度；相关词自身大小表示相关词自身搜索指数大小，红色代表搜索指数上升，绿色代表搜索指数下降。

■图表 量化机构在反驳量化有害时，公众对其观点认同度



数据资料来源：《2023 中国量化投资白皮书》问卷调研

■图表 2023 年 1 月–2024 年，量化机构辟谣不完全列表

信息谣言型	2 月 20 日，市场传言再起，“大几千亿产品清盘、自营盘巨亏 15 亿、DMA 欠款 60 亿、公司破产”等涉及幻方、明汯、九坤等多家头部量化私募相关内容在业内传开。
	2 月，北京涵德出现“巨额赎回”的消息引发市场关注。对此，公司第一时间发文回应，本次赎回金额 500 余万元，赎回金额较小。
	2024 年 3 月，有市场消息称，“各家券商及子公司全面梳理排查本单位离职员工是否有到宁波灵均工作的人员”。对此，宁波灵均投资官方表示，消息属冒充国家机关，恶意造谣、涉嫌违法。
	近日，有传言称“商品期货市场量化交易被叫停”，证监会有关部门负责人 1 月 4 日表示，相关市场传闻为不实消息，
	关于媒体报道的相关部门将召集量化基金去青岛开会，相关制度也将在 10 月份之前出台的消息，记者对此向数家百亿量化机构核实，得到的回复均为“未接到相关通知”。一家百亿量化私募机构相关人士对记者表示，我们没有接到通知，而且为什么召集大家跑去青岛开会呢。
	8 月 28 日，印花税减半征收落地后，当日 A 股市场各主要指数呈现高开低走震荡局面，与此同时，一则传闻在市场中流传，“考虑到印花税减半政策对市场流动性的影响，防止对客户投资造成不利影响，与策略供应商沟通 8 月 28 日暂停量化 T0 算法交易”。就在多家量化私募否认并表示，头部量化几乎都是自建系统，很少采购外部供应商系统，该消息或许为部分券商出于安全考虑的小范围暂停 T0 交易的可能。
情绪爽文型	经济学家任泽平指出，量化交易利用了散户所不具备的信息优势、融券优势、交易优势、大规模操纵个股优势等等，降维收割散户，不仅不为市场创造价值，还一年吸走 1000 多亿，抽血 A 股，量化交易只是利用了 A 股散户多、波动大的市场特征而已。个别量化“大佬”不仅不为市场和社会做贡献，还高调炫富，纸醉金迷，造成了很不好的社会影响。
	《量化基金，今天砸盘的主力军》：量化交易策略所鼓吹的纪律严明，不受基金经理的心理影响，计算机自动化买卖，让那规模庞大的量化基金在今天成了最大空军。而国内的量化策略同质化极其严重，大部分基金的策略都大同小异，直接导致其交易出现趋同效应，成了今天坚决砸盘的主力军，把节奏给带起来了。昨天四箭齐发的政策，是无法直接带来增量资金的，一天之内最大的效果就只能是提振信心而已。结果有超过 1 万亿的对手盘丝毫不受政策影响，机械按照股价高低执行了坚决卖出做空的命令。你鼓舞再多的场外资金，短期内也压不住这群量化基金啊。利好带来的单日新增买入资金没有超过量化基金的单日新增卖出资金，所以出现了单边下跌的情况。
花边新闻型	百亿量化私募被罚，最新回应！掌门人刚拍得“沪上第一豪宅”…
	百亿私募稳博投资前员工唐宇在个人直播间向老东家讨薪
	上海锐天投资起诉幻方量化侵害技术秘密纠纷
	概率投资前员工在个人社交平台对公司“口诛笔伐”

数据资料来源：编写组网络综合

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 2023 年 8 月底主观与量化论战主要文章

日期	时间	文章标题	日期	时间	文章标题
8 月 28 日	19:03	量化基金，今天砸盘的主力军	8 月 29 日	16:06	大佬放狠话：量化基金才是 A 股脊梁
	21:02	砸你妹的盘，量化才是大 A 的脊梁！		18:41	昨天低走，今天大涨，量化，外资，谁在搅屎棍?
	21:52	听说，今天有人黑量化～		21:19	打不过量化，就想毁了量化
8 月 29 日	10:18	对不起，是量化的错！		22:22	主观和量化，谁才是大跌下的软骨头?
	16:11	讨伐量化，理解量化，加入量化	8 月 30 日	02:15	没有砸盘！头部量化掌门人“抱团澄清”
	10:53	量化基金：对不起，我又错了		07:18	吵翻天！都是量化的错?
	11:42	勿谓言之不预！对量化砸盘鸡精的最后警告！		15:08	限制量化交易应从制度性不公入手
	16:01	这根阳线！太重要了！昨天“黑”量化的，今天你又怎么说呢?		15:11	传说“量化基金”将被规范！
	16:03	量化投资在舆论中天生就是弱者		16:23	量化：披着科技外衣的追涨杀跌，和成建制的消灭市场有生力量

数据资料来源：公众号《量化投资与机器学习》

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

02 量化机构间的存量博弈

1.行业：扩张趋缓 格局渐稳

2023 年，A 股量化进入存量博弈。这种存量表现在：

1.增量资金陷入停滞，从 2021 年后，量化私募规模就停留在 1.5 万亿左右，几乎没有较大变化。

2.量化私募数量进入滞缓，新备案私募门槛增高，2022 年以后成立的私募共 299 家，成长到 10 亿以上私募 11 家，占比不到 4%。纯量化策略机构 4 家。

3.规模分布和结构稳定性增强，逐渐拉出梯次，机构管理规模更多体现为阶梯内排名的变化。2021 年左右，行业主要关注 100 亿以上机构，但机构进出频繁，目前 100 亿以上已经分出多个区段，200 亿以上机构基本站稳脚跟，主要都是“熟面孔”，**形成了越来越快的“飞轮效应”：业绩好的机构形成口碑和规模，吸引了人力、资金、资源的大幅进入，同时算力、数据、投研等方面不断积累，资管规模不断扩大，而资源的增加又促成了良好业绩的产生。2023 年，规模 50 亿以上机构从 41 家增加至 50 家，数量占量化私募整体约 5%。**

这也与《2022 年中国量化投资白皮书》问卷调研结果接近，在描述未来 3 年量化行业竞争状态时，受访者意见分歧虽大，但接近 40%的受访者都认为未来量化行业将越来越集中，大型机构将会因规模经济和更强的资金吸引力而继续扩大其管理规模。

大型机构维持护城河，其投资风格上会更为稳健，更加倾向于追求超额的稳定性，对交易频率依赖相对较低。如上章数据显示，过去一年，规模 50 亿以上管理人的平均值和中位数收益更为平稳，在微盘股上投资较慎重，年内在 30 亿流通市值以下的股票投资占比提升在 1%附近，部分管理人会有更加明确的风控限制。反之，50 亿以下的管理人则更激进，有更好的灵活度追求高赔率，增加超额弹性，15-30 亿流通市值的股票最新的投资占比接近 17%，相较年初提升了 8 个百分点。

4.但这样的集中度上升，又未达到排斥新晋者的程度，仅有 7.61%的观点认为量化行业的格局将比较稳定，对于规模在 100 亿以下的机构而言，未来将更难以取得显著的突破。

■图表 2022 年至今新晋 10 亿以上量化私募

公司简称	成立时间	管理规模	办公城市	核心策略	核心人物
阿尔法喵私募	2022/9/6	20～50 亿	广州	FOF	袁义
上海波克私募	2022/7/25	50～100 亿	上海	复合策略	杨秋悦、陆思横
上海睿量私募	2022/7/7	10 亿～20 亿	上海	量化多头	王博、唐嵩松
磐松资产	2022/6/29	10～20 亿	上海	量化多头	张潇、吴确、李光、黄大洲、王栋梁
北京正定私募	2022/2/16	10～20 亿	北京	量化 CTA	李晓潇、蒋林浩
齐家私募	2022/2/15	100 亿以上	北京	量化多头	钟晓婷、陈韬、李薇、申宏威、齐俊杰
亦赋私募	2022/1/5	20～50 亿	杭州	复合策略	陶晓东、罗韬

数据资料来源：私募排排网

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 2023 年一季度及 2024 年一季度量化核心机构分布

资产规模	数量	2023 第一季度	资产规模	数量	2024 第一季度
50–100 亿	15	艾方资产、涵德投资、赫富投资、宽投资产、龙旗科技、洛书投资、蒙玺投资、平方和、千宜投资、乾象投资、锐天投资、思晔投资、仲阳天王星、象限投资、展弘投资	50–100 亿	19	艾方资产、白鹭资管、呈瑞投资、国恩资本、赫富投资、宽投资产、量派投资、洛书投资、蒙玺投资、磐松资产、平方和、千宜投资、申毅投资、盛冠达资产、思晔投资、天算量化、图灵基金、仲阳天王星、象限投资
100–150 亿	8	白鹭资管、博普资产、茂源资本、念空念觉、千象投资、稳博投资、致诚卓远、卓识基金	100–150 亿	14	博普资产、进化论资产、龙旗科技、鸣石基金、念空念觉、千象投资、千宜投资、乾象投资、锐天投资、思勰投资、通怡投资、顽岩投资、展弘投资、致诚卓远
150～200 亿	6	黑翼资产、金戈量锐、进化论资产、鸣石投资、世纪前沿、思勰投资	150–200 亿	6	黑翼资产、金戈量锐、茂源资本、信弘天和、因诺资产、卓识基金
200–400 亿	5	佳期投资、宽德私募、启林投资、天演资本、因诺资产	200–300 亿	5	启林投资、世纪前沿、稳博投资
			300–400 亿		宽德私募、天演资本
400–600 亿	3	诚奇资产、衍复投资、金锝资产	400–500 亿	5	诚奇资产
			500–600 亿		九坤量化、灵均投资、明汯投资、衍复投资
600 亿+	4	幻方量化、九坤量化、灵均投资、明汯投资	600 亿+	1	幻方量化

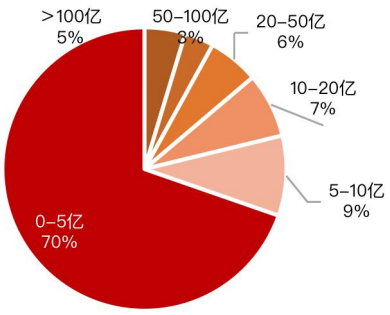
数据来源：公众号《量化投资与机器学习》

■图表 未来 3 年量化行业竞争状态



数据资料来源：《2023 中国量化投资白皮书》问卷调研

■图表 量化私募资管规模分布



数据资料来源：海通证券《2023 年度私募基金年度报告》2024-2-8

公募基金量化是不同景色。近年来，公募量化基金数量、规模稳步增长，截至 2023 年三季度，公募量化基金共 497 只。存续公募量化基金共 3112.51 亿元。当前，发行量化产品的公募机构共 109 家。其中，规模在 100 亿以上的共 9 家，50-100 亿的有 12 家。

规模头部公募主要分为两类，一类是 2015-2019 年规模发展起来的老牌公募；一类是 2021 年以来后来居上的量化新锐。包括博道、西部利得、浙商、万家、中欧、国金、国君资管等。

■图表 量化产品规模排名前列的机构（2023Q3）

基金管理人	合计规模（亿元）	各类型规模（亿元）				规模增长时期
		指数增强	主动量化	对冲基金	量化固收+	
富国基金	318.5	308.7	6.9	3.0	0.0	2015–2019 年
国金基金	299.6	22.7	276.9	0.0	0.0	2023 年

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

易方达基金	247.6	244.2	3.4	0.0	0.0	2016-2020 年
万家基金	191.2	158.7	32.5	0.0	0.0	2020 年
汇添富基金	177.2	134.9	0.7	41.6	0.0	2018-2021 年
华夏基金	140.5	112.1	16.3	12.2	0.0	2020-2023 年
西部利得基金	120.8	72.0	48.8	0.0	0.0	2020-2023 年
招商基金	111.5	77.5	34.0	0.0	0.0	2020-2023 年
景顺长城基金	109.8	66.8	42.0	1.0	0.0	2015-2019 年
博道基金	80.9	29.5	51.5	0.0	0.0	2019-2021 年
建信基金	79.7	79.1	0.1	0.0	0.5	2016-2019 年，2021-2023 年
天弘基金	75.9	75.9	0.0	0.0	0.0	2020-2021 年
中欧基金	73.2	25.8	47.4	0.0	0.0	2021-2023 年
兴证全球基金	64.5	64.5	0.0	0.0	0.0	2017-2021 年
嘉实基金	60.1	36.4	19.7	3.0	1.1	2017-2021 年
国泰基金	59.6	17.4	42.1	0.0	0.0	2020 年
申万菱信基金	58.1	41.8	15.7	0.6	0.0	2015-2020 年
长信基金	57.0	7.4	49.6	0.0	0.0	2015-2016 年
国泰君安资管	54.8	33.4	21.4	0.0	0.0	2021-2023 年
浙商基金	54.6	16.9	37.7	0.0	0.0	2019-2021 年
华泰柏瑞基金	54.1	14.1	39.1	0.9	0.0	2014-2017 年

数据资料来源：《量化这一年：希望、考验、突破》海通证券 金融工程团队 2013 年 12 月 11 日

如平方和投资演讲所说，当前量化策略面临三大挑战：

1.量化策略的超额收益不断下降。2022 年大量指数增强产品呈现出负收益，2021 年四季度中性产品在集体回撤。长期来看，整个行业的超额收益在持续下跌衰减，策略之间同质化现象明显，但“长期”多长是一个问号。目前来看国内 500、1000 指增超额还能做到 20%、30%以上，A 股依然是具有吸引力的 Alpha 市场，只是变得更困难。超额下降趋势变化缓慢但不会消失。对于量化私募而言，需要对赛道的选择和投研方法进行一些改变。横向对比美股，Alpha 衰减到个位数可能是量化的最终归属。

2.量化策略存在规模容量挑战。随着管理规模突破万亿，很多策略容量上限被撑破，对策略造成负面影响，降频和业绩下滑，导致大规模赎回。

■图表 量化策略的三个挑战

策略挑战	技术提升	产品设计
超额下降	超额下降倒逼技术进步	个股多空策略能显著增加收益
量化策略同质化现象明显	同涨同跌问题似是而非	打新红利仍然存在
规模容量的影响	策略容量魔咒可以破解	增强策略的配置

数据资料来源：平方和投资

超额的下降倒逼了机构技术的进步以延缓下降。目前市场平均水平与技术领先的管理人会有较大差距。

针对于同质性问题，排除掉指增 Beta 因素或者中性基差可以发现，不同频段、不同策略的 Alpha 会有相对较低的相关性，高中低频的 Alpha 策略，其涨跌路径、依赖条件并不一样。通常来说，例如市场低迷时，波动率策略会艰难；当市场出现小市值行情的时候，基本面策略会比较困难。

策略的容量、稳定性、收益率不可能三角互相制约、此消彼涨，所以 2021 年行业规模快速增长之后，量化机构多次出现

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

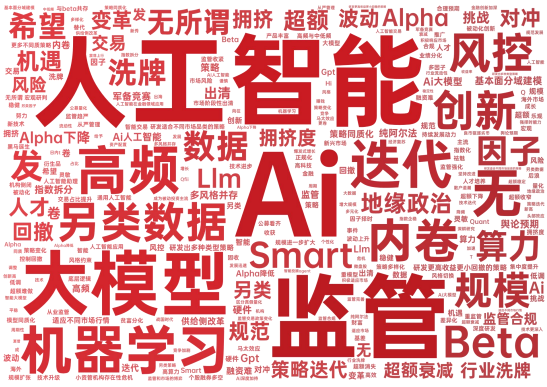
封盘，避免超额收益显著锐减，包括幻方、进化论、金戈量锐、启林、明汯、星阔等，2023 年机构陆续解封，或者是部分机构的策略有所升级，自认可以容纳更多资金，所以展开募集。或者认定未来市场会更趋于活跃，或交易成本更低，给交易策略更多空间。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

2.投研：AI、高频数据 汇聚主线共识

2021 年年初大盘股抱团瓦解，资金重新分配，随后进入了量化掌握定价权的小市值赛场。2023 年行业热点及风险聚集也就主要体现在小市值上。当任何一个投研新方法论或者风格暴露会带来 Alpha 甚至 Smart Beta 之时，机构就会蜂拥而上，涌向同一处 Alpha 源，甚至买盘成为 Alpha 本身。但实际上过去三年，人工智能、高频数据以及另类数据才是 Pure Alpha 真正的投研主线。

■图表 2023 年展望未来三年，行业三大关键词



■图表 2022 年展望未来三年，行业三大关键词



数据资料来源：《2022-2023 中国量化投资白皮书》问卷调研

■图表 2023H2 券商人工智能选股研究

券商	标题	主要方法	券商	标题	主要方法
建投	基于循环神经网络的选股模型初探	GRU	国盛	基于深度学习的指数增强策略	不同数据+多个 GRU+等权
民生	深度学习如何利用公募持仓网络优化选股效果?	基金持仓矩阵分解（微软亚研院）	东方	DFQTRA, 多交易模式学习因子挖掘系统	TRA（微软亚研院）
建投	资金流向相似性图谱与图神经网络	图神经网络	长江	神经网络因子挖掘-TCN 日频量价因子	TCN
建投	基于领域知识生成的基本面因子挖掘框架	遗传规划+枚举	建投	ModelZoo	Transformer 变式
广发	基于深度学习的高频数据因子挖掘	全连接	华泰	自适应网络：从削足适履到量体裁衣	门控自适应网络
东方	基于残差网络的端到端因子挖掘模型	ResNet	浙商	基于 Transformer 因子挖掘的指增策略	Transformer
建投	基于循环神经网络的选股模型改进	MOE、SkipConnection、多任务	华创	GRU 网络在风格自适应中的创新与应用	GRU
海通	多颗粒度特征的深度学习模型，探索和对比	多粒度网络（微软亚研院）+BiAGRU	华泰	基于全频段量价特征的选股模型	不同数据+多个 GRU+多任务学习
民生	从传统策略到深度学习的可转债投资	GRU 转债版	建投	基于时序神经网络的选股模型初探	多类时序模型
民生	基于可见性图嵌入的沪深 300 深度学习增强策略	对时间序列做 GraphEmbedding（微软亚研院）	东方	基于抗噪的 AI 量价模型改进方案	对抗训练、异常信号因子

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

建投	基于 TiDE 及其改进的因子融合模型	TiDE(Google)
招商	多模型集成量价 Alpha 策略	相同数据+多个模型+ICIR 集成

数据资料来源：《中国量化投资季刊•2023 冬》

■图表 公私募量化对比

类别		公募基金	私募基金
规则	交易限制	交易限制多，主要依托恒生 O32 下单，执行速度慢、日内反向限制。	交易限制较少，执行速度较快。
	交易成本	原先一般为万 8，降佣后为 4。	相对较低，万 1 至万 2。
	指增成分股	要求持仓 80%为成分股，日均跟踪误差通常要求不超过 0.5%。	无成分股和跟踪误差限制，可以在全市场范围内选股，或者大比例买入深度贴水的股指期货。
	风控要求	风控要求较高，对跟踪误差、回撤、因子敞口等通常有较严格的控制。有时，可购买的股票通常需要在公司的股票池内。	风控要求较低。
	先天优势	调研和产业资源较丰富，在基本面研究上具有优势。	数据和系统配置上的资源投入较大，技术具有领先优势。
	交易限制	交易限制多，无法做到纯自动化，主要依托交易员下单，执行速度慢、日内反向限制。	交易限制极少，执行速度极快。
投研	因子	以基本面因子为主，以人工挖掘为主	以量价因子为主，部分机构用机器挖掘
	交易	较少使用算法	算法应用度高
	指增做法	围绕指数成分股做收益增厚	无成分股要求，全市场范围内选股，通过风险敞口控制跟踪误差
	卫星策略	打新策略	T0 策略
	选股域	偏向大中盘，产品最初也以 300 增强为主逐渐拓展到 500 增强、1000 增强	偏向中小盘，产品以 500 增强为主，近年来 1000 增强发行较多

数据资料来源：《量化这一年：希望、考验、突破》海通证券，编写组

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

3.产品：寻路出海 产品创新

从产品结构层面来看，2023 年市场讨论最多的是股票多空和 DMA 产品。

多空策略被称为“增强型市场中性”，原因在于对冲精度、资金使用效率更高、成本更低、自由度更广、机构间差异更大：

1.融券对冲的对冲精度高：相比股指期货一揽子对冲，融券对冲可以在全市场选股，通过买卖个股实现多空组合的完全匹配，更好消除市场风险，具有更高的 Alpha 能力。

2.融券成本相对固定：不受股指期货贴水波动的影响，使得组合波动率、夏普率等收益指标更为优秀；股指期货只有三个标的，所有策略都在此体现自身供需，所以波动非常大，甚至一年之内能相差 10 个点。但融券券源有几千个，每只个股的需求都不一样，反映的是更真实的供需。

3.自由度更大：中性策略本质上做空股指期货一揽子股票，拆开之后也是个股，但个股之间的关系被确定了，可选择范围小，导致产品同质化相对严重；多空策略则突破了这种强制绑定的关系，自由度就会更大，不同管理人的选股模型差别较大，例如有的多空两端为同一个模型，有的则做不同模型，信号越强，能够腾挪的空间就越多。同时打破要对标基准指数而产生的行业、风格约束，可以全市场选股，收益空间更大。

4.融券对冲的资金使用效率更高：仓位可达到 100%，甚至有 1-1.2 倍杠杆，综合考虑收益与成本（成本由融券成本、个股折算率、融资融券比例限制、产品费率结构四部分构成），多头年化收益率约 15%。进一步考虑空头端市值增强作用，能够产生的增强收益约为总资产的 5%左右，费后约 3.5%。

然而，因为券源的限制，策略模型难度、组合优化难度、交易难度、券源实时管理能力均高于量化指增和市场中性策略，使得策略门槛较高，同时 2024 年 1 月 28 日，证监会重磅出手。为了进一步优化融券机制，加强对限售股出借的监管，全面暂停限售股出借，将转融券市场化约定申报由实时可用调整为次日可用，对融券效率进行限制。沪深交易所也同步决定，暂停战略投资者在承诺的持有期限内出借获配股票。【更多内容详见《2023 中国量化投资季刊·冬刊》-多空新元年】

同时 2023 年 10 月 14 日，监管发布新规，规定每家券商的 DMA 头寸不能超过 10 月 8 日的头寸，虽未禁止做 DMA，但不能再扩大规模。另外政策也对场内融券有较大限制，限制了保证金比例和大股东出借，随着 2024 年的 2 月风暴，DMA 与多空策略更是渐渐淡出主流观察视野。

出海也是市场热议的主题，很多国内团队来自全球最顶尖的对冲基金，同时拥有全球视野以及本土的多年深耕与积累，所以即便面对全球化的国际竞争者在投研的广度和时间上的历史沉淀，国内量化机构也总是想与国际机构掰一掰手腕。几乎每家头部量化机构都把“建立国际一流的量化基金”写进官网。另外一方面，海外市场由于良好的 Beta、低相关的 Alpha 以及人才战略储备的目的，诸多量化机构通过获取牌照、设立办公室、投资海外标的、招聘人才等具体行动开启了海外征程，尤其是头部、具备海外资源背景的机构更有余力。一般情况下，中国香港是境内管理人开展离岸业务的典型目的地，其次为新加坡和美国。

本问卷调研数据显示：分别有 17.14%和 13.71%的比例正在研究或已涉足港美股标的，积极探索全球多元化投资机会。从量化私募委托金融阶的数据来看，当前招聘海外市场主要为海外市场/渠道、产品经理、资深策略投研等岗位。

中国在量化投资领域的优势在于更深刻地了解本土国情、产业趋势和政策，在进行海外投资时，新的司法管辖区会面临新的监管框架，销售运营、办公空间和人才招聘等现实状况也会存有差异，所以战略和扩张计划一般比国内缓慢，且面临诸多挑战：

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

1.整个市场宏观的分析：不要仅仅关注于投资基本面，思维方式比认知更加重要。在不熟悉的市场中，细节会对投资收益形成影响，因此需要不断调整思维方式，规避风险。

2.策略投研本身：投资者结构的不同会导致投资偏好最终影响价格特征。目前海外市场优势在于工具更齐全，但机构化严重也使得定价误差更小 Alpha 更难获取，目前从国内跨境投资的基金评估来看，在港基金较难获取超额收益。

3.监管框架：涉及境外投资合规、资金安全退出等运营问题。另外还包括投资杠杆比例，海外的中性产品通常会加杠杆。虽然管理的是平层资金，但底层资金可能自带杠杆，这就在交易和波动的限制上需要做很大让步。

4.负债端管理：国内私募如果业绩亮眼可以迅速起规模。但海外投资人尽调要求更为复杂，除了投资策略之外还注重公司治理、内控机制和运营流程。导致尽调的周期和复杂程度远超国内，尽调和考察周期从半年到一年不等，因此规模扩张更为循序渐进。

5.技术框架：根据阿里云材料显示，量化机构通常需要将各种数据传输回本地进行投研，同时海内外员工也有加速访问办公网络的需求，快速实现境内 IT 环境与海外之间的互联互通成为了量化机构最为关注的解决方案。主要用于办公投研、实盘交易等的网络互联场景。量化机构尤其关注新加坡、中国香港、日本、美国等跨境链路网络之间的网络时延、建设周期、组网操作便捷性三个技术指标。基于以上现状，目前在出海业务中，大多数量化私募会优先考虑基于云供应商的骨干网进行快速的网络建设。以便捷解决全球组网复杂度、互联高可用、传输低延时的问题。通过构建跨境网络互联，实现云下办公与云上网络的互联；通过对应用加速，保障全球化金融商品市场的数据访问稳定性。

■图表 国内量化机构纽约扩招岗位表

机构	办公室	资管规模	机构	办公室	资管规模
明汯投资	上海/纽约/北京/香港	600+亿	宽德投资	上海/北京/深圳/珠海/纽约	300–400 亿
佳期投资	上海/北京/纽约	300–400 亿	信弘天禾	上海/北京/纽约	100–150 亿

数据资料来源：编写组整理

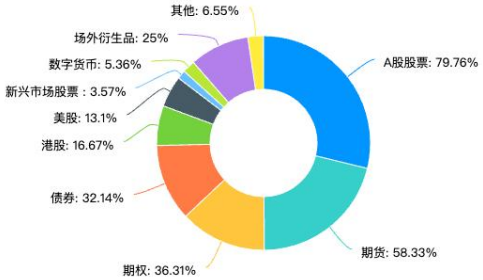
■图表 部分机构香港 9 号牌照获取情况

机构名称	获得牌照日期	机构名称	获得牌照日期
天马资产	2009 年 9 号牌照	鸣石投资	2021 年 9 号牌照
金锝资产	2016 年 9 号牌照 2021 年 2 月提交美国证监会	天演资本	2021 年 9 号牌照
富善投资	2016 年 9 号牌照	进化论资产	2021 年 9 号牌照
喜岳投资	2017 年 9 号牌照	佳期投资	2021 年 9 号牌照
幻方量化	2019 年 9 号牌照	金戈量锐	2023 年 9 号牌照
明汯投资	2020 年 9 号牌照	千象资产	2023 年 9 号牌照
九坤投资	2021 年 9 号牌照		

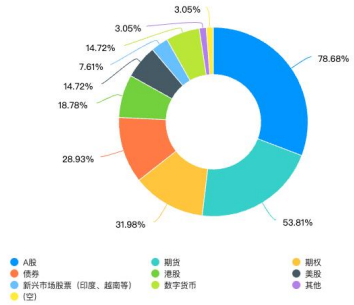
数据来源：香港证券及期货事务监察委员会、量化投资与机器学习

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 量化机构 2023 年投资研究标的



■图表 量化机构 2022 年投资研究标的



数据资料来源：《2022-2023 中国量化投资白皮书》问卷调研

■图表 量化机构 2023 年投资研究标的

选项	2022 年比例	选项	2023 年比例
A 股	78.68%	A 股股票	79.76%
期货	53.81%	期货	58.33%
期权	31.98%	期权	36.31%
债券	28.93%	债券	32.14%
港股	18.78%	港股	16.67%
美股	14.72%	美股	13.1%
新兴市场股票（印度、越南等）	7.61%	新兴市场股票	3.57%
数字货币	14.72%	数字货币	5.36%
其他	3.05%	场外衍生品	25%
（空）	3.05%	其他	6.55%

数据资料来源：《2022-2023 中国量化投资白皮书》问卷调研

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

4.市场：周度披露 持续内卷

从 2022 年，周度的各类信息披露就成为量化机构的紧箍咒，这些信息颗粒度极细，主要在业绩以及相关度比较，极少涉及策略解释。从表面上看，对于投资人而言，这种披露使投资人得以在及时性、透明度方面获得控制权，更高频的披露数据可以增加数据丰富度，投资者可以更快地了解自己投资的表现，有助于及时作出调整或决策，或更好地监测和管理底层基金的风险。

但这对于量化管理人来说亦有矛盾：一方面，周度表现卓越的量化基金可以更快崭露头角，也可以躲避投资人对于策略细节的追求，指导投资人直接查看净值投资。

但高频度的信息披露也使得市场更加短视，机构为了维持短期表现，可能会采取更加积极甚至激进的交易策略。对于投资人过于频繁的报告可能导致投资者关注短期波动，对其过度反应，而非关注长期表现，尤其是目前量化机构长周期负债端/资金端，高净值个人投资者占比仍然较多的情况下。在难以理解策略的情况下，投资人情绪较不成熟，而波动的市场也会放大投资人和管理人的情绪。

国内部分头部机构选择通过多年锁定期的产品设计，形成一个更长的投资观测期，尽量避免投资人因无序择时最终又踏空的情况。也有机构认为国内量化机构非常需要向海外学习负债端的管理，目前很多海外机构只会进行月评，从而保护管理人自身也优化资产端，减少博弈带来的摩擦成本，也减少客户因短期情绪波动而影响负债端。

2023 年 11 月 14 日，一则“百亿 FOF 私募华软新动力踩雷”的传闻引起轩然大波，使得市场业绩披露发生巨大变化。证监会表示，相关人员控制杭州瑜瑶、深圳汇盛等多家机构，多层嵌套投资、虚假宣传、报送虚假信息、违规信披等情形，还可能涉嫌违法犯罪行为，使得行业遭受重创。此事件引发了全市场对 FOF 和量化私募的信任危机，自此事件之后，许多 FOF 机构要求底层估值表从原来的二级改为要求穿透至四级，除了各类标的建仓成本和损益，还包括股票明细，并且要求托管方盖章。对此，许多资方与管理人仍存在分歧。有机构主动上交四级估值，例如磐松资产就直接在微信公众号公示披露政策：1.常规四级估值表，支持由机构客服递延一个月发送；2.数据脱敏版四级估值表，可由支持此功能的托管平台生成后即时发送。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

CHAPTER

第二章

阿尔法的生命力

```
alpha_growth = lambda r: r if r > 0 else 0
alpha = sum(alpha_growth(r) for r in [0.03, 0.05, 0.02, -0.01, 0.04])
print(f"Alpha thrives: {alpha:.2%}")
```



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

量化行业的发展迭代经历了一段先慢后快的过程，中间经历几个关键发展节点：

2008 年，美国金融危机，中国推出 4 万亿的财政政策，迎来华尔街量化人才归国浪潮，主要加入公募量化团队。

2010 年，沪深 300 股指期货推出。

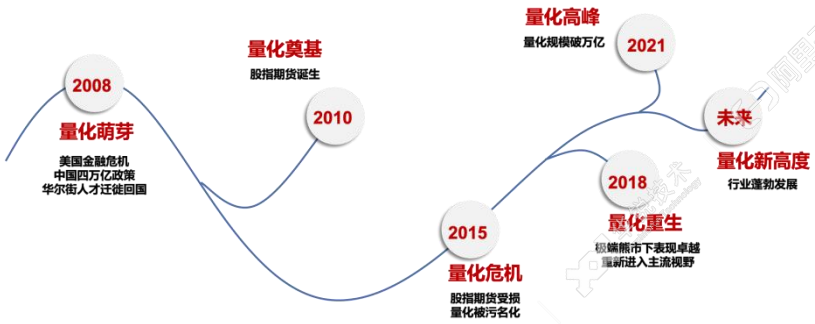
2015 年，股指期货受到限制，量化遭遇首次危机，之后负基差较大，量化机构发展受阻。

2018 年后，量化与主观多头表现一正一反，让市场意识到量化策略的优势和配置可能。同时 2018 年股指期货政策转向，量化进入一个逐渐放松的轨道。

2021 年，量化私募行业规模破万亿。

2024 年初，量化踩踏。

■图表 量化行业发展迭代重要时间点



数据来源：平方和投资

不同的市场环境之下需要不同的策略和技术，总体而言行业与策略经历了三个不同的阶段：

量化 1.0（2010 年-2015 年）：股指期货面世后，机构陆续布局量化，此时市场竞争较少，策略收益又高又稳。整体来说机构做多小盘、做空大盘就能够有一个非常好的超额收益。如九坤王琛所说，彼时中国量化行业的痛点是如何把‘知道赚钱的策略’变成‘能够赚钱的执行’。当时机构主要面临的挑战是数据和基础设施问题：1.拥有策略后，需要拿到服务器、对接接口，并解决机构账户、募资等问题，因为此时许多投资人对于量化还相对陌生。2.策略层面 A 股历史数据偏少，基于不够充分的样本进行统计决策，做出的决策受到数据限制，存在偏差。

策略层面，机构主要根据量价数据、技术指标、财务指标，建立经典因子库，做因子检验、行业市值中性化、因子正交、因子合成模型构建。

量化 2.0（2016 年-2018 年间）：2016 年开始，市场风格就彻底切换成大盘股行情了，叠加股指期货限仓及高额贴水，Alpha 必须扛过负基差才有正收益，量化市场进入艰难阶段，市场上出现了“Pure Alpha”的呼声，即量化必须控制住大小盘、行业风险，并且获得稳健的高水平超额收益。2017 年之后大量高频、偏 T0 策略出现，量化进入“高频指增时代”，量化管理人也开始新老交替，2018 年，随着“北九坤南幻方”以及“量化四大天王”出现进入了新一轮快速扩张期。

策略层面，机构开始涉及高频量价及另类因子、分析师预期 ，同时由线性回归逐步到机器学习，特征选择算法自动做特征检验，超越传统因子检验流程，便于超高维因子库。

量化 3.0（2019 年-2021 年）：量化行业飞速发展，规模不断增加，领域技术加速更新迭代，典型的表现就是 AI 化。对量化机构技术路径、技术栈、业务组织架构、团队构成等产生很大影响。量价策略和机器学习技术的红利得到释放，人工智能技术在量化行业的应用更是催生了很多的小型量化私募，形成弯道超车。2020 年之后量化私募规模暴涨，明汯、灵均、幻方均率先突破千亿规模，但遭遇规模“诅咒”后回撤。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

策略方面，机构深入应用 AI，包括自动化特征工程、神经网络结构设计、端到端学习、因子挖掘和因子合成使用同一目标函数进行优化、从维护因子库进化为维护模型库，有机构通过深度学习构建高维高频模型库，自动化挖掘量价规律

量化 4.0（2022 年至今）：量化进入了分化、再平衡的阶段。行业规模突破万亿策略容量趋近饱和，量化私募开始思考怎么消解多余的需求，其投研趋势之一体现在多资产、多信号、多频段融合。高频策略在降频提升容量，中频机构在中频策略基础上，不断向高频和低频扩展，提升业绩稳定性和策略容量，同时在量价、基本面、另类策略齐头并进。

■图表 量化私募机构的典型策略迭代

时间段	阶段	特征	某量化私募 1	某量化私募 2
2014 年	量化 1.0 时代	从传统多因子模型出发，采用线性因子组合，主要以挖掘中低频因子为主，在这一阶段奠定了公募基金更注重基本面因子，私募基金更注重挖掘量价因子的格局	——	线性模型
2015 年			——	机器学习应用
2016 年			日线级别数据开发线性多因子量价类策略。	分钟线级别数据
2017 年	量化 2.0 时代	中高频交易，量化市场拥挤度增加，高频因子在拥控度和超额稳定性方面均有一定提升，主动封盘控规模+开发低频或超高频策略	基于 Tick 级数据人工、机器挖掘多因子量价类策略。	中小盘市值中性
2018 年			算法交易、机器挖掘非线性多因子量价类、舆情类策略。	Tick 级数据
2019 年			迭代到机器学习中高频策略，储备高频，低频、舆情类策略。	T0 和 Alpha 独立运作
2020 年	量化 3.0 时代	人工智能算法和更强的 IT 软硬件以支持更高的策略迭代速度；人工智能算法在数据挖掘和开发非线性模型方面都有用武之地	机器学习、深度学习、多因子方法论研发多信号、多频段策略。	程序化 T0 和 Alpha 策略有机融合 300+500 轮动策略因子丰富度提升
2023 年			数据、底层策略开发迭代、机器学习策略组合，持续迭代中.....	

资料数据来源：海通证券研究所整理、量化机构材料

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

01 趋势 1 多资产：多策略

多资产指的是投资组合不局限于单一市场或资产类别，包括股票、债券、商品、期货、期权、ETF 等高流动性、多样化的工具。通过不同资产的组合来分散风险、优化回报、提高资本效率和适应不同的市场环境。

调研数据显示：2022-2023 年，A 股股票、期货、期权、债券仍然是量化机构投资最主流的资产类别，另外场外衍生品占比 25%。小众如新兴市场股票（印度、越南等）等量化机构也有所涉猎。对比两年变化来看，2023 年投资者对更传统和稳健的投资（如 A 股、债券）的覆盖度更高，而对高风险或波动较大的投资（如数字货币、新兴市场股票）的覆盖度耕地。此外，场外衍生品的出现占比显著，表明量化投资机构在寻求更多样化和定制化的投资工具。

需要注意的是，覆盖比例的减少并不绝对意味着投研该类别的机构减少，一般而言，量化机构在投入实盘交易之前会进行严格的历史回测和市场研究，一旦确定某个标的符合其策略且具有长期盈利潜力之后，机构会持续监控和调整策略，并不会轻易退出，更多是资金分配上的差异，所以机构研究标的数据可能是样本由于投资偏好的偏差，并非实际的投资额变化。

■图表 量化机构 2023 年投资研究标的

选项	2022 年比例	选项	2023 年比例
A 股	78.68%	A 股股票	79.76%
期货	53.81%	期货	58.33%
期权	31.98%	期权	36.31%
债券	28.93%	债券	32.14%
港股	18.78%	港股	16.67%
美股	14.72%	美股	13.1%
新兴市场股票（印度、越南等）	7.61%	新兴市场股票 [详细]	3.57%
数字货币	14.72%	数字货币	5.36%
其他	3.05%	场外衍生品	25%
（空）	3.05%	其他	6.55%

数据资料来源：《2022-2023 中国量化投资白皮书》问卷调研

多资产研究也并不一定意味着实盘，面对比较复杂的建模任务时，使用全量数据进行训练可能会带来意想不到的效果。例如金戈量锐、衍复等机构在路演或其材料中都提到，公司策略研发虽然投资 A 股市场，但实际在进行新策略开发时，并不是仅仅盯着 A 股市场做研究，而是研究全球股市，希望从不同国家的股市里研发出新策略，并行之有效地运用到 A 股市场。或者 A 股策略挖掘之后在全球多个国家市场进行测试，以提升模型普适性。以促进不同资产类别相互启发，资产之间存有共性。例如期货会给股票市场提供信息，例如包括黑色系、山东发现金矿金价下跌等对资源类股票有很大的帮助。

当前许多机构涉足多资产投研，然后在产品层进行权重投资，或者做反向波动投资组合，主要通过均衡风险配置来降低不确定性。

国内量化机构许多都已涉足多策略多资产，以因诺为例，除了包含套利、Alpha、选股、CTA、债券 5 大类标的子策略。还会根据实际情况分配权重，以期在保持中低波动的前提下，灵活配置多种策略，提升收益风险比。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 多策略配置比例变化

2015 年	70%套利+5%Alpha+15%选股+5%CTA+5%债券	2019 年	0%套利+80%Alpha+15%选股+0%CTA+5%债券
2016 年	5%套利+5%Alpha+20%选股+20%CTA+50%债券	2020 年	0%套利+70%Alpha+20%选股+0%CTA+10%债券
2017 年	5%套利+35%Alpha+20%选股+20%CTA+20%债券	2021 年	0%套利+60%Alpha+20%选股+5%CTA+15%债券
2018 年	5%套利+70%Alpha+5%选股+5%CTA+15%债券	2022 年	0%套利+55%Alpha+20%选股+5%CTA+20%债券

数据资料来源：因诺投资

多配置主要还是为了抵抗极端尾部风险。同样以因诺为例，其极端环境下的策略归因如下：

■图表 极端环境下的收益归因

时间段	关键指数表现	策略归因
15/6/12–15/7/8（股 1.0）	上证：–32.1%	套利策略弥补其他损失
15/8/17–15/8/26（股灾 2.0）	上证：–26.7%	套利策略弥补其他损失
15/12/25–16/1/28（股灾 3.0 熔断）	上证：–26.8%	选股策略亏损，套利和 Alpha 基本无法运行，CTA 盈利弥补所有亏损
16/4/21–16/5/24（商品大反转）	申万商品：–8.6%	商品连续暴涨后突然暴跌，造成 CTA 趋势策略大幅回撤，套利和 Alpha 策略无法运行，选股策略也没有好机会，产品出现回撤，但幅度小可控
16/11/14（商品黑色周五）	申万商品：–4.1%	商品连续暴涨后 1 小时暴跌，诸多品种砸至跌停 CTA 趋势策略损失严重，但 CTA 套利策略和选股策略盈利，当日盈利
16/10/28–16/12/20（债灾）	中证全债：–3.4%	产品债券头寸不高，几乎无影响
20/2/3（新冠疫情暴跌）	上证：–7.7%	资产均匀配置几乎无影响

数据资料来源：因诺投资

但国内量化机构较少从宏观层面进行配置，因为一般宏观量化对冲难度大、耗费人力，具有一定壁垒。

海外目前许多机构也会探索复杂的数学模型来识别价格形成/关系和市场趋势的关键驱动因素，构成投资组合并生成交易信号。重点往往放在宏观经济、地缘政治、政策变化和系统性风险等因素上，如 GDP、通货膨胀、利率、汇率、进出口、增长、资本流动、市场数据等，将以上因子纳入模型和风险管理框架中。

美国宏观量化长期平均夏普比率在 0.8-0.9 左右。仓位往往更偏向于相对价值，盈利与亏损交易的比率往往非常低。宏观量化会随着市场动态变化而动态调整投资组合规模。其他更成熟的基金可能会寻求采取各种市场状态识别措施来调节交易信号和总风险偏好。例如流量数据、天气分析、关键指数的波动性、跨资产关系/相关性。典型如 Bridgewater，总共 1500 亿美金的管理规模分布在两个策略里，一个是 1000 亿的 Pure Alpha，目标波动率为 18%，一个是 500 亿的 All Weather，目标波动率为 12%。其全天候模型也类似地将经济划分为不同的周期，并根据资产表现来优化资产配置。

半夏投资作为国内宏观对冲代表，主要方式是通过宏观环境择时判断，灵活配置股票、债券和商品，寻求在任何市场环境里都能获得超额回报。其宏观对冲方式为主要通过利率、盈利和风险偏好这三个宏观因子去判断是做债券、商品还是股票策略：

(1) 利率因子可以看市场资金供应和需求从而影响债券/股票的价格。

(2) 盈利因子看实体供应和需求影响股票/商品价格。

(3) 风险偏好因子看实体经济增长预期、金融风险预期影响股票价格。

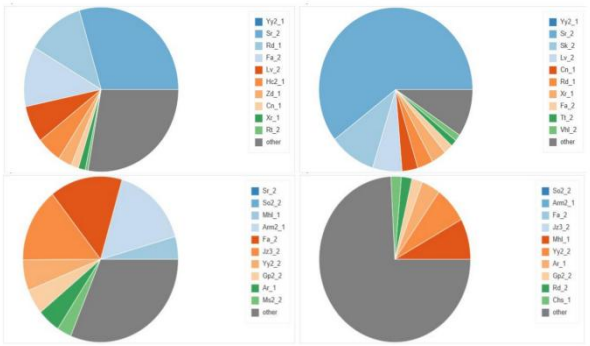
(4) 通过这三个因子指标组合观察，可以大致分析出全年的资产价格走势。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

02 趋势 2 多信号：大数据、强因子

多信号策略是指在进行资产的买卖决策时，不依赖单一的市场指标或模型，而是结合多个信号（或因子）来驱动投资决策。量化机构普遍储备多个低相关子策略，通过模型动态调节子策略权重，以应对不同的市场行情。目前国内主要使用股票多因子策略为核心，搭配其他量化策略以提供多种风险收益组合选择，以此形成更为完整的策略体系。

■图表 子策略数量储备：10+个



■图表 子策略相关系数：0.40 以内

	A	B	C	D
A	0.00	0.34	-0.11	0.35
B	0.34	0.00	0.05	0.11
C	-0.11	0.05	0.00	0.17
D	0.35	0.11	0.17	0.00

数据资料来源：复熙资产

再如天演资本其投资策略体系如下：

■图表 量化机构投资策略体系

日内回转策略	通过全市场选股策略构建股票底仓，提供券源供日内策略交易进行收益增厚
统计套利策略	用数理统计方法对投资标的历史数据构建模型，捕捉市场失效带来的价差套利机会
事件驱动策略	预测市场中事件性机会发生的时点和市场可能的反应，提前布局获取收益
新股申购策略	对新发行的股票价格进行研究，参与新股网下申购以增厚收益
管理期货策略	通过基本面和趋势类量化方法，判断中短期期货价格走势，进而配置多空仓位

数据资料来源：天演资本

很多机构也会将以上部分策略定义为卫星策略来为主策略做收益增厚。典型就是打新、日内、算法交易策略。

1.数据

经过几个代际的发展，量化已经打磨了策略各个流程的细节，形成规范。通常而言，量化投研步骤分为数据、因子、选股模型、策略、交易、风控、归因 7 大关键步骤。同时大型机构越来越多地采用精细化的研发生产流水线模式来实现策略，例如宽德投资形容其研发流程即如下：

第一步：大数据管理团队（识别获取海量数据+存储预处理原始数据）

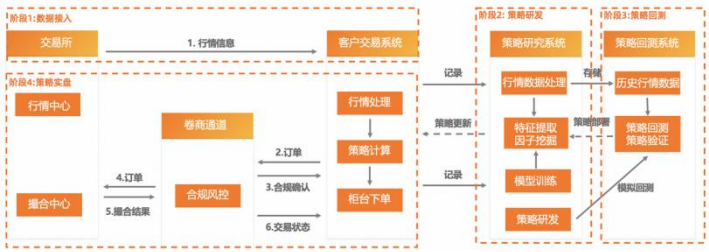
第二步：因子团队挖掘建模（基于技术面和基本面提炼数据+挖掘大量因子，测算因子收益贡献+因子叠加，AI 训练多因子模型）

第三步：投资团队构建资产组合（设计交易信号，构建并持续优化因子组合，策略叠加+策略团队回测历史胜率）

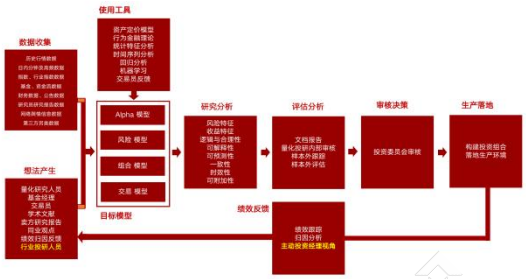
第四步：交易算法团队优化严控（采用中央优化器寻找最优价：超低成本极速交易+亚微秒级别超低延迟策略交易执行平台+异常人工介入）

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 量化策略基础流程



数据资料来源：阿里云



数据资料来源：玄元投资

其中，数据是量化策略寻找市场预测性信号的基础，量化机构使用历史数据来构建和回测其策略，预测未来的市场表现。数据的质量、数量和多样性直接影响模型的准确性和策略的有效性。金融数据与其他行业的数据不同，较为复杂。主要体现为低信噪比，意味着数据中噪音比例高，提取有效信号的难度比较大。模型在学习过程中，如果调整不够得当，就容易学习“噪音”。由于这一特性，在模型环节尤其强调避免过拟合。同时处理金融数据时要非常注重逻辑。以 A 股为例，不同的股票具有不同的涨跌停规则，另外新股上市以及复牌等行为，这些都要特殊处理。所以要在逻辑的基础上，对信息进行合理挖掘、过滤、组合。

■图表 常见数据处理流程



■图表 常见数据主要分类



数据资料来源：稳博投资

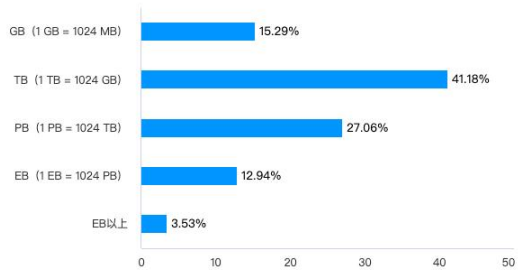
《问卷调研》显示：数据存储量级的差异主要由数据的更新频率、历史深度、细节程度和多样性影响。从数据量级上来看，调查数据显示，**TB 级别（41.07%）和 PB 级别（27.38%）是当前量化机构的主流存储规模，PB 级别（27.38%）才算真正的“大数据”量化机构。**

随着机构的资产管理规模增长，对于高级别数据储存的需求也呈现出增长的趋势。大型机构需要大量的数据来支持其复杂、多元化的投资策略，并且为了保持竞争力，不断增加对高频、多源和高维度数据的利用。

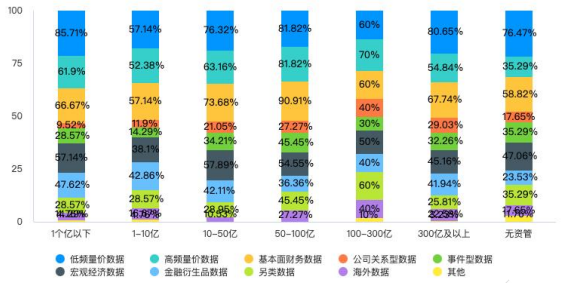
目前来看，1-10 亿级别机构就会渐渐从 TB 级别过渡到 PB 级别，数据分析上的深度和宽度有所增加。随着资产规模的增长，机构可能需要处理更多的数据以支持更复杂的交易策略。100-300 亿规模区间，大多数机构（60%）都在使用 PB 级别的数据存储。而到 300 亿及以上 EB 级别及以上的占比相对较高（总计 33.34%）。在数据获取、存储和处理上的能力远远超过绝大多数竞争对手。这些机构能够分析广泛的数据集，包括高频数据、另类数据等，来发现和挖掘 Alpha。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 量化机构目前数据储存量级



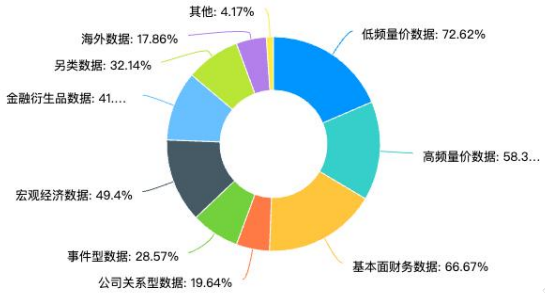
■图表 资产管理规模与数据储存量级比对



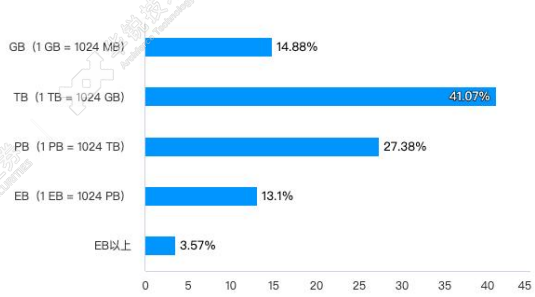
数据资料来源：《2022–2023 中国量化投资白皮书》问卷调研

机构当前使用的数据类型多样，调研数据的使用覆盖度表明，低频量价数据的使用覆盖度最高（72.62%），其次是基本面财务数据（66.67%）和高频量价数据（58.33%）。这些数据被广泛应用于量化投资模型中，因为它们能够提供市场趋势、公司表现和宏观经济状况的关键指标。

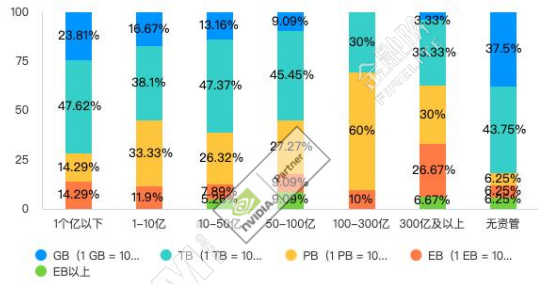
■图表 量化机构目前使用的数据：



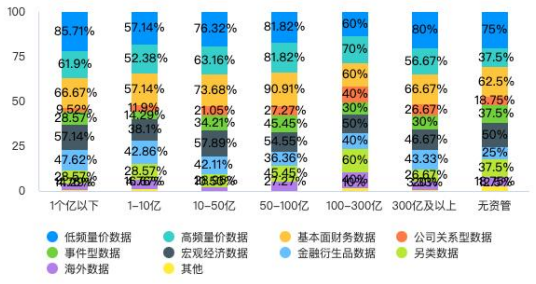
■图表 量化机构目前数据储存量级：



■图表 随着量化机构规模增长，数据储存量增加



■图表 资产规模与数据使用



数据资料来源：《中国量化投资白皮书》问卷调研

■图表 量化数据分类和规模

市场量价类数据	交易所市场价格、成交量等	
	日内订单簿	
基本面数据	上市公司公告	数千万条记录
	财务报告数据	500 万条时间序列
	研究报告	100–200 万条
另类数据	个股新闻	数千万条记录
	商品相关数据	20 万条

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

	宏观数据	35 万条
	产业相关数据	140 万条
	个股指标数据	10 万条
	物流数据	0 万条
	电商数据	数万条
	供应链数据	十万条
	持续增长中...	

数据资料来源：九坤投资

基本面数据（66.67%）用于评估公司的内在价值和经营状况，是许多策略的辅助支撑，更多是从业绩预期，或者是股价空间进行博弈。量价则是基于市场波动，更偏短期、市场情绪研究，两者相关度较低。量化机构将基本面数据提取为多类因子，例如盈利能力、成长能力、营运效率、盈余质量、安全性因子、公司治理、估值因子、分析师因子、股东和规模因子等，基本面量化因子目前在量化许多机构策略中占比约为 20%-30%，也有机构深耕此赛道，称之为基本面量化。

基本面数据虽然在 2010 年代就开始在国内被量化使用，但其作用远不如量价稳定、效果好。

原因在于第一代基本面量化：

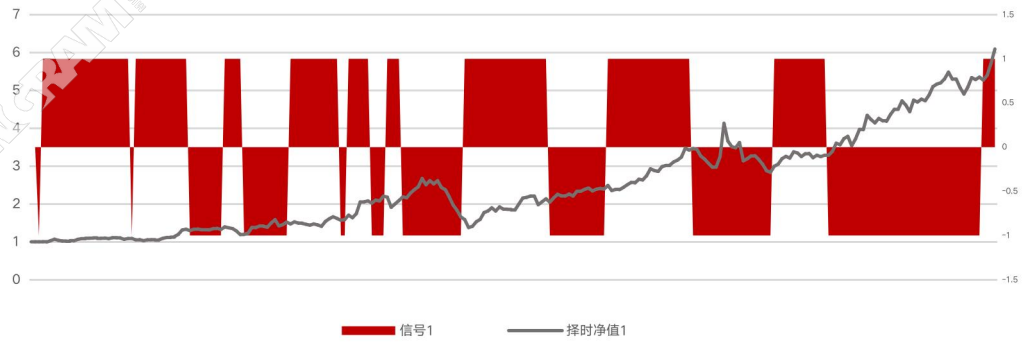
1.信息时效性较差：右侧交易是基本面量化天生难以克服的问题，基本面量化则更多受数据的驱动，当交易机会出现时，无论是财报、盈利预期、舆情、资金等层面，数据均滞后于主观认知，早已被市场消化。因为信息已经体现在价格中，当某个研究员调高了预期时，市场上的主观研究者可能早就察觉到了变化。

2.用法粗糙：很多机构把基本面数据直接用因子的研究方法放到研究体系里面，但纯粹基本面因子堆砌难以获得效果，所得结果既与主观认知背离，也难有收益弹性，难以向市场提供收益预期差，需要更为系统性的分析，找到驱动的核心逻辑。

所以量化交易需要通过结合主动研究、宏观研究以及行业配置层面的研究，尽可能地弥补右侧弱势的缺陷。**目前基本面量化的实践方式主要有：**

1.基金经理深度推票：由基本面研究员推票形成股票池，再通过量化模型从中择股。这方面公募量化拥有相当优势，例如中欧基金量化投资部身靠基本面研究团队，提供深度研究，全年实地调研总计 912000 公里，全年上市公司上门路演专家访谈、实地调研等总计超 20000 场，公司 FIELD 深度研究覆盖 A 股市场 40%的公司（按市值计算）。这是普通量化私募难以启及的信息优势。

■图表 专家洞见：某行业周期，经验投资人员提出行业核心指标，通过回测模型，印证指标的有效性



信号 1 取值正负 1，1 代表看多，-1 代表看空。择时净值代表模型在指数上进行多空择时后得到的累计收益率，统计区间：2000 年 1 月-2020 年 10 月。

数据资料来源：中欧基金量化投资部

2.另类数据引入：大数据部分使得信号更左侧，随着基本面数据的密度、丰富度、覆盖度不断提高，信息更加及时，公

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

司公告、业绩预告、分析师报告、新闻报道以及行业研究报告中包含大量非结构化数据，能体现公司基本面的边际变化、业绩驱动以及未来风险。经过重新加工降低无效换手，即可提高策略容量，同时带来更丰富的差额，提高超额的确定性。

3.构建垂类模型：将量化聚焦范围缩小到细分行业上，分行业构造垂类模型。

但这种方式同样存有挑战：1.有效且持续迭代基本面逻辑难度高，需要具备权益投资禀赋。2.成本很高，单条另类数据的年成本需要几到几十万，需要资源持续投入。

另外，对于相对低频的基本面数据和相对高频的量价数据的处理，一种思路是将高频量价数据生成的高频因子低频化；另一种思路则是对低频的基本面数据进行填充，从而保证在一个截面上可以用到研究需要的多类型数据。

■图表 当前基本面数据的使用

	传统基于财务报表的基本面量化	基本面量化 2.0
数据来源	公司的财务报表相关 结构化数据	公司公告、业绩预告、分析师报告、盈利预期、新闻报道、行业研究等非结构化数据
频率和时效性	季度报表数据 时效性较弱往往已经被市场消化无增量信息	日频、日内，丰富度和覆盖度越来越好 更加及时更加本质
分析方法及建模方式	依据传统财务分析方法构建财务因子，使用线性回归或多因子模型等建模	采用自然语言处理技术对文本信息进行分析和结构化处理，采用深度学习模型，从大量未标注的数据中学习数据的规律和特征，并且通过多模态融合识别复杂关系，通过自学习、自升级和演化的方式不断学习市场变化、提高模型的性能

数据资料来源：思源量化

■图表 基本面与量价的差异对比

	传统基于财务报表的基本面量化	基本面量化 2.0
数据来源	公司的财务报表相关 结构化数据	公司公告、业绩预告、分析师报告、盈利预期、新闻报道、行业研究等非结构化数据
频率和时效性	季度报表数据 时效性较弱往往已经被市场消化无增量信息	日频、日内，丰富度和覆盖度越来越好 更加及时更加本质
分析方法及建模方式	依据传统财务分析方法构建财务因子，使用线性回归或多因子模型等建模	采用自然语言处理技术对文本信息进行分析和结构化处理，采用深度学习模型，从大量未标注的数据中学习数据的规律和特征，并且通过多模态融合识别复杂关系，通过自学习、自升级和演化的方式不断学习市场变化、提高模型的性能

		量价	基本面
容量	持仓周期	日内到周频	周频到月频
	持仓数量	数百只至上千只	300 只至 500 只
收益	策略逻辑	重视统计规律	重视主观逻辑
	收益来源	情绪&波动	好公司&好价格
波动	风险控制	敞口控制	策略配置
	超额稳定性	相对较低	接近于“空气”指增，介于主观和指增之间

数据来源：煜德投资

调研数据显示：量价数据方面，**低频量价数据（72.62%）**最为基本，包括日 K 线、周 K 线、月 K 线数据，**高频量价数据（58.33%）**是近几年的主要赛道，研究中主要使用的是沪深交易所提供的 Level2 行情数据，即快照类数据和逐笔类数据。快

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

照类数据每隔 3s 对目前市场交易状况“拍照”，记录这一时刻的市场。逐笔类数据则是会对每一笔订单进行逐笔记录，刻画的是更细腻的交易结构和交易行为，从 2016 年开始就已经有大量机构开始实践，并结合人工智能算法一起使用。

原始的高频行情数据一般不能直接用作选股因子，在处理和分析之前，需要进行大量的数据清洗工作。在实际使用中，高频量价数据还会面临诸多挑战，例如：如何利用增量行情自动拼接全量行情，利用逐笔数据自动还原出频率更低的快照数据重构订单簿，从而获得任意时刻的全部买卖盘口信息，以及如何实时跟踪到订单簿的全貌，以此“对齐数据的颗粒度”等。同时，机构为了降低因子换手率，会对所有的高频因子进行月度频率的平滑处理。

逐笔成交数据是关于股票每笔成交明细的数据，包括成交量、成交价格、买方编号、卖方编号、买卖标志等信息。重构订单簿利用编号将相同买方或卖方编号的成交数据进行合并，可以还原得到完整的买卖单信息，并由此识别大单，构建大单成交金额占比、早盘大单买入占比等大单因子；根据买卖标志，可判断每笔成交属于买方或卖方主动成交，从而构建开盘主动买入占比、净主动买入与滞后或超前收益率相关性等因子。

高频量价数据优势主要体现在：体量明显大于低频数据，因此拥挤度相对较低、相关性较低、因子更具多样性。高频数据开发的因子一般调仓周期较短，意味着在检验因子有效性的时候，同一段测试期内具有更多的独立样本。例如，在一年的测试期内，只有 12 个独立的样本段用于检验月频调仓的因子，与之相比，有约 50 个独立的时段用于检验周频调仓因子，有超过 240 个独立的时段用于检验日频调仓的因子，而独立样本的增多有助于检验高频因子的有效性。

但这样也就带来诸多挑战：1.有些高频量价加工的因子的逻辑比较难以解释，只能经过严谨的统计检验之后方能使用；2.数据维度大、噪声高，可能掩盖真正的市场信号，导致误判。3.量大且复杂，容易出现数据缺失、不一致或错误。4.同时储存及调用如此大体量数据也面临技术挑战，如从 2018 年以来的全市场 level2 数据来说，所需储存空间就可达几十 TB，这相对于日度指标来说有约至少万倍的体量差距，对信息处理能力和处理效率的要求较高。

■图表 Level2 行情数据

数据类型	更新频率	数据描述	包含的具体数据	数据类型	更新频率	数据描述	包含的具体数据
快照类数据	更新频率为 3 秒；若数据无变化，则顺延至下一个周期发布。	也即行情切片数据，记录证券交易所某一时刻的静态数据信息，包括现货（股票、基金、债券等）快照、期权快照、大宗交易快照、指数快照等。	开盘价、收盘价； 当前时刻的现价； 截至当前时刻的最高价、最低价	逐笔成交类数据	实时推送	记录交易撮合系统中成交的逐笔记录。	连续的成交序号 买方订单号、卖方订单号 成交价格、成交数量、成交类型、成交金额等
			区间成交笔数； 截至当前时刻的成交量、成交额、委托买入量、委托卖出量、买入撤单量、卖出撤单量等 十档行情深度信息：买卖前 10 档的委托量、委托价、委托单数等	逐笔委托类数据		记录市场中实际订单的委托情况。	委托序号 委托时间，委托类型等 委托价格，委托数量，委托方向等
			买一档、卖一档的前 50 笔委托明细	逐笔合并类数据		仅上交所使用，将单只股票的所有委托和成交信息合并推送。	消息序号 逐笔成交类数据 逐笔委托类数据

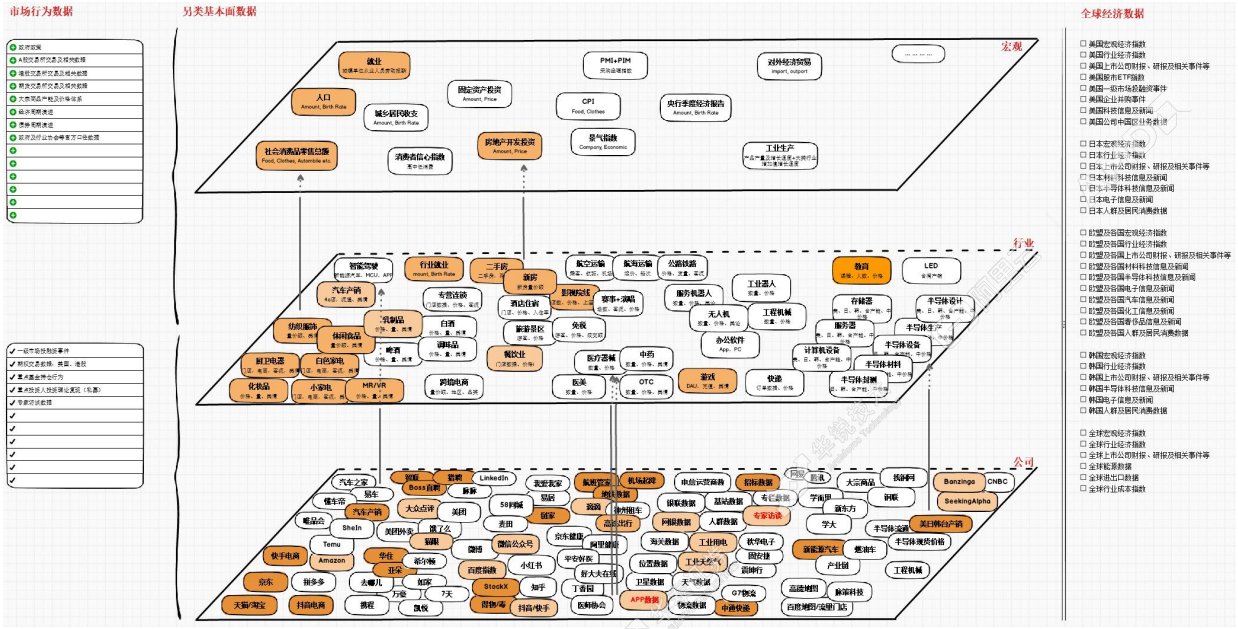
注：上交所和深交所推送的信息在委托类型、委托价格、业务序号、撤单记录等方面的数据定义都有所差异，这也需要在数据接收之后进行及时有效的处理。

资料数据来源：龙旗科技

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

宏观经济数据（49.4%）一般根据经济增长、通胀、利率、信用、汇率、期限利差等构建宏观风险因子，因子的上/下行是影响资产收益率的主要矛盾。宏观风险因子之间一般存在联动关系，许多机构从美林时钟出发识别大类资产的波动规律。但挑战在于：因子动量法在宏观环境发生切换时反应过于迟钝，而且宏观视角需要解决的难题是建立动态的宏观-行业映射关系。

■图表 宏观、产业、另类数据的映射



数据资料来源：中欧基金量化投资部

■图表 初始宏观指标库详细构成

增长：实体经济的“量”	通胀：实体经济的“价”
经济总量：GDP、工业增加值、PMI.	价格指数：CPI、PPI、RPI、CGPI
生产：发电量、汽车产量、货物周转量...	商品指数：CRB 现货指数、南华商品指数、柯桥纺织指数、Mylpic 矿价
投资：房地产开发投资、基建投资...	指数：综合...
消费：社会消费品零售、叉车销量...	单一品种：大宗价：猪肉、平均价：豆油、平仓价：动力煤、螺纹价格指
进出口：进口、出口、港口货物吞吐量...	数、期货结算价：布伦特原油、平均价：铜...
收入：公共财政收入、工业企业利润总额	
信用：广义流动性	**货币：狭义流动性**
货币政策目标：M1、M2、社融存量...	货币市场利率：SHIBOR、银行间质押式回购加权利率、回购定盘利率...
货币政策手段：央行逆回购净投放...	资本市场利率：国债到期收益率、企业债到期收益率、国开债到期收益率...
货币政策传导：M1-M2、M1-PPI、金融机构：新增人民币贷款、金	利差：期限利差、信用利差、中美利差...
融机构：各项贷款余额、金融机构：企业存款余额...	交易：银行间质押式回购成交量...
交易：沪深两市成交额、商品房销售额.	
其他	
资产价格：上证综指、深证成指、恒生指数、70 个大中城市二手住宅价格指数...	
外汇：官方储备资产：外汇储备、中间价：美元兑人民币、人民币：实际有效汇率指数	

资料来源：华泰金工

问卷调研显示：金融衍生品数据（41%）、事件型数据（28.57%）、另类数据（32.14%）和公司关系型数据（19.64%）使用场景也较为广泛，只是有的场景更为特定。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

许多机构把衍生品当成高杠杆或者对冲工具，或者将其当作独立或者混合产品收益增厚当中的一部分。数据显示金融衍生品数据（41%）使用较为广泛，但目前指导股票投资的公开资料较少。

另类数据广义上理解就是区别于量价和财务报表上的数据以外的新型数据，目前使用面最广的为分析师一致性预期及舆情数据。

分析师一致性预期数据的分析主要有以下几种方法：

方法 1（Alpha-Capture）：直接找到所有券商的行业研究员，让其定期给量化机构发送行业研究数据，投研报告业绩驱动、数据清晰，适合量化管理。所以大量的对冲基金和资产管理公司涌入，用人工决策或量化分析管理这些投研报告，人来挖掘 Alpha，而不是挖掘股票。Two Sigma， Mashall Wace， GSA Capital 等都是其中代表。

方法 2（卖方预期评价）：通过评估卖方研究员对于公司评价、预期盈利、预期增长等相关预期的变化以及相关预期的一致性和差异性，建立研究员评价体系，用以分配不同权重；区别于简单市场预期，研究员预期以及预期变化（边际变化）对于 A 股短期会有明显催化效果。

■图表 卖方分析师预期评价

预期盈利	研究员预期未来 1-3 年盈利均值 研究员预期未来 1-3 年盈利方差 研究员预期未来 1 年盈利均值变化：1 周/1 月/2 月/3 月变化情况	证券评级	研究员覆盖个数 研究员评级均值 研究员评级方差 研究员评级变化：1 周/1 月/2 月/3 月变化情况 研究员评级调整加权均值	预期成长	按照研究员预期未来 1-3 年盈利计算成长率均值 按照研究员预期未来 1-3 年盈利计算成长率方差 按照研究员预期未来 1-3 年盈利计算成长率均值变化：1 周/1 月/2 月/3 月变化情况
预期估值	按照研究员预期未来 1-3 年每股收益计算市盈率均值 按照研究员预期未来 1-3 年每股收益计算市盈率方差 按照研究员预期未来 1-3 年每股收益计算市盈率均值变化：1 周/1 月/2 月/3 月变化情况	目标价格	研究员预期目标价格均值 研究员预期目标价格方差 研究员预期目标价格变化：1 周/1 月/2 月/3 月变化情况 研究员预期目标价格调整加权均值		

数据资料来源：玄元投资

方法 3（情绪打分）：直接用原始的语料（如研报文本）送入 NLP 模型（如 BERT）中预测情感得分，并在历史上进行衰减加权，从而获得情绪因子。用于预测未来收益率。其背后的逻辑是，分析师报告中的一些信息，虽然可能并不是通过数字传递出来，但会从文本中一些词汇和结构中隐晦地体现出来。因为词向量中包含词频以及其他 TF-IDF 信息，模型可以利用这些信息进行预测，并将预测结果用于未来收益率的估计。但是由于年报中正面情绪的出现频率远高于负面情绪，因此在情感分析方案中，模型的偏向性会比较大。

方法 4（知识图谱构建）：对量化来说最缺的是对一个行业的知识图谱的构建，通过 NLP 对研报进行分析，然后把所有的信息抽取出来，得到在时间序列上研究报告在历史上所出现的二维结构。基于这样的二维结构可以对其中很多指标进行验证，看其是否对上市公司盈利的结果或者股价的结果有影响，是长期的影响还是短期的影响。第一步，如果把左边的研究报告变成右边的样式，就是非常典型的自然语言处理手段，左图很多是 PDF 格式的文件，要对它进行结构分段处理。在这个过程中，会得到一个与所有行业相关的研究指标或者研究框架体系。在时间序列上把各种多模态的另类数据来源进行交织，最后形成对价值的预测能力。

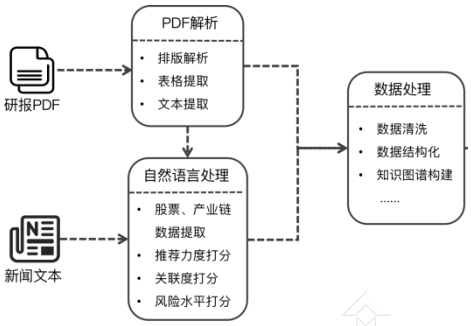
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 行业的知识图谱的构建



数据资料来源：九坤投资

■图表 研报及新闻文本的处理



数据资料来源：思源量化

2022 年发现市场上大量的分析师超预期因子失效，引发市场质疑，许多机构认为这部分的研究需要机构更为深入依托宏观层面的认知，以及配置层面的框架，形成管控能力。

事件、情绪维度的数据主要包括股东、管理层增减持、定增发行、解禁、高送转、分红除权、可转债发行、大宗交易、核心指数成份股调入、调出、权重变化、机构持仓动态、北向资金动态、舆情等等，由于信息的滞后性，这类数据也会需要关注市场反身性，例如九坤投资就曾指出：重大重组、重大合同签署等利好信息发布后一般会迎来下跌。因为重大利好之前股价已经经过了一波快速上涨，所以新闻数据最终结果可能是一个反指。

■图表 文本数据的事件提取，不同市场环境下的事件影响力

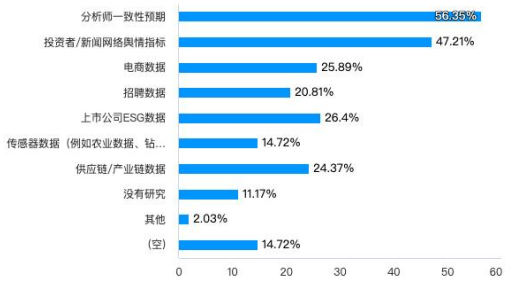


数据资料来源：九坤投资

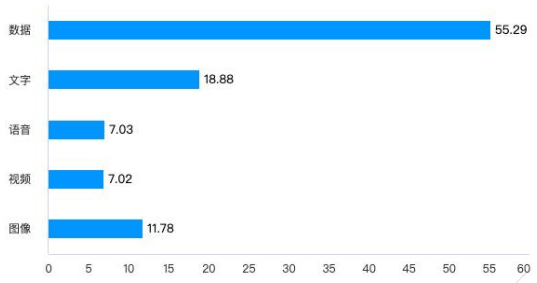
另类数据具备强即时性、覆盖样本广、信息量大等多种优势。目前，行业内另类数据的使用正在增多，宽德投资就购买了全球 10 万+品种行情数据库，还有 1000+另类数据源；另外例如龙旗科技 2022 年累计产出超过 1400 个基于另类数据的因子，数据源主要为：热度、招聘、新闻、专利等。这类数据更新频率不太一致，部分机构用做 5-20 天的中长周期预测。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 另类数据使用类别



■图表 另类数据使用介质



公司关系型数据（19.64%）一般用来做更前置的投资指引。例如亘禹资产就谈到，对 A 股来说不同板块的标的表现是有特色、有规律的。如果以宽基指数作为一个整体去回测，就会发现有很多数据和研究出来的结果对比看似有效，但是表征性不强。所以亘禹利用过渡因子的方法，先于市场判断当前是处于趋势行情还是动量反转，在不同的市场结构下（上升、下降或是震荡）对市场的表征因子做研究，找到当下的最优解。其因子挖掘的方式是首先根据行业把股票分类，通过产业链框架地图辅助细化全市场股票的分组，分组后的产业图谱目前积累 40 多个；随后再对分组后的数据进行研究来挖掘因子。

另外通过卫星、无人机拍摄、热成像、船舶 AIS 信号等现代科技手段获取的数据也部分应用于投资，尤其是 CTA 策略中。例如原油产业与贸易链条较为复杂，因此投研与交易对于数据广度和深度的要求极高。传统方式下对于石油市场基本面（供应、需求、库存、贸易）的跟踪主要倚仗官方机构的统计，虽然具有较好的权威性与准确性，但存在更新频率低、较为滞后、且数据精细度不足等显著缺陷。机构借助新手段从上游开采、中游运输仓储、下游炼化、终端消费等一系列活动进行跟踪，能够提供更加精细和高频的信息支持。例如华泰期货通过 AIS 信号、卫星数据等方式来跟踪全球海上航行油轮的动向，具体到该船目前所处的位置（海域/经纬度）、当前的航速、当前装载的货物种类和数量、始发港/目的地、过去的行动轨迹等。对这些数据进行整合后，可以推算出油轮未来的运动轨迹，从而预测油轮抵达目的港的具体时间。此外借助卫星实时拍摄库中浮顶罐的照片，结合模型可以获取实时的石油储存量，从而提升商品供需分析的精细度和即时性，更好地进行价格发现。

以上高频量价数据中“高频”容易引起争议，尤其是高频量价数据、因子、策略、交易以及程序化交易。其实五者大有不同。【更多内容详见《中国量化投资季刊 2022 冬刊》-高频交易篇】

■图表 高频相关定义辨析

高频数据	高频量价数据是指以非常短的时间间隔（如秒或毫秒级）记录的市场交易数据，包括价格变化（买入价和卖出价）和交易量的信息。高频量价数据可以揭示市场在非常短的时间范围内的动态变化，全市场可通过公开购买方式获得。
高频因子	高频因子是指在高频数据上构建的预测性信号或属性，用于捕捉短时间内价格变动的规律。例如，某个高频因子可能是基于最近几秒或几分钟内的价格动量或交易量的变化计算得出。但这些因子在交易时通常也会进行低频化处理，如前所述，例如通过逐笔数据重构订单簿，使用快照数据来合成分钟级别的 K 线数据。
高频策略	交易次数多，单笔盈利少。最终盈利取决于股票的日内波动和流动性。周期短、胜率高、盈利稳定，基本不受宏观和大盘影响。主要分为趋势性策略、价差策略、做市策略等。 **趋势策略：**往往投资于一个证券品种，运用技术分析或数学工具预测其未来价格走势，并据此确定建仓和平仓时点。只要预测方法能够保证一定的准确率并能抓住大的价格波动，那么这样的策略就有可能获得较好的累积收益。该类策略在商品期货和股指期货市场上已经得到广泛运用。 **价差策略：**往往会投资多个具备某种共性的证券品种，并认为这些证券之间的价差应该维持在一个均衡水平，不会偏离太多。如果发现市场上某些证券之间的价差过大，那么该策略便会做空那些高估证券，同时买入低估证券，直至价差回复到均衡水平再同时清掉多、空头仓位，赚取价差变动的收益。 **做市策略**：是因为其采用了类似于做市商提供买/卖报价（bid/ask price）方式，赚取买卖价差（bid-askspread），但其本身的目的

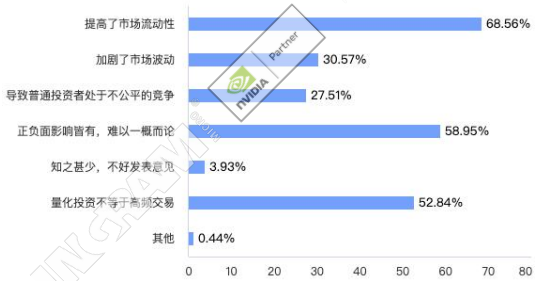
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

	并不是要做市，给市场提供流动性。这种策略通常需要对逐笔成交和挂单报价做建模分析，从海量数据中挖掘定量模式，挂单和撤单之间的时间间隔可能在毫秒之间（1 毫秒=0.000001 秒），因此也是技术要求最高的一种交易模式，通常也称作超高频交易(UHFT,Ultra HighFrequency Trading)。鉴于国内法规、硬件设施、交易费用的限制，该类策略在国内尚且无法实施
高频交易	高频交易是一种算法交易的形式，其中交易指令以极快的速度被执行，通常是毫秒甚至微秒级别。高频交易者利用先进的技术和算法来捕捉市场上的瞬息机会，并且通常会在一天内进行大量的买卖。目前，国内证券交易所对高频交易的筛选标准是，每秒最高申报速率 300 笔以上，或者单日最高申报 2 万笔以上。从筛选结果看，高频交易账户数量总体不多，但交易金额较大，约占程序化交易额的 60%。 接下来政策导向为：1.额外报告机制；2.重点监管；2 差异化收费：对行情使用费、流量费等相关费用做出差异化安排，适当提高交易收费标准，并可能收取撤单费等其他费用。
程序化交易	根据 2024-05-15《证券市场程序化交易管理规定（试行）》，程序化交易为通过计算机程序自动生成或者下达交易指令在证券交易所进行证券交易的行为。 《规定》要求：证券交易所依法建立程序化交易报告制度。程序化交易投资者应当真实、准确、完整、及时报告以下信息：（一）账户基本信息，包括投资者名称、证券账户代码、指定交易或托管的证券公司、产品管理人等；（二）账户资金信息，包括账户的资金规模及来源，杠杆资金规模及来源、杠杆率等；（三）交易信息，包括交易策略类型及主要内容、交易指令执行方式、最高申报速率、单日最高申报笔数等；（四）交易软件信息，包括软件名称及版本号、开发主体等；（五）证券交易所规定的其他信息，包括证券公司、投资者联络人及联系方式等。程序化交易投资者报告信息发生重大变更的，应当及时进行变更报告。 证券交易所对程序化交易实行实时监测监控，对下列可能影响证券交易所系统安全或者交易秩序的异常交易行为予以重点监控：（一）短时间内申报、撤单的笔数、频率达到一定标准，或者日内申报、撤单的笔数达到一定标准；（二）短时间内大笔、连续或密集申报并成交，导致多只证券交易价格或交易量出现明显异常；（三）短时间内大笔、连续或密集申报并成交，导致证券市场整体运行出现明显异常；（四）证券交易所认为需要重点监控的其他情形。程序化交易异常交易监控标准由证券交易所规定。 截至去年底，全市场报告程序化交易账户 11.9 万个，程序化交易不一定是高频的；它既可以应用于高频交易，也可以应用于日内交易甚至长期投资策略。

数据资料来源：编写组整理

问卷调研显示：**公众对上述这些概念的理解正变得更加细致和准确。**虽然有一定比例的公众认为高频交易提高了市场流动性，但也有不少人担忧它可能会带来市场波动和公平性问题。同时，认识到量化投资与高频交易之间区别的人数占比也较高，公众多元化的看法表明对于高频交易的影响，需要更全面和平衡地讨论。

■图表 公众如何看待高频交易



■图表 如何看待做空策略



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

2.因子

量化机构通过研究经验、卖方研报、信息搜集等获取研究灵感，生成因子。因子是量化投资策略的基础，用于模型构建和风险管理，并能够帮助量化投资者识别和利用市场中的系统性预测性信号。多因子，顾名思义，是庞杂的一个因子体系。多种类型、多个低相关的因子组合在一起能平滑收益曲线，最大化收益来源。当前因子主要有以下几种分类方法：

■图表 量化因子分类

因子种类	低频量价、高频量价、基本面、另类因子
因子预测周期	日内：日内高频（1-3min）、日内波段（10-15min）、日内长线（30-60min） 日间：高频 1-3 天；中频 3-10 天；低频 10 天以上
因子产生背景	人工挖掘（投资逻辑、数理统计、研究报告） 机器学习挖掘
Barra	市值因子等

Alpha 因子与风险因子容易倒换，常见的 Alpha 因子已逐渐演变为风险因子。例如 2019 年以来，无论是成长类基本面因子和价值类基本面因子的多空收益都不稳定。2019-2020 年，市场偏向成长风格，成长型基本面因子在这个阶段表现较优，价值型较差；近 3 年，随着市场风格转向，两类因子的收益出现反转。

按因子种类，在实操中，量化机构一般会最终在根据 300 指增、500 指增、1000 指增、中性等各类产品分配不同的因子比重，最常见的为 70-80%量价类因子+20-30%（基本面因子 + 另类因子）。

■图表 头部私募机构量化策略的因子构成

机构	因子组成	因子挖掘方式
机构 1	量价 70%、基本面 20%、另类因子 10%。	以机器挖掘为主。
机构 2	传统量价（中低频）70%，其余 30%。	传统量价均为人工挖掘的理性因子。AI 对挖掘因子、组合优化贡献度约 40%。
机构 3	量价因子 90%、基本面 5%、事件驱动因子 5%	以机器学习为主。
机构 4	包括中低频价量因子（60%）、事件驱动因子、基本面因子、日内因子和高频因子	以人工挖掘为主。
机构 5	中低频量价 70%、基本面及另类因子不到 20%。	因子挖掘人工为主。机器学习少量参与，AI 挖掘的因子会进行人为归因。
机构 6	以中低频量价因子为主，基本面因子和另类因子起辅助作用。	以人工挖掘为主。
机构 7	以量价为主，量价中中低频占比高	AI 挖掘因子占比在 50%以上
机构 8	量价 75%、基本面 25%。量价以中低频为主。	因子挖掘为纯机器学习，因子逻辑偏统计学，金融学逻辑较弱。
机构 9	量价因子 70%左右，其余为基本面、另类因子。	主要由机器学习挖掘因子。
机构 10	1000 指增，量价因子 70%、基本面因子+另类因子 30%； 500 指增，量价因子 60%，基本面因子+另类因子 40%。	因子 80%以上为机器学习挖掘，其余为人工挖掘

资料来源：私募排排网，Wind，海通证券研究所

高频量价因子是量化机构研究的重点，各家机构分类较多：

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 高频因子分类情况

成交分布类	主要包括成交量或成交额日内分布统计指标，用于反映成交日内分布情况；不同条件下成交占比用于刻画投资者交易行为或交易意图。
收益分布类	主要包括价格或收益率的日内分布统计指标，用于反映股价日内分布、买卖压力等情况。
波动跳跃类	利用高频数据刻画股价日内波动和跳跃情况。刻画了股票价格或股票收益在过去一段时间的不确定性程度，高波动率通常反映其不确定性程度较高，未来收益表现可能相对较弱。
动量反转类	主要对应日内动量和日内反转，利用高频数据增强动量效应和反转效应。包含三个细分类型：分段动量、量能动量和 QRS 动量因子。
资金流类	对成交单数据进行大中小单、主动买入卖出、资金流入流出等维度的划分，通过计算不同维度下成交占比等方式刻画资金流向、买卖压力、买卖意愿、追涨杀跌羊群行为、知情交易等情况。主要可分为以下三类：交易单量资金流向因子、开盘/尾盘资金流向因子和收益率成交占比因子。资金流因子反映了不同类型资金流向的信息，该类因子的核心思路是捕捉“聪明资金”的流向，采取跟随策略，以期获得基本面改善带来的超额收益。
流动性类	主要利用 Tick 数据，从宽度、深度、弹性等方面衡量市场流动性，追踪投资者交易意图。包含两类：价格弹性因子和集合竞价因子。
量价相关性类	主要基于成交量和股价的相关关系刻画成交量和价格双方走势的背离或一致程度，以及一些量价复合因子。可分为两类：价量同步因子和领先滞后因子。量能领先阶段往往是信息扩散的初期，但日内量价相关因子基本与未来收益负相关，
高阶特征	通过分钟级别价格数据计算出日内收益率的偏度与峰度，该类因子刻画日内价格分布以及快速变化的特征。
筹码分布	可分为以下两类：筹码分布形状因子和筹码占比因子。筹码分布旨在刻画股票持有人的持仓成本分布情况。
拥挤度	使用 30s 快照成交量序列构建了傅里叶变换机构拥挤度指标，原理是机构拆单下单形成的周期性可以被傅里叶变换后的更显著波形振幅占比捕捉，使用周期性成交占比表征机构参与度占比。
订单侵略性	订单薄性质的变化，体现了多空力量的动态变化，是未来价格走势的重要体现，一个订单的执行概率和订单薄的厚度、参与者对即将到来的订单的预期有关；买盘越厚，一个潜在的买家下 market order 的概率更大；这套说法对卖方同样适用；bid ask 的厚度体现了看涨和看跌者的相对力量。买家挂单越接近 bid1，越激进；卖家挂单越接近 ask1，越激进；体现了买家/卖家完成交易的迫切程度。通过整个订单薄，可以知道所有买家整体的激进程度和所有卖家整体的激进程度，从而构建一系列因子了。此外，买卖侵略性的差异，也是一系列因子。
订单薄形状	订单薄平均形态是有一块隆起的地方，离不开几种情况。1.顶点不同；2.可能是双峰的，甚至是多峰；3.可能是没有峰，就是矩形分布 4.可能就在某些档位有分布；5.可能是单调的，比如从第一档往后几乎单调上升或者下降，单调的函数可以是线性的，也可以是 concave 或者 convex；顶点的位置可以作为一个因子。比如，买单的顶点位置，是位于 0–10%、10–20%等的哪个位置；此外，还需要关注这个顶点位置随时间的迁移。例如订单薄斜率，描述订单薄上的价格和该价格处的挂单量之间的关系。
撤单	撤单行为，不仅反映了交易者的观点的变化，还会影响其他交易者；根据买卖双方各自的撤单量、撤单金额、撤单价位、大单超大单的撤单情况，可以构建很多因子。除了正常交易情况下撤单去追跑掉的价格外，还有订单为欺骗对手的作用，挂单者想制造假象，以诱导其他投资者。
事件聚集	例如撤单、挂买一单、超大买卖单出现，然后在一段时间内统计事件发生的频率，可以构建很多个系列的因子。这是高频数据低频化的常见思路。一些事件具有聚集的特性原因有：交易者拆分大单来掩盖交易意图；交易者之间的模仿；不同交易者对新闻的先后反应；为了在竞价上击败其他交易者。
逐笔数据主动买卖	判断一笔交易属于主动买还是主动卖，通常以一笔交易的买卖双方的订单到底先后顺序来定，生成日内累计主买率、日内累计资金净流入、日内累计大单资金流入率、日内累计小单资金流入率等等；

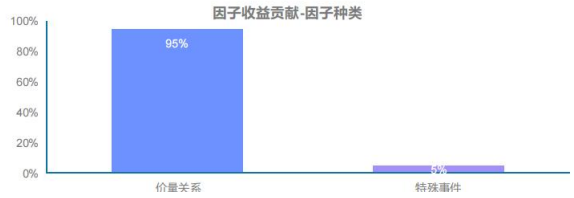
数据资料来源：量化投资与机器学习、中金多因子手册、中环狙击手

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 稳博投资日间因子分类及其收益占比



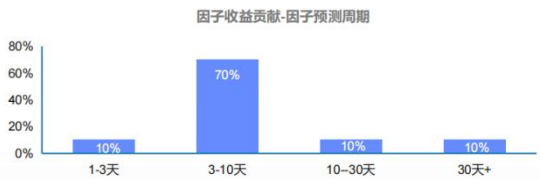
■图表 稳博投资日内因子分类及其收益占比



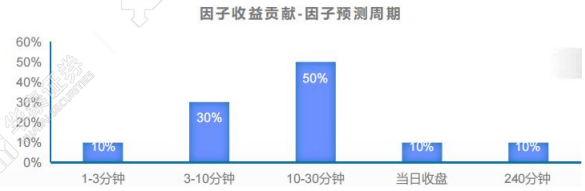
数据资料来源：稳博投资

日内量价类因子对于提升超额收益的稳定性贡献较大，但是在成交量低迷的时期日内因子可能会受到影响；然而长线的基本面、新闻和另类数据因为换手率相对较低，不受市场成交量影响，在交易量和波动率较低的市场环境中其对超额收益有较大的贡献，与量价类因子形成互补。因子多样性使得在不同的市场环境下，都能够有相对稳定的超额收益，同时具备较好的控制超额回撤效果。如赫富投资 2022 年做了大型升级，从 85%量价依赖降低到 70%，同时提高基本面和另类的比例，从 15%提高到 30%。部分机构会对因子进行重新定义，例如灵均的因子包含 30%基本面因子+35%传统交易逻辑因子+35%新型价量类因子；石锋量化因子库中 15%为概念\题材类因子，以把握短期主题概念机会。但本质上仍基本分为以上三大类，并且比例也较为接近。

■图表 稳博投资日间因子分类及其收益占比



■图表 稳博投资日内因子分类及其收益占比



数据资料来源：稳博投资

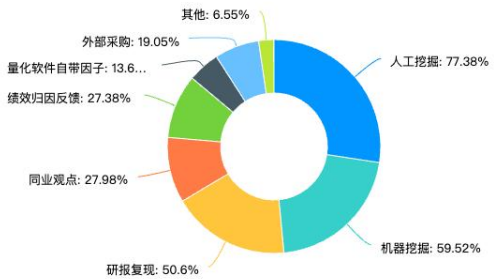
按预测周期：许多机构从分钟到季度都有分布，例如世纪前沿直接根据频率将因子分为高中低频三大类因子：

其中高频最重要的作用是交易执行，能够低成本甚至负成本实现中低频信号的超额收益，高频因子收益贡献占比 10-15%。

中频的价值主要体现在相对艰难的市场环境下，能够在风格快速切换时更快地适应变化，中频因子收益贡献占比 40-45%。

根据因子产生的背景：目前量化机构在因子挖掘上采用了多元化的方法，结合人工智能和机器学习技术的机器挖掘与传统的人工挖掘相结合，以在竞争激烈的市场中寻找到独特的 Alpha 源。

■图表 量化机构因子挖掘方式



数据资料来源：《2023 中国量化投资白皮书》问卷调研

人工挖掘（77.38%）和机器挖掘（59.52%）是最常见的因子挖掘方式。一般而言，前者以逻辑性因子为主，依赖于研究人员的专业知识和判断，包括可能影响资产价格的各种宏观和微观变量。其中典型代表如 World Quant，该机构进行工厂流水

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

线式的生产，在全球雇佣了成千上万人来做策略，然后由总部 100 个左右基金经理将策略进行融合，在过去 10 年积累了超过 1000 万个金融信号，但随后出现了边际效用递减。

机器挖掘（59.52%）以统计性因子为主，利用机器学习算法自动识别可能有预测能力的因子，发现人工方法难以识别的复杂模式和关系。有的机构将之称为 Quant 3.0 时代。机器挖掘出的因子也不会直接喂入模型，一般同样会通过人工进行逻辑可解释性审核。

■图表 统计性因子及逻辑性因子的使用条件

统计性因子	逻辑性因子
基本面因子统计时长不低于 10 年 T0 因子统计时长不低于 6 个月 兼顾长期历史有效和当下市场环境有效	检测每次逻辑成立时，对应的胜率以及极端情况下胜率和表现 因子提交后，在预设时间内通过样本外数据验证，统计学上显著有效，最终被采纳 因子通过检验时长 1 个月–1 年不等

数据资料来源：稳博投资

目前因子挖掘主要分为两种方式：启发式算法和深度学习。启发式算法即在一些既定规则的基础上，不断生成一些新的规则，但就像生物进化、基因变异不一定是向更好的方向，相对来说不那么高效。深度学习的特点是明确围绕损失函数最小化的目标，用梯度向量技术去不断迭代参数的最优化学习方法。最优化算法的目标明确，不需要人再去定义一些规则，网络可以做得高度复杂，搜索空间很大，效率也相对较高。这两类技术都有各自的生存土壤，启发式算法最后会给出一些既定因子、规则，所以可能更适合追求可解释性的资管机构，而基于深度学习的方法，可能会更加激进。

研报复现（50.6%）：研报复现主要是通过查阅公开的券商卖方、学术研究报告，尝试复现报告中提到的因子。

同业观点（27.98%）、绩效归因反馈（27.38%）、量化软件自带因子（13.69%）、外部采购（19.05%）也能帮助量化机构不断拓展因子的类别及数量。

■图表 稳博投资日间因子分类及其收益占比



■图表 稳博投资日内因子分类及其收益占比



数据资料来源：稳博投资

■图表 2024 年 Q1 卖方金融工程热点研报

东北证券	基于 CNN～Transformer 的深度学习模型探究	海通证券	深度学习因子的“模型动物园”
东方证券	基于异构图神经网络的股票关联因子挖掘	华安证券	加速换手因子，“适逢其时”的换手奥秘
	DFQ～HIST，添加图信息的选股因子挖掘系统		基于残差因子分布预测的投资组合优化
	周频多因子行业轮动模型		高成交量回报溢价与经济基本面
	基本面因子的重构		探索股价动态关联，捕捉属性敏感的动量溢出
	自适应时空图网络周频 Alpha 模型		基于端到端神经网络的风险预算与组合优化
东吴证券	换手率切割刀 CTR 因子 CutletsofTurnoverRate，换手率的异质信念	华泰证券	经济周期实证、理论及应用
	交易者结构对动量因子的改进		GPT 因子工厂，多智能体与因子挖掘

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

	换手率分布均匀度，基于分钟成交量的选股因子		如何捕捉长时间序列量价数据的规律
	宏观风险因子构建与大类资产配置应用	华鑫证券	基于技术面量化的指数基金绝对收益策略
	重拾自信选股因子更新～从过度自信到重拾自信	开源证券	订单流系列，撤单行为规律初探
广发证券	基于融资融券因子研究		KyFactor 特色因子体系与应用
	可转债高频因子探索		大小单资金流为核心的综合行业轮动方案
国海证券	DMA 策略带来的小微盘股风险释放及前景展望		行业动力学模型的周频应用
国金证券	基于日内高频博弈信息的商品 CTA 策略		百亿量化私募 2023 年业绩盘点
	追上投资热点～基于 LLM 的产业链图谱智能化生成	民生证券	红利策略优化，“高股息陷阱”与股息预测
	高频因子跟踪 Sora 模型有望进一步提升投研生产力		从增量学习到元学习，深度学习训练新框架
	机器学习全流程重构～细节对比与测试		迈向价值之路，竞争壁垒分析框架下的选股逻辑
国联证券	雪球敲入对市场影响研究	太平洋证券	基于交易异常的广义反转因子
	基于遗传规划的对传统量价因子的优化思路	西南证券	聚沙成塔，微盘股行情解析与微盘股投资策略
国泰君安	行业指数如何择时，通过估值、流动性和拥挤度构建量化择时策略		多维探究黄金定价逻辑，定量择时配置有章可循
	大类资产配置策略落地方法研究	信达证券	雪球敲入是否给市场带来压力?
国信证券	高频订单成交数据蕴含的 Alpha 信息		雪球陆续敲入，基差全面贴水
	个股与行业的共振，联合动量因子	兴业证券	年度复盘&展望三选股策略篇
浙商证券	多因子量化投资框架梳理	长江证券	交易行情高频因子收益来源

数据资料来源：公众号《量化投资与机器学习》

通过梳理上述 2024 年的热门研报，可以提炼出量化投研领域的几个共性趋势：

1.人工智能技术的广泛应用：多个研报提到了基于深度学习模型，如 CNN、Transformer 等。异构图神经网络、图信息的加入、自适应时空图网络等在研报中的频繁出现，显示了量化研究者在探索和利用复杂网络结构和关系数据方面的兴趣，以寻找公司之间的关系或资产之间的相互作用。

2.多因子策略与选股因子的持续创新：报告中提到了因子的重构和优化，量化研究者在寻找更有效的因子组合和更新的选股逻辑。

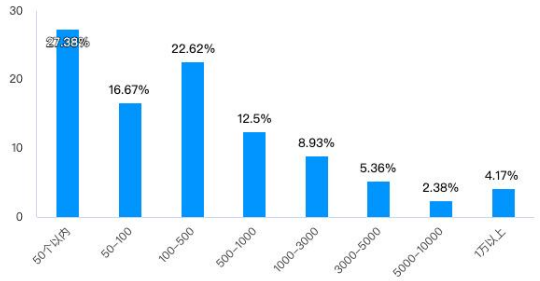
3.宏观量化与大类资产配置：涵盖了从宏观风险因子构建到具体的资产配置应用等多个方面。

4 高频数据的深入研究与利用：高频因子、高频订单成交数据等数据。

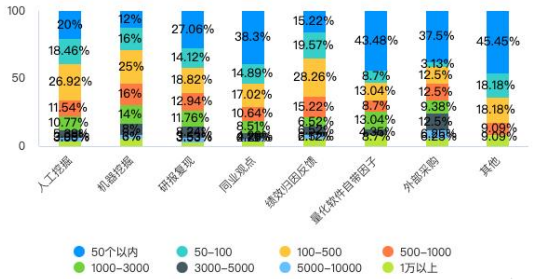
更多、更快、更强，因子发展总不免围绕此路，但比起因子池，实盘池可能更能代表机构目前风格。50 个以内（27.38%）实盘因子占比最高，表明有相当一部分量化机构倾向于维持一个较小、较为精细的因子库，更加注重因子的质量而非数量。当前的实盘因子组中 1000 因子以上的占比略少（20.58%），但代表的是截然不同的方法论与投资哲学，机构进行大规模的机器挖掘，利用机器学习和人工智能技术来生成和测试大量因子，比如茂源投资就在其资料介绍道因子数量超过 10 万个，并且每月增加或更新 10%的因子。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 量化机构目前的实盘因子数量：



■图表 因子挖掘方式与实盘因子数量关联关系



数据资料来源：《2023 中国量化投资白皮书》问卷调研

需要注意的是，因子的数量并不直接等同于投资策略的成功。一个精心挑选和优化的较小因子集合也可能比一个大而不精的因子库更有效。因此，量化机构在因子挖掘和实盘应用时需要平衡因子广度和纯粹性，确保因子的可解释性和稳定性。避免加入非常多的因子之后，反而发现组合噪音不可控。

■图表 因子检验方法比较

方法		优势	劣势
ic 信息系数	IC IR Rank-IC Rank-ICIR	含义简明；计算简便。	只能反映因子值与股票收益的线性相关性，未考虑非线性部分：无法从选股角度反映因子头部选股能力。
分组测试	分十组测试	直接考察组合收益与波动分布： 反映因子多头与空头收益 辅助判断因子在时序上的表现。	用回归取残差的方式无法做到完全中性； 受市场风险及其他因子的影响：
	市值行业分层打分	直接考察组合收益与波动分布： 反映因子多头与空头收益； 保持市值和行业中性； 辅助判断因子在时序上的表现	受行业股票个数影响，分组组数受限； 无法剔除其他α因子和风险因子的影响。
回归法	因子收益 FM-检验	检验因子在时间序列上是否有显著收益； 辅助判断因子在时序上的表现。	无法反映因子在组合中的表现； 受回测时间窗口选取的影响
	纯因子组合	能灵活控制风险因子对待检验因子的影响，提纯因子表现： 与风险模型、组合优化联系紧密。	无法判断收益来自多头还是空头： 纯因子组合不具可投资性，与实际投资组合存在差别； 严格控制其他风险暴露为 0，可能因子走势常有出入

数据资料来源：华泰研究

因子构建完成之后，机构会先观察其收益、夏普率与其他因子的相关性，同时进行有效性和逻辑性校验，使得因子可解释性强、特征性强、新增低相关。同时许多机构也会分观察库、实盘库等多个优先级别，同时在因子失效之后剔除，建立剔除标准。

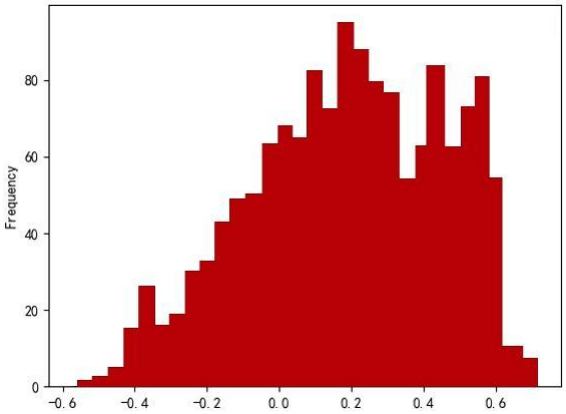
1.边际贡献度：包括 IC、累计收益率、IC_Rank 等。新因子一般会需要半年左右的时间去检验有效性，如果因子确实表现好就纳入实盘，对于长时间表现不佳的旧因子会适时予以剔除。

2.逻辑性检验：包括通过逻辑进行假设，并验证假设，进一步确认信号的合理性或者改进信号。也以此明确信号赚钱的逻辑，在信号失效时做出准确判断。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

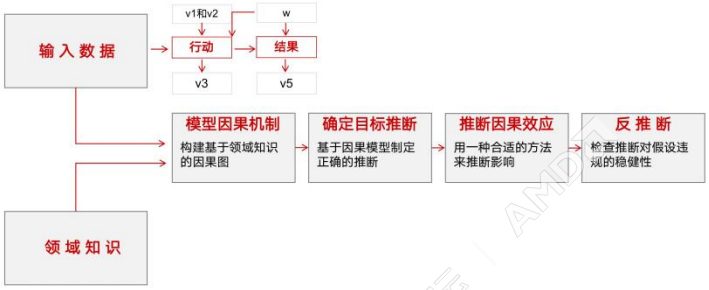
3.相关性检验：有的因子之间相关性较低、放在一起使用预测效果更好，同时有的因子之间相关性很高、即使放在一起也和单个因子差别不大。若新开发的因子与已有因子相关度较高时，将对已有因子进行融合升级，若该因子与现有因子相关性较低，则将其纳入因子库。

■图表 复熙资产因子在相关性频谱分布



数据资料来源：复熙资产

■图表 因子因果推断



数据资料来源：配邦私募

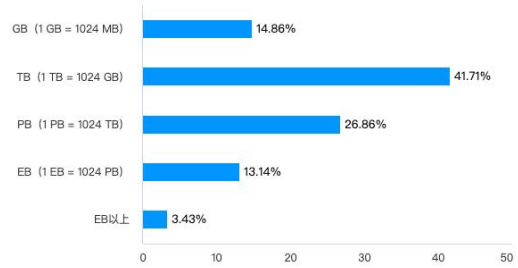
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

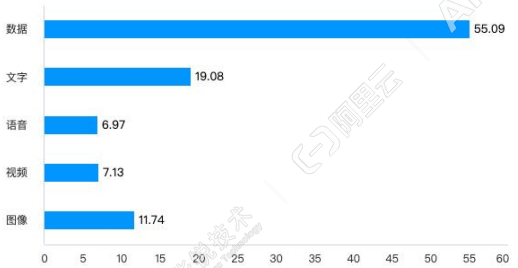
3.算力

“算力是设备根据内部状态的改变，每秒可处理的信息数据量。”2018 年诺贝尔经济学奖获得者 Willam D.N Ordhaus 在《计算过程》一文中对算力进行如上定义。

一个策略回测任务往往需要遍历多个标的、多个参数、多个周期。随着标的的增加、从 level1 到 level2 高频数据的应用、人工智能算法的复杂度增强、另类数据的引入并清洗，量化进入 3.0 时代之后，对算力带来了更高的要求和挑战，以便驾驭更加复杂的模型和海量参数，同时提升实时大样本信息处理能力。这些挑战包括技术框架、算力部署以及其他硬件系统。

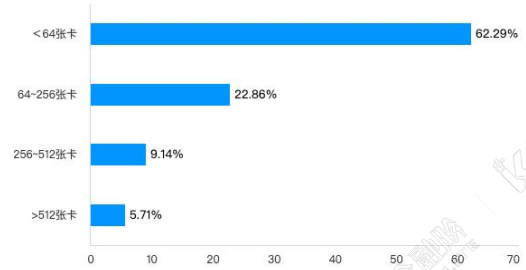
■图表 大量数据储存、多模态数据、复杂的算法催生了量化机构的算力军备竞赛



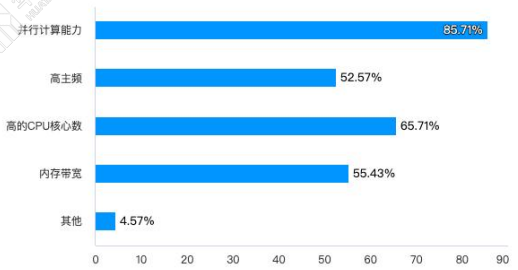


■图表 小型机构及经济学派别算力部署更为克制

人工智能和短周期量化机构派别对数据则更为积极



■图表 量化策略和数据处理中的首要需求



数据资料来源：《2023 中国量化投资白皮书》问卷调研

AMD 也在对客户的走访调查中发现，量化机构对服务器的性能要求多种多样，并行计算能力（85.71%）是量化策略和数据处理中的首要需求。并行化计算将计算任务分割成若干个子任务，同时在多个计算单元上进行计算，最后将计算结果合并得到最终的结果。相对于串行计算，它具有计算速度快、效率高、可扩展性好等优势。同时拥有更高的计算精度，在处理大量数据和复杂计算任务时非常关键。

高的 CPU 核心数（65.71%）在于更多的核心数可以支持更多的并行任务，包括多个交易策略、风险管理计算以及多个市场的同时监控等。在高频交易场景中，处理器主频往往优先于核心数，而在中低频交易或复杂策略计算中，多核并行能力也非常重要。

过半数的受访者认为内存带宽（55.43%）也是一个重要的因素。高内存带宽可以确保数据在处理过程中快速传输，减少因内存瓶颈导致的延迟，有助于快速处理大量数据结构（如订单簿、K 线数据、指标计算结果等），尤其是在进行大量数据筛选、排序、聚合等操作时。内存性能对于实盘交易系统的稳定性、响应速度和策略计算效率也至关重要。而 NLP 自然语言

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

处理以及因子挖掘等人工智能类业务， 它们对内存带宽的要求则更高， 其中涉及复杂的统计计算、相关性分析、特征工程等操作，这些任务都需要在内存中进行大规模的数据移动， 内存带宽的高低直接影响着数据处理的速度。

高主频（52.57%）尽管在调研中的占比相对较低一些，但高主频的处理器可以在单线程任务中提供更快的计算速度，能够在短时间内完成更多的计算任务，这对于快速计算交易信号、执行交易逻辑、处理市场数据流至关重要。

此外，支持向量指令集（如 AVX、AVX2、AVX-512 等）的处理器能够加速文本处理中的向量化操作，如批量字符编码、矩阵乘法等，进一步提升计算效率。

核心技术框架优化方面，例如 2017 年初，念空开始进行深度学习和机器学习的测试与迭代，终于在 2019 年 3 月推出了第一个稳定版的机器学习策略模型。与此同时，念空科技也建立了一个具有跨时代意义的模型平台，从一个封装化的多因子挖掘平台转变到一个封装化的全流程策略平台——超级信号工厂（TFOT）。这是一个一站式体系化的策略研发闭环。整个工厂由以下四个部分组成，包括特征工程（Terms）、拟合工程（Fitting）、优化器（Optimization）及交易算法（Trade）。策略在各个流程化标准化模块的配合下，经过层层检验，被生产出来。念空科技做过一个这样的测试，将过去 5 年全部 4000 只 A 股的特征值，即 24 万亿个数据点保存在一个文件里，一个未经优化的算法给出的时间是 18 小时，而在并没有增加任何服务器的情况下，念空通过算法重构后，给出答案的时间是 1.5 小时。

再如数据方面，任何模型的训练都是基于海量的历史数据。而大模型由于训练成本较高，需要的数据量巨大，无法做到实时或准实时地将最新数据纳入模型的迭代更新中。因此，NVIDIA 使用了检索增强生成 (RAG, Retrieval Augmented Generation) 技术，通过外挂知识库的方式，可以将最新的实时数据导入到 LLM 的应用当中。而除了基于文本的 RAG，还出现了基于 Graph 的 RAG 技术，利用 Graph 对非结构数据的信息抽取能力，以及跨文档的关系建立能力，再结合常规的向量和关键词检索，进一步提升了检索增强生成的能力。

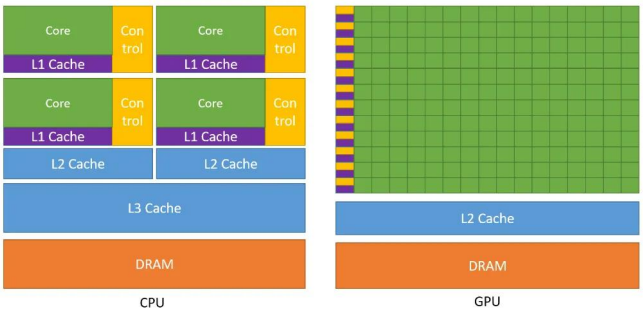
算力部署方面，一般而言，策略方向影响了量化机构的基础设施主张，大型算力在中短期、中高频大量股票的投资上，更有优势。人工智能派别机构会在算力部署上更为积极，经济学意义派别则会将其掌握在合适范围之内，认为机构之间算力的差距微乎其微，更重要的是找到更好的 Alpha 来源，算力配备只是解决策略投研人员的后顾之忧。

对于人工智能和短周期量化机构派别而言，传统的技术框架无法满足量化行业对海量数据存储、分析计算、策略迭代的性能要求，为方便一次性加载更多历史数据到内存，量化机构往往需要不断采购、扩容内存服务器满足需求。近年来，随着量化私募在国内的蓬勃发展，许多百亿量化私募逐渐加大在硬件上面的投入。一些公司在算力方面的开销呈几何级增长，每年近千万的算力和硬件支出使其产生了自建超级计算集群的想法。一些公司则配备了高性能 GPU 集群的服务器，满足庞大的数据处理需求。明汯就谈到，国内量化从简单的几台 CPU 服务器，到目前大规模高性能计算集群，只经历了不到 10 年时间。

与 CPU 相比，GPU 拥有更多的计算核心，一般在几百个到上千个之间。因而 GPU 可以通过并行计算的方式同时处理多个任务，相比于 CPU 有着更加特殊的并行计算层次结构，因此在处理大规模数据时比 CPU 更快、更高效。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 CPU 和 GPU 的芯片资源分布示例



资料来源：《CUDA C Programming Guide》，NVIDIA

■图表 当前部分量化机构算力存储

公司简称	设置
九坤	2020 年，九坤搭建了人工智能方向的技术团队，成立了 AI 实验室，建立起北溟超算集群， 完成在人工智能方向上的技术突破。2021 年，九坤设立了三大实验室，数据实验室（DATA LAB）、人工智能实验室（AI LAB）、水滴实验室（主要用于优化一些软硬件基础架构的性 能优化）。
幻方	2019 年，幻方量化成立 AI 公司，投资 2 亿元自主研发深度学习训练平台“萤火一号”，搭载了 1100 块 GPU。投入近 20 位工程师历时 2 年自建完成。主要解决 AI 投资的两大难题：模型更大、更复杂——支持超大型神经网络训练，一个 4 亿参数的大型经济分析模型，在幻方上一代高性能集群（4 台 DGX-2， 100G-IB 互联）上训练需要 2 个月时间，而“萤火一号”仅需 84 小时。数据量更大——其算力匹敌 4 万台个人电脑。幻方“萤火一号”超算占地面积相当于一个篮球场，功耗 400KW，由一个存储集群和一个计算集群组成。存储集群提供每秒 1.3 亿次 IO 响应、4.1Tbps 读写带宽以及 1.2PB 容量。计算集群搭载 1100 张高端显卡，每秒可以进行 1.84 亿亿次浮点运算（18.4PFLOPS， 32 位精度），相当于 4 万台个人电脑算力。所有节点通过 200Gbps Infiniband 交换机进行互联。2021 年，幻方量化对“萤火二号”的投入增加到 10 亿元，并且搭载了约 1 万张英伟达 A100 显卡。以分时调度共享 AI 算力，弹性运行超大规模深度学习训练，目前 96 %集群使用率，85 %GPU 使用率，8.0TB/s 读，500GB/s 写。
明汯	2021 年，明汯自有的高性能计算集群已经位居世界超算排名 TOP500 榜单前列。预计到 2023 年底，计算集群将拥有 1500 张 GPU 卡，3 万 CPU 核、1Pb 内存以及 5Pb 磁盘存储，在金融数据的应用场景下 AI 算力达到 400P Flops（每秒浮点运算次数）。
灵均投资	2020 年大力发展超算集群。近三年每年灵均投资在人才引进和硬件设施建设（比如机房建设、GPU、服务器、数据库）等方面的资金投入超 1.5 亿人民币。2022 年，筹建数据异地双中心，引入世界先进算力资源。
上海天演	建大数据平台，每秒能完成 10 的 13 次方次计算，每年科技投入经费超 1 亿元。
宽德投资	已建立高性能存储系统（存储速度可以达到 200Gb/s）和高并发策略研发平台。
启林	启林近一年投入 1 亿元进行超算集群的搭建，构建较为强大的算力支撑体系。在算力中心迭代之后预计在股票上可做到每五分钟动态配置，未来计划做到每一分钟动态配置，且结合短期基本面不同策略冲击以调整策略。

数据资料来源：编写组整理

目前算力部署机构会采取本地部署和上云两种方式:

国内量化机构并未全面使用分布式计算框架，投研人员更多采用单机并行的方式进行回测。受限于单台服务器配置上限等主要原因，量化行业普遍出现回测效率慢，投研人员排队等资源的痛点场景。借助云上无限算力，越来越多量化机构在云上开始逐渐引入HPC、Kubernetes 等技术，构建分布式架构集群。类似技术方案可以帮助机构显著缩短策略回测时间，提高投

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

研团队工作效率。解决了庞大数据量对传统量化 IT 架构的挑战与冲击。目前多家私募量化机构已经在策略回测、策略研发（AI 模型训练）等场景下探索出云上成功落地的最佳实践。

据阿里云透露，某量化机构在阿里云上按需运行回测任务，每日最高任务数量可达上万个。借助阿里云的弹性实例，投研团队可以快速获得回测结果。阿里云数字金融解决方案负责人赵庆元提及，“当前量化机构普遍存在基于单机和 slurm 调度的回测框架；这与几年前自动驾驶等行业极其相似，但这类客户最后都逐渐过渡到了 kubernetes 等方案上。市场会看到越来越多的量化机构拥抱云计算，继续向更先进的 IT 架构演进，提高投研效率。”

本地部署方面，多家机构开始逐渐布局适配中国市场的算力芯片。例如 AMD EPYC 系列处理器，采用了先进的制程工艺，在单颗处理器上集成最多 128 个物理核心，主频高达 4.1GHz；在内存支持方面，AMD 处理器支持多达 12 通道的 DDR5 4800MHz 内存，提供远超市场平均水平的内存带宽．同时能够支持完整的 AVX256 / AVX512 指令集，助力量化机构各类业务快速发展。

特别是针对大模型的任务场景，在业务上线后，量化机构会需要更多的算力来支撑高并发和低延迟的业务推理要求。具体来说 GPU 的显存大小，决定了其可以承载的模型大小和能够支持的上下文长度。当模型大小超过单一 GPU 的显存大小时，就需要使用模型并行的方式对模型进行拆分，并通过多 GPU 之间的协作，进行分布式推理。而进行分布式推理时，上层的推理框架需要对底层的算力设备具备很好的适配程度。

目前主流的推理框架包括 NVIDIA TensorRT LLM 以及 vLLM 等，这些推理框架都包含了对 LLM 计算过程中的算子优化，并且 TensorRT LLM 还支持 FP8 的计算精度。这些优化都需要建立在对底层硬件设备的架构与特性完美适配的情况下，才能发挥出其应有的威力与价值。在 LLM 训练方面，NVIDIA NeMo 和 Deepspeed 通过模型并行和 Zero 等策略，可以支持 LLM 模型训练过程中较大的参数量，并优化模型训练过程中的显存占用。此外，NVIDIA MeMo 还支持 sequence 并行和选择性激活重计算，来进一步优化训练过程中的中间变量的划分与存储。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

4.组合

量化模型一般通过因子及权重设置后对预测收益率进行从高到低的排序。模型在不断经历复杂度逐渐递增的迭代过程，在早期简单的线性模型即可表现出色。随后，随机森林、Boosting 算法用于量化投资，其模型复杂度、参数相比传统统计学习模型有了很大的提升，预测效果也更好，基于决策树的集成模型已经成为量化机构的标配，很多机构更是把 Boosting 模型作为研究员拿到超额奖金必须战胜的基准线。再到后来的深度学习模型，复杂度进一步扩展。

量化机构的模型会根据实际情况混合使用并保持更新迭代，例如锋滔鑫起点 2 号为 50%线性模型+50%非线性模型（800 只票左右），锋滔指数增强 2 号为 100%非线性模型（600 只票左右），行业偏离度均控制在 15%以内。为尽力避免机器学习带来的过度拟合风险，其还在非线性模型中增加了一些惩罚项。再如金戈量锐模型每半年迭代 15%左右，力争适应多变的市场环境。

有的量化机构对多个目标进行建模预测，例如磐松就会对个股、行业、市场同时建模，从其实践来看，个股模型因子预测值分布广，对产品业绩的贡献度最大，约占 80%；行业因子预测值离散度较小，行业配置对业绩贡献约 20%；市场对整体市场涨跌的预测误差大，需要严控择时。预测目标除了收益率，近年也有机构将最终预测目标设置为夏普比率，以期在收益与回撤之间取得平衡。

过往三年问卷调研数据显示，当前量化模型：

1.技术呈现多样性：2021 年数据显示 MLP（多层感知机，47.29%）和 LSTM（长短期记忆网络，46.93%）的使用率最高。但总体而言量化投资领域需要多种技术的组合来应对不同问题。虽然一些技术在某些环节中比较突出，但量化投资领域仍然没有形成一个关于哪种方法最优的普遍共识。这可能是因为市场本身在不断变化，策略和技术需要适应这些变化才能维持效果。不同的团队和个人可能会根据他们的经验、资源和投资理念采用不同的方法。乾象投资在其资料里面就强调了当前其模型使用的多样性，包括强化学习、自动机器学习（AutoML）、监督学习、微分动态规划（Differentiable programming）、transformer。

2.实践中的复杂性：在实际操作中，一个量化策略可能需要结合多种技术来处理不同的问题。例如，在数据预处理和清洗阶段使用聚类模型来识别异常值，在特征选择阶段使用树模型来确定重要的输入变量，在预测阶段使用神经网络来建立预测模型，在执行阶段使用强化学习来优化交易决策。数据显示，在高频交易这一要求实时处理能力的环节中，所有技术的应用比例都相对较低，这可能表明在该领域仍存在技术上的挑战。

3.高比例的应用通常出现在数据处理相关的环节（如数据清洗、特征提取、模型构建），而在直接的决策制定环节（如策略优化、高频交易）则较为分散。具体而言：

Ⅰ.集成学习和树模型在几乎所有环节中都表现出较高的应用频率，这可能是因为它们的鲁棒性和泛化能力使它们在多种场景下都非常有效。

Ⅱ.神经网络则在收益预测、市场模式识别、衍生品定价、组合构建、策略优化、高频交易等多个环节中的应用率较高，尤其是在模型构建和收益预测中，使用比例超过 40%。

Ⅲ.聚类模型在数据清洗、另类数据处理和产业链聚类中的应用较为突出，这表明它们适合于探索性的数据分析和模式识别任务。

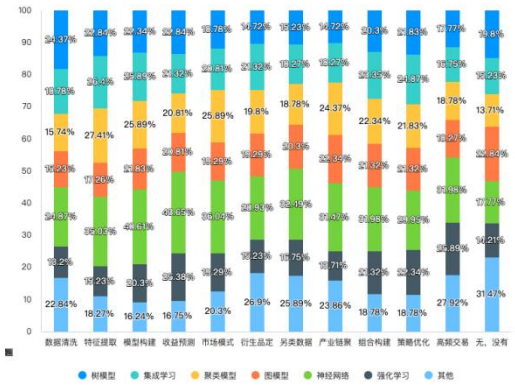
Ⅳ.图模型在市场模式识别环节中较为常用，可能是因为它们能够很好地捕捉市场参与者之间的相互作用。

Ⅴ.强化学习在策略优化和高频交易中的应用率相对较高，在策略优化和高频交易中的应用比例分别为 22.34%和 25.89%。

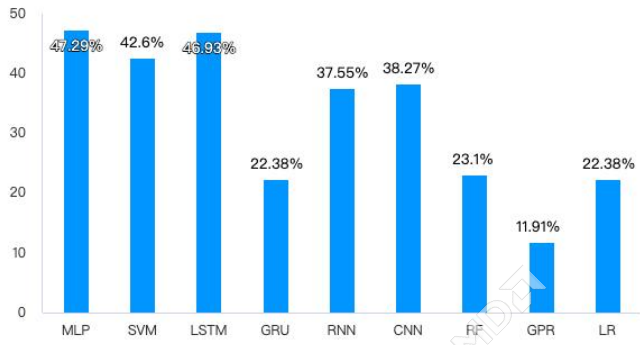
但过去两年最火热的 Transformer 和 Diffusion 模型因故未进入统计量。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 2022 年人工智能模型在各个投研环节的使用



■图表 2021 年机构主要使用的人工智能模型



数据资料来源：《2023 中国量化投资白皮书》问卷调研

上述模型很多来自人工智能领域，在量化投资的实际应用过程中可能存在缺陷，需要做进一步的改进。例如玄元投资 2023 年在《中国量化投资白皮书》研讨会北京场就讨论过：目前 AI 训练的主要优化方式是梯度下降，其基本思想是在每个点处通过计算函数梯度来确定下一个移动方向，以逐步接近最优解。这种方法虽经典，但实际上也会存在问题。仔细分析后发现它们会对训练产生很大的影响，一是可能找不到全局最优解，二是花费的成本高昂。

使用梯度下降方法可能找不到全局最优解的原因在于 1.目标函数通常是非凸的；2.目标函数可能不可导；3.在现实中，目标函数经常存在大片导数为 0 的区域；4.使用梯度下降会遭遇维度诅咒；5.GPU 太贵。

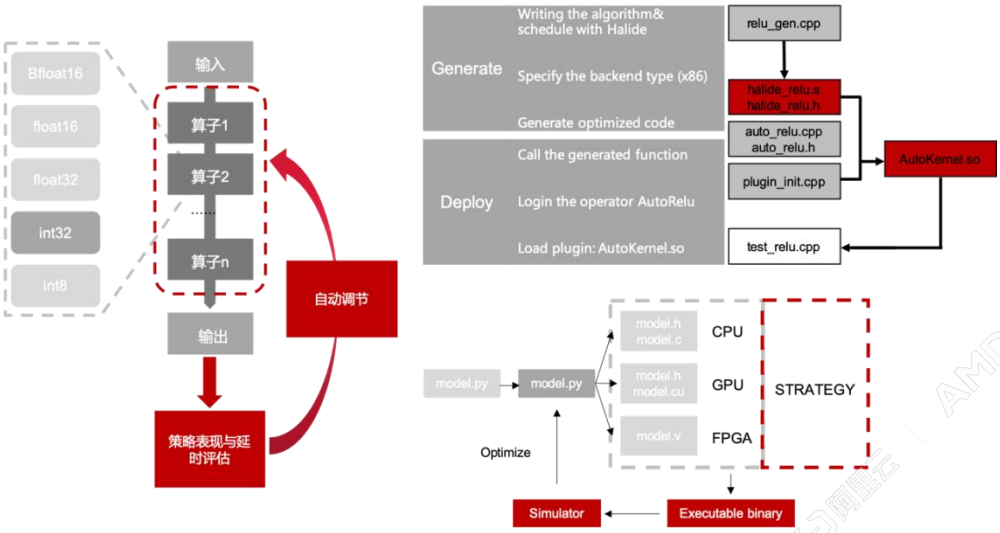
所以公司考虑利用不基于梯度下降的全局算法 MARS（Model-based Annealing Random Search）来弥补梯度下降算法的缺陷。MARS 算法是一种基于模型的随机搜索算法，它使用分布采样来寻找最优解。与传统的随机搜索算法不同，MARS 算法不仅局限于局部搜索。简单来说，MARS 算法通过实验来考察样本点的表现，然后更新采样函数，利用样本构造一个 Boltzmann 分布，以寻找更优的点。如果某个样本点表现良好，附近领域的点的表现也应该不错，领域里含有全局最优点的概率更大。通过不断迭代，MARS 算法最终找到最优解，即$x^* \in \underset{x \in \mathbb{X}}{arg\,max}\ H(x)$。

也有机构认为人工智能模型的成功实践对量化最大的启示不是可以产生更多因子，而是其建模的过程和思路。在倍漾的投研方法论中，数据、预测、决策、执行是最核心的四件事，在内部搭建较为通用的底座，在底座的支撑上对于四个环节使用机器学习来自动化。

使用机器学习方法进行因子挖掘以及预测，而后决策的核心在于形成对未来的某种预测后，如何进行资产的组合优化以及风险管理。从算法复杂度的角度来看，在有限的时间内很难找到确定的解决方案。然而，机器学习可以实现实时向相对理想状态的逼近。执行环节中拆单以及代码的执行效率变得至关重要。AI 编译器可以用来把源码变成机器代码执行，但是目前所有传统的机器学习、神经网络，或者各种树模型的编译过程都面临着一些问题。编译器比较注重通用性，对于一些特定场合的支持不够好，比如面对低延时需求或者流式数据的时候，传统的 AI 编译器很难满足高性能代码的编译要求。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

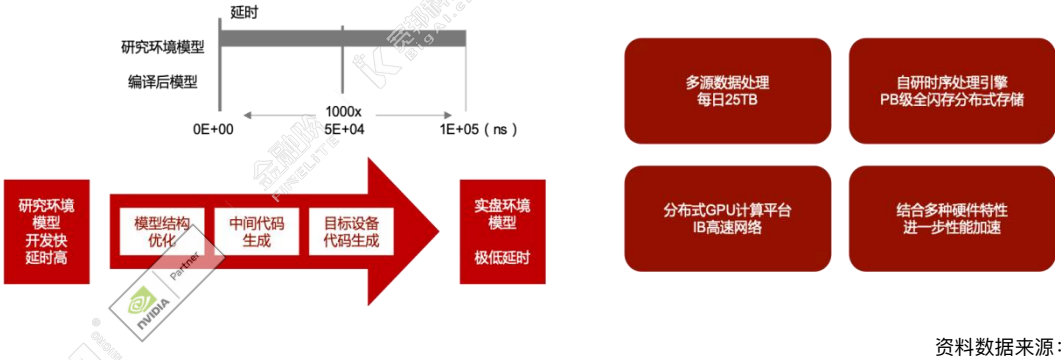
■图表 倍漾低延时 AI 编译器关键技术



资料来源：倍漾量化

倍漾自主开发的波塞冬 AI 编译器，第一个特点是能够实现自动混合精度加速。模型从输入到输出的过程中包含大量算子，不同算子在不同设备上所需的最优精度各异，有时需要 Bfloat16，而有时仅需 int8。精度选择对于提速至关重要，可以利用强化学习技术等各种数据驱动的方法进行选择。第二个特点是端到端的延时优化。传统 AI 编译器需要预先载入许多并非立即需要的程序，会占用一定的内存空间。波塞冬 AI 编译器则不要求预先载入，从而实现内存占用为零，此外还采用了一些强化学习技术，能够实时自适应地在编译过程中对问题进行优化。

■图表 倍漾低延时 AI 编译器关键技术



资料数据来源：倍漾量化

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

03 趋势 3 多频段：组合、风控

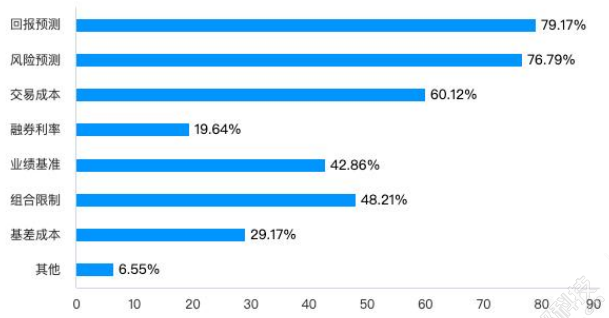
1.组合优化

量化模型一般可以给出成百上千只股票的预测收益率，并对预测收益率进行从高到低的排序。但仅有排序是不够的，还需要把模型预测的结果转化成实际的股票仓位，这就会涉及组合优化，具体来说需要在优化模型里加入各种各样的约束条件，在收益、风险与成本之间进行取舍调整。

以风险管理条件为例，包括刚性约束与柔性约束。刚性约束即出于交易规则和合规要求所必须遵循的原则，例如将最高申报速率限制在每秒 300 笔以下。

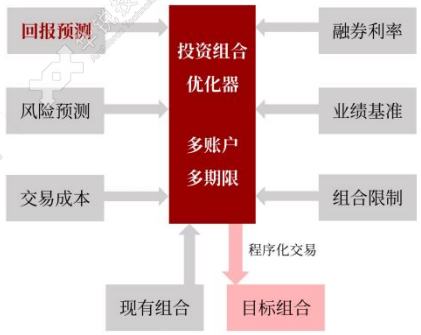
另一类柔性约束，例如指数增强策略会根据对标指数对行业或个股的偏离度、持仓集中度给出限制等。在一系列风险、成本的约束限制条件之下，优化器便会根据预测收益率排序，选择买入排序靠前的股票，这就生成了实际交易所需的股票仓位。

■图表 目前量化机构所使用的组合优化器功能



数据资料来源：《2023 中国量化投资白皮书》问卷调研

■图表 磐松组合优化器功能示意图



数据资料来源：磐松资产

目前量化机构的优化器呈现更快、更复杂的趋势。

更复杂表现在：除了回报 (79.17%) 和风险预测 (76.79%) 是优化的核心常见功能，交易成本 (60.12%) 、组合限制 (48.21%) 、业绩基准 (42.86%) 等、融券利率 (19.64%) 及基差成本 (29.17%) 等也有机构将其纳入了优化器功能之中。

磐松资产公开介绍过其组合优化器逻辑：目前量化机构都是多产品同时运行，每个产品都有各自的投资目标，而不同的产品账户之间存在流动性竞争，由于机构几乎只有一套预测模型，假设一只股票预期回报非常高，每个产品都会有很强的动机买入这只股票，如果每个产品独立优化，那么都可能大量买入这只股票，严重低估所有产品共同造成的市场冲击。如果简单分配流动性限制，每个产品对单个股票的交易，再对每个产品独立优化，则无法将流动性物尽其用。对于产品共用的融券额度也是类似的逻辑，因此优化器要求每个产品既考虑到单一产品需要怎么样交易，又能看到其他产品的影响，从而控制对市场造成的整体冲击。综合考虑不同产品的投资目标，公平合理地分配流动性和融券额度。所以投资组合优化的本质就是在限制范围内找出让效用函数最大的交易决策，而效用函数是一个多账户、多周期的函数，会综合考虑每个账户每个周期的超额收益、风险、风格偏离和交易成本等，在合理的限制下追求稳健的超额收益。多期限优化，则能够让优化器在投资过程中智能地做出更有远见，更符合全局利益的交易决策，多账户多期限的优化系统能更加深刻合理地刻画投资问题，而代价则是庞大的问题规模和缓慢的求解速度。

更快求解要求体现在：随着账户数量、投资标的和投资逻辑的不断增加，优化器所面临的问题规模也会日益增长。目前

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

磐松的优化问题中有超过 1000 万个变量和超过 1000 万个组合限制，在不添加任何加速逻辑的情况下，求解这样巨大的数学优化问题可能需要几天时间。因此其研究团队投入了大量精力，在几乎不损失求解精度的情况下，大幅提高求解速度，精准缩小可行域，深度利用并行计算等方式方法，将问题的求解时间控制在分钟级别，让调仓具有时效性，确保优化器能够稳定承载盘中不断增加的管理规模。

启林投资也介绍过，其优化器可实现盘中 5 分钟出交易清单。乾象也有很多不同频率的信号，短周期的量价类信号大约到分钟频，长周期的基本面类信号大约可以到年频，每 10-30 分钟乾象会跑一次流程，包括从产生信号到投资组合优化器再到最后输出目标仓位，所产生的短信号与长信号进行多信号周期叠加从而实现收益。

风险预测（76.79%）方面，2023 年尾至 2024 年初，量化机构的风控及风险预测被放在了市场极高关注的位置。目前几乎所有量化机构都将风控分为事前、事中、事后三个环节，以下是量价机构的典型风控设置：

■图表 机构典型风控环节设置

	机构 1	机构 2
事前	结合历史回测结果，对仓位、交易量、浮动盈亏、可用资金及风险敞口等关键指标设置阈值，严格遵循交易所和券商合规要求	1）策略内嵌风险模块，依据过往十余年历史数据统计归纳，包含风险预警指标，自主研判市场是否处于常态。 2）合理评价策略的适应度、风险度、预期收益和替换效益，在此基础上动态优化
事中	系统性地通过在报单前植入检测逻辑，确保所有委托价格，数量均在合理范围，避免报错单。 系统实时对账，不断从交易所查询盈亏，检查是否和交易系统自算信息完全匹配，确认交易系统运行的正确性	1）自建含极限亏损比例、最大下行风险、流动性、回撤幅度等核心指标的风险管理系统，有效实现盘中实时、程序化自动监控。 2）过程管理、损益管理、部位审核相结合的多层次风险管理系统
事后	通过复盘交易流水方式，对交易系统每日表现进行系统性分析，确认每日无异常交易行为。	1）自建完善的绩效评估体系，每日复盘。 2）体系包括策略效果评估和交易管理评估，重点考察投资结果与设计目标的差异、投资结果与理论模型的差异

在风险控制环节，绝大部分机构都称不做主观风格暴露、不做择时，强调在更长的时间尺度而不针对某单一年份去做优化，不主动开放风险敞口。流动性及 Barra 体系中行业敞口、市值敞口、成分股占比最能量度当前 A 股市场风险。机构主要做以下敞口控制：

市值敞口控制：A 股目前 5000 多只股票，量化管理人在用模型打分之前，通常会根据一些条件剔除掉一部分股票，例如 ST 股票、市值较小、流动性较差的股票，股票池跨度的覆盖范围差异比较大，多则 4500 少则 3500，大部分量化管理人是在 4000 只左右股票里选择。一般选股数量少的，通常市值选择下限就高一些，例如将前置筛选条件设置为 30 亿市值或者前 5% 分位以上。

也有机构会根据实际情况调整风险暴露，如大岩量化会将模型分为低波动和高波动，其中低波动模型的风格偏离 0.35；高波动模型风格偏离 0.75；另外行业偏离+/-1.5%，个股偏离+/-0.5%；流动性<2.5%全市场成交量；换手率：50-100 倍；分散持仓：500-800 只股票；个股波动率：<5%。

以量价为主的量化机构大多都会持有一定程度的小票，一方面是基于分散投资的考虑，另一方面也是因为近年来小市值风格相对大市值在 Beta 上更强势，加之小票波动性大、流动性也有显著上升，机构投资人参与较少，获取错误定价产生的超额能力也相对较强，这也是市场化趋势选择的结果。

流动性敞口控制：许多机构会根据不同的指增产品做不同的持仓流动性配置。例如星阔投资空气指增为了使策略充分获取放松风格约束后带来的收益提升，在市场上行阶段有更强的收益弹性，其流动性放得最开；但其他指增产品 9 成持仓则在

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

全市场流动性最好的 50%股票中。

■图表 不同指增产品流动性分配

产品	持仓流动性分配	产品	持仓流动性分配
300 指增	接近满仓在全市场流动性最好的 50%股票中 9 成持仓在流动性最好的 20%股票中	1000 指增	9 成持仓在全市场流动性最好的 50%股票中 5 成持仓在流动性最好的 20%股票中
500 指增	9 成持仓在全市场流动性最好的 50%股票中 6 成持仓在流动性最好的 20%股票中	空气指增	7 成持仓在全市场流动性最好的 50%股票中 3 成持仓在流动性最好的 20%股票中

数据资料来源：星阔投资

分散度控制：机构避免在低流动性股票中暴露自身策略，也会对持股分散度进行限制，一般而言持股数在 500-1000 只，多的能达到 2500-3000 只，最大个股持仓偏离度 1%-2%左右。持股分散有利于进一步降低交易冲击成本，提升策略容量。

成分股控制：成分外选股是否为风险一直存在争议，公募指增持仓要求 80%以上成份股，私募指增一般实为全市场选股，极少机构严格限制其挂钩指数的成分股，或对主要指数成分股进行限制，例如衍复会对三大指数 1800 只成分股进行总的要求。

■图表 不同指增的成分股分配

策略产品	持股数	成分股	超额归因
300 指增	500-1000	沪深 300 指数成分股占比约 70% 中证 500 和中证 1000 指数成分股占比约 20% 1800 以外股票占比 10%。	量价类因子占比 50%，基本面贡献度占比 30%，新闻舆情和另类数据各占 10%
1000 指增	2500-3000	沪深 300、中证 500 指数成分股占比约 20%， 中证 1000 指数成分股占比约 40%，三大指数 1800 只成分股以外的股票占比为约 40%	主要是量价类因子贡献，占比在 80%以上，因为中证 1000 以小盘股为主，波动弹性更好，更利于量价类策略盈利。

数据资料来源：衍复投资

Barra 模型是目前国内的主流风险控制模型，但在实际使用中存在许多局限性，例如预测周期偏长，没有考虑特异性股池，A 股的特点结合性有待提升等。所以很多机构除了根据其控制因子敞口，对标基准进行行业和风格约束之外，还会进行例如周期改动等措施，以期在短周期组合构建中更好地发挥作用。

除了固定 Barra 约束，也有机构采用滚动约束，如天算量化在近期策略迭代了三个方向：

严控行业敞口-组合优化器对风险敞口的动态调整，风控组合优化器严格控制行业暴露在 4%以内，大部分都在 2%以内。

行业轮动策略-策略严控行业风险，同时动态暴露有限敞口捕捉轮动收益。

市值轮动策略-策略严控行业风险概念，同时动态暴露有限敞口捕捉轮动收益。c

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

2.中性对冲

相较于敞口风险，所有机构面临最大的风险实际为市场风险。

2010 年，首个股指期货 IF 合约（沪深 300 股指期货）在中金所上市，量化基金具备了可行的对冲工具，开启国内股票量化对冲基金的时代，但量化与对冲并非绝对相等，量化进化至今代表整个投研方式与体系，而对冲仅是其面对市场波动风险所采取的避险措施。

我国股指存在明显的“大涨大跌”现象，短期内涨跌行情切换频繁，上涨的收益和下跌的风险并存，机构将这部分风险定义为 Beta 风险。为了使得量化策略收益曲线较为平滑，若整个收益 = 量化全市场选股多头 Alpha-市场 Beta/空头 Alpha-融券成本，则波动较小、夏普率较高。一般而言，量化机构会采用股指期货、融券、期权、收益互换等单一或结合的多种对冲方式，持续跟踪成本差异，以求得对冲成本、对冲精度的最优解。

■图表 不同对冲方式优劣势

对冲方式	优势	劣势
期货对冲	期货市场交易量大，流动性好，可以灵活调仓。	波动性较高，过去多年贴水较深，对冲成本较高（2022 年下半年开始贴水逐渐收敛，近期贴水较浅）。
融券对冲	对冲成本均摊至每期匀速支付，对产品净值波动影响较小，降低对冲端对股指期货基差波动的敏感程度。具备对冲精度、资金使用效率更高、成本更低廉、自由度更广、机构间差异更大等优势。	融券市场交易量相对较小，流动性较差、融券的额度也往往有限；融券操作需要满足一定的资产净值和交易规模要求；管理人需要具备较强的风险和流动性管理能力。
收益互换	让中性产品净值更加平稳。	收益互换空头交易存在一定的流动性风险。收益互换交易需要与机构投资者进行协作，不同的对手方可能存在不同的信用评级和偿还能力，需要严格控制对手方风险。
期权对冲	期权交易流动性较好，操作灵活、可以通过期权市场实现灵活的对冲策略，另外期权具有非线性的特点，期权的买方和卖方权利义务不对等，有类似保险的功能。用期权盈亏不对称的特点，替代股指期货对冲股票现货可以一定程度节省成本，甚至带来额外收益。	期权市场交易成本较高，需要支付权利金和手续费；期权价格受到市场波动、时间价值等因素的影响，存在一定的价格风险；需要对管理人对于期权市场有一定的了解和操作能力，对专业要求更高。

数据资料来源：衍盛投资，编写组整理

股票量化对冲策略的收益存在一定的周期性，也受市场环境（例如一九行情、抱团行情、成交活跃度、严重负基差）等的影响，但其中大部分因素都难以准确的预判，只能作事后的解释；即使专业投资者也很难持续准确择时，在对冲比例上，目前市场分完全对冲（即市场中性）、部分对冲和灵活对冲，其中完全对冲类产品的规模占绝对主导，不留多空敞口。

少数机构会加入择时，例如 2018 年，九坤投资发布灵活对冲多空产品，其目的是在高风险、高收益的指数增强策略，与低风险、低收益的中性策略之间，找到集两种策略优势于一体的平衡点。所以在策略加入择时因子使得空头端敞口更为灵活，降低对冲成本的同时收获上涨收益；从收益及回撤来看，九坤多空也确实介于指增与中性产品之间。多头端选择 85%的基金满仓持有 500 指增，持仓股票几千只，对股票流动性进行风控，如单只股票近 20 日流动性交易均值的 5%作为持仓上限。

空头端则采用 15%资金用股指期货做杠杆交易，空头敞口按照指数走势、波动率、基差成本三个方向做监控，根据模型择时信号决策不断调整风险敞口做“不完全对冲”，以使得净多头敞口保持在 0-80%之间浮动，例如在上涨市中降低期货空头的头寸，获得市场上涨收益和超额收益；在下跌市中提高期货空头头寸，控制风险，获得对冲市场风险后的超额收益，并利用量化择时模型以规避市场风险赚取双重超额收益。适合预期收益高于中性产品，但是不想承担指增高波动的客户。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 九坤灵活对冲多空产品介绍

夏普比率	1.63
波动率	10%～15%
收益	2018 年–2023 年 6 月，累计收益 123.27%，年化收益 18.12%
风险特征	中高风险，定位风险度低于指增策略，但收益预期高于对冲策略
策略逻辑	利用量化择时模型不断调整风险敞口，承担市场波动基础上力争稳健超额收益
多头构造	85%资金做 500 指增股票多头，不择时满仓持有
空头构造	15%资金利用股指期货做杠杆交易，控制空头敞口比例上下浮动 多头部分减掉空头部分所谓的净多头敞口可以在 0~80 之间浮动
对冲敞口择时指标	指数走势、波动率、基差成本

数据资料来源：九坤投资

艾方的做法是以指数增强为“多”，以看跌期权保护为“空”，既有上涨进攻性的同时又兼具尾部风险保护。通过持有看跌期权截断权益资产的尾部风险，从而降低组合风险和波动率，主要实现方式是自建市场风险评分系统，进行期权择时对冲，其对冲水平根据市场环境动态调整，期权年化对冲成本＜3%。

■图表 期权策略对冲





注：数据截止 2023.3.31. 数据来源于 wind，艾方资产

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

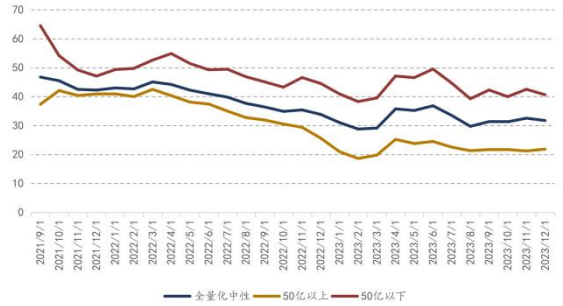
3.全频段融合

过去量化策略以高频段为主，即通过较高频率地买入卖出捕捉资产短时间内的买卖价差。高频的好处在于，对未来资产价格的预判周期相对较短、独立预测的次数多，策略整体胜率相对较高，尤其适用于高波动或高成交量的市场行情。

高频的劣势在于其需要进行频繁交易，加上参与者逐渐增加，投资机会转瞬即逝，策略容量相对较小。从 2021 年 9 月开始，因量化规模增长带来的巨大冲击成本，促使量化私募开始降频以扩充策略容量，2023 年 9 月，沪深交易所发布公告，将重点监控最高申报速率达到每秒 300 笔以上，或者单日最高申报笔数达到 2 万笔以上的交易行为，这对于高频交易（尤其是持仓极为分散的高频量化策略）造成了较大冲击。

以量化中性策略为例，平均年化单边换手由 47 倍降至 2024 年 2 月份的低点 28 倍，50 亿以上的头部管理人更是低至 20 倍以下。目前一些主流的股票量化私募策略的年化换手率已经逐步下降至 30-60 倍左右，细分子策略里偏短周期量价策略的占比也相应下降，相反，基本面策略的占比相应提升。

■图表 股票市场中性年化单边换手变化趋势



■图表 不同换手产品收益表现



数据来源：招商私募指数数据库

■图表 量化私募换手率

标签	机构	年化双边换手率	年化换手率预估区间	机构占比	标签	机构	年化双边换手率	年化换手率预估区间	机构占比
低频	机构 A	40–80 倍	50 倍以下（以30 倍–50 倍居多）	7.92%	中频	机构 E	50 倍左右	50–80 倍	24.32%
	机构 B	40–50 倍				机构 F	80 倍		
	机构 C	30–50 倍				机构 G	70 倍以上		
	机构 D	30 倍				机构 H	30–50 倍		
高频	机构 I	80–120 倍	80–200 倍	18.03%	超高频	机构 L	200–300 倍	200 倍以上	2.19%
	机构 J	80–100 倍				机构 M	600–1000 倍		
	机构 K	150 倍							
				多种频率					47.54%

数据资料来源：国泰君安研究

一方面量化私募在降频，但另外一方面，随着交易费率的降低，AI 及相关技术的兴起，公募量化开始采用基于 AI 的算法，交易频率从传统的月频提升到周频甚至日间级别。从方法论层面和交易层面来看，公募量化产品私募化已经呈现出一定的苗头，具体表现在提高了机器学习和高频因子的应用。据统计，公募量化产品的换手近几年也出现了逐年提升的趋势；此

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

外，一些公司的量化产品交易获得了对反向交易的豁免，在交易层面进一步给公募量化产品松绑。整体公募量化呈现出了一定的私募化趋势，涌现出了一批业绩不输量化私募的量化公募产品。

总的信号频段只是结果，目前量化领域竞争越来越激烈，不同的交易频率都有其适应的市场行情，如果仅仅是擅长某一类因子或者某一个频段的预测，在多变的市场环境下可能会遭遇逆周期时出现回撤，更全面的频段覆盖有利于降低超额收益的波动，例如成交量萎缩对于换手率较高的高频策略影响就会更大一些，可以依靠中低频交易实现收益。再如 T0 策略适合日内波动高、交易活跃的股票，但是拉长周期来看，这类股票相对指数有明显的负 Alpha。所以机构需要捕捉市场不同周期的收益来源，逐步丰富覆盖低、中、高多个频段，从分钟级别到月频、季频等，使策略变得更丰富以及稳定。

同时机构可以结合不同信号产生不同策略和产品，但在信号割裂的体系中，可能会发生信号互搏。以日内信号为例，A 股处于 T+1 的交易环境下，日内与日间通常具有反转效应。所以一般用日间选择多空，而用日内选择入场时间点。一般日间与日间例如 1 天与 5 天相对更容易结合，可以认作 5 天的底仓、1 天做择时。但这样的融合，局限性在于非常依赖长周期的判断，以长周期做决定短周期做微调。这也就类似图表界的实用口诀：月线看趋势、周线看方向、日线看买卖。

再如开盘之前日频的模型就确定好了当天要买入的仓位，要买入 A 股票、卖出 B 股票。结果 A 高开 5 个点，B 低开 5 个点，割裂的策略会追高买入 A、杀跌卖出 B，没有用到实时信息形成损失。

策略融合现阶段市场上比较主流的组合模型是 Alpha+T0 模型，可拆解为：Alpha 模型是通过大量因子组合筛选出具备超额的股票组合，旨在跑赢指数。在 Alpha 模型调仓完成后，T0 模型再在底仓的基础上进行高抛低吸，增厚收益。但这种模式当前也面临挑战，需要在 Alpha 收益和 T0 收益上做出取舍。即选择 Alpha 收益高，但是 T0 表现不好的底仓；还是 Alpha 收益低，T0 收益高的底仓。过往 A 股波动率较高，T0 收益足够丰厚，足以弥补割裂体系带来的超额损失。但近年来 T0 收益衰减严重，传统模式下的 T0 收益在超额中的贡献越来越微薄，T0 的收益并不能覆盖掉高波动带来的 Alpha 的损失。

■图表 T0 策略基本概况

定义	T0 策略又称日内交易策略，持仓时间较短，基于对未来短期股价走势的判断，低位买入，高位卖出，以此来获得价差收益，买入卖出交易在日内完成。
分类	按策略逻辑 1：融券 T0 利用融资融券的方式变相实现 T0 交易，管理人首先在券商创立两融账号，并在券商的两融名单基础之上选取适合进行日内交易的股票名单进行交易，买入股票的同时利用融券卖出，以此来变相实现 T0 交易，所以融券 T0 的主要成本在于两融利率。
	按策略逻辑 2：底仓 T0 则完全复制或部分复制指数构造出底仓，之后利用底仓进行日内交易，而为了避免底仓的净值波动导致产品净值出现波动，底仓 T0 利用股指期货或指数 ETF 融券等方式进行对冲，因此底仓 T0 对基差走势较为敏感，投资者在选择底仓 T0 时也需要对基差进行合理的判断。
	按交易方式 1：人工 T0 则聚焦于几十只股票，收益弹性更强，但是会受到交易员的影响，成熟的交易员往往收益更加稳定。人工 T0 策略成本更高，对股票的要求较高，收益弹性更强，规模受制于交易团队人员数量。
	按交易方式 2；机器 T0 又称为程序化 T0，通过提前设定好的规则，采用程序化方式进行日内交易，优势在于利用高频数据可以对全市场股票进行监测，成交次数多，持仓分散，但是相较于交易员，缺乏对市场的灵活应对。一般来说，程序化 T0 策略收益较为稳定，但是弹性较低，策略运营成本更低，可覆盖股票数量更多，规模上限更高，
适宜行情	市场换手率市场波动率日内振幅当市场处于高换手状态时，T0 策略收益较好当市场波动率处于高波状态时，T0 策略整体表现更好当日内振幅处于高位时，T0 策略收益表现最好
风险点	融券 T0 对券源的依赖程度较高，因此建议关注以下 3 个方面： 1-券源的稳定程度：融券 T0 对券源的要求较高，因此需要明确管理人与券源提供方是否有长期的合作关系。按策略逻辑分类 2-融券的数量和规模：融券的数量对策略的容量有较为直接的影响，一般而言，融券的数量和规模越高， 3-3-差源的利率水平：适合做日内交易的票息较高，因此需要交易团队根据自身能力特点选择特定的票

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

	底仓 T0 由于需要使用股指期货或 ETF 融券进行对冲，建议关注以下 3 个方面： 1-基差：如果使用股指期货对底仓对冲，则需要关注基差的变化情况，避免因基差剧烈波动导致策略收益出现较大程度波动。关注的风险点 2-底仓构成：有些管理人会采取部分复制底仓，以此来选取更多适合进行日内交易的股票，部分复制底仓的产品相较于完全复制底仓的产品净值波动也会更大。 3-收益拆解：一般而言，底仓 T0 的收益由三个部分组成，对冲成本、跟踪误差以及 T^0收益，其中 T0 策略的收益才是对管理人日内交易能力的直观反映
	其中人工 T0 依赖于交易员的交易，且多采用融券形式进行日内交易，可以从以下 3 个方面进行考虑： 1-产品间的业绩差异：部分私募管理人，可能不同的交易员负责不同的产品，产品间的收益存在差距。 2-交易员数量：交易员数量对 T0 策略规模会有较大的影响，一般而言一个交易员管理的资金规模不会很大，因此需要多位交易员一起进行交易来扩大产品的规模。 3-交易员限额：交易员限额对 T0 策略的风控有重要影响。一般而言，管理人会对交易员的交易规模以及最大回撤进行限制，比如：1%亏损提醒，2%亏损强制平仓。
	机器 T0 由于完全由系统对全市场个股进行监控与下单，多为程序 T0，因此应当关注量化系统，可以从以下 3 个方面进行了解： 1-交易系统：交易系统是机器 T0 日内交易能力的基础，一般而言，单独 T0 策略的产品会有自主研发的交易系统，而部分将 T0 作为收益叠加的产品采用的可能是交易商提供的交易系统，相比较而言，自主研发的交易系统综合能力更强。 2-预测周期：按照预测周期的不同，可以分为 Tick 级别、分钟级别、小时级别等等，一般而言预测周期越短，胜率越高，策略容量也越小。 3-风险管理：机器 T0 的风险管理也同样重要，主要涉及对于风控以及换仓频率的要求，当市场环境不佳时是否会减少开仓频率，而当市场环境表现较好时增加开仓的频率，降低一风险的同时增加收益。”

数据资料来源：公众号《白小白的学习笔记》

■图表 T0 策略和市场中性策略业绩表现情况

		2018 年	2019 年	2020 年	2021 年	2022 年
T0 策略	年化收益率	8.81%	10.00%	17.28%	8.87%	5.70%
	最大回撤	-0.25%	-0.35%	-0.15%	-0.23%	-0.21%
	夏普比率	7.53	5.01	6.27	5.76	4.56
	卡玛比率	35.15	28.41	112.71	38.12	26.70
市场中性策略	年化收益率	8.44%	9.79%	8.15%	3.42%	4.19%
	最大回撤	-1.28%	-1.07%	-1.60%	-3.43%	-1.57%
	夏普比率	2.52	2.76	2.46	0.57	1.00
	卡玛比率	6.59	9.13	5.11	1.00	2.66

数据资料来源：私募排排网，华宝证券研究创新部

注：2023 年截至 6 月 25 日

全频段覆盖的难点在于：首先将不同频段的策略进行融合得无明显频段短板，市场上很多其他管理人仍处于超额有“瘸腿”的状态，即专门擅长某部分频段，部分频段收益一般；其次如何平衡、分配不同预测周期的权重是其中攻克难点。

所以现在量化机构部分会做信号融合，做法百花齐放。有的是从特征层面融合，有的从优化器层面融合，还有的是直接从组合层面融合。白鹭投资在 2022 年底到 2023 年初即将手头的所有信号频段以及频段内所有信号进行整合，从而实现最优仓位，其是从组合层面融合。

卓识也在不断的策略迭代过程中，形成了“多频融合调仓+微趋势算法”的高频交易特色。多频段融合调仓通过不同预测频段，结合日内连续调仓，充分挖掘超额收益。

概率投资的多频段融合系统则将信号简单分为日内和日间两部分，日内信号本质上是在为日间信号做增强或削弱，即增

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

强同向的信号，削弱反向的信号。最终融合得到的调仓信号同时用到了日内和日间的信息，更能将底层的预测能力发挥到极致。

■图表 概率投资多频段融合模型



数据资料来源：概率投资

■图表 组合优化多信号

	日内信号	日间信号
预测周期	30 分钟、2 小时	5 天、20 天
常用数据	截面数据、逐笔数据、资金流数据	eod 数据、基本面数据、另类数据
因子类型	日内量价因子、复合因子	日间量价因子、基本面因子、另类因子
常用模型	神经网络模型	树模型

数据资料来源：概率投资

半鞅接下来的研发目标也是所有信号频段以及频段内所有信号进行整合，从而实现最优仓位。从 2018 年开始，公司即开始机器学习模型测试，是行业内最早运用机器学习取代传统多因子模型的机构之一。现在半鞅已实现工业流水线级投研流程，端到端全流程优化，以提升整体性能为目标，输入为原始 Level2 量价数据，使用批流一体引擎以毫秒级延迟聚合成不同尺度的特征，通过统一的端到端深度学习模型进行多周期、多目标的建模，具备全天候任意时间点的毫秒级预测和决策能力，不受限于定时定点决策，更好适应市场变化。

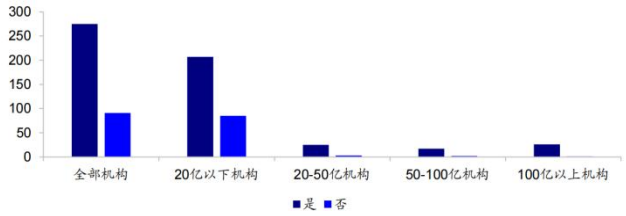
端到端除了省去因子层这一环节，半鞅在其他环节也在不断优化流程，例如机器学习方面做多尺度、多周期动态建模，大规模分布式训练，自动化机器学习；策略方面做多周期融合、多决策点融合。同时 Alpha 策略进行多策略自适应叠加，每个子策略从预测周期、数据输入、模型类型、预测目标 4 个维度打标签。上层策略配置模型根据市场环境、产品特性来配置子策略权重，综合考虑收益和风险，灵活换手，最终实现稳健的收益输出。单策略的深度决定了收益率，策略的广度决定了波动率，半鞅力求在单个策略的深度和广度上都做到最优，以提供给客户更好的风险调整后的收益。同时公司持续灰度上线新的端到端模型，抵抗原有策略的衰减。【更多内容详见本书第六章最佳实践-半鞅投资】

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

4.交易执行

从调查问卷的结果来看，量化私募大部分采用自建的交易系统（275/366），占比 75%。百亿私募中（27 家有相关标签）也仅 1 家表示采用外部采购的交易系统。相比私募，公募绝大部分还是采用外部采购的交易系统。低延迟或高频受市场重点关注，实际上目前有两个应用方向：1.高频策略交易；2.算法交易下单。

■图表 是否采用自建交易系统的量化私募的数量对比（家，2023 年）



资料来源：私募排排网，海通证券研究所

低时延的追求是一个持续不断的过程，全链路时延竞争正在行业内全面展开。国际交易所、对冲基金和投行都在不断推进其在低时延领域的步伐，与此同时，国内的交易所和券商也在不断提升系统的交易延迟，以提升市场竞争力。

交易所层面，以深交所为例，2016 年上线的新一代交易系统，即采用流式报盘，交易回路时延可以到 1.5 毫秒左右。即使后来深交所将机房从从竹子林数据中心迁移至南方中心，增加了链路的物理延时，后续也通过优化控制在了 2-3 毫秒。而上交所此时仍采用轮询的报盘方式，50 毫秒一轮询，交易回路时延达 100 多毫秒，与深交所存在一定差距。直到 2020 年，上交所从期权系统开始，改用流式报盘网关，期权交易回路时延降低到 10-20 毫秒。到 2021 年的时候有了更多的变化，上交所新债系统上线，采用流式报盘，并在底层采用分布式架构，债券交易回路时延降到了 1.5 毫秒左右。再到 2022 年，上交所现货竞价平台进行了内部优化，通过数据库报盘节省流失，并重新规划内部撮合，现货交易回路时延从 100 多毫秒降低到 20 多毫秒。

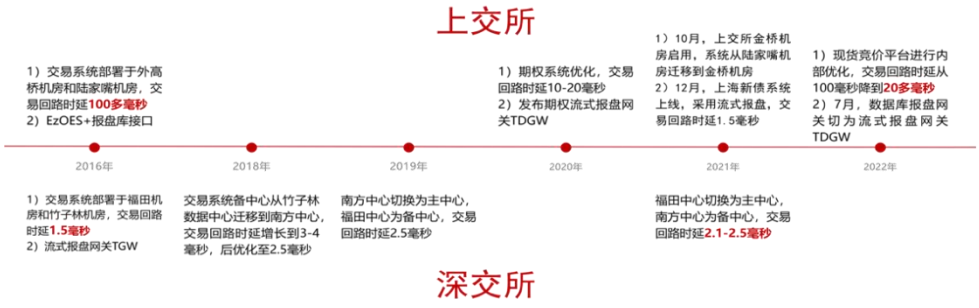
近几年，国内越来越多的券商投入到极速交易的军备竞赛。少数券商采用自研的方式构建极速交易系统，例如中泰证券、华鑫证券等，大部分券商则通过采购技术厂商的系统构建自身极速交易的技术服务能力。华锐技术作为该领域市占领先的技术厂商，已为国内 30 多家券商提供相关系统，并实现了头部前 20 券商的全面覆盖。

全链路时延竞争在行业内全面展开，从交易所到券商再到量化机构，在基础设施、技术架构和应用系统三个维度上持续优化提升。基础设施层面，全球最新的硬件设备，例如超频服务器、FPGA 卡、低时延交换机等，已在行业内实现了成熟的应用。技术架构层面，底层基本采用分布式架构为主。因其可以提供高可靠、高并发、低时延的消息传输能力，并可以支持上层应用轻松实现水平扩展。应用系统层面，在操作系统、资源占用等细节上也有多种手段来进行跟踪优化。

但在追求低时延的过程中，保障系统的安全可靠仍是一切手段的前提。2022 年 7 月，中国证券投资者保护基金有限责任公司向行业通报了 3 起证券经营机构极速交易系统和集中交易系统出现异常的案例，主要包括“虚增”客户资金、客户使用“虚增”资金后账户出现透支、多扣收客户融券红利资金等问题。其中部分是由于极速交易系统与集中交易系统信息交互出现异常导致。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 国内证券交易所的低时延演进之路

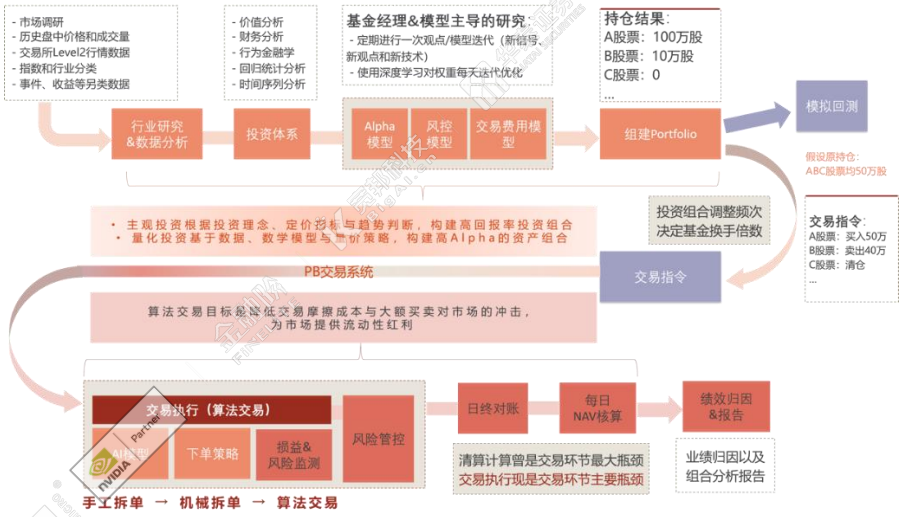


数据资料来源：华锐金融科技研究所

第一代传统算法在上世纪八十年代于海外产生，直到 2000 年海外三方算法公司方才进入中国市场。而第一代智能算法于 2011 年左右诞生，从被动、机械的算法，到智能化算法，中间相隔了近三十年。2017 年国内第三方技术厂商开始涉猎智能化算法，飞速拉近了国内与海外的技术差距。随着近几年市场格局发生变化，赛道热度、认可度、规范化程度进步良多。

算法交易的赛道发展可以分为几个阶段：从最早阶段的人工交易，发展到以 TWAP/VWAP 为核心的第一代被动机械算法，追求执行效果和降低市场冲击；第二代主动机械算法，以执行价差为核心，关注交易成本和风险；第三代交易算法——Iceberg 算法，主要目标聚焦在寻找流动性、隐藏交易意图。AI 技术的应用，正在为算法交易带来新技术范式。

■图表 算法交链易在产业的位置及作用



数据资料来源：卡方科技

高频策略交易当前主要是基于 Level2 行情数据，通过多路优选、硬件加速等形式，拿到交易所低延迟的逐笔和盘口切片数据，且按照交易所的撮合规则按需生成高频切片数据，挖掘到市场上较全较新的量价指标。结合实时市场行情和交易员选择的策略及参数等，利用计算机程序和数学模型来决定交易下单的时机、价格和数量。

其主要目的在于：1.降低交易成本，通过拆母单，拟合市场成交量分布，战胜市场均价。2.减少对市场的影响，采用特殊目的算法，有效保护交易意图，避免引起市场异动。3.降低人力成本，使用自动化交易算法交易策略，大幅提高交易员效率。4.降低监管及合规性风险，内嵌风控阈值检查令交易更安全，严格符合监管要求。5.减少手工误差，减少因市场发生快速变化或交易任务过多造成的人工失误。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

其使用场景，除了服务量化机构一次交易成百上千只股票，减少频繁交易的滑价、提高交易效率之外；对于主观基金、减持解禁等交易场景，也可以实时追踪市场流动性情况，合理地拆分交易指令，提升大金额场景下的调仓效率，减少执行难度。

几种常见算法有 TWAP、VWAP、POV。

TWAP：在设定的时间范围内匀速下单，降低市场冲击，最小化与市场 TWAP 的偏差。

VWAP：在设定的时间范围内对根据对市场成交量分布的预测进行下单，降低市场冲击，最小化与市场 VWAP 的偏差。

POV：多用于减持相关，价格不敏感，市场冲击可控的情况下，按照市场参与度，尽可能快而多地完成交易。

在这一环节，海外机构经历过人工手动交易研究，程序化交易再变成现在的 AI 程序化交易。国内则省略中间步骤，直接从手动到现在 AI 下单。第三方算法交易平台提供了丰富的算法交易类型，提升了算法执行效率。

量化机构有的是自建算法交易，也有的使用第三方算法，直至 2023 年市场仍然不断有新的算法厂商崭露头角，年内某券商招标，共有 23 家机构报名，包括自诚、皓兴、启能达、非凸、迅投、卡方、丽海弘金、金纳、跃然等。

一般情况下，智能算法能帮助量化机构降低 3-5BP（基点）的交易损耗。参考《上海股市流动性和市场冲击成本分析》中的数据——沪市平均人工交易成本是 60 个基点（60/10000），普通算法的冲击成本是 10 个基点（10/10000），而主动智能算法拆单交易目前甚至能做到高于平均价净赚 6 个基点（6/10000）。

按照 10 亿规模、一年 100 倍年化换手率的私募产品来估算：

第一种情况——人工交易，会产生 6 亿的冲击成本。10 亿 × 100 ×（60/10000）=6 亿；

第二种情况——普通算法拆单交易，冲击成本是 10 个基点，也就是 10 亿 × 100 ×（10/10000）=1 亿，相对于人工交易可以节省 5 个亿的冲击成本，每年节约 5 亿，大约可以提升 50%的产品业绩；

第三种情况——主动智能算法拆单交易，目前的算法下不仅没有冲击成本还能额外创造出 6 个基点的业绩，也就是 10 亿 × 100 ×（-6/10000）=-0.6 亿的冲击成本，相对于人工交易来说节约 6 亿之后还赚了 0.6 亿，大约可以提升 66%的产品业绩。使用人工智能方法的核心优势是，当数据量足够大、有很好的性价比的时候，更复杂模型的非线性性质能够更好地挖掘数据价值，更好地服务于流动性提升、增效市场反馈与价格发现。人工智能方法通过类似于人脑学习的方式不停迭代反馈，不断优化模型，对于不同主体、不同交易量、不同订单不会千篇一律地做选择。

■图表 不同特征对收益的影响

特征因素				绩效
调仓时间	全天	上午	9:30–10:30	正常
			10:30–11:30	负影响（一般）
		下午	13:00–14:00	负影响（严重）
			14:00–15:00	负影响（一般）
	周期<30 分钟			负影响
成交额占比	<3%			正常
	3%–5%			负影响（一般）
	5%–10%			负影响（严重）
高低价股	低价股<5 元			负影响（严重）
个股特征	中证 500、沪深 300、上证 50			正常

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

	换手率低、成交额低的个股	负影响

正常：指的是某因素对绩效不造成影响；负影响：指的是某因素会降低绩效（程度：一般、较大、严重）。

资料来源：非凸科技

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

5.业绩归因

许多时候量化机构会将业绩归因为各类因子为业绩所带来的贡献，但从 2021 年开始，各类 FOF 机构等开始质疑量化机构创造 Alpha 的能力，尤其是到达一定规模之后。问卷调研显示，投资人的质疑与量化机构的自我认知其实较为接近。

Pure Alpha（平均分 4.25）：Pure Alpha 是指策略相对于市场基准的超额回报，与市场波动无关。数据表明，量化机构普遍认为 Pure Alpha 对业绩的贡献是有限的，这可能是由于市场效率提高，导致纯粹的 Alpha 机会变得更加稀缺。

Smart Beta（平均分 4.79）：Smart Beta 评分较 Pure Alpha 略高，量化机构认为自身通过 Beta 因子暴露获得了一定回报。从 2014 年小市值暴露，到 2021 年严控风格到逐渐放开，至 2023 年小市值暴露的回归，量化的风格暴露呈现脉冲式，量化机构这部分收益也在逐渐走向成熟。在进行风格研究时，机构会深入理解暴露的风格是什么，在市场上的表现如何，以及暴露风险是否可控，在可控范围内是否可以带来更多的 Alpha 收益等。此外，机构还需考虑客户对暴露风险的接受程度。

Pure Beta（平均分 4.2）：Pure Beta 通常与市场指数的表现挂钩，反映了市场整体的回报。量化机构给出的评分表明，市场回报依然是整体业绩的一个重要组成部分。

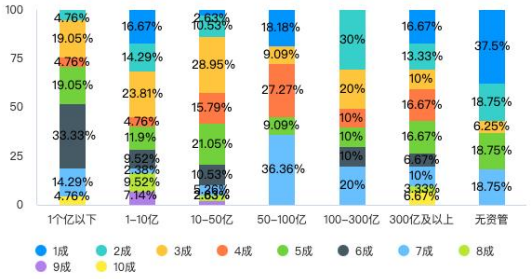
值得注意的是，1 个亿以下规模的机构（平均分 5.29）和 50-100 亿规模的机构（平均分 4.55）对自己创造 Pure Alpha 的能力给出了相对较高的评分。一般而言，随着机构规模的增加，对持仓收益的依赖增强，而对 Pure Alpha 和交易收益的依赖减弱。反映了量化机构在不同规模下的策略调整和风险管理的差异。

2022 年中国量化投资问卷调查发现，现在量化机构资产管理规模/最优容量比例的平均值为：47.27%，最优容量指的是量化策略能有效管理并保持预期收益的最大资本量。平均值为 47.27%意味着在对所有参与调查的量化机构进行平均后，他们当前的资产管理规模大约为最优容量的一半左右。整体对于规模和收益之间的平衡持谨慎态度。这意味着，市场整体仍然存在增长空间，同时量化机构在风险管理方面持谨慎态度，倾向于保持较低的资产规模，以避免因超出最优容量而带来的额外风险。

■图表 量化认为当前行业业绩总体归因应为



■图表 各规模区间自认为其创造 Pure Alpha 的能力



数据资料来源：《2023 中国量化投资白皮书》问卷调研

■图表 各规模区间以及业绩归因

规模	Pure Alpha	Smart Beta	持仓收益	交易收益	规模	Pure Alpha	Smart Beta	持仓收益	交易收益
1 个亿以下	5.29	5.48	5.67	6.10	50-100 亿	4.55	4	6.64	4.18
1-10 亿	4.10	4.55	5.95	4.86	100-300 亿	4.1	5.7	7.3	4.8
10-50 亿	4.24	4.71	6.16	4.42	300 亿及以上	4.27	5.03	6.03	4.63

数据资料来源：《2023 中国量化投资白皮书》问卷调研

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

如前所述，近年量化机构在不断降频，基于规模扩大、监管环境变化、交易成本影响等，传统认知上，量化机构会通过短期买卖价差获取收益，但问卷调研数据显示并非如此，体现在量化机构对持仓收益、交易收益和配置对业绩贡献的评估上：

目前持仓收益（平均分 5.99）：指的是资产长期持有产生的收益。数据显示，量化机构认为持仓收益对业绩的贡献是最高的，这可能意味着机构更加重视长期投资而非频繁交易，且可能倾向于更加稳健的投资策略来实现收益。

交易收益（平均分 4.75）：来自短期的买卖操作。相对于持仓收益，量化机构认为交易收益对业绩的贡献较低，这可能反映出机构正在减少频繁交易的策略，转向更少频率的交易以降低成本和市场冲击。

配置（平均分 4.84）：配置收益来自资产间分配的优化决策。量化机构认为配置对业绩的贡献比交易收益高，这可能表明机构在资产间的分配和调整上花费了更多精力，以提高整体投资组合的效率。

择时一向被认为是量化投资中比较难的部分，预测市场最佳时机颇具挑战，许多机构在公开资料中表示不参与择时，但从业绩归因看，量化机构仍然给予了选股（平均分 5.53）择时（平均分 4.22）一定比重。目前来看少量机构在灵活对冲产品中会进行择时，这种策略并非在多头仓位择时，而是在空头仓位进行择时对冲。

大多机构在做的是空头端择时，极少机构在 Alpha 端做择时。思勰的策略模型有择时信号，会在上涨下跌环境采用不同 Alpha 模型：当市场上涨时，通过分散投资，做稳定 Alpha，会持仓 700-1000 只票；在市场下跌时，更倾向去做超额 Alpha，控制回撤，持有 300-500 只票。这也是因为思勰为高频 CTA 起家，擅长做市场趋势方向判断，后沿用到了股票策略里，以用量价数据做择时信号触发，做更灵活的操作。北京丰润恒道拥有 40 多个择时模型，独立发出择时多空信号；根据策略逻辑性以及历史表现赋予相应权重；根据策略多空信号以及权重得到总体市场判断，分别判断上证 50、沪深 300 和中证 500 多空。

■图表 业绩归因与机构关注的策略特征之间的对比数据

策略特征	Pure Alpha	Smart Beta	持仓收益	交易收益	策略特征	Pure Alpha	Smart Beta	持仓收益	交易收益
成单率	4.33	4.89	5.95	4.91	胜率	4.03	4.70	5.92	4.79
跟踪误差	4.35	4.92	5.97	4.80	与其他机构的异质性	4.28	4.96	6.27	4.78
收益率	4.25	4.82	5.96	4.84	在不同市场环境下的适应性	4.22	5	6.17	4.79
波动率	4.19	4.88	5.81	4.98	抵抗回撤与极端风险的能力	4.06	4.72	5.95	4.69
容量	4.27	4.78	6.07	4.76	模型的可解释性	4.26	5.03	6.11	4.78
拥挤度	4.33	5.11	6.09	4.93	其他	3.55	4.45	5.18	6

数据资料来源：《2023 中国量化投资白皮书》问卷调研

在归因时，策略失效是量化机构需要面临的永恒命题，目前许多基金经理都表示过，策略失效越来越快。稳博投资使用归纳法和演绎法对其进行解释。归纳法使用风险解释模型来解释超额波动原因，后者则寻求市场或者策略失效的本质、外因、内因。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 策略失效控制方法

策略失效原因	风险解释因子：行业、成长性、估值、动量、市值、分析师等	本质，新的市场环境和历史规律有较大变化 内因，超出容量，因子拥挤 外因，大体量资金投资偏好转变
策略失效办法	求解优化方程： 最大化： 优化目标=组合目标收益−组合风险暴露扣分−交易成本 约束边界：单因子暴露边界 最大个股持仓偏离度 个股最大换手规模	风控方式： 核心逻辑：不干预策略运行，争取屏蔽确定性高的回撤。 规避风险：因子拥挤度检测和剔除，滚动训练，控制管理规模。观测市场，主动降低风险度。 风险处理：出现无法解释的回撤幅度/时间长度，优先降低各类已知的风险敞口，防范风险扩大。

数据资料来源：稳博投资

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

04 市场竞争

1.机构竞争

“百亿扩容，新老分化”2022 年问卷调研有人如此形容接下来的行业竞争态势。

如第一章所述：2023 年，A 股量化进入存量博弈。这种存量表现在：1.增量资金陷入停滞；2.新备案私募门槛增高；3.从规模分布和结构上看逐渐拉出梯次，量化私募行业在过去几年已经形成了越来越快的“飞轮效应”：业绩好的机构形成口碑和规模，吸引了人力、资金、资源的大幅进入，同时算力、数据、投研等方面不断积累，资管规模不断扩大，而资源的增加又促成了良好业绩的产生。大型机构维持护城河，其投资风格上会更为稳健。

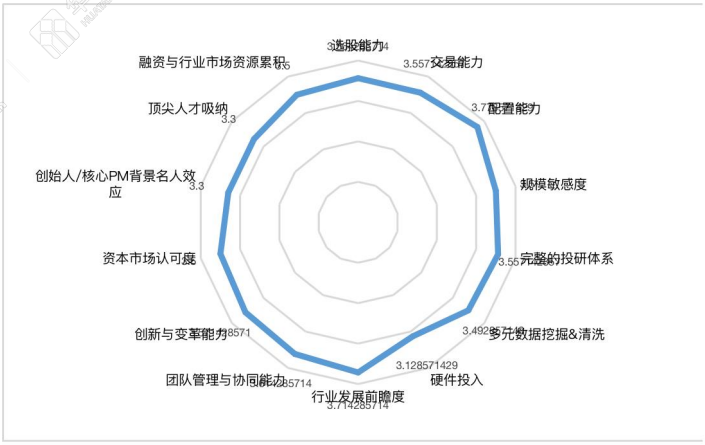
机构竞争力影响市场募资。而量化投资机构的核心竞争力是多方面的，需要在技术、人才、研究、风险管理等多个领域综合施策。随着市场的发展和技术的进步，这些竞争力因素也在不断地变化和升级。核心竞争力通常体现在以下几个关键领域：数据处理能力、研发和模型构建、技术和算法优势、风险管理、投资研究、策略多样化、适应性和迭代更新、人才队伍。

但作为一家商业机构又体现得更为多元，包括行业发展前瞻度、资本市场认可度以及创始人/核心 PM 自身的影响力。综合来看，量化机构普遍认为自己在配置能力、市场交易经验、迭代能力和团队管理与协同能力方面具有较强的竞争力。同时，数据也揭示了量化机构在产业资源、顶尖人才吸纳和资本市场认可度方面可能需要进一步提升。这些竞争力评分可能与机构的业务焦点、市场定位和内部资源配置有关。

■图表 量化机构竞争力（2023）



■图表 量化机构竞争力（2022）



2023		2022	
配置能力	3.49	选股能力	3.56
产业资源	2.89	交易能力	3.56
数据广度及深度	3.25	配置能力	3.78
策略丰富度	3.15	规模敏感度	3.50
市场交易经验	3.44	完整的投研体系	3.56
迭代能力	3.45	多元数据挖掘&清洗	3.49
规模红利	3.2	硬件投入	3.13
行业发展前瞻度	3.47	行业发展前瞻度	3.71

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

团队管理与协同能力	3.49	团队管理与协同能力	3.61
资本市场认可度	3.17	创新与变革能力	3.57
创始人/核心 PM	3.44	资本市场认可度	3.50
顶尖人才吸纳	3.03	创始人/核心 PM 背景名人效应	3.30

数据资料来源：《2022-2023 中国量化投资白皮书》问卷调研

接下来，机构为了维持竞争力，不断投入精力进行策略研究，如果把所有的研究分为进攻端和防守端，前者寻找和利用市场机会以创造收益；防守端则是管理和控制风险，确保资本的安全和策略的持续性。据数据显示，进攻是最好的防守，机构仍然更重进攻而非防守。

防守端（风控和避免类似事件）的综合得分低于进攻端（创造收益）。可能源于风险管理复杂性、收益驱动偏好、资源分配偏差、合规和监管环境的不确定性等，尽管风控至关重要，但在竞争激烈的市场中，机构需要展示其能够创造收益的能力，以吸引和保留投资者。而监管环境的快速变化可能使得量化机构在监管层面的研究得分较低，因为监管政策的不确定性使得机构难以准确预测和适应。

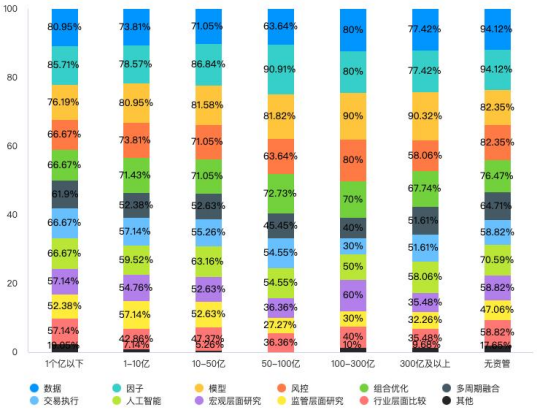
进攻端（创造收益）：因子（综合得分 8.68）：模型（综合得分 8.3）：数据（综合得分 8.02）：人工智能（综合得分 4.6）；交易执行（综合得分 3.72）：多周期融合（综合得分 3.66）。

防守端（风控和避免类似事件）：风控（综合得分 6.2）：组合优化（综合得分 6.13）：监管层面研究（综合得分 2.26）：行业层面比较（综合得分 2.18）。

■图表 未来一年，预计大幅精力投入及研究的环节【排序题】



■图表 未来一年，预计大幅精力投入及研究的环节为与规模交叉比较



数据资料来源：《2023 中国量化投资白皮书》问卷调研

以 50 亿为分界线，将数据合并为两个组，数据显示不同规模的量化机构在研究重点上存在差异，其中 50 亿以下的机构可能更倾向于在进攻端投入，寻找提升收益的新机会和策略，而 50 亿以上的机构可能更加聚焦于防守端，通过风控和组合优化来保护资产并提高策略稳健性。这种分布可能反映出机构规模对研究方向和资源分配决策的影响。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

2.机构榜样

连续两年，问卷都邀请量化机构填写其心中的海内外的三家量化机构榜样，这些机构作为隐含了其行业标准制定者、技术创新引领者、市场影响力塑造者身份。

声誉建设是一个多维度的过程，涉及业绩、创新、品牌建设、市场参与和个人影响力等多个方面。但业绩与规模仍然是行业内部声誉第一性的存在，部分机构即便在过去 3 年深陷社会负面舆情，但作为榜样力量仍然被多次提及，2022 年最为突出的机构为九坤与幻方，2023 年新增的机构为衍复投资，三家机构在行业中以断档式存在，在 455 次有效提及中，幻方（68）、衍复（63）、九坤（58）、宽德（36）、明汰（33）、灵均（13）、佳期（12）、稳博（8）、诚奇（6）、鸣石（6）名列前茅，目前这些机构均为 150 亿以上。

在社会公众方面，658 次有效提及 117 家机构，最为知名的机构分别为：九坤（118）、幻方（110）、衍复（73）、明汰（54）、灵均（45）、宽德（44）、稳博（20）、黑翼（14）、聚宽（9）、天演（8）、世纪前沿（7）、启林（7）、诚奇（6）、鸣石（6）。

2.未来挑战

将量化机构面临的挑战为监管与市场环境、策略与研究、投资者关系与人才三大类，数据显示：

监管政策（平均分 3.82）：被放在了首要位置，甚至超越策略研究（平均分 3.61），反映出量化机构对监管环境的不确定性。担心监管政策的改变可能导致交易成本增加，限制某些交易策略，或要求机构增加合规和风控的投入。

市场风格（平均分 3.59）的极端变化可能会导致量化模型失效，以及随着行业的成熟行业竞赛（平均分 3.51）走向更高的差异化和专业化使得量化机构更多压力。

面向未来一年，总体来看，量化投研人员对未来市场的预期较为中性，部分因素略偏向乐观（如市场行情 Beta 走势、成交量和流动性、科技创新），而对于监管政策及舆论导向和交易、策略集中度则较为担忧。量化机构可能需要在维持创新和适应市场变化的同时，加强风险管理和合规能力，以应对潜在的市场挑战。

量化人心理预期积极的类目为：市场行情 Beta 走势、成交量和流动性、科技创新

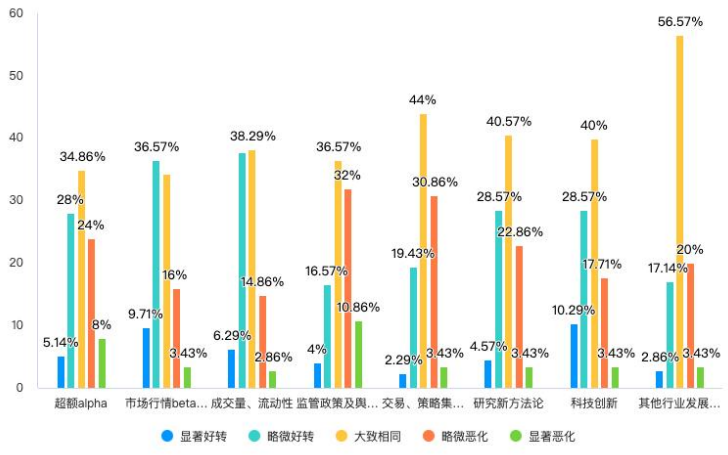
中性预期：超额 Alpha、研究新方法论、其他行业发展因素

悲观预期：监管政策及舆论导向、交易和策略集中度

■图表 未来一年量化市场包含哪些挑战

类目	平均分
流动性枯竭：市场极少增量资金	3.25
监管政策：交易规则修改及监管与合规风控	3.81
市场风格：市场风格极端演绎、变化，超越模型	3.59
全球经济：衰退、货币政策、地缘政治等	3.31
策略研究：方法论过于集中、差异化较小、Alpha 获取难度加大等	3.6
行业竞赛：行业到达一定阶段，后发机构难以突围等	3.52
净值压力：募资压力以及投后净值压力	3.46
投资人信念丧失：投资人对量化持怀疑态度	3.29

■图表 量化人员对下一年度行业前景的预期



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

人才招聘：招聘及人才培养的投入过多	2.98
市场 Beta：指增挂钩宽基指数大幅回撤	3.09
基差波动：对冲成本较难控制	3.08
小计	3.3

数据资料来源：《2023 中国量化投资白皮书》问卷调研

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

CHAPTER

第三章

风暴中的我们

```
storm = lambda: [1] * 1000
anchor = storm().count(1)
print(f"Us in the {anchor}")
```

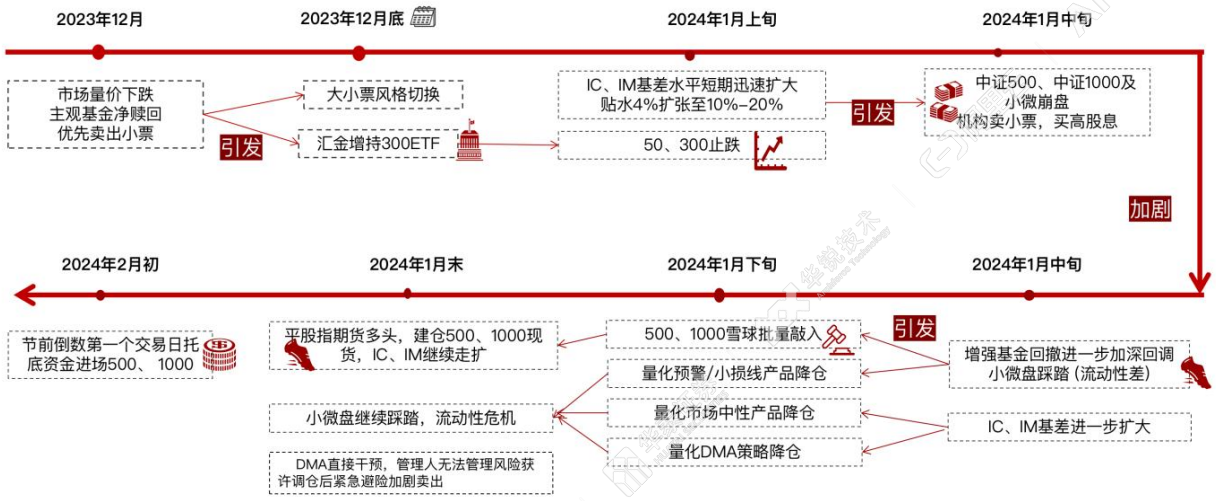


官方下载及活动链接—https://BigQuant.com/quantpaper/2023

01 回顾：雪球敲入 引发 Beta 对冲 Alpha 连续三杀

本书虽定位回顾 2023 中国量化投资市场，但几乎很难对 2024 年春节前的量化崩塌视而不见：2024 年 1 月下旬-2 月初，雪球大量的敲入诱发负面情绪，之后市场经历“市场量价下跌，主观基金净赎回优先卖出小票——雪球集中敲入引发基差快速扩大——对冲成本快速增加导致 DMA 亏损平仓——DMA 多头部分平仓导致小微盘股踩踏——暴露在小微盘的主动增强基金回撤进一步加深回调——资金入场拉升 500/1000ETF 导致市值分布不均的指增超额收益回调——DMA 直接干预，管理人无法管理风险获许调仓后紧急避险加剧卖出”连锁反应多个阶段。

■图表 2023 年 12 月–2024 年 2 月量化流动危机演变

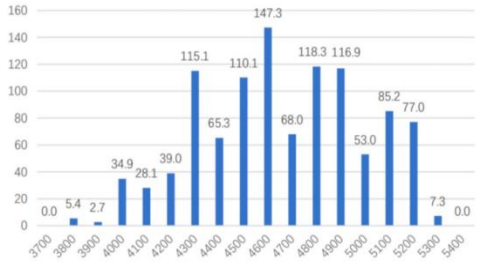


世纪前沿的情况说明梳理了时间节点的情况：3 年前，也就是 2020 年末-2021 年年初，市场多次出现单日千亿资金抢购一只基金首发的疯狂盛况。但经过三年大熊市的煎熬，其中部分三年期封闭基金在 12 月底面临开放，叠加市场持续负反馈状态，以大盘成长股为首的主观股票基金持续净赎回，跑得越晚亏得越多的心理在投资者心中逐渐蔓延。在流动性递减的情况下，规模较大的管理人通常会选择卖出小票以应对负债端的赎回压力。市场情绪进一步脆弱化，避险情绪提升，指数的下跌逐渐蔓延到中小市值。

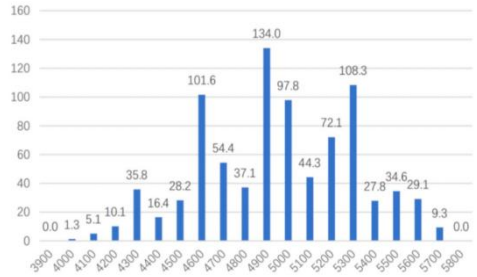
一月首两周的市场普跌后，500 和 1000 指数在第三周（1 月 15 至 19 日）接近雪球的敲入密集区（编者注：据测算，中证 500 雪球的集中敲入区间在 4300 点至 4900 点之间，中证 1000 雪球的集中敲入区间在 4600 点至 5300 点之间）。为应对雪球结构敲入，券商交易台需要减仓其作为对手方持有的股指期货。与此同时，国有银行、中字头国企等避险板块在资金支持下企稳拉升，对市场产生虹吸效应，进一步加重市场避险情绪，中小市值指数继续加速下跌。大量的集中敲入导致中证 500 和中证 1000 股指期货基差贴水快速持续扩大，基差扩张到年化 20 点以上的罕见水平。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 中证 500 雪球在不同指数点位的敲入合约规模估算（亿）



■图表 中证 1000 雪球在不同指数点位的敲入合约规模估算（亿）



数据资料来源：信达研究

国内的量化产品由于调整被限制，无法进行应对。而在海外的外资量化资金则利用此次机会平仓中性产品，导致股票仓位被卖出，但此时流出量没有特别大，小票流动性尚存，影响有限。该周量化私募的 500 指增超额基本在正负 1%以内，中位数约为-0.3%。

自 1 月 18 日开始大量资金开始买入沪深 300ETF，使得沪深 300 稳定在 3200 点以上水平，而 500 和 1000 指数则因为雪球敲入带来的卖盘影响下加速下跌。此时的市场已在恐慌情绪中。市场参与者基于观察到的情况普遍形成了 300 指数是安全的共识，大量抛售小票而买入 300 成分股。这个共识在 22 至 26 日逐渐形成，并在 1 月 29 日至 2 月 2 日迅速强化。

在大小盘风格出现极致分化的情况下，量化私募的 500 指增超额在前一周表现还较稳定，中位数约为 0%。而后一周的超额中位数则到了-4%，基本破了各家机构单周回撤的历史记录。从以上信息可见，量化产品自 1 月中起就一直面对资金面的不利影响，而且是逐渐加速的。量化策略本身有吸收资金面波动，稳定市场截面的作用，因此前两周（1 月 15 日-26 日）的超额影响有限。但当市场风格一直走极致的单边行情时，量化的吸收能力被耗尽，随后市场的流动性迅速消失，量化超额也开始出现加速亏损。

市场整体进入极度悲观甚至恐慌的状态，小微盘股票流动性开始枯竭，这些压力在最后一周内变成现实。2 月 5 日-7 日，大量资金买入 500 和 1000 的指数 ETF，正好对应了绝大部分量化资金的对冲或对标端，指数相对非成分股票大幅跑赢，导致量化超额进一步出现大幅亏损。同时再进一步引发中证 2000 等小市值股票的流动性缺失，市场的微观结构出现失衡，成分股持仓比例不足成为超额回撤的核心原因。在此期间，部分避险资金会选择将大/小市值股票切换为中市值（中证 500 及中证 1000 指数）成分股，进一步形成资金踩踏。

■图表 1 月 22 日-2 月 2 日沪深 300 与沪深 300-中证 2000 指数收益差

指数涨跌幅	沪深 300	中证 2000	中证 500	中证 1000	沪深 300-中证 2000 指数收益差
1 月 22 日-1 月 26 日	1.96%	-1.41%	0.17%	-0.70%	3.37%
1 月 29 日-2 月 2 日	-4.63%	-16.78%	-9.23%	-13.19%	12.15%

数据资料来源：世纪前沿

■图表 节前两周的市场涨跌幅数量和指数变化情况

	上涨数量	下跌数量	主板跌停数	上证 50	沪深 300	中证 500	中证 2000
1 月 29 日	474	4591	83	-0.43%	-0.90%	-1.97%	-3.49%
1 月 30 日	228	4852	87	-1.45%	-1.78%	-2.29%	-3.19%
1 月 31 日	306	4765	184	-0.73%	-0.91%	-2.45%	-5.26%
2 月 1 日	1078	3947	99	-0.27%	0.07%	-0.39%	-1.06%
2 月 2 日	368	4704	108	-0.86%	-1.18%	-2.47%	-4.47%

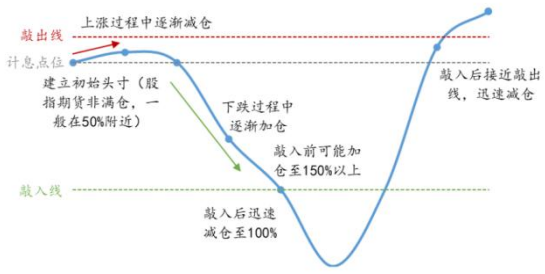
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

2 月 5 日	378	4704	1368	1.18%	0.65%	-2.26%	-9.49%
2 月 6 日	3586	1464	135	2.82%	3.48%	7.75%	2.02%
2 月 7 日	2132	2932	282	1.09%	0.96%	6.32%	-2.80%
2 月 8 日	4671	400	7	0.36%	0.64%	0.80%	8.75%

数据资料来源：灵均投资

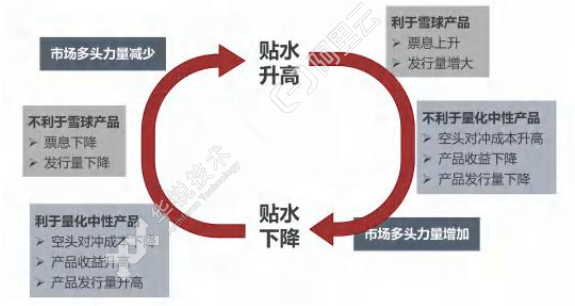
连续经历以上 Beta 端、对冲端、Alpha 端“三杀”，启林也谈到 2024 年 2 月 5 日至 2024 年 2 月 8 日，国家队开始买入 500ETF 救市，市场迎来反弹，但偏中小市值股票依旧承压，量化超额进一步大幅亏损无法卖出，部分管理人的 DMA 产品被追保，或被迫减仓甚至穿仓被强平，同时不少量化管理人为避免超额进一步回撤，通过买入成分股来平衡组合市值因子暴露，给权重股带来新的买盘增量，其他个股则越来越弱，导致超额亏损进一步扩大，这些操作造成了量化行业的挤兑和恐慌。

■图表 雪球发行方股指期货仓位运作示意图



注：蓝色曲线为雪球产品对应标的物走势模拟图

■图表 雪球与对冲交易的相互影响#



数据资料来源：《中国量化投资季刊·2022 夏》《穿透雪球》

数据资料来源：国泰君安证券研究

■图表 风暴时期主要宽基指数 ETF 成交额

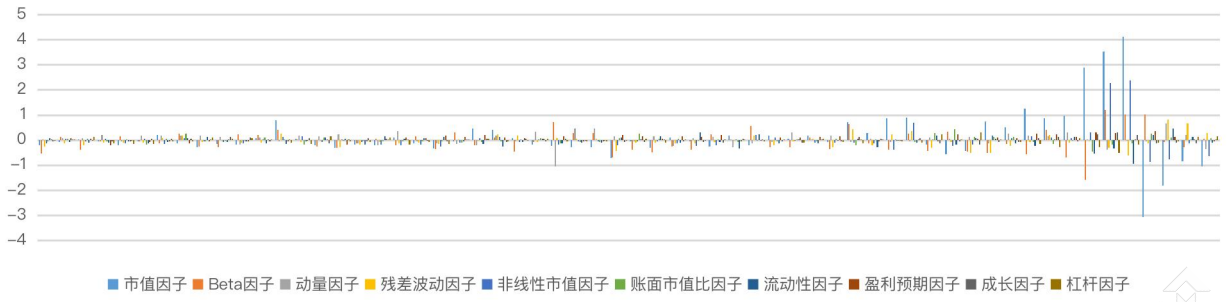
日期	沪深 300ETF	中证 500ETF	中证 1000ETF	中证 2000ETF	日期	沪深 300ETF	中证 500ETF	中证 1000ETF	中证 2000ETF
1/2	32.7	17.1	9.5	3.1	1/22	151.1	18.6	6.6	0.9
1/3	36.5	12.4	9	2.2	1/23	44.8	16.2	6.8	1.2
1/4	56.7	15	8.6	1.6	1/24	73.7	17	6.9	0.8
1/5	58	11	7.3	1.3	1/25	54.5	17.6	7.6	1
1/8	58.2	15.5	67	1	1/26	72.7	15	6.9	1
1/9	37.9	11.7	6.7	1	1/29	33.6	9.9	6.2	0.8
1/10	32.1	11	5	0.8	1/30	61.9	9.4	5.5	0.6
1/11	42.5	14.3	7.8	0.8	1/31	61.6	16.8	7.4	0.8
1/12	28.3	12.7	67	1	2/1	60.9	16.7	10.6	1
1/15	30.3	8.3	4.9	0.5	2/2	80.7	18	11.3	1.2
1/16	59.3	12.6	6.3	1.6	2/5	84.4	160.8	87	2.3
1/17	42.8	10.3	4.7	0.7	2/6	65.6	48.6	54	4.2
1/18	152.6	24.1	9.4	0.9	2/7	115.1	67.9	39.2	6.7
1/19	120.8	16.4	9.8	0.6	2/8	80.2	86.8	29.6	55.7

*沪深 300ETF 为（510300）；中证 500ETF 为（510500）；中证 1000ETF 为（512100）；中证 2000ETF 为（563300）

数据资料来源：龙旗

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 风暴期因子脉冲式变化



数据资料来源：火富牛

■图表 量化私募总结 2024 年初行情

宽德	1 月中下旬随着雪球产品敲入，市场结构出现了反转。2 月 5 日起，群体恐慌中很多非理性的操作出现。行业挤兑踩踏，流动性崩毁。
世纪前沿	挤兑性卖出和调仓引发的短期流动性危机等多重市场风险事件的连锁反应导致
衍复	雪球大量的敲入诱发负面情绪，导致指数下跌、基差扩张，间接导致中小市值股票被抛售，随后指数 ETF 净买入量异常放大，超额亏损严重继续造成量化机构恐慌式挤兑
卓识	市场大幅波动，风格极致反转、流动性踩踏、宏观限制，就像一场未有预警的风暴，中证 2000 及之后股票的流动性缺失、资金踩踏等多种情况，多空力量开始失衡
灵均	雪球集中敲入，融资盘爆仓，中小盘杀跌导致许多管理人调仓，引发踩踏行情。
黑翼	非指数成分股流动性严重短缺，个股分化严重，同时伴随着对冲成本提升，在多种不利因素影响下，量化股票策略承受较大压力。
九坤	A 股市场走出了连续的、超过历史极值的大小盘极端分化行情。春节前一周市场迎来反弹，但主要涨幅集中在几大宽基指数成分股内、偏中小市值股票依旧承压。
幻方	维稳资金进行强力干预，量化 DMA 平仓调仓导致小市值股票剧烈下跌，出现流动性危机，量化行业强化了风控。收紧小市值持仓比例。加剧了资金从小盘流向中盘股票。
启林	各指数板块间极端分化，雪球敲入导致股指平仓。资金轮番买入各宽基成分股，量化触发风控，换成分股，引起市场的流动性踩踏和非成分股和成分股之间的分化。
稳博	宽基大幅走强，引发量化对冲绝对收益/指数增强的超额出现大幅回撤，不少管理人更换成分股，加剧中小市值走弱。
明汯	小票急跌波动率放大
因诺	500 和 1000 指数出现反常暴涨。量化机构普遍大幅回撤，迫使各机构纷纷抛售小盘股，买入 500 和 1000 成分股，从而形成了负向反馈。
乾象	短促且剧烈的大小盘分化，叠加流动性影响，使得全市场选股的量化策略难以获得正超额。
金戈量锐	微盘股经历跌停潮，同时 500 及 1000 指数成分股大幅走强。回撤较大的产品可能会追加权重股、成分股，微盘流动性困境，股指期货基差巨幅波动，大小盘走出剧烈的剪刀差风格。
致诚卓远	1 月中下旬雪球大量敲入诱发了负面舆情，中小盘股票遭到抛售，股指期货出现大幅贴水，政府大额买入 500 和 1000ETF，部分 DMA 产品被迫减仓并调至成分股内，加剧超额亏损。
蒙玺	市场风格发生急剧风格切换，尤其在大市值和小市值之间的分化。由于指数的严重分化，叠加上市场流动性风险，市场环境的不利因素使得产品业绩上均遭遇一定的回撤。
锋滔	1 月下旬中小盘下跌引发大量雪球产品敲入，基差扩大到历史极值。中性产品、DMA 类产品为锁定基差进行大规模平仓，诱发了负面情绪，2 月 5 日宽基指数反弹，中小市值依旧承压，部分管理人增持权重股，其他个股走弱
信弘	主力资金入场拉升指数，市值严重分化，全市场杠杆性产品线回撤中小股票踩踏引发流动性风险。而后量化资金逐渐开始布局成分股。
龙旗	从 0118 开始，在市场一路下跌后主力资金开始救市。0129–0202，市场的小票开始持续走低，超额出现回撤。0205 开始，主力救市资金，转移到了 500 和 1000，小票发生踩踏，出现流动性危机。

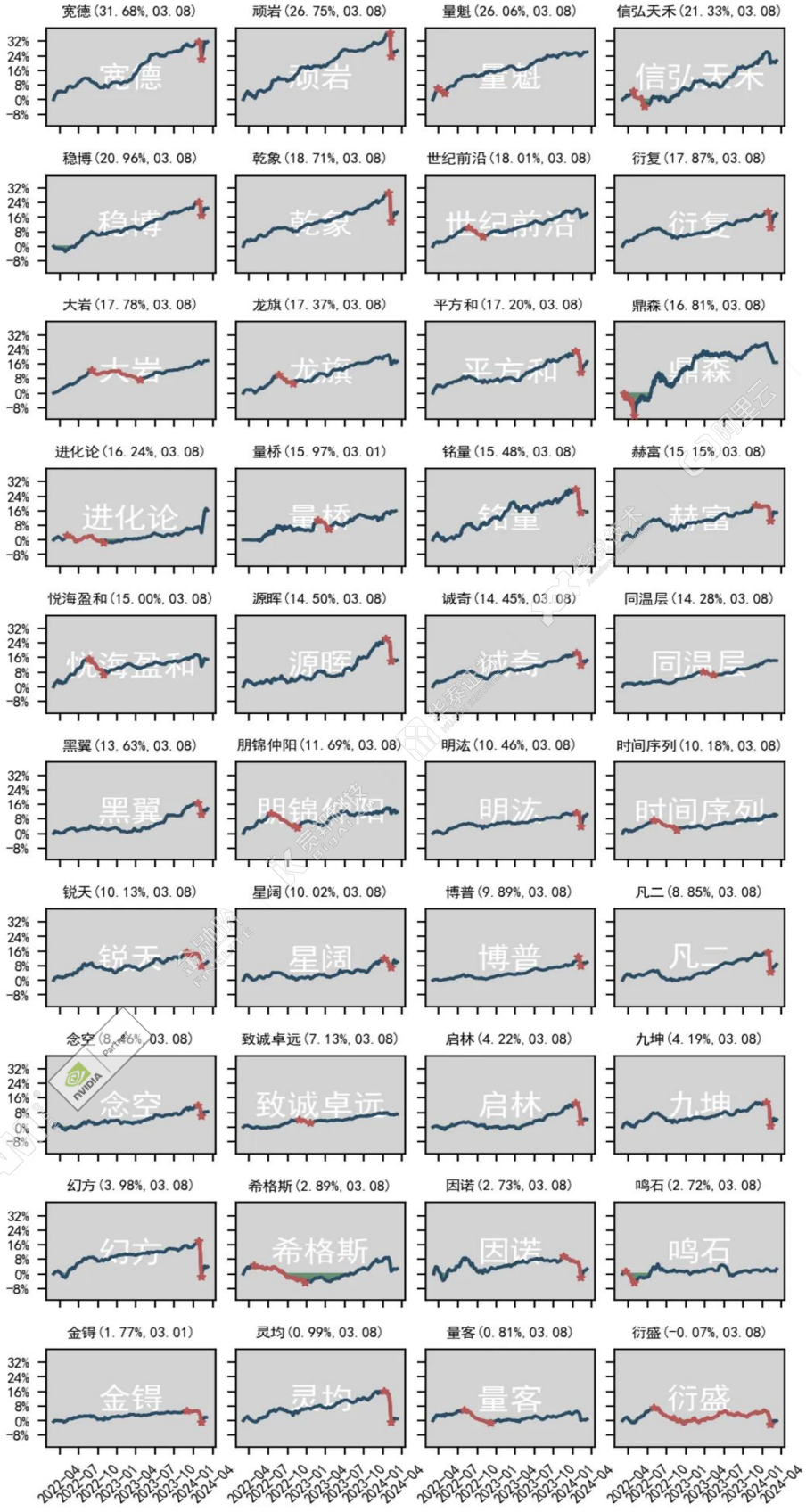
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

念空	市场不准净卖出且国家队大举托市，狂拉沪深 300 和中证 500 指数最终造成了中小市值板块的流动性挤兑进而引发了本次量化指增的 Beta 和 Alpha 的双暴跌。
宽投	面对持续下跌爆仓风险，各机构加速收紧敞口，非理性竞相卖出，加速多空力量的失衡。
诚奇	雪球敲入导致股指期货贴水上升小微盘股下跌，叠加杠杆的影响出现了踩踏迹象。
聚宽	市场风格急转直下，流动性受到严重考验，宽基指数快速拉升，中盘股整体收益率高于其他股票。
概率	雪球敲入、股指期货巨额贴水，负面情绪导致中小股集体抛售，大盘宽基大幅拉升。
同温层	股指深度贴水，市场交易结构极度分化，大市值股票表现显著优于小市值股票。微盘股抱团瓦解等因素引发市场流动性风险。
仲阳	雪球集中敲入，各策略减仓，抛售小微盘，小市值股票流动性枯竭。
玄信	此次市场环境，是汇聚了长期下跌，雪球、基差、大小盘极致价差、流动性等多重因素的结果
橡杉	政策压力以及市场的各种哀嚎此起彼伏，中小盘流动性枯竭以及监管政策叠加市场情绪导致的踩踏，致使众多量化资产出现较大回撤。
鸣熙	大小风格极度偏离，存在很强非线性，触发了流动性崩溃。
尊和	随着雪球产品陆续敲入，市场波动率加剧，出现了结构性反转的行情。指数分化明显，风格转换的情况，微盘股遇到流动性风险，出现了连续数天的踩踏行情。
半鞅	宽基成分股开始爆拉，股指期货贴水急剧收缩，外部的突发力量改变了 A 股的微观结构、交易习惯。
罗维盈安	指数下跌触发雪球敲入、DMA 产品爆仓，中证 500 和中证 1000 指数因救市措施急剧拉升。
无隅	除市值风格暴露外，指数大幅上涨，DMA 资金的踩踏也起到一定作用。
俊丹	小微盘股交易过度拥挤，赛道脆弱，外部因素打破小票抱团，必然出现踩踏式离场。
天算	雪球集中敲入，使得股指期货基差扩大，进一步引起量化中性以及 DMA 产品的平仓，导致了中小盘股票流动性的集中踩踏。
茂源	市场各板块极端分化，小微市值流动性危机，多空力量失衡。大量资金为了避险，踩踏性地把大市值和小市值的股票转换为 500 和 1000 的成分股，链式过程不断加剧。
子午	市场增量资金匮乏，尾部风险事件主推，量化同质策略遭遇基差和负超额共振后的流动性踩踏。

数据资料来源：各量化投资管理人信

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 2024 年初部分量化机构回撤



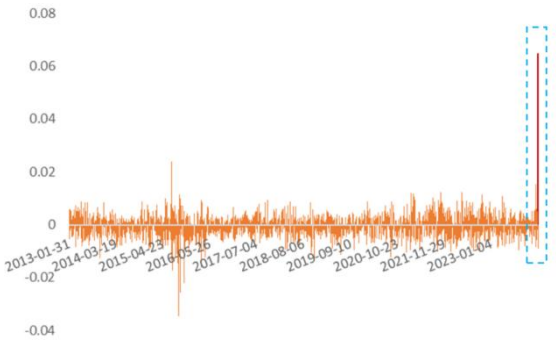
数据资料来源：公众号大白话

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

02 应对：模型失效 调整派坚守派各占一半

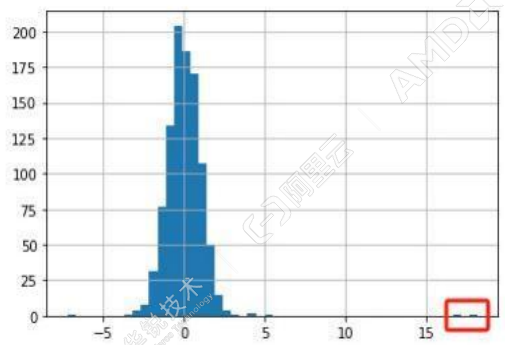
如第二章所述，量化机构一般执行较为严格的、系统性的风控体系，通常不会主动人为介入。但由于 2 月初市场行情极端，中证 2000、中证 1000 与沪深 300 指数之间的收益率差距多次超过 3 个标准差，尤其是在 2 月 5 日达到了 7 个标准差，春节前最后一周非线性市值因子收益超过了历史均值+15 倍标准差，属于非常典型的极端行情。在这种情形下，成分股的权重占比已然超过了所有因子的贡献。

■图表 春节前最后一周非线性市值因子收益超过了历史均值+15 倍标准差



数据资料来源：方正研究

■图表 非线性市值因子波动日度标准差情况



注：横坐标为标准差，纵坐标为频次（一天为一频次），红框处为本周非线性市值因子情况，可以看出近 4 年该因子波动都在±5 标准差以内，该因子到了 15–20 标准差，极为罕见。

数据资料来源：灵均投资

由于量化投资是基于历史数据挖掘和统计学规律来开发策略，挖掘正常市场行情下的错误定价，为市场有效性作出修正。短时间内，市场环境的大幅变化，股票之间的价格关系持续被打乱，过往数据就不再具有指导意义，模型只能失效且持续发出错误信号。

世纪前沿指出，这种失效具体在：以量价为主的量化策略，在小票极端下跌的情况下会判断其后续反转概率较大，同时由于市值的波动，在历史上并不突出，二者叠加使得组合在极短的时间内，会重新偏向小票风格。所以世纪前沿对风控模型做出了调整，适当地收紧了模型的波动率约束，调整迅速地反映到了各个产品线上，使得超额回撤保持在了相对可以接受的范围内。

本问卷调研了 2024 年初极端行情期间，量化机构策略进行了何种调整。数据显示：调整策略与未调整的机构各占一半，调整的类目最主要的是未明其他（66.47%），另外包括因子权重（51.18%）、杠杆比率（49.41%）、交易执行（48.82%）、指数成分（45.88%）。

■图表 2024 年初极端行情期间，量化机构的策略调整

题目\选项	没有调整	调整	题目\选项	没有调整	调整
交易仓位	62(36.47%)	108(63.53%)	模型优化	68(40%)	102(60%)
指数成分	78(45.88%)	92(54.12%)	交易执行	83(47.43%)	87(51.18%)
杠杆比率	84(49.41%)	86(50.59%)	其他	113(66.47%)	57(33.53%)
因子权重	87(51.18%)	83(48.82%)	小计	575(48.32%)	615(51.68%)

数据资料来源：《2023 中国量化投资白皮书》问卷调研

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

例如灵均就在投资人信中分享了其截至 2024 年 2 月 7 日，策略持仓成分股分布、Barra 风格控制及收益来源拆解情况如下：

■图表 灵均投资持股变化情况

	中证 500 指增	中证 1000 指增	风格特征：	中证 500 指增	中证 1000 指增
持仓数量	约 2500 只	约 2500 只	市值因子	0.2 个标准差	0.1 个标准差
权重占比：			收益来源占比：		
沪深 300 成分股	40%	10%	纯超额 Alpha 贡献占比	65%	65%
中证 500 成分股	16%	14%	小市值贡献占比	20%	15%
中证 1000 成分股	14%	40%	其他贡献占比	15%	20%
中证 2000 成分股及之后	30%	36%			

数据资料来源：灵均投资

是否调整策略是一个关键又沉重的选择，决策过程是对管理人的极致考验，市场快速演绎甚至使得有的机构被小市值反复打脸。

进入到 2 月 5 日，龙旗科技继续评估策略，发现部分子模型正在愈发抄底超跌的小票，于是降低了偏量价因子和高波动的子策略权重，提高了在极端行情下回撤更小、更稳健的子策略权重，收紧了流动性和市值的约束，希望通过市值因子的约束降低超额的回撤。

本着少做人工干预的原则，龙旗科技继续保留了让模型在全市场选股，并未做太多选股域与期货端的切换。但春节最后几天的行情远比预想的更为猛烈，从 2 月 5 日开始，主力救市资金转移到了 500 和 1000。当周前三个交易日中证 500 拉涨 11.97%，大量的资金抛售小市值股票，转而持有 500、1000 成分股。小票发生踩踏，出现流动性危机，个股中位数跌幅达-6.82%。中证 500 指数和中证 2000 指数的收益差（用除法计算相对走势后再按复利计算）急剧反转，从 0.7 迅速降到了 0.2，在几天之内回到了 21 年初的低点。

这是过去十多年从来没发生过的情况，也导致原有的风控体系无法起到控制超额回撤的效果。从归因结果来看，NLSize 非线性市值这个风格在 2 月 6 日和 7 日达到了史无前例的负向显著性，偏离度达到了 40 多个标准差，造成了超过 8%的风格负收益。在这种情况下，只有把选股域限制在成分股内，才能起到控制回撤的效果。但龙旗在 2 月 5 日到 7 日一直坚持没有人工干预模型的选股域，导致龙旗的超额出现了巨大的回撤。

在坚持了 3 天之后，龙旗科技在 2 月 7 日下午评估当前的行情过于极端，出于风控的考虑不得已收紧了一部分的选股域，同时对风格因子做了非常紧的约束。却没想到市场的风格切换如此急剧，8 日主力资金开始拉中证 2000ETF，中小票报复性反弹，中位数上涨 7.35%，而前期强势的 500 指数与 300 指数反而处于尾部 10%分位数，前一日的风控措施让其错过了大幅的反弹空间，最后导致 2 月 8 日当周的超额表现非常差。

回顾春节前两周，龙旗认为公司预判到行情的极端程度，导致在最被动的时候才不得不人工干预，出现了大幅的负超额。应该在风控上更加主动地去预警，更早地切换选股域去应对结构化行情。在这种十年不遇的行情下，应对非常不佳，也不得不非常被动地做一些调整，最终造成了非常差的结果。

调整策略的机构大多并非一蹴而就，天算量化总结了风暴集中期的操作：

周一（2 月 5 日）：流动性危机开始，市值因子控制到 0.1 附近，量化对冲继续降低到 30%-40%仓位。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

周二（2 月 6 日）：流动性危机继续，国家队入场拉指数，天算调整 30%策略仓位为成分内避险。

周三（2 月 7 日）：流动性危机加剧，非成分股无差别大幅跑输指数，天算继续调整 50%仓位到 80%成分股。

周四（2 月 8 日）：周三收盘后证监会领导更替，周四资金流入中小盘股，流动性危机暂时解除天算成分内选股策略降低到 60%。

受调问卷中几近一半的机构几乎未做任何调整，部分机构认为即便再大的冲击也将会暂时性的。

衍复选择不调整策略的原因与底气在于：根据美股历史上的几次 Alpha 流动性危机和 A 股 2014 年底 Alpha 集体大幅回撤的经验，只要产品不因为爆仓、触及止损线等原因被动终止运作，且不人为干预量化模型，最终超额将会在不长的时期内自然修复。当流动性危机一旦解除，市场恢复正常状态后，前期因恐慌而不惜代价夺路而出的交易者（主观和量化）又将重新调回正常模型或正常交易行为，届时将加速推升超额的修复。类似 2 月 6 日、2 月 7 日两个指数的反向逼空行情也会出现在其他的股票身上，2 月 8 日中证 2000 指数及其他小市值风格指数的表现就是例证。因此公司深知在这种情况下不进行人工干预是正确选择。

宽德 2 月 5 日当周制定了相应的计划，但并未触及设置的干预阈值，所以没有做过多的人工干预。宽德认为主观调整持仓对于点位的判断要求较高，会带来一定的人为干扰，本质上也增加了不确定性，因此 2 月 5 日那周内没有做选股范围的切换。对于中性策略，由于外部的限制和主观考量，也没有在基差扩张的时候减仓。

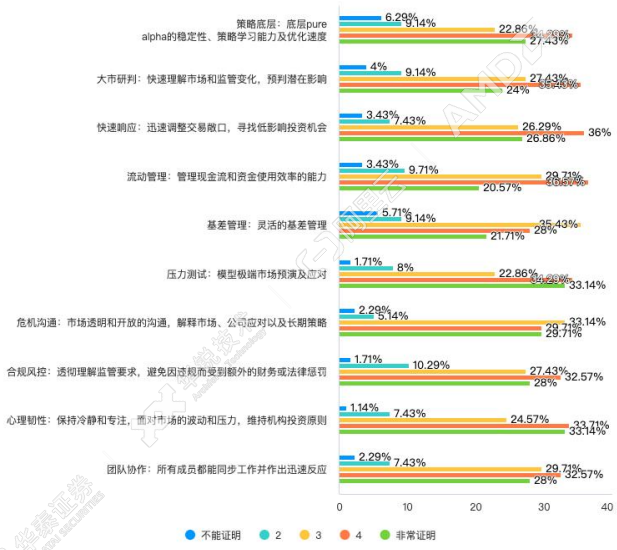
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

03 判断：风暴之中 软性实力凸显韧性

极端时期究竟是否能展现量化机构的特殊能力？问卷数据显示，量化投资机构认为，本轮回撤在很大程度上可以证明机构在策略底层、大市研判、快速响应、流动管理、基差管理、压力测试等投研底层硬性能力。但可以发现本轮危机中，一些软性的实力被放在了更显著的位置，排名前 5 的为心理韧性、危机沟通能力、团队协作、快速响应、合规风控。

■图表 本轮回撤多大程度能证明量化机构的能力

策略底层：底层 Pure Alpha 的稳定性、策略学习能力及优化速度	3.67
大市研判：快速理解市场和监管变化，预判潜在影响	3.66
快速响应：迅速调整交易敞口，寻找低影响投资机会	3.75
流动管理：管理现金流和资金使用效率的能力	3.61
基差管理：灵活的基差管理	3.51
压力测试：模型极端市场预演及应对	3.89
危机沟通：市场透明和开放的沟通，解释市场、公司应对以及长期策略	3.79
合规风控：透彻理解监管要求，避免因违规而受到额外的财务或法律惩罚	3.75
心理韧性：保持冷静和专注，面对市场的波动和压力，维持机构投资原则	3.9
团队协作：所有成员都能同步工作并作出迅速反应	3.77



数据资料来源：《2023 中国量化投资白皮书》问卷调研

在所有选项中，**心理韧性得到了最高的评价（平均分 3.91）。这表明在市场波动和压力下，保持冷静和专注，坚持投资原则被认为是机构强大实力的最重要体现。**

例如涵德投资合伙人顾小军就提到，其 2006 年入行加入 WorldQuant，2007 年美国经历了量化崩塌，WorldQuant 创始人 Igor Tulchinsky 迅速砍仓，三天内将所有仓位全部清完，两周之内又全部加回来，决策又快又准。但开始交易后发现这样的决策其实极其艰难。【更多内容详见第六章涵德投资最佳实践】

市场透明和开放的危机沟通（平均分 3.79）排名第二，这其实更大层面表现投资者关系的处理，在危机时期与投资者保持清晰沟通，有效传达机构的应对措施和长期策略，是非常关键的能力。

历史上，Citadel 经历的最大回撤为 2008 雷曼倒闭的那一个月，最大回撤 20%、全年回撤 55%，Citadel 当即要求暂停投资者赎回，2008 年之后 Citadel 全面改变了风控，所有的交易都考虑到因子暴露，并坚持分散投资，以达到杠杆率、收益和波动之间的稳态，第二年基金即反弹 62%。【更多内容详见《中国量化投资季刊 2023.冬刊》】

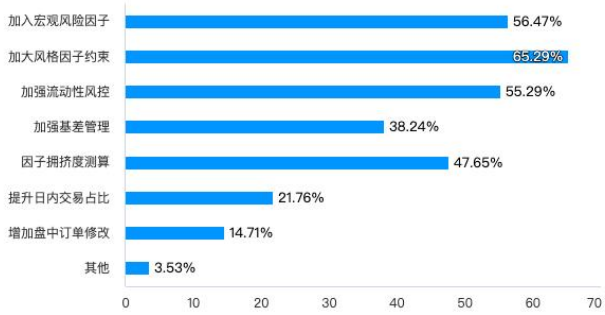
事实上，本轮教训中还最需关注的有合规风控，透彻理解监管要求并避免违规而受到额外的财务或法律惩罚。2 月 20 日，上交所和深交所发布公告称，灵均投资集中大量下单卖出，影响了正常交易秩序，从 2 月 20 日起至 2 月 22 日暂停其交易，并启动公开谴责纪律处分的程序。随着监管趋严，不少量化机构便开放了合规相关岗位招聘，配合落实监管部门对私募基金的合规管理工作要求。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

04 优化：因子约束 宏观判断

2024 年春节，量化机构无人能眠。诸多机构在春节期间加班加点上线了控制风险的子模块，从优化措施的排序中可以看出，量化投资机构在极端行情之后更加注重风险管理和策略的稳健性。包括加大风格因子约束（65.29%）、宏观风险因子（56.47%）、加强流动性风控（55.29%）、因子拥挤度测算（47.65%）以及加强基差管理（38.24%）。

■图表 极端行情之后，策略将进行何种优化?



■图表 替代本轮踩踏之中，量化行业最该学习成长的#



数据资料来源：《2023 中国量化投资白皮书》问卷调研

问卷以“回顾本轮踩踏，您认为量化行业最该学习成长的是”为题进行了文本类调研，要求列出三个关键词，数据显示量化机构认为最该学习成长的领域围绕着对风险的敏感性、对市场变化的适应性、对策略的创新性，以及对监管环境的理解度等方面展开。因其信息含量较大，编写组作以下全文刊载：

关键词 1	关键词 2	关键词 3	关键词 1	关键词 2	关键词 3
舆论	监管	合规性	对商业模式的理解	对产品的理解	对风控的理解
提高风控模型的重要性	增加超额来源，提高分散度	加强与监管和投资者的沟通	风险管理	技术技能	量化模型和策略
策略多样性	因子暴露管理	流动性监测	宏观	流动性	国家政策
对超额是属于 Alpha 还是 Smart Beta 的判断	对极端情况下市场的前瞻判断	和投资者的沟通	风控	风控	风控
方法论的更新迭代	对于 Alpha 和超额的理解	听妈妈的话	风格因子暴露度的控制	杠杆的合理使用	交易拥挤度的把控
如何应对监管	如何应对风控	如何应对科技变化	减少抱团	跟着趋势	减少操作
行业整体的风格集中度	加深监管理解	无	科技提升风控能力	科技提升策略计算和迭代能力	科技管理团队绩效
风格暴露	数据迭代	同质化问题	快速响应	及时止损	风险控制
数据	行业标的	策略	避免过度押小盘股	控制好尾部风险	关注流动性风险
做好自己	超越自己	顶峰相见	风控	执行	人性
踩踏因子的研究	与监管的沟通机制	与公众的沟通机制	宏观理解	监管和政策导向	策略和模型的稳定性
风险控制	策略多样	迭代	严格风控	模型严谨	不盲目跟风
宏观风险分析	风格因子约束	极端流动性风险的防范	风控约束风格	同行拥挤因子储备	杠杆不用
紧盯政策导向	加大策略的个性化研究，防范趋同性一致导致的	无	对黑天鹅时刻保持警觉	极端情况下的应对	对模型恢复的信心

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

	踩踏				
提高预警风险评估能力	做好仓位灵活管理	做好投资品种的风险对冲	风格暴露度	情绪控制	市场理解
政策	敏感度	人工干预模型	做好双边分析	模型优化	风险预警能力
对于政策的学习和执行	更灵敏地优化模型	低调	盈亏同源	高杠杆只适用高风险收益偏好资金	卷超额可能最终把行业卷没了
人工智能	数学	政治立场	拥挤度	杠杆率	投研能力
监管要求	拥挤度	市场主流	衍复	宽德	佳期
风控	市场风险	监管政策	市场压力稳定	合规变化	大市判断
对宏观风险及政策的研判和解读	极端行情下的交易模式革新	杠杆的限制	关注政策	时机	风险
极端行情应对	反思量化本质追逐的是什么	更灵活优质的策略	注意因子暴露	控制杠杆、对冲等工具	配合监管进一步规范
算法的有效性	重视监管	控制仓位	宏观研判	趋势跟踪	情绪跟踪
主观思维	长期价值	监管	策略多样化	投资渠道多样化	流动性约束方面需要加强
衍生品对冲	宏观经济/因子	大类资产配置	更贴近市场，监控风险	储备策略	风险事件应急机制
对监管的理解	如何避免策略同质化	如何做好投资者教育	关注监管	关注宏观	分散策略
加强政策学习	关注模型通过极端行情考验的能力	强化风险管理	加强与投资人交流沟通	重视封控合规	探索不同市场可能性
合规风控	策略多样化	危机应对管理	减少市值风格暴露	加强人才培养	加强风控
不要过分追风格，几年一轮动	好好做 Pure Alpha	好好约束各种持仓风格，不仅仅是 Barra 和行业	策略	风控	探索差异化
别赌	别赌	别赌	控制风险	政策研究	黑天鹅事件概率
组合优化	风险模型	Alpha 因子的适配度	不能抱有侥幸心理	行业轮动是一定存在的	顺势而为
风控	监管政策	舆论	模型的把控	监管的风控	技术的调优
人工干预是否必要?	如何引导投资人不过度比较?	暂时没想到	极端行情风控	监管政策感知	流动性监控
风控	控制敞口	提升 Alpha 能力	及时操作	择良	组合优化
宏观把控能力	板块轮动的特异化分析	中频交易的新策略	明确策略风格定位	敬畏市场	规范化管理
流动性	拥挤度	监管	宏观	择时	衍生品
对于国内政策底层逻辑的理解	策略差异化	与监管的沟通	风控	投研方法论	跟踪新技术
模型优化	人才引进	数据分析	策略同质化	交易拥挤	模型风控
政策研究能力	风险管理能力	交易拥挤度监控能力	敞口管理	压力测试	交易计划
策略同质化严重	保持初心	深度理解市场肥尾效应	观察因子拥挤度	对市场作出快速反应	监管预判
负债端管理能力	策略相似度	Pure Alpha 能力	风险控制	策略创新	风格暴露
风格约束	对监管影响判断	极端环境流动性估算	政策法规	流动性	风控
投研	服务	投资者教育	适应监管	收紧风控	控制管理规模
降低收益预期，控制风格偏离	重视因子模型的逻辑及回撤控制	避免异常交易	认识流动性风险	不要过于集中	产品不要设置清盘线
监管	风控	负债端管理	对政策环境的学习和预判	对模型的迭代	对市场的敬畏

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

关注监管动向	关注同行超额相关性	关注策略的趋同性	舆论监控	风控	择时
控制风险敞口	行情失效时止损	多学习下极端行情	迭代	总结	继续
对行情的宏观把控	对历史行情的细致复盘	更好的风控和抗压能力	危机应对	拥挤度研究	舆论
因子择时	模型漂移情况	因子拥挤度	择时	择时	择时
合规性	大势判断	防御型策略研究	风险管理	市场沟通	合理超额
流动性风险评估	因子暴露控制	慎用杠杆	监管政策导向	中国特色金融文化	一致预期的感知
不要卷业绩	要敬畏市场	不要闭门造车	策略底层	基差管理	其它
交易拥挤度监测	交易行为学习	风控	宏观风险	流动性枯竭时的应急预案	提供情绪价值
因子差异	关注基差	多策略	风控	监管动向	模型迭代
对底层策略的认识度	对交易细节的把握度	对宏观政策的了解度	政治周期	监管	策略拥挤
拥抱亏损	拥抱市场	拥抱自己	风格	相关性	危机管理预案
极端行情应对措施	拥挤度	相关性	永远保持对市场的敬畏	差异化发展	风险管理
流动性预判	监管严格程度要重视	严格执行模型不要人工干预	回撤	风控	因子
数据、特征、因子、模型的强可解释性和强因果性	行业舆情的快速有效发现和反应	人工智能在行业中应用的进一步发展	策略的把控	监管动态的掌控	风控的调节
对敏感性的风格加强监控	无	无	尾部风险识别	非人云亦云	做好自己
风险控制	极端应对	模型调整	风控深入到模型	人多的地方少去	不要加杠杆
快速反应	风险第一	策略储备	策略趋同	风控	N/A
市场微观结构衍生风险	风控模型	对于量化策略的监管态度及导向	哪怕市场β很弱，也别过于集中在某几类策略上	增强策略的差异化	异构策略百花齐放
流动性监控	宏观风险	资产配置	交易风险控制	因子拥挤度	宏观风险控制
调整的及时性	投资的心态	市场信息梳理	风控手段	预测因子	行业动态掌握
宏观周期配置风险	交易底层逻辑	事前风险控制	对风格暴露的敬畏	模型继续优化	多策略叠加
自身能力提升和持续迭代	过度的风险暴露需要谨慎对待	敬畏市场	宏观尾部风险管理	市场交易通道	指数负相关
严谨	独立	客观	面对极端事件的应对	宏观风险的把控	仓位的管理
尊重科学	重新认识 Alpha 和 Beta	减少虚假宣传	注重配置的力量，低相关和可解释性的重要意义	防止抱团，注重机构异质	不要一边承担过度的风险赚取短期相对业绩的同时，加大销售，加剧抱团和后续的踩踏，影响行业
合规	对投资者透明	注重风险管理	监管政策影响	风格暴露控制	尾部风险模型
政策	风控	模型调整	合规风控	压力测试	策略调整
丰富子策略	加强风控	极端环境下策略回撤控制	模型加入做空时融券费率和约束	对市场主流产品（雪球）的潜在影响分析包括市场情绪预判	资产分散化全天候策略研发
不要过多暴露于某风格	基差管理模型设计	极端行情应急操作指南	策略管理做细做精	团队精细化管理体系	对策略加速优胜劣汰的制度
极端行情下的研判与处	超额暴露的丰富度与分	衍生品在极端行情下的	预测	预测	预测

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

置能力	散	保护运用			
差异化	言行一致	勇于认错	不要反复横跳	说到做到	知行合一
极端行情下的风险管理和控制	因子轮动转换的把控	无	加强风格因子约束	提高策略丰富程度	提高监管政策变动敏感度
风控	风格	模型优化	流动性管理	策略多元化	走出国门
政策	政策	政策	风格均衡	心态	因子广度
基本面因子	政策风险	普及	风控	减少人为干预	分散底层资产
流动性的保持	策略底层 Pure Alpha 的优化	交易仓位的把控	分散投资	市场交易时点观察	区分 Alpha 和 Beta

数据资料来源：《2023 中国量化投资白皮书》问卷调研

对于后市的判断，有量化机构认为 2 月超额回撤事件有多项催化剂，包括中小市值股票在春节前流动性枯竭，国家队大量买入 500ETF，市场交易环境变化导致不少量化机构出现恐慌式挤兑等。以上催化剂缺一不可，才会导致超额的历史性回撤。因此这一回撤在未来几乎不会复现。但有几点高确定性变化属于行业共识：

1.行业出清。经历这次行情后，量化行业的整体规模会有所收缩。一些量化资金的赎回、DMA 规模的大幅下降，都会降低量化策略之间的竞争，让超额比之前更好做。

2.风险释放。随着节前量化的踩踏和超额的回撤，量化策略在过去多年里积累的风险已经大幅释放，甚至超跌。未来量化策略的运作将会轻装上阵，迎来相对较好的市场环境。

从中长期看，随着政策发力预期提升、市场风险逐步出清，成交量有望逐步回暖，情绪逐渐回归理性，机构坚定看好后市指数和超额的表现。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

05 因子拥挤度测算：中周期因子与日内因子

行情极致演绎、模型的必然选择，使得 2023 年小微盘强势。根据招商数据统计，30 亿以下微盘股贡献了 12%的收益，整个 75 亿以下小市值贡献 43%。使得一部分行情适应程度高、小微盘配比高的量化管理人获取极为亮眼的超额收益，也引发了市场担忧，如果市场风格发生转向，低流通市值的股票容易发生踩踏，市场对于量化业绩尤其是小微盘高配比管理人的潜在风险的担忧也愈加浓烈。最后在 2024 年 1 月底 2 月初风险集中兑现爆发。

■图表 量化私募各流通市值股票平均配比情况

时间	<15	15–30	30–50	50–75	75–110	110–150	150–200	200–500	500–1000	1000–2000	2000+
年初	3.40%	10.30%	13.00%	10.50%	9.10%	7.40%	5.90%	17.70%	9.20%	6.20%	7.20%
12 月底	4.70%	13.30%	17.10%	12.30%	8.60%	6.00%	6.00%	13.10%	7.70%	5.90%	5.10%
变化	1.30%	3.00%	4.10%	1.80%	–0.50%	–1.40%	0.10%	–4.60%	–1.50%	–0.30%	–2.10%

资料来源：招商私募指数数据库

■图表 市值分布因子分年度表现情况

统计年份	累计收益率	回撤幅度	最大回撤开始时间	结束时间	最大回撤持续天数	夏普比率
2009	7.69%	–3.56%	2009–01–05	2009–02–23	49	0.78
2010	1.21%	–2.99%	2010–07–05	2010–09–14	71	–0.20
2011	10.20%	–1.95%	2011–01–25	2011–02–23	29	2.86
2012	7.18%	–1.42%	2012–02–01	2012–03–13	41	1.90
2013	3.00%	–1.14%	2013–03–13	2013–05–31	79	0.43
2014	8.86%	–1.99%	2014–01–13	2014–02–18	36	2.37
2015	15.88%	–3.99%	2015–01–05	2015–06–02	148	2.28
2016	20.10%	–1.61%	2016–03–16	2016–04–06	21	4.72
2017	17.78%	–0.84%	2017–07–17	2017–07–31	14	8.32
2018	11.02%	–1.09%	2018–02–01	2018–03–05	32	4.72
2019	2.93%	–2.52%	2019–01–31	2019–03–12	40	0.41
2020	4.08%	–3.69%	2020–04–01	2020–07–13	103	0.67
2021	3.30%	–1.72%	2021–10–13	2021–11–29	47	0.50
2022	8.87%	–3.13%	2022–04–26	2022–08–18	114	2.09
2023	6.06%	–1.64%	2023–10–23	2023–11–20	28	1.67
2024	1.78%	–3.88%	2024–02–05	2024–02–08	3	1.66

数据资料来源：国泰君安研究

由于中盘股，也就是非线性市值因子，历史表现较为一般，量化基金持仓市值一般分布在大小两端，较少买入中间市值部分股票，例如组合等权配置沪深 300 与中证 2000，并不会等于配置了中证 500，50% 300+ 50% 2000 ≠ 100% 500 市值，市值因子暴露不大，但是在极端市场情况下，也会形成超额回撤，节前资金在底部通过 ETF 进场，500ETF 和 1000ETF 属于中市值股票，与量化基金持仓相悖，市场结构对量化影响极大形成“牛跌”。量化机构通过控制 Size 敞口是无法应对这样的市场行情的。

例如，卓识投资在复盘 1 月市场大跌之时，就谈到其风控体系主要依赖于历史数据规律，追求投资组合在各个特征的数学解角度和对标基准保持类似。以市值为例，过往通过市值加权来控制组合敞口暴露，在持有一定大市值股票的同时，持有一定小市值股票，在数学角度是可以同时持有大小市值股票的方式来对冲市值风格风险做到中性，这个风控措施在历时 4 年

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

多的实盘中并未暴露出很大风险，在更长数据中也没有显示出过高的波动，且在美国市场是长期有效的。但是，在大量资金集中买入中证 500ETF 和中证 1000ETF 的行情里，市场的风险已经变成了是否持有足够多的成分股，放大了成分股持股这个风险因素的影响，如果持有的 500、1000 成分股不够高，那么就会带来巨大的超额回撤。而与市值因素低相关的其他超额来源在这两周的极端踩踏环境中也没能带来其他 Alpha 收益。

据招商证券测算，2023 年底，量化策略投资的股票流通市值范围集中在 30-50 亿，占比达 17.1%，市值区间提升显著，超 4%；其次是 15-30 亿占比 13.3%，整个 75 亿以下的股票投资占比达 48%，小微盘几乎占据了一半，2023 年 8 月中旬市场下行趋势明显，策略在小微盘占比尤为抬升，11 月有所回落，12 月再次回升。

其中：50 亿以上头部的量化机构投资市值分布变化会更为平滑，小市值票增长幅度更小，大市值的缩减幅度更缓慢，微盘股的投资上更慎重，年内 30 亿流通市值以下的股票投资占比提升在 2%附近，部分管理人会有明确的风控限制。相反 50 亿以下管理人相对更激进一些，15-30 亿流通市值的股票最新投资占比 16%，比年初提高了 7.6%。

■图表 不同规模区间量化私募各流通市值股票平均配比情况

组别	<15	15–30	50–75	75–110	110–150	150–200	200–500	500–1000	1000–2000	2000+
全量	1.4%	10.7%	15.1%	11.0%	8.2%	9.9%	14.2%	6.1%	3.8%	3.7%
50+	1.7%	10.4%	14.9%	11.0%	8.2%	10.0%	14.3%	6.2%	3.9%	3.8%
50–	–21.1%	34.1%	25.6%	9.2%	7.2%	1.7%	8.8%	6.7%	3.5%	–3.6%

数据来源：招商私募指数数据库

细分不同业绩比较基准的指增来看：

1.30 亿以下的微盘股，300、500、1000 指增的平均投资占比在 6.5%，14.5%，23.7%；

2.75 亿以下的小盘股，300、500、1000 指增的平均投资占比在 18.9%，40.7%，61.6%；

3.300 指增大幅缩减了 2000 亿以上流通市值股票的配比，增配了部分小市值股票。相比于 500 指增，1000 指增对标指数的天然优势，使其在小市值上的可控增配幅度更高。

不管诱因是什么，但将踩踏走向极致的是拥挤的因子。很多量化机构回顾本次风暴，认为严重程度超越 2007 年的美国与 2014 年的 A 股。并反复提及 2007 年 8 月美国量化地震（Quant Quake），甚至 2019 年 9 月美国动量崩溃（Momentum Crash）以及 2020 年 3 月美国量化爆炸（Quant Bust），通过分析其原因、演变过程以及对市场的影响，帮助 A 股理解究竟什么是量化的尾部风险，进而思考如何去应对。

有机构总结美国这三次量化危机的共同点：

1.市场极端波动、模型失效：每次量化危机都伴随着市场的极端波动，这些波动打破了市场常态。量化策略通常基于统计模型和历史数据来预测市场行为。然而，这些模型可能无法捕捉到极端事件或市场结构的变化，导致在危机期间预测失准。

2.策略拥挤：在危机之前，一些风格因子，或者有一类风格因子会相对比较拥挤。这导致市场对某些交易信号过度依赖，一旦市场条件改变，这些策略可能会同时遭遇损失，引发连锁反应。

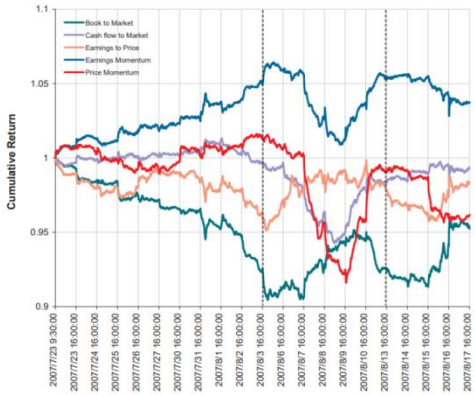
3.流动性干涸：在市场压力增大时，流动性往往会干涸。在踩踏的结果下，流动性就陷入恶性循环，价差的变化也会相应地增大，使危机自我演绎到对整个市场都有比较恶劣的影响。

4.杠杆使用：许多量化基金为了提高收益率会使用杠杆。在市场下跌时，杠杆会放大损失，对整个市场的冲击更加剧烈，迫使基金进行强制平仓，进一步加剧市场波动。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

这些共同点似乎都能够映射到 A 股在 2024 年 2 月经历的量化回撤，虽然情况跟美国些许不同，但是本质其实是一样的，在美国量化地震发生一个多月以后，MIT 的 Amir Khandani 和 Andew Lo 发表了一篇研究“What happened to the Quants in August 2007”中推演了因子回撤的路径，前期表现较好的因子在 2007 年 8 月 7 日至 9 日发生快速反转，此后回归正常。

■图表 经典因子（包括动量、价值）在股票多空头寸的模拟净值表现



数据来源：Khandani A E， Lo A W. What happened to the quants in August 2007: Evidence from factors and transactions data[R]. National Bureau of Economic Research， 2008.

研究机构 Extract Alpha 也发现，量化地震期间受损最深的因子，都是之前最赚钱、最拥挤的因子。包括 earnings yield，12 月价格动能，5 日价格反转（5-day pricere versal）等。作为对比，ExtractAlpha 也列出了当时没那么拥挤的三个因子：股价的季节性因子、Volume（看跌对看涨期权、期权对股票的量比）、Skew（OOM 看跌期权的隐含波动率）。下面是两组因子历史（2001-2007）回报，以及量化地震期间回报的比较。

■图表 美国 8 个因子表现

	拥挤因子				非拥挤因子			
	earnings yield	动量	反转	平均	季节因子	Volume	Skew	平均
2001–2007 日均收益	11.00%	14.76%	35.09%	20.28%	8.64%	3.60%	17.10%	9.78%
日收益								
7–Aug–07	–1.06%	–0.11%	–0.34%	–0.50%	–0.06%	0.33%	–0.85%	–0.19%
8–Aug–07	–7.76%	–4.19%	0.23%	–2.24%	–0.21%	–0.04%	0.21%	–0.01%
9–Aug–07	–1.66%	–3.36%	–3.41%	–2.81%	–0.29%	–1.27%	–0.23%	–0.60%
10–Aug–07	3.91%	4.09%	12.45%	6.82%	0.71%	0.01%	1.70%	0.80%

数据资料来源：Extract Alpha

上表显示，**在量化地震之前的七年时间里，平均而言拥挤因子表现更好，非拥挤因子平均年化回报率为 10%，约为拥挤因子（年化 20%）的一半。但在地震期间非拥挤因子的投资组合中应该有更多的流动性。**

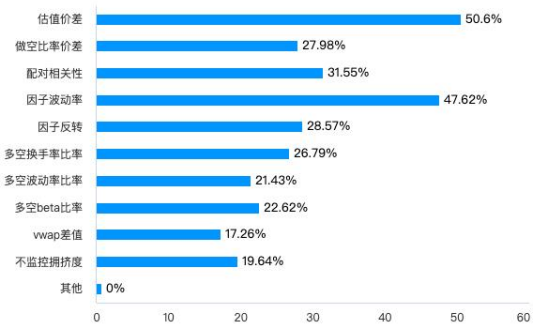
本次问卷调研也调查了国内机构目前度量因子拥挤度的不同方法。这些测算方式的背后逻辑原理通常是基于对市场行为和参与者行动的洞察，以及对金融市场结构和效率的理解。

数据显示，当前量化机构在评估策略时，首先考虑的是能够带来高收益的策略（收益率综合得分 8.88、波动率 7.34）），同时也非常重视风险控制和策略稳健性（综合得分 5.51）。仅 2 成的机构不测算因子拥挤度，但策略拥挤度（综合得分 3.6）是机构最不关注的特征之一，目前测算方式呈现多样性，例如，一些机构可能更关注估值驱动的策略，而另一些机构可能更

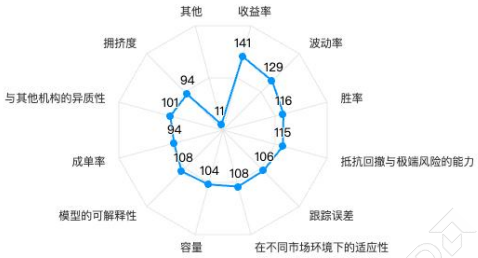
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

侧重于动量或其他类型的策略。

■图表 量化机构当前对于因子拥挤度的测算方式



■图表 量化机构最注重策略的特征[排序题]



数据资料来源：《2023 中国量化投资白皮书》问卷调研

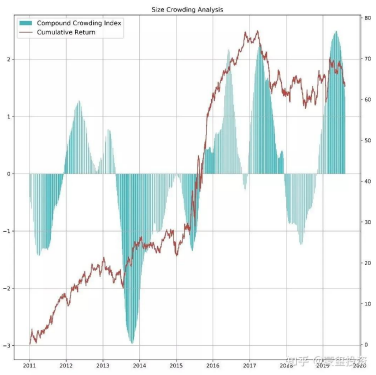
■图表 因子拥挤度测算方式及意义

方式	定义	意义
估值价差	因子多头仓位股票平均估值和因子空头仓位股票平均估值之比，估值指标为 B/P，Sales/P 以及预期 E/P	当越来越多的投资者采用相同的因子进行投资时，这些因子多头仓位的股票可能会被推高估值，而空头仓位的股票可能会被低估。估值价差的扩大可能表明市场对于这些股票的分歧加大，因子可能处于拥挤状态。
做空比率价差	因子空头仓位股票平均做空比率和因子多头仓位股票平均做空比率之差	做空比率高的股票可能表明市场上有更多看空的观点，如果空头仓位股票的做空比率显著高于多头仓位股票，这可能表明市场对于空头仓位的负面情绪加剧，因子可能过度拥挤。
配对相关性	因子多头仓位股票间收益率配对相关性和因子空头仓位股票间收益率的配对相关性之比	如果多头或空头仓位股票之间的收益率相关性增加，可能表明因子驱动的市场行为趋同化。当市场波动或逆转时，高配对相关性可能导致集体大幅度的价格调整。
因子波动率	MSCIBarra 模型预测的因子波动率和预测的市场波动率之比	因子波动率的提高可能意味着市场对于该因子的反应更加敏感，这可能是由于过度交易或拥挤引起的。高波动率可能导致因子在短期内性能不稳定。
因子反转	因子过去三年的累计收益	因子在一段时间内表现良好之后可能会遭遇反转，尤其是当市场参与者普遍认识到这一趋势并在其上过度投资时。因子反转可以作为因子过度拥挤的一个预警信号。
多空换手率比率	因子分组下，多空两组的平均个股换手率比值；换手率为个股过去 3 个月的日均成交换手	换手率较高的股票可能表明它们受到了更多的关注和交易。如果多空两组的换手率差异显著，这可能意味着市场对这些股票的关注度不均衡，可能存在拥挤风险。
多空波动率比率	因子分组下，多空两组的平均个股波动率比值；波动率为个股过去 3 个月的收益率标准差	波动率可以反映股票价格变动的不确定性。如果多头或空头仓位的股票波动率显著不同，这可能表明市场对于这些股票的预期存在分歧，因子可能面临拥挤风险。
多空 Beta 比率	因子分组下，多空两组的平均个股 Beta 比值；Beta 为过去 3 个月个股收益与万得全 A 指数收益回归的 Beta	Beta 衡量的是股票相对于整个市场的系统性风险。如果多空组合的 Beta 显著不同，这可能意味着因子策略与市场的同步性不一致，可能是因子拥挤的一个标志。
vwap 差值	自定义成交回测每日收益减去 vwap 回测每日收益，再进行 250 日求和。	使用成交量加权平均价格（vwap）进行交易的策略，如果与实际成交价格差异较大，可能意味着市场流动性或交易执行效率的问题，这可能与因子拥挤度有关。

蒙玺投资曾经对市值、盈利、长期反转三个因子做拥挤度分析，从表现结果来看，三大因子具有明显的周期性。

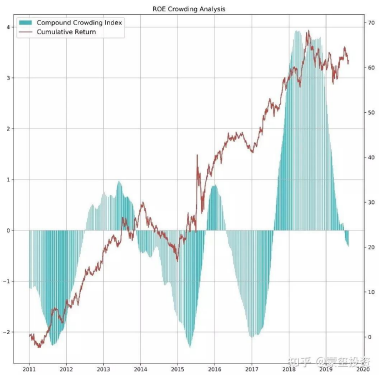
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 市值因子拥挤周期



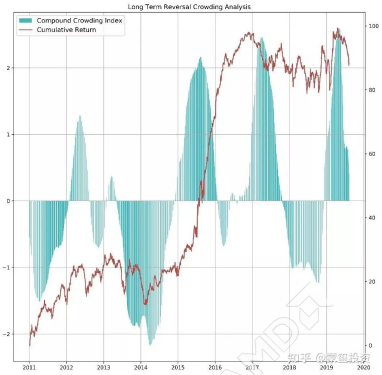
定义：按照股票的流通市值从小到大排序，做多前 10%的股票，做空后 10%的股票。

■图表 盈利因子拥挤周期



定义：按照股票的 ROE 指标从小到大排序，做空前 10%的股票，做多后 10%的股票。

■图表 长期反转因子拥挤周期



定义：按照股票过去 120 天的累计收益率从小到大排序，做多前 10%的股票，做空后 10%的股票。

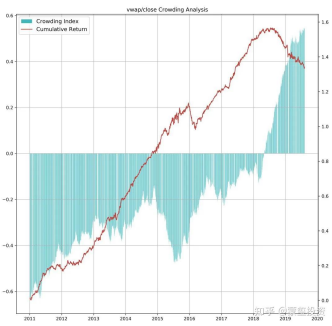
选股范围限定为中证 500 成分股中。在计算因子的拥挤度时，保留因子的空头；但是在计算因子的收益率曲线时，只保留多头部分，同时用同等仓位的 IC 做对冲。这样做的原因是因子的空头部分收益实际很难获取，计算多空收益意义不大。红色曲线代表因子累积收益情况，蓝色柱状图代表拥挤度（正值越大，代表越拥挤）。

短周期因子（高频因子）方面，蒙玺投资则用订单执行效果衡量拥挤度。有鉴于目前行业内基本上都能做到 90%以上的成交率，所以成交率的区分度实在不高。此处订单执行不只是低延迟行业传统意义上的成交率，而是指成交价格。对于隔日 Alpha 策略（当天晚上生成目标仓位，第二天开仓或者调仓），第二天的执行结果都是去和 vwap（成交量加权平均价）进行比较，所以一般机构都是希望接近这个基准或者超越它。各量化基金如果使用类似的因子，会同时买到相同的股票。大多数基金下单方式为 vwap/twap，如果使用这类因子的资金量足够大，也就是足够拥挤时，那么买的当天会缓慢推升股票价格（卖的当天会推低价格）。这时候，用一种“简单的成交方式”进行成交，效果就会比 vwap 成交价格好（这个“方式”是由蒙玺投资研究得出，暂无法公开）。所以用“简单的成交方式”回测收益要高于用 vwap 回测收益，而这个差值越高就代表拥挤度越高。

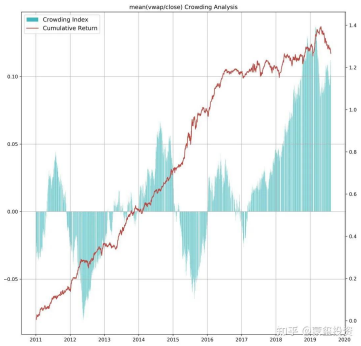
a.回测条件：在中证 500 成分股的范围内，每日选出因子值最大的 50 只股票，等资金持仓，默认成交价为当天的 vwap。回测曲线为组合相对于中证 500 指数的超额收益。

b.拥挤度指标：拥挤度指标等于用“简单的成交方式”回测每日收益减去 vwap 回测每日收益，再进行 250 日求和。这样得到的是因子使用两种成交方式导致的年化收益率的差距，衡量执行交易的拥挤度。

■图表 因子 1:vwap/close 拥挤周期



■图表 因子 2：MA(vwap/Close，20）拥挤周期



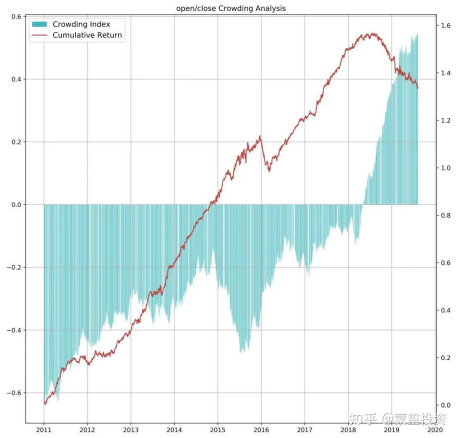
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

vwap/close 因子，反映的是一种短期反转的效应。

在 2018 年变得异常拥挤，因此使用 vwap 成交的回测曲线也发生了历史最大回测。

MA(vwap/close，20）因子，是 vwap/close 的 20 日平均，反映的是一种中短期反转的效应。2017 年之前，拥挤度指标时而正时而负值，没有一个固定的变化趋势。在 2017 年后变成明显的正值，因子的回测表现也不如之前稳定。

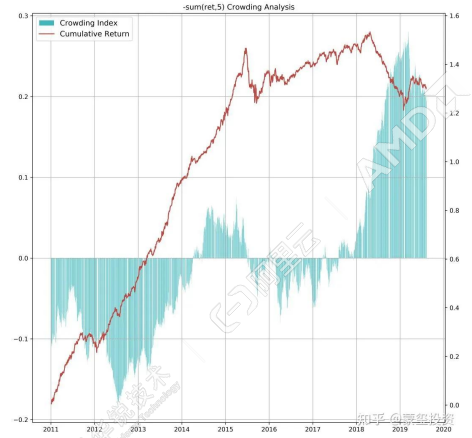
■图表 因子 3：open/close 拥挤周期



open/close 因子，反映的是一种短期反转的效应。

在 2018 年变成明显的正值，因子使用 vwap 成交的回测曲线也发生了历史最大回测。

■图表 因子 4：-5 日涨幅拥挤周期



-5 日涨幅因子，反映的是一种短期反转的效应。

2017 年之前，拥挤度指标时而正时而负值，整体是一个向上变化的趋势。

在 2017 年后变成明显的正值，因子的回测表现也不如之前稳定。

通过成交情况分析短周期（高频）因子的拥挤度，可以明显看出，自从 2017 年以来，短周期因子拥挤度显著上升，简单的短周期因子效果也明显大不如前。这很符合我们直观上的感觉。而且相比较于中长周期因子，短周期因子收益情况随着拥挤度的变化反应得更灵敏。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

CHAPTER

第四章

市场与他者

```
market_signal = lambda s, n: s + n
signal_strength = market_signal(1, 0.1*np.random.randn())
print(f"Signal amidst noise: {signal_strength:.2f}")
```



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

行为金融学表明人们在投资中存在很多误区存、非理性的抉择与行为，包括情绪波动、选择记忆、过度自信、自我确认、羊群效应、未知恐惧、赌徒心理、锚定效应等。而量化遵循可验证的方法，为投资人理解投资提供另外一种思考角度，对于引导投资人认知方面具有天然优势。

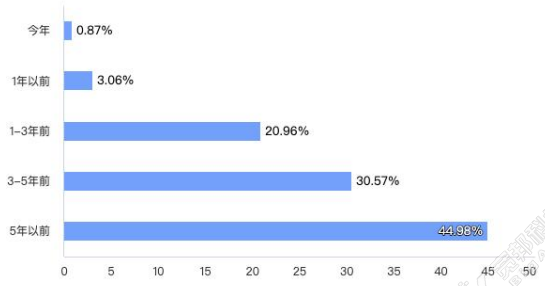
本轮危机中，市场透明和开放的危机沟通（平均分 3.79）是量化机构学到的课中排名第二，但“藏好自己，做好清理”仿佛也是打在量化机构投研思想中的钢印，这就形成了与公众沟通之间的矛盾。

本调研组同时调研回收了 224 份公众问卷样本，总体来看，受访公众对量化投资并非完全不解，他们认为量化投资能提升市场效率并增强流动性，但对可能引起的市场波动和系统性风险持有一定的担忧。同时，受访者对于量化投资的监管和未来发展保持着开放但谨慎的态度。

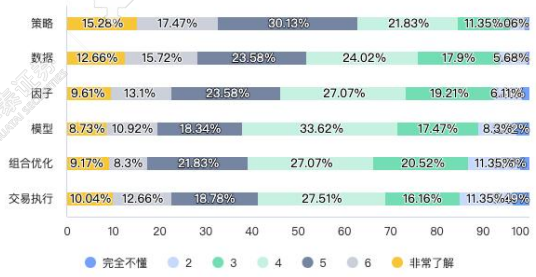
01 投资认知

第一批国内量化私募机构大多数刚好成立 10 年，到 2021 年行业迎来爆发，数据显示这也与公众的认知相符，7 成机构 3 年之前已经开始接触到量化投资这一词汇，有相对较大一部分受访者（约 44.98%）表示他们至少在 5 年前就已经了解量化投资，这表明量化投资在过去已经得到了一定程度的关注和讨论。

■图表 公众第一次听到量化投资这个词的时间



■图表 公众对于量化投资的各个流程理解程度



数据资料来源：《2023 中国量化投资白皮书》问卷调研

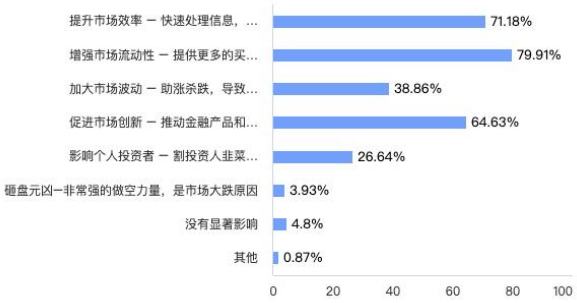
随着时间累积，公众对量化投资各个流程因其普遍性和接触频率、复杂性和专业性、实用性和实际操作等因素有不同的认知，总体而言公众已经认为自己比较懂（4.4 分/7 分）。将量化投资各个流程的理解程度进行平均分排序，如上所示：策略-平均分 4.91；数据-平均分 4.62；因子-平均分 4.43，交易执行-平均分 4.25；模型-平均分 4.24；组合优化-平均分 4.17。

策略和数据这两个词是量化投资中最基础且普遍接触的部分。几乎所有投资者都需要了解策略的概念和如何处理数据。因此，这些流程的平均分较高。但实际上具体实践的时候，这些分数就逐步降低，例如，因子和模型可能需要更高的数学和统计学知识，而组合优化则需要深入的风险管理和资产配置知识。这些领域的专业门槛较高，因此平均分较低。

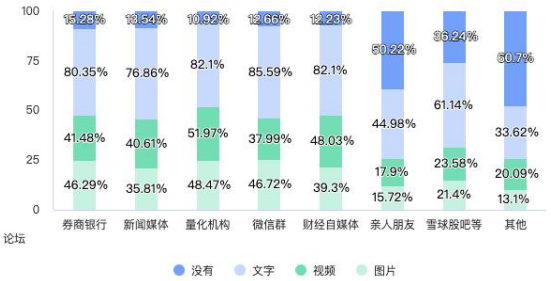
量化交易对 A 股市场影响方面，大多数公众认为量化交易对 A 股市场有积极的影响，特别是在提升市场效率（71.18%）、促进市场创新（64.63%）和增强流动性方面（79.91%）。同时，也有一定比例的公众担心量化交易可能会增加市场波动，反映了对量化交易影响的复杂性和双刃剑特性的认识。尽管存在一些担忧，但整体看法倾向于认为量化交易对市场是正面或至少不是负面的影响。只有极少数受访者（3.93%）认为量化交易对 A 股市场没有显著影响或是砸盘的原因。另有 26.64%的受访者表示量化交易影响了个人投资者，例如被称为“割韭菜”。

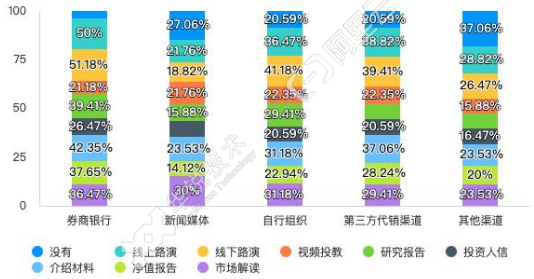
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 公众认为量化交易的发展会对 A 股市场的影响



■图表 公众了解量化的渠道





数据资料来源：《2023 中国量化投资白皮书》问卷调研

微信群、量化机构和财经自媒体是投资人了解量化投资的主要渠道，这些渠道的共同特点是便捷性和互动性，容易获取和分享信息。文字是最主要的信息传播方式，几乎在所有渠道中都有最高的认知比例，这可能是因为文字信息易于传达详细内容，且方便快速阅读和参考。视频和图片作为辅助性的信息传播方式，在某些渠道也有相对较高的比例，能够直观展示量化投资的概念和案例。雪球股吧等网络论坛虽然是传统金融知识的交流场所，但在帮助公众了解量化投资方面的作用相对较小。

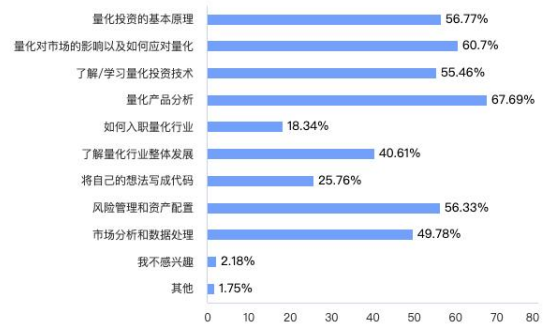
目前来看，**量化信息传播者（量化机构）和接收者（投资人）在了解量化投资方面的主要渠道和介质存在差异。**

线上线下路演 vs.文字和视频：量化机构倾向于使用线上线下路演作为沟通方式，然而，投资人在了解量化投资时，主要通过文字和视频这些可以自主控制学习节奏的方式。原因可能在于量化机构基于合规，主要面向符合资格的投资人进行募资路演，投教并不是其主要工作内容。视频投教的使用：尽管量化机构在视频投教上的使用比例不是特别高，但投资人通过视频了解量化的比例相对较高。

量化机构可能需要考虑更多利用视频和社交媒体这些投资人偏好的渠道来提升沟通效果。同时，投资人在自主获取信息时，更倾向于使用具有高度可访问性和易于理解的介质。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 公众最希望从量化投资相关的教育课程中学到的内容



■图表 不同公众最希望从量化投资相关的教育课程中学到的内容



数据资料来源：《2023 中国量化投资白皮书》问卷调研

或许由于幸存者偏差，样本数据普遍显示对各类量化相关议题表示高度的兴趣，仅有 2.18%公众表示对量化投资不感兴趣。感兴趣的人中有的是打不过就加入型，将量化作为投资类别，或将方法论化为己用。有的将量化作为对手方型，想要了解量化对市场的影响以及如何应对量化。

数据显示，公众在参与量化投资相关教育课程时，最希望学到的内容为量化产品分析（67.69%），其次就为量化对市场的影响以及如何应对量化（60.7%）。

如以细分类别来看，专业机构投资人在多数领域的兴趣略高于个人投资者，但个人投资者在量化产品分析方面的兴趣略高于专业机构投资人。

■图表 不同公众最希望从量化投资相关的教育课程中学到的内容

	量化投资的基本原理	量化对市场的影响以及如何应对量化	了解/学习量化投资技术	量化产品分析	如何入职量化行业	了解量化行业整体发展	将自己的想法写成代码	风险管理和资产配置	市场分析和数据处理
主观类专业机构	61.82%	63.64%	65.45%	65.45%	27.27%	49.09%	27.27%	61.82%	49.09%
个人投资者	50%	59.70%	50%	66.42%	16.42%	35.82%	24.63%	57.46%	49.25%
企业&上市公司	100%	66.67%	33.33%	66.67%	33.33%	50%	16.67%	66.67%	66.67%
行业卖方分析师	57.14%	42.86%	57.14%	57.14%	14.29%	42.86%	0.00%	85.71%	85.71%
FOF	62.96%	55.56%	55.56%	77.78%	14.81%	55.56%	40.74%	70.37%	66.67%
代销机构	56%	68%	62%	82%	14%	46%	24%	60%	54%

数据资料来源：《2023 中国量化投资白皮书》问卷调研

如前所述，**公众对高频与做空最容易产生误解，但样本池内数据显示，量化机构对公众也可能存在刻板化偏见**。相反，数据显示出公众对做空策略有相对较为成熟和均衡的看法，能够认识到做空在市场中的作用（占比 83.84%），认为做空是市场正常机制的一部分，有助于发现价格泡沫和过高估值。只要不是恶意操纵市场，做空策略应该被允许（占比 65.5%），虽

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

然不是绝大多数，但有相当一部分的投资人（占比 36.24%）理解到做空并非量化投资独有的策略，也被传统的主观投资策略所使用。

如第二章所述，我国股指存在明显的“大涨大跌”现象，短期内涨跌行情切换频繁，上涨的收益和下跌的风险并存，机构将这部分风险定义为 Beta 风险，为了使得量化策略收益曲线较为平滑，量化机构会采用股指期货、融券、期权、收益互换等多种对冲方式或结合的方式，持续跟踪成本差异，以求得对冲成本、对冲精度的最优解。而股指期货是最为主要的对冲方式，并非单纯地做空某只股票，而其存在也有明显的市场价值：

1.抑制单边市，完善股票市场内在稳定机制。股指期货为市场提供了做空手段和双向交易机制，增加了市场平衡制约力量，有助于降低市场波动幅度。机构投资者运用股指期货可以替代在现货市场的频繁操作，增强持股信心。同时，股指期货具有一定的远期价格发现功能可在一定程度上引导现货交易，稳定市场预期，减少股市波动频率。

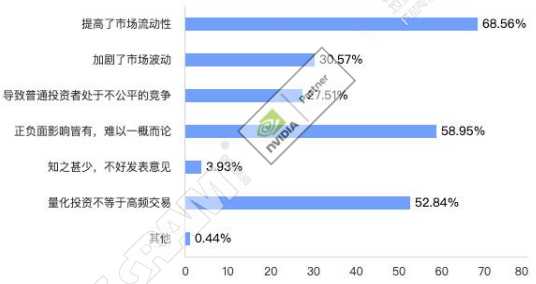
2.提供避险工具，培育市场避险文化。股指期货市场是一个专业化、高效的风险管理市场。股指期货不消除股市风险，但它使得股市风险变得可表征、可分割、可交易可转移，起到优化市场风险结构、促进股市平稳运行的作用。

3.完善金融产品体系，增加市场的广度和深度，改善股市生态。发展股指期货等简单的基础性风险管理工具，不仅能够完善金融产品体系，增加市场创新功能，提高市场运行质量，同时也有助于保障金融资源配置的主动权，实现国家金融安全战略的重要选择。

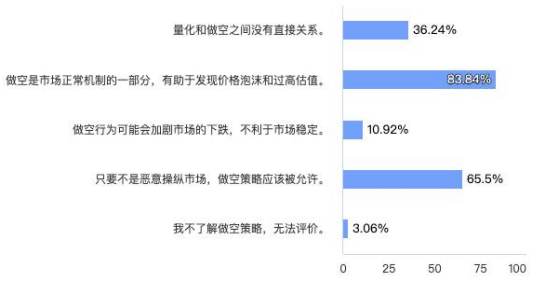
同样，公众对高频交易也有相对复杂和多样的看法，并不全然是误解。公众在认识到高频交易提高市场流动性的同时，也意识到了其可能导致的市场波动和不公平竞争等问题。此外，超过一半的受访者能够区分量化投资和高频交易，表明他们对这一领域有较为准确地理解。

大多数受访者（68.56%）认为高频交易提高了市场流动性。这是对高频交易最积极的评价，表明公众认识到高频交易在提供市场买卖挂单方面的作用。超半数的受访者（58.95%）认为高频交易的影响是双面的，难以简单定义为好或坏。反而较少的受访者（30.57%）认为高频交易加剧了市场波动。或者认为高频导致普通投资者处于不公平的竞争（27.51%）。

■图表 公众如何看待高频交易



■图表 如何看待做空策略



数据资料来源：《2023 中国量化投资白皮书》问卷调研

略微矛盾的是：一方面，有 52.84%的受访者能够明确区分量化投资和高频交易，表明他们理解这两者是不同的概念。而另一方面，“高频”作为公众听到“量化”二字时第二常想到的词汇，表明在公众的认知中，量化投资和高频交易之间仍然存在一定程度的联系。

这种看似矛盾的情况可能由以下几个因素导致：**1.量化投资中包含高频交易策略：量化投资是一个宽泛的概念，它包括了多种不同的投资策略，其中包括高频交易。因此，一些人在听到“量化”二字时会想到其中的一个子集“高频交易”。**

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

2.媒体和公众讨论：媒体报道往往会聚焦在某些特定的话题上，而高频交易因其争议性和复杂性，可能在公众讨论中占据了较多的篇幅，导致公众更容易在听到“量化”时联想到“高频”。

3.公众对量化投资概念的简化理解：在尝试理解一个复杂的概念时，人们往往会寻找最容易抓住的特征作为记忆点，而“高频交易”可能因为其特点明显而成了一个突出的记忆点。

因此，这两个数据点实际上揭示了公众对量化投资认知的复杂性。这也指出了对公众进行金融教育和普及工作的重要性，特别是在阐释量化投资的多样性和细微差别方面。

编写组邀请公众写出听到“量化”二字时，脑海中首先浮现的三个名词，最为高频的词汇分别为：因子（48）、高频（38）、模型（38）、算法（22）、数据（19）、机器学习（16）、程序化（16）、策略（14）、数学（12）、超额（11）、对冲（10）、高频交易（9）、计算机（9）、程序化交易（9）、技术（8）、机器（8）、算力（7）、程序（7）、私募（7）、分散（6）。

词汇中蕴含大量名词和动词，基本形容量化投资中的某些特征或者属性，较为中性，没有明显的情感色彩，但也有少量的词汇，包含负面情绪或对量化投资持批评态度的词汇，如黑箱、做空、割韭菜、爆雷、胡搞、吹牛逼、客观、量价、风险、失效、波动、搅屎棍等。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

02 投资感受

公众在听到“量化”二字时，脑海中首先浮现的情绪可以分为正面、中性和负面情绪。

其中正面情绪可细分为：兴奋：40.17%、安心：10.92%、骄傲：4.37%、感激：4.37%、希望：24.89%、信心：28.38%、平静：38.86%。

中性情绪可细分为：无感：9.17%、平静：38.86%（平静可以视为中性情绪，但在此上下文中，也可以被认为是积极的）

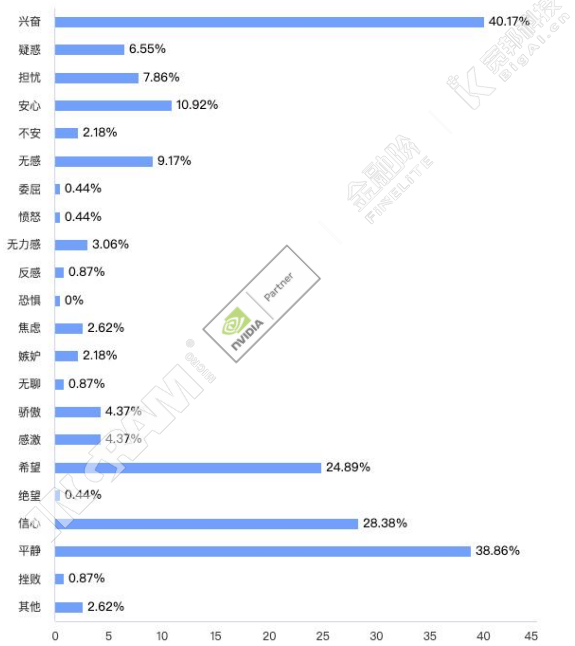
负面情绪可分为：疑惑：6.55%、担忧：7.86%、不安：2.18%、委屈：0.44%、愤怒：0.44%、无力感：3.06%、反感：0.87%、焦虑：2.62%、嫉妒：2.18%、无聊：0.87%、绝望：0.44%、挫败：0.87%。

总体占比显示：正面情绪的总占比较高，显示出公众对“量化”这个概念有较多积极的联想。如果将平静情绪也算作积极，则积极情绪的总占比达到 152.06%（因为多选题的总比例可能超过 100%）。中性情绪：无感和平静（如果单独考虑）的比例加起来是 47.03%，表明有一部分公众对量化持中立态度。负面情绪：负面情绪的总占比相对较低，总和为 27.78%，显示出负面联想相对较少。

综合来看，受访公众对量化投资持有积极或至少不是消极的态度。

针对特定量化相关名词的好恶程度。题目采用了 1-7 的量表，其中 1 表示“非常讨厌”，7 表示“非常喜欢”，4 表示“无感”。受访问卷总体名词评分为 5.15 分，尤其对于人工智能量化、策略和量化培训等概念表现出较高好感。而对于杠杆、做空和融券等可能存在较高风险和争议的概念，公众的好感度相对较低，但仍然没有表现出整体的负面态度。

■图表 当听说“量化”二字时，公众首先浮现的情绪



■图表 公众对特定名词的好恶程度

较高好感度	人工智能量化	5.37
	多空策略	5.38
	量化培训	5.43
	策略	5.56
	量化投资	5.56
中等好感度	高频交易	5.06
	股指期货	5.09
	另类数据	5.17
	因子	5.34
	算法交易	5.48
	量化基金	5.37
较低好感度	杠杆	4.28
	做空	4.54
	融券	4.43

数据资料来源：《2023 中国量化投资白皮书》问卷调研

量化投资机构认为公众对行业的误解**主要来自媒体报道的简化和曲解，以及行业内负面新闻的放大。此外，信息不对称和量化复杂性也导致了公众难以充分理解行业。**

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

其中媒体报道（平均分 4.16）：媒体报道是被量化从业认为最重要的误解来源。媒体在报道量化行业时，可能会出于简化信息的需要，忽略一些细节或者不准确地传达信息，导致公众片面的理解。

负面新闻（平均分 4.01）：包括个别机构的不当行为或失败案例，通常会受到公众和媒体的关注，这可能会对整个行业的形象产生不利影响。

信息不对称（平均分 3.93）：量化机构可能没有充分地与公众沟通，导致了信息不对称的问题。这表明机构需要更积极主动地与投资者和公众沟通，提高透明度。

然而，实际上，4 成的受访者对量化机构没有负面情绪，剩下的有情绪的公众中，大家归因的主要方向是，**信息不对称（31.88%）、量化复杂性（27.51%）、媒体报道（28.82%）**，公众更加关注信息不对称和量化复杂性的问题，这表明量化机构在信息披露和教育投资者方面可能需要做得更多。

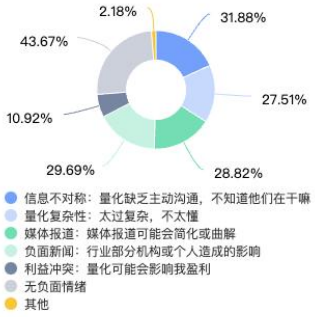
同时公众认为媒体的报道可能会简化或曲解量化行业的事实，这与量化机构自身的认知相符，都认为媒体报道是误解的来源之一。

为缓解公众的负面情绪，量化机构需要加强与公众的沟通，提高信息的透明度，简化解释量化投资的复杂性，并积极正面地影响媒体报道。同时，机构也需要从负面新闻中学习，提升整个行业的品质和形象，减少对公众信心的负面影响。通过这些努力，量化投资机构可以帮助公众更好地理解行业，从而改善公众对量化投资的整体看法。

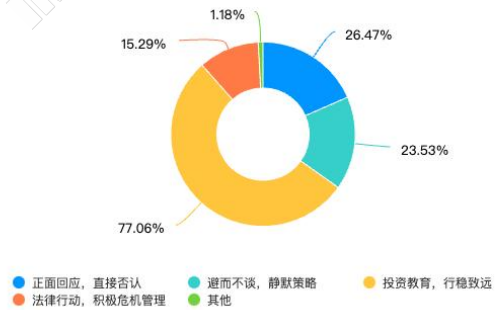
编写组邀请公众写出其印象最为深刻的三个量化的新闻/事件：负面新闻/事件占据了较大比重，绝大多数事件反映了公众对量化投资的担忧和负面看法，特别是与市场波动、砸盘、做空、回撤和监管处罚相关的内容。包括“微盘股危机”，“限制做空”，“光大乌龙指”、“灵均被处罚”、“明汯别墅”、“幻方瓜”等。负面事件往往会引起更强烈的公众反应和更广泛的媒体报道。

■图表 您认为公众对量化的误解主要原因是什么?





■图表 面对公众及市场对量化的误解，贵机构采取过的态度是



■图表 公众印象最为深刻的三个量化的新闻/事件#



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 公众印象最为深刻的三个量化的新闻/事件

单词	词频	单词	词频	单词	词频	单词	词频
量化	147	处罚	25	私募	19	市场	13
灵均	93	事件	24	春节	16	年初	13
交易	62	融券	22	别墅	15	基金	12
限制	54	幻方	20	策略	15	投资	12
监管	50	高频	20	大跌	14	集体	11
DMA	69	超额	20	老板	14	崩盘	11
回撤	36	交易所	19	机构	14	雪球	11

数据资料来源：《2023 中国量化投资白皮书》问卷调研

量化行业看上去是技术驱动的行业，但无法离开健康的行业生态，在利他的过程中实现了自我的价值。技术无法解决所有问题，如果投资者没有长期主义的坚持和获得感，公众对于行业没有持续的支持而是一味反对，行业就无法发展。

面对公众的质疑，7 成机构选择了投资教育作为应对公众误解的主要策略。希望通过教育和传播正确的信息，帮助公众更好地理解量化投资，这是一种合理的长期解决方案，有助于建立公众的信任和理解。但从过去 10 年的多次舆论危机来看，这样的传播与教育效果并未如预期，公众教育甚至成为机构名为保持“低调”，实为并不沟通的安慰剂，本质上是一个静默策略，离“投资者不是用来教育的，是用来服务的”更是遥远。

例如实际执行过程中，仅有 5 成机构参与过券商银行组织的线上线下路演，自行组织视频投教仅为 22.35%，部分机构甚至未与任何渠道做过任何沟通。投资者教育活动可能未达到预期的覆盖面或深度，或者未能针对公众的真实需要和疑虑；实为表面工作，并非真正的沟通和传导；在公共场合宣称重视公众教育和透明度，但实际上没有采取相应的行动，都可能会导致公众对机构的不信任。

■图表 量化机构与市场主动沟通渠道及形式

选项	没有	线上路演	线下路演	视频投教	研究报告	投资人信	介绍材料	净值报告	市场解读
券商银行	10.59%	50%	51.18%	21.18%	39.41%	26.47%	42.35%	37.65%	36.47%
新闻媒体	27.06%	21.76%	18.82%	21.76%	15.88%	15.88%	23.53%	14.12%	30%
自行组织	20.59%	36.47%	41.18%	22.35%	29.41%	20.59%	31.18%	22.94%	31.18%
第三方代销渠道	20.59%	38.82%	39.41%	22.35%	20.59%	20.59%	37.06%	28.24%	29.41%
其他渠道	37.06%	28.82%	26.47%	15.88%	16.47%	16.47%	23.53%	20%	23.53%

数据资料来源：《2023 中国量化投资白皮书》问卷调研

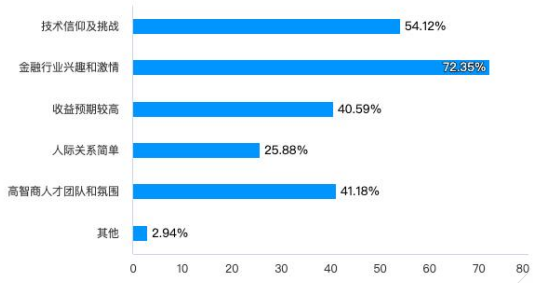
公众及市场对量化的误解可能会导致一系列负面后果，这些后果不仅影响量化投资机构本身，还可能影响整个金融市场的健康发展。形成严重的连锁反应，行业共识公众误解带来最大的是：监管压力上升（85.88%）、投资者信心下降（78.82%）甚至赎回撤资（64.12%）。从金融阶委托招聘的情况看，2023 年不管是百亿还是非百亿量化私募，一些量化私募都新增了品牌公关的招聘需求。随着行业发展到新阶段，量化机构也开始重视企业的品牌建设，规模化的管理和品牌效应也在逐渐形成。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 公众及市场对量化的误解，可能形成的后果：



■图表 从业者选择以量化作为职业的主要原因是：



数据资料来源：《2023 中国量化投资白皮书》问卷调研

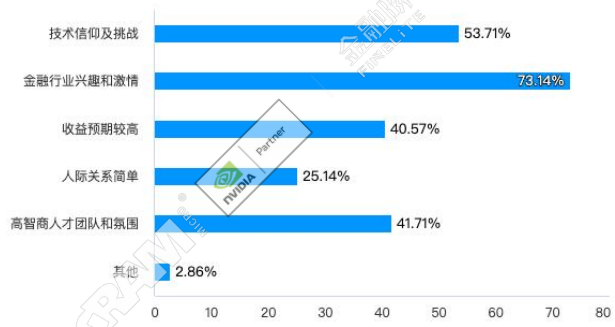
民意不是静态的，会随着事件和市场状况的变化而变化。当市场出现剧烈变化时，可能会加剧公众对量化的矛盾，在金融市场中，当公众对特定市场行为或事件有强烈看法时，监管机构可能会感受到压力，需要响应这些关切。在民意和专业的市场分析之间找到平衡。如果公众对量化概念或市场机制的理解不足，可能会被误导，导致对监管政策的不恰当要求。

而且如果公众误解量化投资，可能会导致投资者对量化策略的信心下降，影响募资和公司扩张的能力。甚至因为误解而选择撤资，这会直接影响量化机构的资金流和运营稳定性。

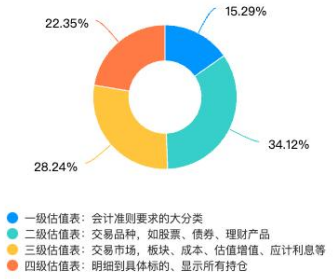
由于担心误解或负面舆论，量化团队在策略创新时可能变得更加谨慎，这可能限制了行业的创新和发展。例如 2023 年火热且极富争议的 DMA 策略及多空策略，几乎是火热元年即成灭亡之年。

公众误解导致的人才的流失和从业者的信念的担忧虽然占比不高，但实际上占比也超过 3 成，当前普通人选择量化投资职业主要是基于 1.金融行业的兴趣和激情（占比 72.35%）；2.对技术的信仰（54.12%）；3.对高智商团队的向往（41.18%），收益预期（40.57%）反而占比在第四位。但如果量化投资被普遍误解，则可能会降低职业吸引力，因为人们通常希望从事受尊敬和理解的工作。

■图表 从业者选择以量化作为职业的主要原因是



■图表 公司目前投后向市场及投资人提供的估值表颗粒度为



数据资料来源：《2023 中国量化投资白皮书》问卷调研

2023 年的投教特殊之处还在于，2023 年 11 月 14 日，一则“百亿 FOF 私募华软新动力踩雷”的传闻引起轩然大波，证监会表示，初步判断，相关人员控制杭州瑜瑶、深圳汇盛等多家机构，多层嵌套投资、虚假宣传、报送虚假信息、违规信披等情形，此事件引发了全市场对 FOF 及量化私募的信任危机，余波持续荡漾，自此事件 FOF 对私募底层估值表从原来的二级很多要求穿透至四级，除了各类标的的建仓成本和损益，还包括股票明细，并且要求托管方盖章，许多资方与管理人争执不休。

也有机构主动上交，如磐松资产就直接在微信公众号公示披露政策：1.常规四级估值表，支持由机构客服递延一个月发送；2.数据脱敏版四级估值表，可由支持此功能的托管平台生成后即时发送。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

仍有不小比例的机构对此保持审慎。主要在于商业机密或市场波动，存在暴露持仓策略被反向的可能性，或者向市场提供了其他增量信息。即便当前各量化机构策略较为复杂，但机构不愿因此暴露风险，所以当前最主流的估值表仍然为二三级，即披露交易品种，例如股票、债券、理财产品（占比 34.12%），或者进一步细化到交易市场、板块、成本、估值增值、应计利息等信息（占比 28.24%）。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

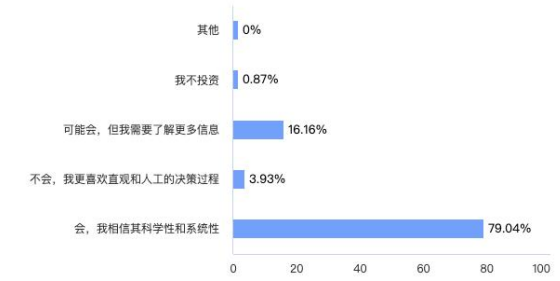
03 投资接纳

投资认知深浅、个人好恶会影响投资结果，但其投研决策并不绝对以此为标准。

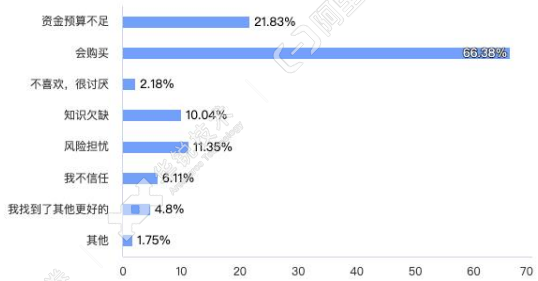
数据显示：公众对于将量化投资策略纳入个人投资组合持积极态度。多数受访者（79.04%）表达了对量化投资的信任，愿意考虑将其作为投资组合的一部分，同时有三分之二的受访者表示会购买。部分受访者（16.16%）虽然对量化投资感兴趣，但他们需要更多的信息来作出决定，了解量化投资的具体内容和潜在风险。只有极少数（0.87%）受访者表示不会考虑使用量化投资策略。

三分之一的受访者不会购买主要在于：资金预算不足（21.83%）、知识欠缺（10.04%）以及风险担忧（11.35%），极少受访者直接表示不信任量化基金（6.11%）或不喜欢（2.18）。

■图表 公众是否会考虑将量化投资策略纳入个人投资组合



■图表 公众对购买量化基金等产品的态度是



数据资料来源：《2023 中国量化投资白皮书》问卷调研

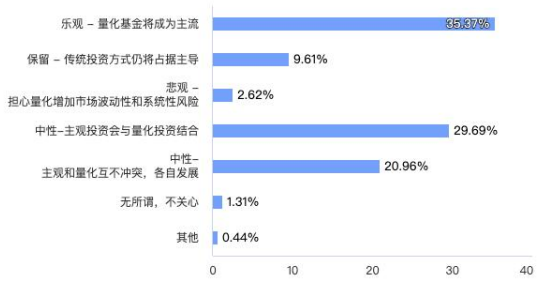
对于政策走向，公众对量化投资的监管持有多元化的看法。他们支持投资教育（58.95%）、提高交易透明度（58.52%）和防范系统性风险（55.02%）这些措施，同时希望监管不要过度（48.91%），以免抑制市场的创新和活力。受访者较少支持对做空（9.17%）和高频交易（14.85%）的严格限制，反而限制杠杆（25.33%）被更多认同。

对于整个行业走向，公众对于量化投资的未来发展保持相对乐观，很多人（35.37%）相信量化投资会随着技术的进步而变得越来越主流。同时，也有一部分人认为，未来投资领域将会见证主观投资和量化投资的结合或并行发展。保留和悲观的观点较少，表明大多数人对量化投资的未来持积极或至少不是消极的看法。这些态度可能会影响个人投资决策、市场参与者的策略选择以及对于量化投资领域的整体期待。

■图表 公众对于量化投资的监管的态度



■图表 公众对于量化的未来发展趋势



数据资料来源：《2023 中国量化投资白皮书》问卷调研

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

在涉及填空题，公众对量化的个人看法补充时，反馈中性偏正向，不乏洞见之言，言辞中肯，以下亦作全文刊录：

■图表 公众对量化的个人看法

投资人 1	认同，支持	投资人 22	监管应创造公平交易环境
投资人 2	势在必行的投资工具	投资人 23	超额会衰减，管理人间分化严重
投资人 3	应加强对量化的监管、引导，使其本身的发展能够推动资本市场发展，而非阻碍、危害市场发展	投资人 24	希望量化的词不要被滥用，量化不代表稳定，也不代表洪水猛兽，就是一种交易方式，与主观相比没有谁比谁更优。但感觉量化机构的收益没有很稳定，没有看到哪家一直很稳定地保持相对较高收益。
投资人 4	祝好	投资人 25	建立适度公平规则
投资人 5	1.监管的政策和规则约束是前提 2.量化私密公司的靠谱度。3.让更多人认识量化，但难点是能投资私募的高净很少，所以舆论的现状是：完全对量化接触不到或者不懂的人来点评量化。4.私募公司路演中减少包装和吹嘘，故意制造神秘感，多宣传点对投资者有用的，通俗易懂的。5.量化提供的流动性是“瞬时流动性”，要辩证看	投资人 26	搞因子确实有水平，但高频有点不道德了，尤其是主要靠塞单割韭菜的，这种方法放美国，分分钟被监管了。希望国内的高频，要提高自己的专业能力，向美国看齐，不要本身具备信息优势和高速条件，一味地用步枪暴打散户原始人，一边割韭菜，一边道貌俨然地吹自己是脊梁，真的很过分。如果真的精进自己，即使在美国那样严监管也能如此用高频盈利，那我真心佩服，我想那时候如果高频被诋毁，也能更有底气地反驳吧
投资人 6	只要公平合法，量化应该提倡	投资人 27	看好量化的发展，但是门槛比较高，自己也想学
投资人 7	最应该做好的是投教工作，向市场上的散户普及相关知识，而不是听风就是雨，无端散播恐慌情绪。同时如何让他们避免成为噪声交易者是一件重要的事情。	投资人 28	目前市场长期 3000 点不是某种交易方式造成的，是另外一个深层次原因让上市企业不去发展自身而是想着上市即套现造成的。
投资人 8	应该有所限制	投资人 29	长期看很好
投资人 9	量化是工具手段，取决于如何使用	投资人 30	希望更多管理人更加关注自己的策略，而不是营销大于策略。不要出现劣币驱逐良币的现象
投资人 10	量化和主观都是一种工具和方法论，本身没有对错	投资人 31	万物皆可量化
投资人 11	曾在量化私募作为量化研究员任职，有幸参加了 2022 北京的量化投资白皮书会议，个人认为量化的管理，特别是对小量化公司来说比较重要	投资人 32	算力不是来替代经验和敏感度的
投资人 12	国内量化策略应该回归模型构建和挖掘有效因子本身不应该为了排名或管理费等过度依赖某类因子或过度暴露某些风格始终追求pureα才是量化投资发展的根基	投资人 33	仍然有较大的发展空间，但未来优胜劣汰，常见策略会继续头部化集中
投资人 13	作为量化新人，坚信这个行业未来的蓬勃发展！	投资人 34	降频，规范做空
投资人 14	长期看会是趋势	投资人 35	如果更好地让投资者了解。
投资人 15	前景广阔，有望渗透更广	投资人 36	工具无对错，需要加强监管和信息披露
投资人 16	应该积极引导量化投资	投资人 37	加强风险敞口管理，避免出现量化抱团现象
投资人 17	应该更广泛普及让更多人学习了解	投资人 38	希望能在合理监管下，对投资人信息充分披露
投资人 18	因为有涨停板制度应该限制量化的交易速率和频率等优先性	投资人 39	量化基于数据和算法，会大力发展，对个人投资有重大帮助
投资人 19	fof 视角如何配置、优选量化基金	投资人 40	需要个人学习
投资人 20	需要加强监管	投资人 41	未来投资趋势
投资人 21	量化无错，券商应给予整个市场公平性；融券无错，到时只有大户可以融券，是不是有问题。		

数据资料来源：《2023 中国量化投资白皮书》问卷调研

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

CHAPTER

第五章

量化星辰大海

```
quant_universe = np.random.uniform(low=-1, high=1, size=(1_000_000, 2))
print(f"Quantum ocean of {quant_universe.s  } stars")
```



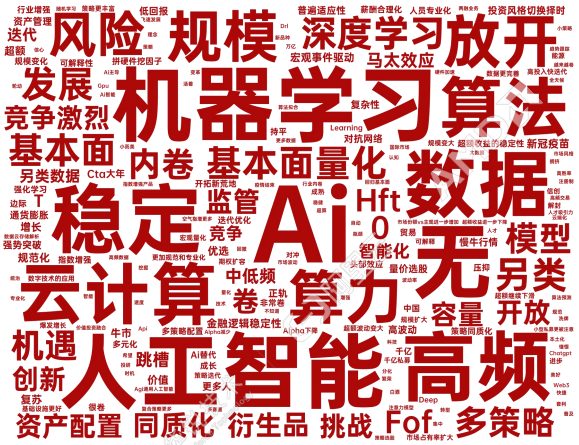
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

展望未来 3 年行业关键词，人工智能、高频数据、另类数据连续三年为行业三大关键词。另外机构提及最多的有技术创新、投资策略、金融产品与市场三个方面，其中技术与创新主要还是聚焦人工智能与机器学习包括大模型，另外还包括、算法交易服务、区块链硬件优化等。投资策略方面最主要的是主动投资、包括高股息主观加量化策略、基本面结合、主观择时、宏观研究、政策研究等。

■图表 2023 年量化机构展望未来 3 年行业关键词



■图表 2022 年量化机构展望未来 3 年行业关键词



数据资料来源：《2022-2023 中国量化投资白皮书》问卷调研

01 人工智能：大模型的想象与工具的应用现状

若因子与模型的经济学意义与人工智能代表两个相反方向，可以大致对量化机构划分为以下三个派别：

传统经济学派：这类机构当前更倾向于采用基于经济学意义的因子和模型，对于人工智能的看法比较审慎，数据显示自评 1 到 3 的机构占比 44.12%，有一小部分机构（8.82%）表示没有使用机器学习或人工智能技术。预计未来三年，这个比例会下降到 20.59%。

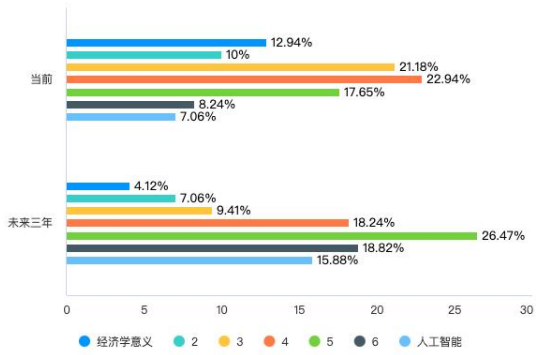
平衡派：这类机构试图在经济学原理和人工智能技术之间找到平衡，结合两者的优势来进行投资决策。当前时点，评分 4 的机构占比 22.94%，而在未来三年，这一比例预计减少到 18.24%。

人工智能派：这类机构行动可能更为积极，占比达到 33.53%，但预计未来三年，这个比例会显著增加到 60.59%。

尽管目前仍有不少机构坚持使用经典的金融理论和方法，但整体趋势表明未来量化投资领域将更多地采用人工智能，以处理大量数据和识别复杂模式的能力，需要注意的是，这种划分并不是固定不变的。随着市场环境、技术发展以及监管政策的变化，机构的策略和倾向可能会发生相应的调整。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 若因子与模型的经济学意义与人工智能代表两个相反方向，量化机构当前所处位置分布



■图表 量化机构人工智能派别划分占比

派别	定义	当前时点占比	未来三年占比
传统经济学派	1–3 分	44.12%	20.59%
平衡派	4 分	22.94%	18.24%
人工智能派	6–7 分	33.53%	60.59%

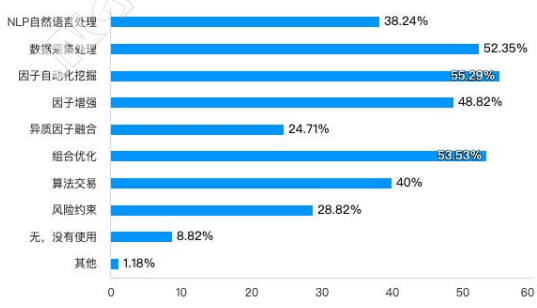
数据资料来源：《2023 中国量化投资白皮书》问卷调研

人工智能主要有四个任务：预测、决策、探索和生成。在具体实施阶段，量化机构在机器学习和人工智能的应用方面已经非常广泛，涉及从数据处理到交易执行的各个环节。**因子自动化挖掘和组合优化是应用最广泛的领域，**自动化挖掘（55.29%）可以帮助机构发现未被充分利用的市场机会，而组合优化（53.53%）是管理投资组合和提高收益的关键，这些领域的应用能够显著提高投资效率和决策质量。

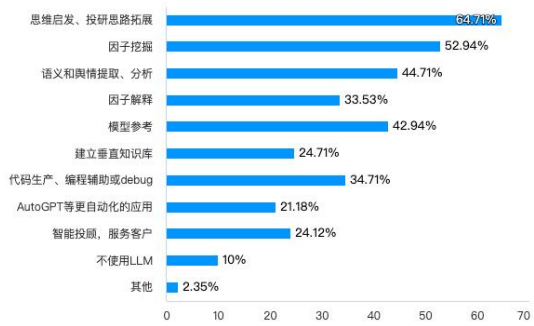
在人工智能领域，2023 被讨论较多的包括大模型、端到端、分域建模。端到端方面，以往常用做法是人脑构建因子，再使用因子合成的模型，到 2022 年和 2023 年，较多的做法是将因子挖掘与因子合成放在同一个网络里进行。分域建模方面，以往训练模型是用一个模型学习所有规律，但因为市场极其复杂，所以有的机构在探索训练多个模型去学习不同的规律。但如果单模型能够训练出来且算力足够，而多模型建模就是一件水到渠成的事情。

大模型是市场最关注的，2023 年年初市场对其量化应用审慎、怀疑、乐观、中性态度不一。从年初到年尾，量化机构从“大忽悠”“大机会”两种极端到经过探索之后逐步收敛至理性区间。因为量化是一个低容错率、低信噪比环境下对于数值类型数据的预测问题，而 GPT 是一个容错率和信噪比都较高的文字类数据的生成问题，两个场景迥然不同。但至年尾其应用也已非常广泛，从增强投研团队的创新能力到提升客户服务水平均有涉及。实际证明，大模型在因子构建、逻辑解释、代码生成等方面均有惊喜，展现出了赋能量化研究的潜力，**编写组主要将其应用分为创造 Alpha、提升效率以及投资者关系三个方面：**

■图表 当前贵机构的机器学习/人工智能量化应用领域



■图表 大模型在贵机构业务中的主要应用领域



数据资料来源：《2023 中国量化投资白皮书》问卷调研

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

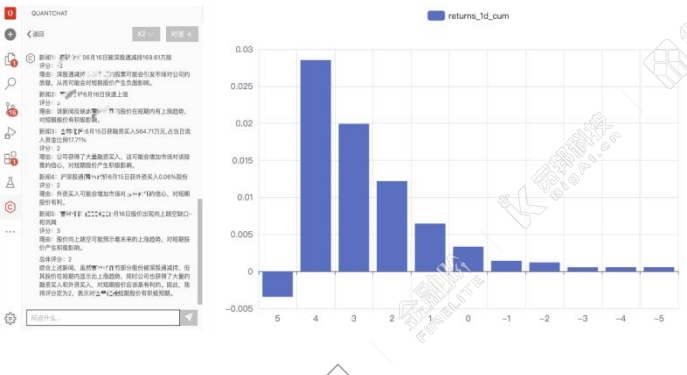
从创造 Alpha 角度来看：其中最明显的作用在于思维启发、投研思路拓展（64.71%），占比最高的应用领域。在寻找新的投资策略和研究方法时，使用大模型来生成新的思路和视角。

落足到具体技术实践层面看，包括因子挖掘、语义和舆情提取、分析、模型参考等也有接近 4-5 成的机构在已经进行实践，量化机构可以利用自然语言处理（NLP）大模型来提取和分析新闻、社交媒体和其他文本数据，以便洞察市场情绪和预测市场动向，实现大量非结构化数据的采集和分析，拓展量化研究的数据源。另外因子解释（33.53%）则帮助投研人员更好地理解模型输出和投资决策的背后逻辑、经济或市场含义。

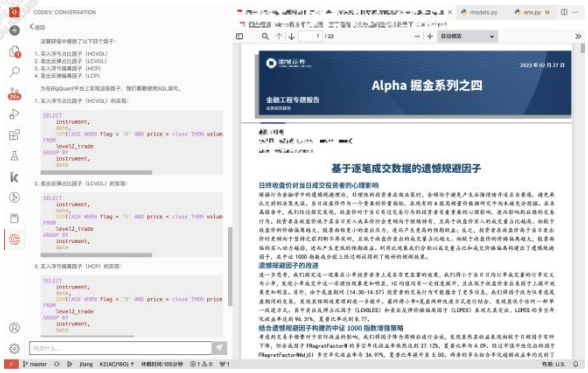
■图表 大模型用于因子生成



■图表 大模型用于舆情解读及股价趋势预测



■图表 大模型用于研报解读及因子摘取



■图表 大模型用于因子解释



■图表 大模型用于因子生成



数据资料来源：宽邦科技，BigQuant

概率投资曾分享过，关注 ChatGPT 本身的实现逻辑和模型结构，**预训练语言模型+奖励模型训练+强化学习模型微调**的三个核

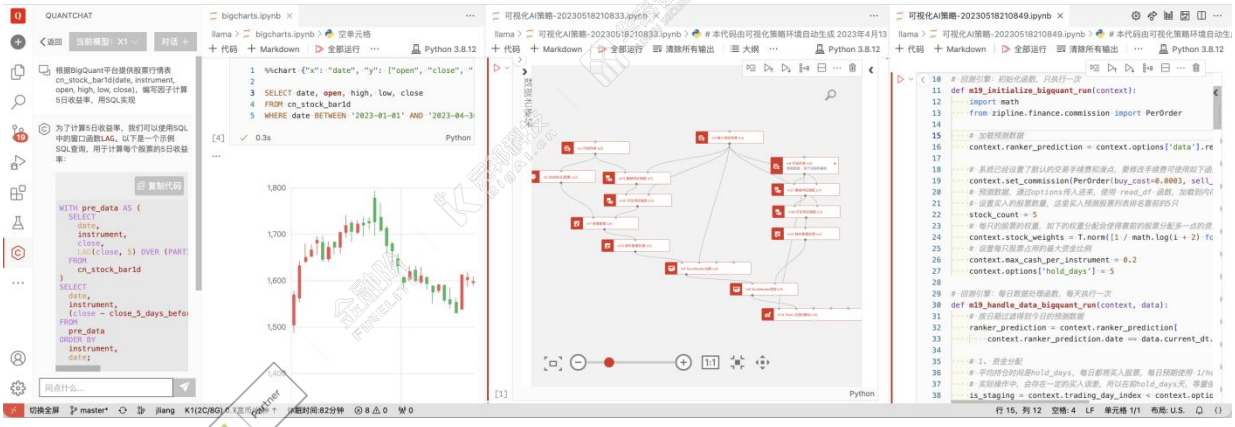
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

心步骤结构是量化可以借鉴的。特别是奖励模型训练中的“人类反馈的强化学习”，其基本目标是构建一个模型或系统，能够通过接受一个文本序列，返回一个反映人类偏好的标量奖励，就像是“知识注入”一样。可以类比到量化投资中的强投资逻辑，或许可以解决现在复杂模型在量化投资领域缺乏解释性、过拟合的硬伤。

■图表 大模型在投研助手中的应用



■图表 大模型应用于量化策略生成各环节



数据资料来源：宽邦科技，BigQuant

华泰研究认为，生成模型可以应用于多种金融数据，包括时间序列、K 线图、截面、订单簿等。使用 AIGC 模型生成订单簿数据，本质上是以上帝视角再造一个虚拟市场，并进一步使用订单簿数据训练强化学习模型，这种互动有利于模型的训练，从而提高投资实践的效果。生成模型在金融投资领域的应用还有很多，例如使用 GAN 模型训练模型，或者使用条件生成对抗网络技术，基于历史上数据对未来进行模拟，以预测收益和风险。因此，生成模型的研究仍然有很大的发展空间，特别是在金融投资领域。虽然扩散模型已经取代了 GAN 在学术界的地位，但是在金融投资领域，使用扩散模型进行投资的研究还相对较少，这也是一个大有可为的研究方向。

从提升效率来看：代码生产、编程辅助或 debug 占比 34.71%，以帮助量化研究人员编写更高质量代码方面的应用。也有 21.18%机构在 AutoGPT 等更自动化的应用，探索如何将自动化技术集成到其业务流程中，例如使用 AutoML（自动机器学习）工具来简化模型开发和部署。

在投资者关系辅助方面：有占比 24.12%的机构探索为客户提供个性化投资建议和资产管理服务时应用大模型。这其实与

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

海外的使用略有差异，根据 Invesco《2023 年全球量化投资研究》，北美把大模型归类到自然语言处理（NLP）处理，聚焦其语言任务方面的潜力。应用于一系列内部流程，最主要的就是投资者关系辅助，包括：1.以简洁的格式总结和消化文字；2.快速解决客户自发的问题；3.将建议转换为销售团队和客户可理解的语言；4.调整不同客户群体的沟通语气。

ChatGPT 从模型层面上并没有划时代的进步，其良好的性能得益于其训练，但每次训练的成本高达上千万美元。因此，极少有机构真正能够训练模型。2023 年 4 月 14 日，幻方量化发表微信公众号，宣告公司进入新征程，将集中资源和力量，投身人工智能技术，成立新的独立研究组织，探索 AGI（通用人工智能）。

后幻方公司推出 MoE 模型，并开源了技术报告和模型权重，其中 DeepSeek-V2 模型引入了 MLA 架构和自研 Sparse 结构 DeepSeekMoE，以 236B 总参数、21B 激活，实际部署在 8 卡 H800 机器上，输入吞吐量超过每秒 10 万 tokens 输出超过每秒 5 万 tokens。

从技术路径来看，DeepSeek-V2 基于高效且轻量级的框架 HAI-LLM 进行训练，采用 16-way zero-bubble pipeline 并行、8-way 专家并行和 ZeRO-1 数据并行。鉴于 DeepSeek-V2 的激活参数相对较少，并且重新计算部分算子以节省激活内存，无需张量并行即可训练，因此 DeepSeek-V2 减少了通信开销并且，DeepSeek-V2 完全开源，可免费上用，开源模型支持 128K 上下文，对话官网/API 支持 32K 上下文（约 24000 个 token），还兼容 OpenAI API 接口。

幻方大模型基于 MoE 大模型（即 Mixture of Experts），该模型是一种基于分治思想的深度学习模型。将整个模型划分为多个子模型（或称专家），每个子模型负责处理一部分任务。在推理时，根据输入数据的特性，选择最合适的子模型进行处理。这种架构使得 MoE 大模型能够同时兼顾模型的复杂性和效率，实现高性能和高效率的统一。

目前 DeepSeek-V2 API 定价为：每百万 token 输入 1 元、输出 2 元（32K 上下文），据该机构计算，在其算力服务力利用率最高的情况下，DeepSeek 每台服务器每小时收益可达 35.4 美元，毛利率在 70%以上。

公募基金方面，一直被市场认为人工智能的使用更晚一些，但从招商证券统计 2023 年的表现来看（*《公募量化基金 2023 年度盘点：激流勇进，再接再厉》招商证券*），以量价为主结合机器学习框架的公募基金产品年度表现相对突出。2023 年在业绩排名靠前的主动量化和指增产品中，或多或少都会涉及类似策略的应用。在投资决策描述中说明引入机器学习策略的基金也越来越多。当前，在因子挖掘和因子组合两个阶段，线性模型已经开始逐渐被非线性模型/机器学习模型所取代。例如，以遗传算法为代表的自动因子挖掘系统对人工量价因子的挖掘起到了很好的补充作用。树模型来做因子组合可以一定程度地缓解 Alpha 的时变特性。

招商证券认为未来以机器学习为方法论的量化策略迭代方向主要包括以下几个：

1.特征工程，通用模型在简单数据集上的表现已经足够出色，人工先验信息的引入是提升策略表现的关键。人工信息的引入，主要体现在特征工程上。

2.算法升级，不同类型的模型在单个数据集的学习结果存在一定的信息增量，多模型集成能够提升整个机器学习策略的表现，更多不同模型的研究也是关键点。

3.目标优化，不同的业务场景，优化目标存在一定的差异。例如因子学习的目标是 IC 最大化。其他不同的学习目标例如降低组合波动率，最大化夏普同样值得探讨。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

实践案例

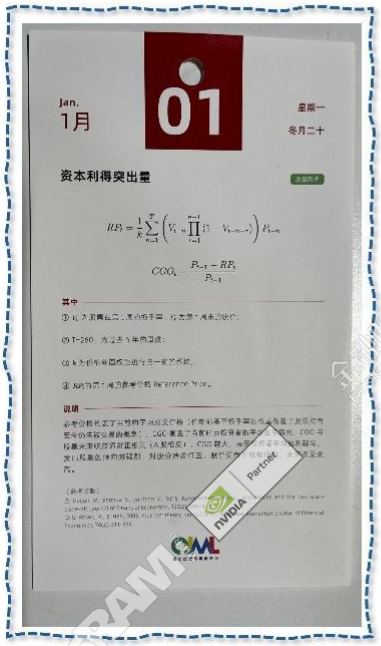
华泰研究更是探索大语言模型在量化研究领域中进行代码复现的实际应用。基于 GPT-4 系列模型和 RAG 技术（Retrieval-Augmented Generation，检索增强生成），构建了一个完善且易用的代码复现框架：“GPT 如海”。在框架内部，GPT 多模态模型提供了提取图片语义信息的能力，RAG 模块则有效支持海量文档的切分与检索，为大模型提供外部知识库，基于此，“GPT 如海”能够根据输入的图片或 PDF 文档，准确提取与代码复现任务相关的信息，并进行代码自动化生成。测试结果显示，“GPT 如海”能较准确地复现因子计算和人工智能量化策略代码。

多模态大语言模型：勾勒图片中蕴含的语义信息

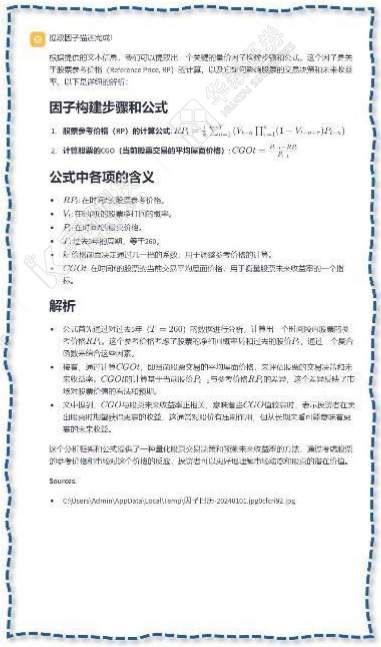
传统的大语言模型仅能处理单一模态的数据，例如文本，而多模态大模型拥有多模态感知和生成能力，例如图像、语音和视频，在交互性上更贴近通用人工智能的愿景。近年来，多模态大模型已经得到了长足的发展，目前多模态大模型已经具备同时感知文本、图像、音频的能力，并且逐渐发展出生成多模态内容的能力。在量化研究领域中，研究报告同样蕴含较为丰富的多模态数据，包括文本和图像数据，在多模态大语言模型图像感知能力的加持下，代码复现系统可能会更具准确性与实用性。

■图表　单因子代码复现（图像文件输入）

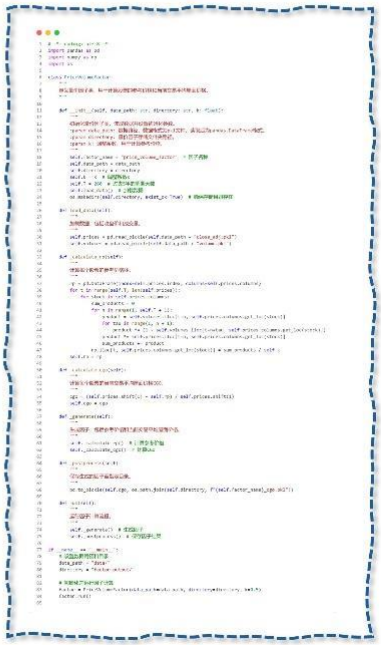
①输入文件为包含因子详细构建信息的图片



②“GPT如海”从图片中准确提取因子公式和构建步骤



③“GPT如海”直接输出因子构建代码



我们可详细观察“GPT 如海”代码复现的特点。对于代码注释而言，注释丰富鲜明，表达上也较为谨慎，有效提升了代码的可读性和可参考性；对于原文核心信息捕捉上，“GPT 如海”复现系统也较为准确，例如原文强调神经网络架构的整体损失为主干网络和门控网络的合成，复现代码中也的确书写正确，但“GPT 如海”也并非没有错误，例如主干网络的损失实际是收益率衰减加权的 mse 损失，但它写为交叉熵损失（代码第 22 行）；在代码准确性上，我们发现“GPT 如海”依赖的 gpt-4-0125-preview 模型能力较强，无论是神经网络的反向传播、类的撰写以及函数的调用等等都较为准确，表明该系统的代码复现水平较高。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

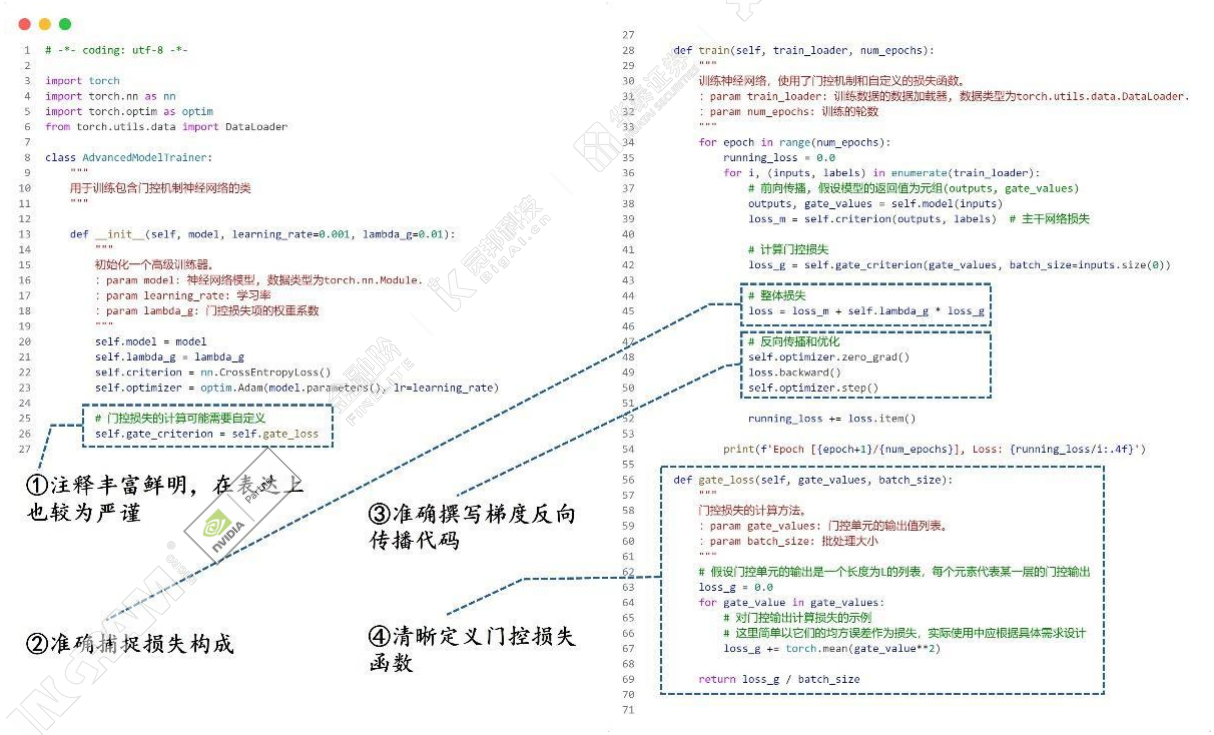
检索增强生成：为大语言模型披上知识之铠

RAG（Retrieval-Augmented Generation，检索增强生成）是一种通过使用外部来源事实构建向量知识库，用于提升生成式人工智能准确性和可靠性的技术。具体而言，在大模型生成内容时，RAG 将首先在外部知识库中检索相关信息，随后依据相关信息大模型将生成更为准确的回答。基于可随时更新的外部知识库，RAG 在有效提升大模型生成内容的时效性与准确性的同时，降低了大模型产生“幻觉”现象的可能性。对于代码复现系统，RAG 能支撑海量文档并提供高效信息检索，为代码生成提供信息依据。

“GPT 如海”可复现因子计算和人工智能量化策略代码

我们针对不同应用场景设计了两套代码复现模板，分别是因子代码复现和人工智能量化策略代码复现。对于因子代码复现场景，测试结果显示“GPT 如海”能够较准确地提取出图片或 PDF 文档中的因子表达式以及因子构建步骤，并能进一步以此为基础构建因子计算代码；对于人工智能量化策略复现场景，测试结果显示，“GPT 如海”能够较准确地提取出人工智能量化策略的构建步骤和细节参数，并依次输出数据集构建、模型架构和模型训练的.py 代码，在代码细节上，我们观察到“GPT 如海”能够敏锐捕捉到策略构建的细节，例如损失函数设计等，尽管这些细节信息散布在文章各处。

■图表　人工智能量化策略代码复现示例一中的代码亮点（以训练代码为例）



多模型复现效果对比：小模型也可“纳须弥于芥子”

在大模型如火如荼的发展进程中，专为响应大模型落地需求的小型大语言模型也渐为兴盛。我们简易对比了 GPT-4、通义千问（Qwen 1.5，7B）和 Llama3（8B）的代码生成效果，在不同评价维度上，三个模型展现出差异性特征，例如 GPT-4 并不严格遵守代码模板，而通义千问和 Llama3 较为遵守代码模板。客观而言，通义千问和 Llama3 以小参数体量博得不错的代码生成效果，体现出小型模型“纳须弥于芥子”的潜力。风险提示：大模型存在幻觉现象，模型生成结果可能不符合事实。大模型生成的代码可能存在错误，使用需谨慎。大模型提取信息可能存在遗漏。大模型训练集广泛，可能存在过拟合风险。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表　多模型代码复现效果对比详细信息

	GPT-4	通义千问	Llama3
模型全称	gpt-4-0125-preview Qwen1.5-7B-Chat	Llama3-8B-Chinese-Chat	
参数量	-	7B	8B
来源	OpenAI	Hugging Face:	Hugging Face:
代码能否直接运行	是	否	否
代码正确情况	少量错误	部分正确	部分正确
是否遵循代码模板	否	是	是
注释丰富程度	丰富	丰富	一般
模型运行环境	联网 api 调用	本地运行	本地运行

——2024 年 5 月 06 日华泰证券研究所金融工程团队《人工智能系列之 77：“GPT 如海”：RAG 与代码复现》

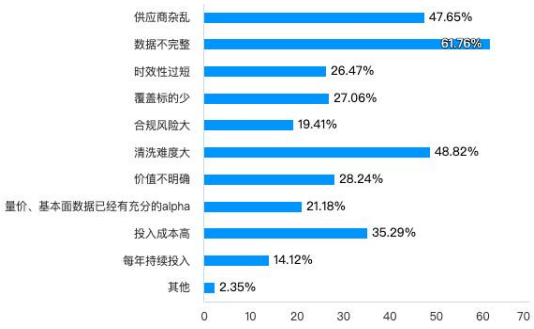
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

02 另类数据：是理念而非机密

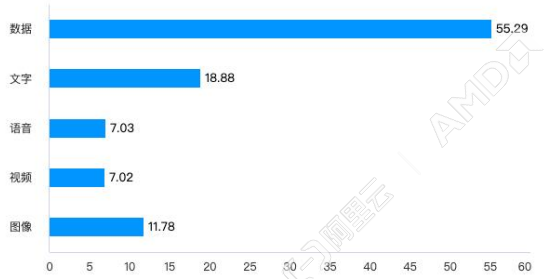
正如蒙玺投资所说，**另类数据是理念而非机密**【更多内容详见第六章-蒙玺投资】。

另类数据的投资道阻且长，目前另类数据的使用受到数据质量、使用成本以及效应存疑等多方面因素的限制，使得量化机构在使用另类数据时需要权衡成本、效益、风险和可行性。

■图表 目前量化机构另类数据使用的限制性因素



■图表 目前量化机构另类数据使用的主要介质



数据资料来源：《2023 中国量化投资白皮书》问卷调研

数据质量仍然是最核心的限制因素，包括数据不完整 (61.76%) 缺少关键信息，影响分析的完整性；清洗难度大 (48.82%)，意味着需要投入更多的时间和资源；供应商杂乱（47.65%）多且标准不一，也会给数据采购和质量控制带来挑战。

在应用这些数据的时候，最核心的还是是否具有强逻辑。量化机构需要寻找逻辑链条较短、逻辑清晰、历史更长的的产业链数据或行业特有数据，有一些数据可能非常有趣，但其逻辑链条非常长，例如医疗行业临床二、三期数据，这涉及对行业背景进行深入研究。而 APP 信息推断用户身份、职业，再找到用户涉及的上市公司，逻辑链条也会比较长。如果数据时间跨度较短，则很难进行有效性检验。当然强逻辑性的数据无需进行过多复杂模型的研究和验证。

目前数据显示，另类数据的清洗难度直接影响使用权重。目前结构化数据仍然是最主要的数据来源（权重占比 55.29%），非结构化数据如文字、语音、视频和图像还需机构自行清理，所以应用范围较少，权重占比分别为 18.88%、7.03%、11.78%，文字（权重占比 18.88%）通常来源于新闻、报告、研究文章等，其相对较高的使用比例表明量化投资机构在利用自然语言处理（NLP）技术来从非结构化的文本数据中提取信息，增加对市场情绪和事件的理解。语音（权重占比 7.03%）与视频（权重占比 7.02%）使用率较低，处理技术相对复杂，对算力要求高，且信息提取的精确度可能低于文字和结构化数据。然而，这些数据类型可能包含丰富的情绪和行为信息，对于社交媒体分析、CEO 讲话分析等领域可能有独特价值。图像 (权重占比 11.78%) 包括如卫星图像分析农作物产量、零售店铺客流量等。尽管图像数据的处理也是技术密集型的，但人工智能领域有非常清晰的技术路径，占比略高于语音、视频。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

03 星辰与圣山：通用人工智能与天体物理

除了量化投资，量化机构还在探索什么星辰大海？“无情的 geek”是量化人的一种简单片面形象，文艺复兴的基础科学、《破局》与圣山、基础教育与慈善是其更多面体。

1974 年，西蒙斯与陈省身共同提出了陈-西蒙斯规范定理（Chern－Simons form），这一理论框架将几何、拓扑学与量子场论相结合，对理论物理学，特别是量子场论和弦理论产生了深远影响，西蒙斯因此在 1976 年获得美国几何界最高奖——美国数学会凡勃伦几何奖。

文艺复兴公司成立之初就聘请了数学家、天体物理学家、密码破译者和其他类型的科学家，共同从一系列数据中发现模型。2016 年，西蒙斯基金会投入巨资启动“熨斗研究院”（Flatiron Institute），以期将“文艺复兴”的的算法分析策略应用于拓宽知识、帮助人类的科研项目等基础科学研究，包括计算生物学、天体物理、量子力学、神经科学等领域。熨斗研究院不会独立开展任何新的实验，在此工作的大多数研究人员与大学合作，从一线实验室——那些装备着培养皿、高压灭菌器和基因改造过的小白鼠的设施——当中采集新的数据。该研究院的算法和计算机模型，旨在帮助研究人员发现已采集数据中所隐含的信息。例如，从时空扭曲的程度推断未发现新行星的位置；从染色体看似无用的部分中识别与突变有关的基因链；意大利小蜂的神经元网络的绘制工作……在成立 2 年后，熨斗研究院就已经产出了无数的分析程序和算法，形成了 280 余篇学术论文。

在国内，许多量化机构也以企业公民的角色积极参与各种社会活动。通过教育和科技赋能帮助需要的社会群体，共同推动技术革新与社会进步。

在精进技术细节的同时，许多机构也做更有温度的量化私募。从幻方 2020-2021 年捐款接近 6 个亿，涉及 28 个项目，其他机构也在通过各类方式支持教育、助力科研、乡村振兴、扶困救灾等。

■图表 部分量化机构公共事务参与

名称	类目	内容
佳期	春蕾计划	“春蕾计划”是在全国妇联领导下，由中国儿童少年基金会于 1989 年发起实施、致力于改善贫困家庭女童受教育状况的儿童公益项目。佳期团队以“助学+陪伴”的帮扶模式，将连续三年资助甘肃省困难春蕾女童，通过多种线上和线下活动，多方位助力贫困女童身心健康、快乐成长。
	Technovation	Technovation 是一个全球性的公益竞赛项目，主要愿景是通过对世界各地有意参赛的女生进行编程等技术能力的辅导，启发她们探索发现社区存在的某一特定问题，并通过技术手段加以解决、最终以科技回报社区。佳期团队成员作为参赛队伍的志愿者导师，对参赛的青少年进行了全方位的指导，给她们提供了学习计算思维和提高解决问题能力的机会，帮助她们在技术驱动的社会环境中实现自我价值。
	THE Hack	THE Hack 是一个国际化的大学生创客马拉松赛事，意在为参赛者提供一个技术交流探讨、以及制定可行解决方案的合作性平台。参赛选手多为人工智能、大数据、机器人设计等领域优秀的年轻人，通过比赛的形式寻求金融、教育、健康等不同行业的解决方案，便利人类生活。佳期团队以资金赞助的方式支持该有意义的赛事，让这些年轻人可以充分挖掘自己的潜能、将他们的智力与热情转化为推动社会进步的动力。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

灵均	乡村振兴	在乡村振兴方面，公司坚持“扶贫必扶智”，为加强新时代乡村教师队伍建设，在北京师范大学设立“灵均优师基金”，投入 10 年时间全力支持北师大“优师计划”“强师工程”，引导“优师”学子毕业后赴中西部 832 个脱贫摘帽县一线执教，到国家最需要的地方建功立业；携手南京农业大学，重点帮扶来自农民工、农村劳动力缺失、因病致困等家庭的学生；携手浙江大学，设立“灵均乡村振兴基金”，用于支持、激励师生开展田野调查、洞察“三农”国情、扎实开展“三农”领域政策和对策研究，积极服务乡村振兴和共同富裕；在清华大学、北京大学、浙江大学等高校设立助学金，进一步帮扶品学兼优、志在报国的优秀学生；公司及时响应中国证券投资基金业协会号召，积极参与到山西汾西精准扶贫工作中，努力提升当地困难群众生活水平。
	教育科研	在教育科研领域，公司一方面持续加大对基础教育阶段贫困适龄学生及家庭的捐赠和帮扶，携手上海真爱梦想公益基金会、北京华远达公益基金会、北京市行远公益基金会等公益慈善组织，促进落后地区儿童的教育公平；另一方面在清华大学、北京大学、浙江大学等国内多所知名高校，发起设立“灵均教育发展基金”，通过奖学金、科研基金、赛事基金等多种方式，在量化行业优才培养、量化研究赋能、困难学子帮扶等方面贡献力量。
	医疗卫生	公司持续资助上海交通大学医学院附属瑞金医院，努力提升科学医疗水平和医护人员工作条件。在疫情防控方面， 2019 年新冠疫情暴发后，第一时间向武汉汉阳医院、武汉市武昌医院、北京协和医院、首都医科大学宣武医院、华中科技大学同济医学院附属协和医院、上海市红十字会等医疗与慈善机构捐款捐物，用于支持一线医护人员防疫，保障医护工作者的健康安全。
幻方【2020–2021 年】	灾情疫情驰援 幻方在行动	1350 万元众志成城，驰援灾情；涉及公益组织：中国红十字会、河南及山西省慈善总会
	幻方新长城自强班	1800 万元 2600 高中生助学纾困；涉及公益组织：中国乡村发展基金会
	幻方童伴妈妈	2000 万元 100 个童伴之家；涉及公益组织：中国乡村发展基金会
	起点工程项目（一二期）	3050 万元提升农村儿童学前教育质量；涉及公益组织：中华少年儿童慈善救助基金会
	希望工程幻方 1+1 助学计划（一二期）	1.5 亿元保障低龄女童受教育权；涉及公益组织：中国青少年发展基金会
	灾情驰援 幻方在行动	1200 万元提升农村儿童学前教育质量；涉及公益组织：中华少年儿童慈善救助基金会
	白血病救助小天使基金	1.5 亿元保障低龄女童受教育权；涉及公益组织：中国青少年发展基金会
	乡村工匠项目	100 万元支持地震灾后救援；涉及公益组织：四川省泸定县红十字会
	赢未来项目	2400 万元帮助困难家庭的白血病患儿；涉及公益组织：中国红十字基金会
	梦想中心项目	1800 万元支持传统技艺传承；涉及公益组织：中国乡村发展基金会
	幸福工程项目	500 万元支持 6 省 9 市 35 所职业院校教育提升；涉及公益组织：中国发展研究基金会
	爱的分贝项目	2000 万元惠及 2100 乡村教师+92000 乡村儿童；涉及公益组织：上海真爱梦想公益基金会
	新长城高中生助学项目	500 万元困境母亲帮扶；涉及公益组织：中国人口福利基金会
	拾穗行动项目	1000 万元听障儿童支持；涉及公益组织：中华思源工程基金会
	天使妈妈项目	1000 万元 1500 高中生助学纾困；涉及公益组织：中国乡村发展基金会
	儿童大病救助项目	800 万元失依儿童关怀；涉及公益组织：中国人口福利基金会
	慧育中国项目	700 万元儿童大病救助；涉及公益组织：中华思源工程基金会
	一村一园项目	800 万元助数百困境大病患儿家庭；涉及公益组织：浙江省慈善联合总会
	e 万行动（孤儿支持）项目	1000 万元山村入户早教；涉及公益组织：中国发展研究基金会

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

隰县兰考汾西乡村振兴项目	1200 万元支持中西部地区 200 余所幼儿园建设；涉及公益组织：中国发展研究基金会
夕阳红山区老人扶持项目	600 万元帮扶 1800 名事实孤儿；涉及公益组织：爱德基金会
关爱老兵项目	2000 万元支持乡村综合发展；涉及公益组织：中国乡村发展基金会
美丽中国支教项目	200 万元支持浙江省山区 26 县养老事业；涉及公益组织：浙江省慈善联合总会
血友病患者焕新计划项目	100 万元老兵生活、医疗关爱；涉及公益组织：爱德基金会
石阡乡村振兴项目	1500 万元乡村支教支持；涉及公益组织：北京立德未来助学公益基金会
春晖妈妈项目	500 万元血友病患者救助；涉及公益组织：北京血友之家罕见病关爱中心

数据资料来源：相关机构官网，编写组整理

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

CHAPTER

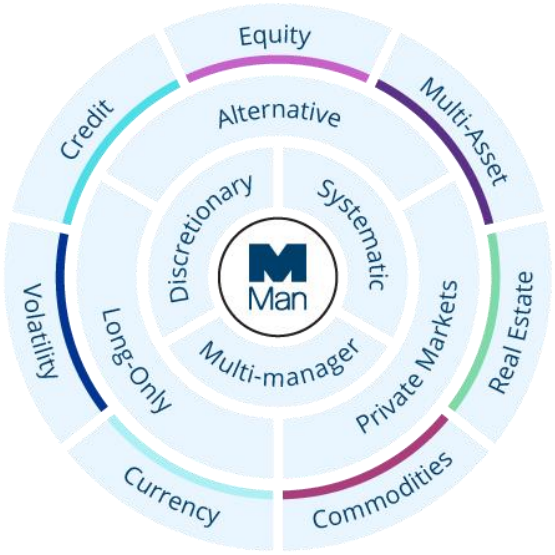
第六章
机构最佳实践

特别说明：本章节所有文字内容的搜集和整理均基于2023年12月至2024年2月期间访问的资料和数据。在此期间引用和参考的各类市场分析和研究报告，均以当时的市场环境为基础。请注意，由于市场环境的动态变化，后续可能会发生较大变化，敬请读者在阅读时尽量避免对本书中的内容和引用资料进行曲解。



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

01 Man Group：我们的本土化创新



1.2012 年，英仕曼就在北京设立办公室，在过去十多年时间，英仕曼在商业、策略研究上都做了哪一些工作？策略与集团旗舰策略有何异同，如何做本地化的创新？

许铭：英仕曼集团最早在中国设立代表处，后来成立了专注于私募基金业务的全资控股子公司。

英仕曼很早就开始关注并投资于国内市场。国内期货市场流动性非常好，品种也非常丰富，很多品种国际上没有，而且国内期货 市场的运作逻辑较为独立，与国际主流市场相关性低。

我们的投资策略本土化改造有几个方面：一是具体参数上的一些调整，使策略框架适用于国内市场；二是管理策略的波动率；三是加入一些本土化期货品种，在投资组合构建过程中，或者是新品种上市时，我们都会根据品种的流动性判断这些品种能否纳入 组合，或者每年哪些时间该品种流动性好，可以进行投资。

2.2023 年，哪个市场事件给您留下最深刻的印象？

许铭：一定是生成式人工智能（Generative AI）。在英仕曼集团，我们对此也做了大量研究投入。在我们看来，生成式人工智能未来有很多方面有潜力应用于资产管理行业，比如协助软件开发和编程、支持面向客户更加具有针对性和互动式设计的营销活动。大语言模型（LLM）在法务领域也有很大潜力，比如加快法律合同的拟写和审查、总结文本信息回答尽职调查问题等。投资方面，一些研究证据表明大语言模型（LLM）可以进行更好的情绪分析，例如财报的文本分析、业绩报告电话会的纪要，以确定其传达的情绪是积极、消极还是中性的。此外，最近的一项研究表明，ChatGPT 能够有效地总结企业财务报告的内容，甚至表现出针对特定主题生成摘要的能力。大语言模型（LLM）在资产管理领域的另一个用例是他们能够识别概念主题和公司描述之间的联系，以创建主题股票篮子。除了大语言模型（LLM）之外，生成式对抗网络（GAN）是另一种生成式人工智能的模型应用，一个例子是创建模拟的金融市场时间序列数据。研究表明，生成式对抗网络（GAN）可以生成更加接近实际数据模式（例如含有肥尾分布特征和波动率变化特征）的价格数据，最先进的应用包括使用 GAN 来估计尾部风险，合成尾部场景。GAN 还被用于投资组合构建和策略参数调整、创建模拟市场交易数据等。资产管理中的其他应用还包括使用大语言模型（LLM）生成宏观经济的模拟情境来构建和评估宏观交易信号，使用 ChatGPT 作为财务顾问等。

3.谈谈英仕曼集团在金融创新和技术投资方面最新的进展？

许铭：集团层面，英仕曼集团始终致力于站在金融创新技术前沿，持续加大人才、技术和研发的相关投资。2023 年 3 月，英仕 曼集团和彭博（Bloomberg）合作，签署一项新的 DataFrame 数据库产品 ArcticDB 多年开源技术开发和产品整合协议。这是英仕曼集团首次开展此类业务，产品落地后将成为彭博分析平台 BQuant 的一部分。ArcticDB 是一个高性能的 Python 数据库。搭建 ArcticDB 数据库的初衷是为了应对英仕曼集团前台研究人员日益增长的数据量和复杂程度，这也是当前许多大

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

型买方和卖方机构所面临的挑战。彭博计划将 ArcticDB 数据库整合进彭博分析平台 BQuant。英仕曼集团每年的交易下单规模达万亿美元，这就对数据科学工具的速度、灵活性和从规模上存储并处理复杂数据的熟悉程度都提出了更高要求。ArcticDB 数据库处理的数据可达数亿行、数十万列，在处理深度历史成交数据、为公司债市场建模等方面均有很强的实用价值。通过 ArcticDB 数据库，英仕曼集团的投资专家和技术人员可以更好地完成稳健的、几近实时的自动化交易。ArcticDB 数据库也提供了研究数据库的实时分析和信号回测功能。

4.我们知道，英仕曼集团是联合国负责任投资原则的签署人，在这方面领先很多机构，公司如何纳入负责任投资原则？是否有负面排查标准？

许铭：在 ESG 投资方面，英仕曼从大框架的两个层面来实现其应用。

第一，设置负面清单，包括不符合规则的公司或者是行业的名单。

第二，有专门负责投票的部门，在公司所有的基金所持有的股票，都会尽量参与公司投票。

另一方面，在股票的方面通过打分制，对环境、社会和治理这三个方面底下分十多个小类逐一进行打分，最终的分数会与其他所有的阿尔法模型的分数做一个加重来达到就是对股票的整体评分，然后进行选股，对该区间里所有的股票在广度上和深度上都能进行 ESG 分析。

在负责任投资基金框架下，英仕曼集团自有的负责任投资负向屏蔽排除清单。负责任投资负向屏蔽排除清单有效反映了国际普遍的准则和传统，支持四个领域的排除：争议性武器和弹药、核武器、烟草和燃煤。

5、坊间有人质疑趋势追踪策略危机 Alpha 的效果，您觉得这些质疑的声音有道理吗？

许铭：突然的市场冲击常常使我们对趋势追踪策略的“危机阿尔法”特性产生疑问，特别是针对它的偏度特性。我认为答案在于我们所观察的时长。简而言之，当市场波动的持续时间以天为单位衡量时，趋势追踪策略的盈利能力就像抛硬币一样。趋势追踪策略的仓位可能正确，也可能相反。策略几乎没有时间做出反应，而收益的分布可能相当左尾。然而，当市场波动持续数周或数月时，这与趋势的持续时间更加一致，策略的仓位也有时间做出相应的反应，随着时间的推移，趋势追踪策略能够通过平掉仓位并进入观望模式，从而在统计上摆脱左侧肥尾的特征。如果市场的抛售持续，趋势追踪策略可能会转移到空头仓位，并将先前的损失转化为收益。如果抛售没有持续，趋势追踪策略可能会恢复其原有的多头仓位头寸。相比之下，对于股票多头策略而言，如果不进行干预且市场环境继续恶化，做多股票的仓位将继续产生亏损。

6.能否介绍一下英仕曼 AHL 的投研团队和投资流程？

许铭：英仕曼 AHL 投研团队大约有 130 人，拥有多元化的学术背景，同时倡导开放的挑战辩论文化。AHL 的投研人员分布在伦敦、纽约、香港、上海等多地，采用全球合作的模式。我们秉持合作透明的理念，在开放的环境中充分交流观点，不从传统金融的视角来观察市场。我们追求超额收益阿尔法没有界限；我们在所有市场品种、所有模型、所有数据中搜寻阿尔法机会。我们这种全球合作架构可以有效地运用世界各地的数据库与资源。

7.数据当中不良数据如何判断与清洗？国内不同板块品种之间的关系是否会与海外的商品有所差异？

许铭：是的，这个问题在另类数据的使用中是尤其需要注意的。另类数据相对传统数据而言，种类更多样，例如网上搜索的信息、网上交易、论坛留言、app 的使用再到卫星成像天气的数据等全部都属于另类数据的范围，更能及时提供第一手信息。这块应用在未来非常有前景。实践中发现，另类数据相对于传统的宏观经济数据而言没有固定的结构，主要原因是另类数据非常新；第二另类数据来源多样化，并且非结构化。需要有一个框架去对于数据质量做判断，然后对它进行清洗，最

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

后才能把它转换成投研可以使用的一个数据。这方面需要很大投入，这几年各家管理人在这方面的竞争也日趋激烈。英仕曼成立了专门的数据科学团队，从数据引入、质量评估等全流程处理后才转化为投研可以使用的数据。

至于第二个问题，会有差异，国内品种相比海外品种有着较低的相关性，从而对国际投资者而言也有着较好的风险分散价值。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

02 半鞅私募：“三个火枪手”的端到端量化之路

01 三个火枪手：Quant+AI+NLP

“大家都希望我们有一些波澜起伏的故事，但是很可惜我们没有。”

半鞅投资总监王芹笑容爽朗，沟通轻快直接，公司主要由她、技术总监仇伟以及研究总监周东组成核心三角，从各自擅长的领域出发，相互补充协同作战。

王芹有超过十年的量化从业经历，2012 年开始在太平资产担任高级量化投资经理，当时主要以人工挖掘的因子做多因子策略为主。

“那时候市场上除了大厂以外，行业金融机构很少涉及机器学习。最为人所知的世坤投资，其多因子策略的创新主要集中在因子挖掘上。而太平资产想通过机器学习来改进选股。”2017 年，太平资产开始推行机器学习方法论与微软亚洲研究院合作，引入机器学习取代传统的多因子模型，并建立了机器学习的框架。

半鞅研究总监周东曾任微软亚洲研究院机器学习组金融团队核心成员，设计并开发了业界开源的 AI 量化研究平台 Qlib。周东与王芹在合作过程中相识，两人共同推进了太平资产研究框架从多因子向机器学习的转变。而技术总监仇伟则是达摩院出身的自然语言处理（NLP）专家，除了在机器学习模型方面的专长，他还具备强大的技术架构能力。三人惺惺相惜，因共同的兴趣和目标而结成了伙伴，随后一同创立了半鞅，开始着手打造私募基金。

对于量化，他们入行有些早，但半鞅的成立又好像有点迟。

公司的成立时错过了 2020-2021 年的量化浪潮，2022 年才开始对外发行产品，但随后又碰上 4 月的上海疫情封控，产品遭受了巨大回撤。CTA 也没能赶上 2020 年 CTA 大年。“使得过去两年做起来非常痛苦。”王芹笑着表示：“尽管我们的发展节奏受到了影响，但我们仍在按照既定计划前进。”

虽然公司取得牌照的时间较晚，但产品推进迅速。王芹认为，这与之前的技术和经验积累有很大的关系。实际上，策略的孵化要远早于公司的成立。团队自 2021 年起引入了端到端的深度学习模型，专注于挖掘高频 Alpha，并自主开发了交易算法和交易系统，形成了完整的投研框架。2022 年更是被视为半鞅的创始年，公司吸引了多位行业顶尖人才的加入，确保了策略的快速迭代。高频 Alpha 策略的储备得以翻倍，新开发的策略在性能上实现了质的飞跃。

目前，半鞅的投研团队由约十人组成，中后台团队大约有四人。合伙人之间的分工明确：王芹主要负责优化和配置，仇伟负责技术架构和算法，尤其是高频端的研究，而周东则负责机器学习端。其余同事主要从事数据处理。所有系统都是内部自主开发，包括 C++低延迟交易系统、批流一体的数据处理平台、模型质量控制与监控系统、以及自主开发的大数据平台等。

“我们与市场其他机器学习类管理人最大的差异，可能是做的时间比较长，踩的坑比较多，因此我们会更加知道怎么去掌控这个东西。”当前，半鞅量化策略超 10 年经验，无论是传统多因子模型还是深度学习模型，都有充分积累，已经历过多轮牛熊的考验。而团队多元化带来了思想的碰撞和高效的行动。他们经常进行头脑风暴，捕捉灵感闪现的时刻，进行尝试、实盘、迭代，并从中获利。

02 端到端 从数据直接提炼特征

2023 年，行业策略热点之一就是端到端的 AI 模型多因子选股，其具体方法是：因子挖掘算法构建海量因子，使用线性或非线性模型做因子合成。2022 年量化逐渐开始使用 CNN、RNN 等神经网络模型，直接学习原始行情数据，或者从原始数据提炼特征，从而将因子挖掘和因子合成放在同一组网络中进行。这种方法黑箱程度更高，但是对于量价信息的挖掘也更加充分。

半鞅从 2018 年起就开始测试机器学习模型，是行业内最早采用机器学习取代传统多因子模型的机构之一。现在，半鞅已实现了工业流

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

水线级的投研流程，端到端优化全流程，以提升总体性能为目标。输入数据为原始 Level2 量价信息，通过批流一体引擎以毫秒级延迟将数据聚合成不同尺度的特征，并通过统一的端到端深度学习模型进行多周期、多目标建模。这使得半鞅具备全天候在任意时间点进行毫秒级预测和决策的能力，不受定时定点决策的限制，从而更好地适应市场的变化。

在传统的量化团队组织架构中，通常会包括数据组、因子组、机器学习组、优化组和交易组等几个核心部分。其中因子组会有大量的人员进行因子的挖掘、测试，得到横截面上的分层效果、以及 IC 和 IR 信息等。然而，随着方法论的进化至机器学习阶段，因子组的任务转变为提取特征。

传统私募可能会提取上万维的特征，更关注特征的体量和存储方式。但半鞅并没有传统机构中的关键组成部分，只有数据组。数据组的主要任务是进行基本的降噪和持续的数据平稳化，“这两项是学习模型的基本前提，而特征提取都是通过网络层实现，省去了人肉提取特征的过程。”王芹解释道。

除了省去了因子层这一环节，半鞅在其他环节也不断优化流程，包括在机器学习方面进行多尺度、多周期的动态建模、大规模分布式训练和自动化机器学习。在策略方面，实现了多周期融合、多决策点融合。同时，Alpha 策略适应了多策略自适应叠加。每个子策略均按照预测周期、数据输入、模型类型和预测目标四个维度进行标签化。半鞅的上层策略配置模型根据市场环境和产品特性来配置子策略权重，综合考虑收益和风险，灵活换手，最终实现稳健的收益输出。单个策略的深度决定了收益率，策略的广度决定了波动率。半鞅力求在单个策略的深度和策略的广度上都做到最优，以提供给客户更好的风险调整后的收益。同时，公司持续灰度上线新的端到端模型，以抵抗原有策略的衰减。

■图表 半鞅 Alpha 策略：高频α端到端深度学习量价模型



•输入为原始 Level2 量价数据

•使用批流一体引擎以毫秒级延迟聚合成不同尺度的特征

•通过统一的端到端深度学习模型进行多周期、多目标的建模

•具备全天候任意时间点的毫秒级预测和决策能力，不受限于定时定点决策，更好适应市场变化

■图表 Alpha 策略：多策略自适应叠加



每个子策略从预测周期、数据输入、模型类型、预测目标 4 个纬度打标签。

半鞅上层策略配置模型根据市场环境、产品特性来配置子策略权重，综合考虑收益和风险，灵活换手，最终实现稳健的收益输出。

单策略的深度决定了收益率，策略的广度决定了波动率，半鞅力求在单个策略的深度和策略的广度上都做到最优，以提供给客户更好的风险调整后的收益。

持续灰度上线新的端到端模型，抵抗原有策略的衰减

数据资料来源：半鞅私募基金

03 端到端人工智能量化

为什么同样是运用机器学习，半鞅和其他机构的净值曲线却有如此大的差异？

“每个机构的策略本来就不应该是相似的。如果完全一样，那就不是真正的 Alpha 了。” 王芹解释说，机器学习方法论天然容易学习到明显的特征，如高波动的小盘股或特别大的股票。但半鞅并没有这些，公司的 Alpha 策略不会随着市场风格的变化而大幅摇摆，尤其是小盘股的影响，与市场上其他管理人的相关性较低——这背后的本质是知道如何控制风险，是时间与人力共同作用的稳健结果。

但像其他量化策略一样，该策略对流动性差、热点概念频繁变化造成的无效换手、大小盘分化造成的交易机会减少或大盘剧烈波动下数据分布异常等情况比较敏感。接下来，半鞅还将继续迭代升级交易系统性能，精进技术细节，合理整合资源，以实现更高效的交易。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

一些管理人在谈到人工智能及其特征时，总会担心人工智能策略会失效得更快。王芹却持有不同态度，她认为：“简单的策略更容易失效，因为大家可以很容易地复制出来，并很快就会把蛋糕分光，快速地将市场消耗殆尽。而人工智能策略的复杂性使得复现变得困难，这可能会延缓策略的失效。”

端到端模型的高适应性并不意味着它可以未经调试就广泛应用于全球市场。由于 A 股和海外市场在市场结构和投资者结构上存在巨大差异，提取信息的方向可能完全相反。例如，在国内市场中，横截面交互特征相对强烈、大盘与小盘的分化非常显著。这就意味着所需特征和特征提取网络结构之间存在显著差异，无法保证在 A 股市场能够获得成功的网络结构在美国市场也能取得同样优异的表现。

半鞅在 DNA 中深植 AI 基因，但他们对于工具的态度实用主义。例如，尽管 AIGC 技术大火，但半鞅并没有急于将其与现有工具结合。王芹认为，像 ChatGPT 这样的技术对于处理另类数据的管理人可能更有用，如帮助识别语音等，但对于主要依赖量价信息的管理人帮助并不大。在半鞅看来，AI 只是一个工具，一个用来赚钱的工具。关键是不要陷入另一种固定思维，反而被局限，而是要真正做到“为我所用”。在这个市场中，最终的竞争可能还是对投资的深刻认知。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

03 黑翼资产：控制回撤是第一位的，足够稳健才是长远目标

在 A 股历史上，从来不缺乏宏大叙事，也从来不缺乏极致行情的演绎。

2008 年，一场金融海啸席卷全球，美国股市一片萧条，华尔街开启了第一代华人量化基金经理归国潮。

在这场金融余波中，当时就职于雷曼兄弟的陈泽浩和就职于巴克莱的邹倚天也相继回国，转战中国资本市场。

随后 2010 年融资融券和股指期货相继推出，中国的私募量化基金也开始生根发芽。之后的几年中，国内的量化私募管理人如雨后春笋般涌现，量化投资如火如荼。

2014 年，私募全面阳光化，陈泽浩和邹倚天合伙创立的黑翼资产，成为了第一批登记备案的量化投资机构之一。

十年过去，黑翼从蹒跚学步到跻身百亿阵营。回望这十年成长之路，陈泽浩和邹倚天多次提到一个词：稳健。



黑翼资产创始合伙人兼 CEO 陈泽浩



黑翼资产创始合伙人兼 CTO 邹倚天

01 理性、严谨的量化老将

All knowledge is，in final analysis，history.在终极的分析中，一切知识都是历史；

All sciences are，in the abstract，mathematics.在抽象的意义下，一切科学都是数学；

All judgements are，in the irrationale，statistic.在理性的世界里，所有的判断都是统计学。”

——Calyampudi Radhakrishna Rao《统计与真理》

如果把量化比作是习武，那么数理思维就是做好量化投资的内功。

邹倚天在受访时表示，人类已经积累了这么多年的知识，能掌握分析的大部分都是历史数据。所以在终极分析中，都是在复盘以前的经验。数学是所有学科的基础，量化运用了前沿的自然科学技术，包括数学、物理学、统计学、计算机等，用来挖掘市场运行的规律，进而形成合乎逻辑的投资方法，实现策略。

数学出身的陈泽浩则认为，量化追求科学，但也需要艺术。“假设某个策略在当前的市场环境中失效，是要等到其失效概率变成 99%，还是在概率达到 80%的时候就做出决策，这就涉及主观判断。这不仅取决于对策略本身逻辑的理解和对策略的信心，更考验自身对市场环境的认知。”

理性、严谨，是黑翼两位创始人身上的共同特质。这些共同特质背后，则源于两位多段重叠的经历。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

把时间的指针拨回到 20 多年前，在美国斯坦福大学的一场迎新晚会上，两位对量化投资都有着浓厚兴趣的理工科学霸就此相识。在统计学博士毕业之后，陈泽浩前往雷曼兄弟从事固定收益衍生品研究工作。邹倚天则在电子工程系硕士毕业之后前往巴克莱全球投资公司 (BGI) 从事量化股票投资策略的研发。

最后，两人在国内的老牌公募博时基金汇合，负责管理股票量化专户产品。2014 年，是国内私募行业阳光化的第一年，个人理想、时代机遇和人生使命在这一年交织，两人搭乘中国量化私募发展的东风一起创立了黑翼资产。

是校友，是同事，也是合作伙伴，这两位斯坦福“双子星”相识并共事十余年，彼此之间的默契早已不言而喻，在战略决策层面往往能够快速达成共识。

02 稳，在极端行情中沉淀

在共同掌舵黑翼的日子里，陈泽浩和邹倚天两位一起经历了多轮 A 股市场周期和极端风格转换。

据邹倚天回忆，公司在 2014 年 5 月刚成立时，规模还很小，尚在蹒跚学步。结果第一年就恰好赶上年尾的蓝筹行情，大小盘风格极速切换，旋即感受到了第一波大型回撤，很多机构单月回撤达到十个点。

“我们当时的策略偏基本面，并未受到小盘大跌的影响。”邹倚天表示，加上当时黑翼有较好的风控模型和组合优化，当月的超额基本没有回撤。虽然这波极端行情对自身冲击不大，但黑翼还是进一步迭代了风控体系。

2015 年市场骤变，同时股指期货贴水还超过 30%，彼时很多量化中性策略都被迫暂停。据邹倚天回忆，当时考虑到对冲工具的流动性依然缺乏，中性产品超额收益又不断遭受成本侵蚀，所以决定先把仓位降下来。

不过，这也促使量化管理人加速策略研发。2017 年底，黑翼就提前开始布局机器学习和 AI 技术，组建机器学习研究团队，这为后来推出的 300 指增和 500 指增产品增强了收益。

稳，见于每一次对极端行情的战略应对，也见于每一刻对市场的感知。

“很多人说量化不需要经验，我个人并不认同，量化很需要对市场的感知度。”陈泽浩表示，黑翼并不是等到规模扩大才加入风控等各种方法，而是在平时就做了严格的控制。所以在遇到类似深幅调整时，黑翼并没有专门临时抱佛脚或修改风控条款，也因此基本上都承受住了市场考验。”

2021 年，量化业绩规模齐飞，行业格局初显，黑翼凭借卓越的 Alpha 以及风控体系，也顺势站上百亿梯队。然而同年 9 月大盘回调，许多量化管理人经历成立以来最大的一轮回撤。黑翼则全面升级了股票量化策略，因子库进一步丰富。

邹倚天认为，每类底层因子都有一定的周期性，单一因子很难长期战胜复杂多变的市场，只有把量价因子、基本面因子、机器学习因子三类主流因子结合起来，才能更有效地应对行情变化。

迭代后，黑翼股票策略的超额夏普提升超过 50%，超额回撤减少了 50%以上。数千万的投入使得整体算力得到跨越式提升：回测速度提升 3 倍以上，训练速度提升 5 倍以上，IO 吞吐速度提升 10 倍以上。

03 有所为，有所不为

经受过回撤考验，管理人才会在风控上考虑得更加周全。在黑翼，控制回撤是第一位的，足够稳健才是长远目标。

黑翼在规模不足 10 亿元时，就已经着手研发优化器，包括风险模型在内的整个投研体系都非常完备。邹倚天表示：“这是因为我们一开始的理念就是走机构化路线，不过度关注规模，所以能够很好地遵循严格风控原则。”

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

稳是基于理性判断，也是有所为有所不为。黑翼一直严格控制各类因子的暴露敞口，不在小市值风格上过度暴露，不依赖暴露程度获取超额收益。

据邹倚天回忆：“在 2014 年之前，小盘因子持续表现较好。如果在小盘风格上暴露便能够产生不错的超额收益，但风险会很高。我们最终选择不追逐热点，不在小盘风格上过多暴露。”

得益于黑翼对风格暴露的冷静和克制，黑翼在 2014 年底的市场波动中业绩保持了相对稳定，这也印证了两位合伙人对于市场风险的预判和认知深度。

目前，黑翼仍然保持稳健的风控体系。指增产品的行业和风格因子偏离依然采取比较严格的限制，单只个股的仓位上限不超过 1%。其中 300 指增和 500 指增的行业偏离度、大小盘风格因子的标准差会更加收敛。

市场周期，循环往复。2024 年 2 月春节前一周，A 股市场再次呈现极致行情，非指数成分股流动性严重短缺，个股分化严重，很多量化管理人都在短期内遭遇了历史级别的净值波动和超额回撤，黑翼也未能幸免。

在多轮行情洗礼下，这两位合伙人认为，这次可能也是短期流动性冲击，只要产品不因为爆仓，触及平仓线等原因被动终止操作，当流动性危机解除，市场参与者模型或交易行为重回正常，超额应该也会很快修复。

黑翼的风控策略是根据实际情况和自身经验“量身定制”。在本轮的极端行情中，黑翼在收紧选股范围和高强度迭代策略的基础上，没有过度调整量化模型和风控参数，并尽量保持了仓位的稳定，所以较为平稳地度过了这次风险期。

未来，极端行情仍会上演，只是不知会以怎样的形式出现。经历过海内外多轮牛熊交替，也见证了 A 股市场的波澜壮阔和量化领域的日新月异，这两位身经百战的量化老将始终坚信一点：稳健才是王道。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

04 量派投资：万物皆存周期 行则将至



与其交流量派的投研细节，不如交流其创始人孙林的底层交易思维。

例如“万物皆周期”“不要相信绝对”。

01 万物皆存周期 行则将至

“道阻且长，行则将至。”2018 年，国内量化逆市全面飘红，有机构如此形容。

行业丰盛的想象力近在眼前：更多的期权和期货新品种上市；量化交易接口的放开；股指期货进一步放松；市场交易限制进一步减少。这些都令身在美国东海岸的孙林“蠢蠢欲动”。

彼时，孙林在 Two Sigma 担任短周期交易自营部门联席主管以及美国股票做市团队主管，带领着一个由奥数奖牌获得者和美国 Top5 高校博士组成的团队，进行短周期策略的研究、开发和交易。

巧合的是，他复旦大学的师兄余航也有意回国发展，两人在聚会中结识。彼时余航正在 Tower Research 自营部门核心团队，担任短周期量化研究员，负责期货、外汇、股票、期权、虚拟货币等市场短周期信号研究以及策略开发，交易涵盖了全球各个地区三十多个交易所的上千种产品。

两人一拍即合，2018 年着手团队筹备。在度过竞业期后，2019 年，两人开始审慎地招募精兵强将，并开始自营业务。2020 年 9 月，量派获得私募基金管理人牌照，也全面转型资管之路。

“我们经常开玩笑，创业最后悔的是没有早几年回来。”对于孙林和余航这种沉淀多年的专业管理人来说，他们早已见惯了盛极必衰、否极泰来。“万物皆周期，任何业务都是如此，包括量化、养猪、汽车制造等等。当前量化处于周期较低的位置，未来一定会有反弹。”孙林以美国股票量化交易为例，最初股票日内交易非常火爆，但是后来美股一直上涨，波动率在下降，机构赚钱能力变弱，到了 2017 年甚至有专门做股票日内交易的头部公司宣布退出这个赛道。

反转很快到来，2018 年日内策略绝地反弹，很多起步较晚的公司反而吃到了这波红利，尤其是 Two Sigma 成为其创纪录之年，“那时候市场觉得遍地都是钱。”

也正因此，孙林和余航一直坚信，如果将量化当作长期事业，那么当前更需要关注的是：持续地投入，将模型、技术做得更好；不断壮大运营团队，为投资人提供足够好的服务质量，余下的就只需等风来。

02 “量化的核心竞争力是迭代的速度”

对于量化行业而言，“更高、更快、更强”是发展道路上始终不变的目标。可以说，在量化行业，没有人能吃老本。

量派最初的方法论和基础架构汇集了创始人孙林和余航多年经验。仅仅 5 年时间，量派策略已几经升级。令孙林非常高兴的是，2023 年之前量派绝大多数的模型和产出来自合伙人。但从 2023 年开始，量派投研团队的产出呈现井喷状态，开发了众多深度学习模型，且不少已投入实盘应用。

目前量派总资管规模超 50 亿元。据孙林介绍，量派深耕于短周期策略，追求纯 Alpha，强调“高夏普、低回撤、低相关性”的特点，并严控风控，强调产品工具属性，以满足专业的机构投资人需求。

此外，量派也拥有丰富且完善的产品线，包括指数增强、量化多头、中性策略、全天候中性策略以及量化 CTA。其中全

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

天候中性策略是量派的亮点产品，该策略为类固收短周期策略组合，其中又分为股指期货的日内策略、可转债的日内策略、股票中性策略。

“高夏普、低回撤、低相关性”，要实现这一点，必须得使模型与他人保持不同。近两年，众多量化私募都在追求多周期信号挖掘与融合，但做法“百花齐放”，有的是从特征和因子层面融合，有的是从优化器层面融合。

量派的做法也独具特色。股票方面，量派并非国内主流 World Quant 多因子体系，而是生成各种维度的预测，最快是秒级别，最慢可以到一天，然后将预测组合优化后生成仓位，最后形成交易指令。“可以将其简单理解为有一个优化器，输入各种模型，模型对应不同的预测周期，或者可能是同样的预测周期，但是来源不同的模型。多周期融合全部发生在优化器中，最终得出交易指令。”孙林自信地表示，这并非教科书或者论文中的内容，而是量派打磨出来的特有方式。

在输入端，量派使用股票、股指期货、可转债，分为秒级/十秒级、分钟级/十分钟级、半小时级/小时级、日级等多个频率纬度，并研究其数据与指标。数据则包括短周期交易数据、盘口数据、量价数据、资金流数据等，指标包括动量指标、成交量指标、波动性指标、多维周期指标、价格指标。

■图表量派策略开发流程



数据资料来源：量派投资

CTA 策略方面，量派同样每日程序化动态调整组合，底层采用Tick 数据，灵活合成多维度、多周期特征，长周期以日度数据和周度数据为主，数据源丰富，包括量价、基本面、期限结构、另类基本面等。此外，在多策略组合时，量派会根据不同品种的特性，采用不同的组合方式动态管理组合，严格限定波动率及最大开仓比例。

值得一提的是，2023 年，CTA 市场低迷，但量派的收益依旧取得了正收益，且位于行业前列。与此同时，在优化器方面，量派也做得相当独特，其中包含大量优秀的模型，例如量价模型、期限模型和基本面模型，预测从 1 天到 10 天都有，所有预测在优化器中，并生成最终指令。“股票的数据足够多，过拟合风险相对 CTA 较小，所以完全可以制作成一个数据驱动的产品。但与股票不同，CTA 确实需要在优化器或者 Alpha 上面有更多想法。”孙林解释道，量派 CTA 策略中绝大多数模型偏截面，即在 40 个左右品种中都会产出一个数字，因此投入的许多模型都是数组和向量。

03 不要相信绝对的事情

“永远不要绝对地认为某一件事行或者不行。”在访谈交流中，孙林说了三次类似的话语。“永远敬畏市场，不要妄图以为掌握真理。”这不仅是孙林的信念，更是量派的发展理念。

十几年前，孙林刚刚毕业，在伦敦面试了瑞士信贷的一个股票交易团队，面试官看了孙林的简历，发现他毕业论文写的是神经网络。“有一位面试官非常确定地告诉我，这条路径永远无法生效。有趣的是十几年后，神经网络已经成为行业通用

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

做法。”这一“反转”的经历，不仅让孙林至今记忆犹新，也让他开始对市场与新技术保持敬畏：“不要相信绝对！许多现在认为无法使用的东西，也许未来某一天会变得有用，甚至会成为有助于保持市场竞争力的、不可或缺的东西。我相信未来这个行业也会出现更多、更复杂的模型。”

在巴克莱资本工作三年后，2012 年，孙林加入了 Two Sigma，一家以强调前沿科学技术运用而闻名世界的量化机构。但其实直到孙林回国创业时，机器学习和深度学习的市场占比仍较小。究其原因，一方面在于彼时的算力水平无法与今相比；另一方面，当时也缺乏较好的网络结构做预测，而且一个不错的网络结构被运用到金融预测也需要不少时间。当时，线性模型是华尔街量化机构最主流的方式，花费大量时间寻找强因子，也是典型的海外路径。

而国内人工智能的研究日行千里，有的公司从无处往有处去，完全按照全新的方式探索出路径，不仅使用到非常前沿的科技，也做出了卓越的成果。这也让孙林为之惊讶并赞叹。“开发出因子之后线性组合的方法有很多，关键是如何决定它的权重。”孙林进一步说明，量派会使用独特的方法组建线性模型，并且赋予这些因子权重，这也是很长一段时间内量派的主导模式。但值得一提的是，自 2021 年以来，随着量派持续发力机器学习和深度学习，如今量派也陆续开发了树模型、深度学习模型和端到端模型，也改良了公司的技术架构。

从美国量化历史可以看到，跑在行业前列的顶尖机构通常“人无我有，人有我优”。因此对于量化机构而言，坚持特点至关重要，同时也必须补充通用打法。“我的经验是不要拘泥于形式。黑猫白猫，抓到老鼠都是好猫。因此，永远不要过于绝对地谈论一件事情。”孙林说道。

当前，量化行业逐渐受到监管及公众的关注，但孙林表示，长期来看，这其实有利于量化行业健康发展。孙林回忆起自己海外工作时，Two Sigma 便是被官方严格监管的公司之一。每季度美国证券交易委员会 (SEC) 或者金融行业监管局 (FINRA) 都会要求其回答问题，孙林还经常是第一责任人。也正是这样的经历，让孙林明白监管对行业规范运作的必要性。“健康的金融市场必然如此。中国量化行业想要长期健康发展，监管介入越早越好，如果任由不规范行为恣意发展，最终结果会毁灭整个量化行业。”

反映在量派的日常中，孙林也始终配合监管的工作。“回答得能有多清楚就多清楚，能有多坦诚就多坦诚。”孙林解释道：“我们非常开心，监管开始了解量化行业的工作。但在我看来，想要让行业长期健康规范，这也需要更多人共同参与，无论是私募量化、券商还是柜台开发软件提供商……大家共同努力让普通大众了解量化行业的工作与用途。只有行业上下游共同发声，促进彼此交流，更多人才能了解、参与量化，甚至有更多人选择、信任量化来为自己的资产增值，同时也让金融市场更完善、更健康，流动性更好。这不仅是一个具有责任感和使命感的管理人希望达到的状态，也是一种从业者、行业、投资人共赢的状态。”

对于中国量化行业未来的发展之路，孙林坚信，未来中国量化定能在世界上占据一席之地，并做出独属于中国的贡献。“国内量化具有充足的人才储备，这也是我回到祖国后最惊喜的一件事情，无论是储备的数量还是质量都非常优秀。同时，当下我们正处于技术革命中，随着技术的发展，会有越来越独特的成果出现。在这些先进技术上，人人平等，我们也要学着拥抱它们，持有开放态度。”

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

05 蒙玺投资：守正出奇 与卷同舒

在获取量化 Alpha 超额的路上，蒙玺投资选择了“另”辟蹊径。

在蒙玺投资看来，在市场热捧的人工智能领域、非线性模型本质是 Smart Beta，Pure Alpha 应来自逻辑性强、可解释性强的单因子。所以公司从全球 200+数据源挖掘不同数据中的 Alpha，在其多因子模型中，另类因子占比达到 60%，基本面及量价占比仅为 30%与 10%。在其看来，布局另类数据，是公司适应市场变化而采取的一种商业战略定力。

另一方面，低延迟在过去两三年间争议不断。蒙玺投资作为国内低延迟领域的先行者，在其看来，技术的发展是非线性的，追求极致的同时，也在推进边界外扩，把不可能变为可能。对蒙玺投资而言，保持在这一细分赛道的差异化优势，是情怀也是图腾。

01 量化 2023：小市值暴露 AI 与 Smart Beta?

问：请您回顾 2023 年投资感受。

蒙玺李骥：2023 年是量化大年，行业在 Alpha 超额获取方面表现良好，尤其是偏量价类因子和小市值因子，我们公司在小市值方面暴露偏低，加上策略大量使用的另类数据表现较为稳健，同时覆盖了各类创新性产品，加上稳定的超额获取，所以产品线整体表现不错。

2023 年整体微盘股表现亮眼，但对于 2024 年而言这是潜在风险。暴露过多在接下来的时间可能面临不确定性，我们对此有一定担忧。

问：您认为 AI 模型更多是一种 Smart Beta，现在很多 AI 应用较多的管理人，在小微盘上有明显暴露，是否 2023 年行情也在某种意义上也印证了这种说法?

蒙玺李骥：过去几周，市场风格切换剧烈且迅速。有些优秀模型或机构能自适应市值偏离度较大的情况，有时可以达到-1，在 2023 年表现较为亮眼，但风险在这周（访谈时间：2024 年 1 月 19 日）已经逐步体现。

问：所以非线性模型本身并非 Smart Beta，而是这个模型本质是否具备学习能力?

蒙玺李骥：对于风格适应能力，非线性模型通常具备 Smart Beta 的特点，即预测性很强，非线性模型能够天然地挖掘出风格动量、风格收益，这本身是好事。有些风格可以被描述例如市值，还有一些风格甚至我们自己都不知道，但本质也是一种风格。我们也观察到一些模型在过去一两周内选择风格越来越偏大票，对这类模型我们不会过度担心，但有些模型预测周期偏长、适应性较慢，风控优化维度上还没有做市值控制，接下来可能会有一定风险。

问：关于大模型，当前国内量化公司大模型运用到哪一步？未来应用前景如何以及对量化行业有哪些影响?

蒙玺李骥：目前大家更多地关注公司流程和效率提升维度，包括 debug 等，投研方面增益不大。我们尝试运用它在数据和信息处理方面突破，但模型精度无法满足专业需求。未来在数据设计端也会有更多运用。我们时刻关注大模型发展。

问：如您所说，这项技术确实是非线性的，也不清楚后续情况。

蒙玺李骥：对技术永远要保持敬畏。

02 另类数据：是理念而非商业机密

问：蒙玺投资在另类数据方面研究非常深入，对超额收益贡献也在较高水平，请您围绕另类数据展开介绍。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

蒙玺李骥：我认为，另类数据更多是理念上问题，而非能力或者商业机密。

另类数据定义非常明确，就是除量价和基本面之外的数据。量价数据是指所有通过交易所推送的数据，目前机构构建数据时主要使用 1、5、15 分钟等 k 线切片数据，或者基于此异构 14 分钟、58 秒等 k 线数据，所有二级市场可交易标的包括股票、期货、期权都存在量价，考虑到目前截面上有 5000 多只股票，量价超额仍然丰厚。

基本面数据则是按照监管要求上市公司定期披露的财务数据，基本面研究门槛和获取价格相比量价都会较高，已经需要第三方人工整理，需要具备财务、会计等领域的基础知识。

量价和基本面数据权威、覆盖度广、获取极其便捷。便捷，首先意味着便宜，只需简单清洗即可运用到投资，其次意味着研究门槛较低，需要数学好或交易直觉强一些，以及编程和统计的技能，对理工科应届生来说都不是问题。

既然量价广泛、低廉、门槛低，就意味着一定是最先被研究机构获取、应用最为广泛，所以最早失效也是它。目前中国市场已经出现量价衰竭迹象，在研究单因子时就能明确看到。

AI 或者非线性模型广泛运用虽然延缓了这个趋势，让量价挖掘得更加充分。但从某种意义层面上来讲，例如金矿、石油等早期开采时，更多的是挖表层露天矿，现在需要更先进的技术和武器。同样我们现在需要更加先进的模型，使得量价方面挖掘得更充分。

而另类数据覆盖度有限，不同数据覆盖度存在差异性。获取也非常困难，需要由专门第三方数据公司第一时间获取、整理和清洗。它也不是定期数据，例如舆情、炒股社区上观点表达等。单个另类数据获取超额较低，但其逻辑非常朴素。

目前，很多机构会让应届生在量价因子方面练手，但如果让应届毕业生做另类因子，则无法取得非常好的结果，还需要学习大量知识才能研究，最后结果发现可能只能提升 1‰，这对研究人员信心打击较大，在这种情况下，提前布局另类数据需要一定勇气，毕竟获取数据成本相当高，而且还有一定的难度。

问：是否可以理解为一类数据获取成本、门槛与其稳定性、衰减速率有关，成反比？单个另类因子收益不高，所以研究另类数据是需要经历量变到质变过程？只有积累足够多数据才能在超额上有明显体现？

蒙玺李骥：我个人比较认同这个观点，人工智能与另类数据，这两类方法论，给大家感受度不完全相同。

坦率地说，现在整个行业对于非线性模型方法论的运用已经非常充分，人工智能包括购买算力等投入很多，但压力并不大。如果投入有差距，也可以迅速赶上，目前还没有掌握人工智能或者还在研究阶段的机构已经濒临淘汰。

我相信很多机构都具备做另类数据的能力，需要注意的是，数据获取门槛和研究门槛高，投入还是持续增量，第三方购买只是使用权，而不是所有权。而且，拥有 2023 年数据并不代表在不努力情况下，一定就一定能获得 2024 年数据，所以很容易得到结论是吃力不讨好。

这需要坚持不懈地投资，它不像非线性模型立竿见影，不论是回测还是实盘都能快速反馈。我们在业绩归因和分析时，明显发现另类数据的投入和产出性价比较低，回撤提升 2‰，对人刺激感很低。而非线性模型更换特征选择方式年化提升 2%-3%，一周甚至能产出 20 个与之前不同模型，几乎没有投入什么成本。

所以不能简单地从投入和产出维度评价另类价值，如果丧失耐心，就无法坚持这个方法论，而我们已经做好准备。

问：量化水平提升与科技水平有关。AI 在量化行业中有明显提升，是因为 AI 在学术界不断引入新思想和信息。另类数据出现是由于全社会数字化程度提高，采集数据能力提高，可以源源不断地将以前大家记录不下来的信息，转换成相对结构化数据，提供到量化行业中，某种程度上，为这个行业注入了新活力。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

蒙玺李骥：新信息提升行业生产力。我们感触非常深刻，无论是 AI 还是其他前沿科技在行业中都运用得非常充分。有研究 AI 的同事需要不定期翻阅前沿学术期刊和文献，才能保证竞争力。

另类数据也是如此，例如现在语音技术已经非常成熟，转化文本效率都在大幅提升。如果需要研究，就可以从全世界范围内寻求许多语音，并将其识别并纳入数据库，这是五六年前完全无法企及的事情。只有建立在科技进步基础上，数据获取才会更加便捷。

另类数据也在急速扩张，甚至可以作为一种生产资料来定义，具有非常多的价值，国家层面也意识到所以成立数据局，而数据局面对绝大部分数据都是非结构数据，即另类数据。我们只是站在前瞻角度，早期投入到数据维度上而已。

问：大家认为国外另类数据发展相对更早，因此国内对另类数据研究能力也会更强。从您观察来看，国内外量化行业在此使用情况如何？

蒙玺李骥：海外另类数据发展比较快，因为资本市场发展相对成熟，量价衰减不赚钱以后就研究基本面，但在过去十年基本面也逐渐不太赚钱。因此选择使用另类数据。也正因此，行业整体发展速度领先于国内数据机构。

近几年，许多国内数据供应商逐渐崭露头角，由于非线性模型获取量价超额仍然非常有效，大家推进另类数据发展急迫性就不会高，所以供应商发展力度不如海外那么大。

问：数据行业发展需要生态数据采集者和使用者，每个环节都必须让他们赚钱，有利可图，这个行业才能有长远发展。

蒙玺李骥：我非常认同这个观点。

问：您访谈中提到另类数据对策略稳定性有帮助，请您详细介绍。除稳定性之外，还有哪些方面有增益？

蒙玺李骥：随着科学技术进步，以前没有收集到或者被忽略掉的数据，现在也能够重新被运用。

首先，单个因子超额不够丰厚意味着大家不愿意花费更多成本获得它，因此竞争不激烈时，一定是稳定且持续的。

其次，而且另类数据维度非常广，维度广意味着单因子之间相关性低，当组合在一起能够提升稳定性。

最后，由于另类数据超额可以通过线性组合方式获取，无需过于担心非线性过强预测能力，导致潜在的风格暴露或非线性过拟合。

综合这三点，另类数据能够降低或者提供比较稳定的超额。

问：大家在分析或使用另类数据时存在痛点：一方面，它覆盖面相对较窄，每类数据仅覆盖特定行业或股票。另一方面，它可追溯历史相对较短。请问您在这方面是否有可以分享经验？

蒙玺李骥：覆盖面窄无法改变，因为有些行业天然不具备另类数据。

覆盖历史时间，其实比我们预期长，一些成熟优秀数据供应商早已经开始收集相关信息（但是没有针对性地处理）。我甚至不想用数据这两个字，它实际上是文本，我们也是购买直接文本，得到更长久的历史数据。

问：另类数据自己挖掘还是从数据商购买？

蒙玺李骥：让专业人做专业的事情。

我们非常乐意与全球数据供应商合作，目前 90%以上数据都是通过第三方提供，他们甚至惊讶于自己过往积累数据能够运用于二级市场交易，由我们自己整理并且使用 NLP 技术自定义处理，这能形成共赢合作。我们也尝试写一些接口抓取或者收集数据，但占比非常低。未来我们希望有一些是自己整理出来或者自己收集数据，从收集端开始参与。

问：如何看待另类数据在投资中未来应用前景，有哪些新趋势？

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

蒙玺李骥：我相信大家会越来越关注另类数据。一方面是量化经过过去的快速发展，大家工具箱、方法论已经趋同。大家挖掘深度和比拼细节维度越来越细，策略有效性下滑。虽然不至于对 AI 方法论过度悲观，在 1-3 年可见短期内方法论仍然超额丰厚，但如果看更长期，另类会像海外成熟市场一样，类似我们经历过的低延迟领域竞争，随着市场成熟，另类数据关注度一定会越来越高。

■图表 蒙玺另类数据运用在量化投资案例

数据种类	数据内容	数据处理	构建模型与参数
文本类数据	研报提取文本	对原始文本贴标签，用 NLP 技术生成结构化列表	生成特征表示模型
招聘数据	员工招聘信息	上市公司招聘数据，反映企业营收状态与经营状况	挖掘招聘数据因子
APP 数据	APP 下载次数、搜索率、点击率等	对下载次数、搜索率、点击率，分别建模预估	生成 APP 热度指数
航运数据	各港口货物进出情况	分析各公司产品货运量的变化和地点分布	生成产品销售指数
搜索引擎数据	对应上市公司或相关产品的搜索	根据搜索次数或频率，用机器学习生成搜索热度	对搜索热度去噪声

数据资料来源：蒙玺投资

03 低延迟：是情怀亦是图腾

问：行业内有说法，即国内交易策略执行与欧美存在差距，例如执行算法逻辑、交易软硬件系统以及网络方面，这是否符合实际？

蒙玺李骥：低延迟本身是非常窄细的领域，集中度相对其他策略更高。因为它对技术要求更全面，具备这个能力的机构较少，大家理解高频主要通过一些不够专业的观点或者臆想出来的状态来解，容易形成妖魔化或者不客观评价。

实际上，低延迟实际上有两个应用方向：1、高频策略交易；2、算法交易下单。

策略方面，高频交易非常中性，甚至具有正面效益效应。目前与海外高频机构有一定差距，但经过多年比拼，差距并没有大众想象的那么大，但也没有达到海外头部高频机构水平。

在算法交易下单方面，海外机构经历过人工手动交易研究，程序化交易再变成现在的 AI 程序化交易。国内则省略中间步骤，直接从手动到现在 AI 下单。第三方算法交易平提供了丰富的算法交易类型，提升了算法执行效率。

问：差距原因是海外发展得更早，还是有更丰厚的市场土壤？

蒙玺李骥：两个方面都有原因，主要是积累。

海外最头部高频交易机构，只需交易某几种标的就足够养活整个公司，甚至发展得不错。而中国机构由于利润极薄，衍生品相对海外市场品种更少、容量也较小，不足以养活太多技术人员。

两个市场差异的最终结果，就是海外头部机构经过丰厚土壤培育、多年技术和人才储备，导致国内外竞争不在同一维度。当然，这种差距正在变小。

一方面，中国市场给予了我们行业一些发展机会，我们也抓住了这些机会；另一方面，国内高频机构创始人相对年轻。大家都更考虑长期和未来，既然扎根在中国资本市场，就希望长期发展，情怀上而言，我们某种程度上在与海外机构竞争是打保卫战。

以低延迟策略为例，除了预测短期价格走势策略外，还有做市商为市场提供流动性策略，这种策略在一定程度上与金融安全挂钩。它可以让交易者或套期保值者用更低廉的价格套期保值，增强市场有效性，价值非常强大。如果我国的本土机构

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

不具备这个能力，那整个市场就会存在风险。目前中国具备高频能力的量化机构普遍都能做市，哪怕海外机构离开，国内量化机构仍然能为绝大部分交易标的提供流动性技术。

同时，国内大多数量化机构的从业者都非常年轻，有一腔热血，都想在各自领域做出成就，造福社会，报效国家。我本科就读于中科大，我们校训是“红专并进、理实交融”。我就所在的物理化学系，有课程讲我们第一届系主任、“两弹一星功勋奖章”获得者郭咏怀院士为何从海外归来。虽然我们自认为所从事的行业的社会价值与国防科研完全无法比拟，但我们依然想在所在的行业发光发热。很多量化机构创始人都有这种情怀。

因此，可以说，虽然海外头部机构在低延迟领域依然领先，但是这个差距肉眼可见在缩小。只要中国还有这个土壤和发展空间，未来就能在可预见范围内从追赶角色逐渐变成并驾齐驱，再逐渐发展到卷海外机构的状态。

问：如何看待市场现在对低延迟交易观点，以及哪些是真实，哪些是误解。

蒙玺李骥：整个低延迟领域，尤其是高频策略领域，观点普遍存在偏颇。因为领域太窄细，大众很难接触到这个领域。面对这些误解，很多管理人也会担心“如果我辩论，会否引起不必要问题”，因此大家选择沉默或者低调。

从业者角度来看，外界容易误解的主要有：

首先，低延迟并非仅通过财力、技术和设备等投入就能提升。还需要很多维度积累，尤其是策略端细节和技术设施层面的积累过程。

其次，大众容易误解高频稳赚不赔。因为样本点和交易信号足够多，高频交易更具统计意义。例如，如果收益分析价值投资，每 5 年拉出曲线都是盈利，而这只相当于高频的一天。实际上，目前由于竞争激烈，高频很难实现每天都盈利。最终能够具备低延迟交易能力的机构相对较少，所以市场份额较大，各家机构也能各分一分，只有最优秀机构能够生存。

第三，大众会有误区高频交易都是一个策略。其实高频策略应用维度很广，包括套利类、短周期类、预测类、做市类和流动性类。虽然生态很小，但是仍然是生态具备价值，其中做市策略提供流动性，短周期预测策略提供更多价格。

第四，低延迟另一应用就是算法交易。随着越来越多算法交易平台出现，普通大众也能享受到算法交易便利，变成一种服务，更具有普惠性。大众可以通过拆单方式降低对整个市场的冲击。正是因为高频算法的存在，操纵市场会需要付出很高的成本，使得不合规机构更不敢轻易扰动市场。

问：公众对高频交易还有交易公平性讨论。例如大家参加奥运会游泳比赛，认为一般人只是赤身游泳，而高频交易是吃违禁药品，或者有人站在高频交易这一方，表示高频是穿鲨鱼皮。您对此有何看法？

蒙玺李骥：奥运比赛对所有参赛者一视同仁，高频交易也是如此，国内很多具备低延迟能力的机构，创始人都是最早个人起家，只是在不断迭代和提升过程中发现了更好的方式。

当然，不排除市场因为不理解导致误会。所有国内机构都在监管下交易。作为从业者，我们一切以监管和规则为准，做好底线思维准备。哪怕整个行业不允许程序化交易介入。我们应该也不会关门，而是倒回到手动下单时代，信号仍然由服务器计算，我们对取得很好的业绩表现有信心。

举个例子，在火车发明初期，大家认为冒着黑烟太危险，不允许火车上路，必须回归到马车时代，所有开火车的和铺火车轨道的都要关闭。维持现状当然没有问题，但伤害的是整个中国资本市场生态。为什么说在回归马车时代后，铁路和炼铁都不会发展，因为它们都是产业链条。

问：量化本质上是工业化生产体系，即使在某个小环节，例如交易执行环节由于某些不可抗力最后关停或者限制。在其他信号数据处理和信号生成环节，量化作为方法论仍然具有优势。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

蒙玺李骥：是的，我们并非完全依靠程序取胜，而是依靠前端研究取胜，研究本质上是工程化过程。

问：超低延迟领域是否存在最优边界？

蒙玺李骥：低延迟并不是最优边界，而是相对优势。例如，有些机构将 FPGA 运用到低延迟维度。在最早时这些技术并不必要，就已经比别人快，但现在大家竞争越来越白热化。

前段时间，我看到一篇文章，非常有共鸣。文章主要讲述人工智能芯片，最早使用 AI 芯片或者显卡的主要买家是游戏行业，大量需求使得芯片供应者不断提升技能，导致算力提升促进大模型产生。科技发展是非线性的，这种循环上升也形成各类科技的迭代和提升。如果最初没有游戏需求，AI 芯片无法发展。

我们追求在技术上保持总是能快人一步，追求极致是我们的理念，这个过程中，必然会用到一些非常前沿的技术。最优不是取决于竞争到哪个层面。

问：因为技术是非线性，无法预知最优边界。我们能确定是大家都在不断努力。

蒙玺李骥：我个人非常喜欢竞技，不仅是奥运会，还有电子竞技等各种竞技。只要规则清晰、一视同仁，大家在同一起跑线上，就能不断推进极限。百米赛跑提升零点几秒，虽然看起来没有太大的实际价值，但是能让人受到鼓舞，因为提供了一种可能。

站在创始人和公司负责人角度来看，考虑商业层面因素，永远追求热点，会永远落后别人半步。大家追求各类极致，在追求极致过程中，推进边界外扩。竞技魅力在于原以为不行，结果又推进。人类进步状态是新技术不断迭代和产生，把不可能变为可能。

问：卷字是写在中国人基因里面。

蒙玺李骥：某些维度的卷，我觉得并非坏事，尤其是技术维度，能让我们看到不同的世界。

问：人类历史上战争最激烈，而战争期间科技也是进步最快。

蒙玺李骥：大家为生存一定会全力以赴。蒙玺在低延迟技术方面投入无止境，无论是人才还是技术维度，只要有价值的投入，我们都会评估，只要有优秀的人才，我们就愿意招揽。在低延迟维度，不断积累，做技术储备，是我们永远的追求，因为它不仅与收益、策略表现挂钩，更是从业者心中图腾般的存在。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

06 涵德投资：慎思之 明辨之 笃行之

审问之，慎思之，明辨之，笃行之。在和涵德投资交流时，时常可以感受到一种谦逊又笃定的态度，如何打磨投资策略，如何践行勤勉尽职，如何在关键时刻做出正确的抉择，涵德始终慎独且思辨。

秦志宇 总经理、合伙人

清华大学自动化系本科、硕士，拥有 10 余年国内外量化投资经验。曾供职于美国千禧管理及其量化团队，历任高级研究员、副总，研究美国股票期货市场的中频和高频量化交易，具体工作包括研发超过 60 个不同持有周期的统计套利模型。现专注于研究中国期货市场的中高频量化模型和投资组合优化方法，开发基于现有量化投资模型的优化系统。

顾小军 投资总监、合伙人

清华大学学士，拥有 10 余年国内外量化投资经验。曾供职于美国千禧管理及其量化团队，历任高级研究员、助理副总，主要研究欧洲和日本股票市场的中低频量化交易，同时深入理解不同数据源对选股的影响，建立不同的统计套利模型。现专注于研究中国期货市场的中低频量化模型，在自主开发的自动交易平台上研发交易执行算法，并建立完善的风险控制模型。

李维佳 研究总监、合伙人

清华大学自动化系本科，计算机系硕士，拥有 10 余年的国内外量化投资经验。曾效力于美国著名对冲基金 WorldQuant，历任研究顾问、全球研究总监、高级基金经理，带领全球超过 250 人团队开展全球市场股票、期货等资产的策略研究和组合优化。现担任涵德投资研究总监，全面负责公司股票策略研发管理。

01 投资之故 在德不在险

涵德投资，成立于 2013 年 8 月，公司始终秉承“投资之固，在德不在险”的理念和价值观，在严格的风控与合规尺度下，利用统计方法分析市场趋势和套利机会并建立交易策略获利。

问：请介绍涵德的创立故事？

涵德顾小军：人生很多时候是机缘巧合，你根本无法预知哪件事会对你的人生产生转折性影响，进入 WorldQuant 亦是如此：之前觉得高中参加竞赛用处不大，去了清华之后很多竞赛学的东西都用不上，没想到后来却在求职的时候发挥了作用（很多企业的笔试面试题和竞赛题的逻辑很接近），最终我选择接受 WorldQuant 抛来的橄榄枝，进入了量化投资行业。

创立涵德也是机缘巧合，我与志宇是 WorldQuant 北京办公室的同事，从 2007 年开始合作策略，一直同进同退。当时公司计划派我们前往美国担任基金经理，我们都不太愿意去，开玩笑地讲是因为难以忍受美国的食物，但其实也是因为公司当时的竞业约束很严苛。所以我们离开了 WorldQuant，创立涵德。

期货交易需要的资金体量比股票交易少一些，我们就从期货开始做起，从 2011 年初推出沪深 300 股指期货便开始交易。2013 年建立公司，当时量化公司都走 CTA 叠加 Alpha 以及做空沪深 300 股指期货的路，公司管理规模迅速达到十几亿。2015 年股灾当头一棒无法再进行对冲，导致很多产品出现问题，一直到 2017 年很多产品处于半空仓状态。资金是有成本的，而且当时固收都能做到年化 12%，于是 2017 年下半年我们在 CTA 上面做了比较大的优化调整，随后一直到 2020 年逐步得到业界认可，拿到了金牛奖。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

2021 年，维佳加入涵德。涵德也从 CTA 驱动变成 CTA+股票双驱动的状态，比仅凭一条腿走路更为有效，辅之自营期权项目，多样化的策略带来了不同的收益来源。

问：2011 年国内做量化的公司并不多，甚至没有私募备案，当时你们从 WorldQuant 离开时是否考虑过其他选择？您之前在 WorldQuant 研究欧洲、日本股票的中低频量化交易，其实 DNA 里面是有股票的，为什么一开始没有两条腿走路？

涵德顾小军：回过头看无论是公司战略还是研究，我们都走过一些弯路，这些弯路也让我们成长。最初我们根本没有考虑过再找一家公司工作，也没有像别人那样规划全面。自己在 CTA 领域做得比较成功，就主观地认为人少也能把股票策略做好。实际上导致了早期在股票的积累不深，发展也有滞后，在股票策略上的投入也陷入了负反馈，更不敢继续扩大投入。维佳的加入是一个转折点，他在美国的成功经验证明了他在股票策略上很强的盈利能力，和维佳的合作打破了投入产出比不佳的负循环。

我们后来反思，创业最关键的是保持开放的心态，和大家多多互动。在策略发展、公司管理等方面可以有更多启发，也许会少走一些弯路。

问：量化是交易自己对世界的认知，涵德有什么投资认知总结？

涵德顾小军：涵德一直强调“投资之固，在德不在险”。先强调“德”之后再强调技术。

早年整个资管、私募以及量化行业缺乏监管，私募基金业协会成立之后才逐步走上正轨。那时借通道是很正常的，不巧第一次出于帮忙性质的借通道就遇到通道产品爆雷。公司内部中层以上干部同事讨论过之后达成共识，选择从正确而非自身利益的角度，及时向监管机构完整地汇报了状况以及涵德在其中的角色。我们甚至激烈地辩论过是否要把这件事情告诉其他客户，因为客户知晓后很有可能选择赎回，但我们义无反顾地在市场出现消息之前就如实披露给投资人。事实是客户赎回了一半，但大部分客户在风波停息后又重新投资甚至扩大投资。正是这件事情让所有客户认识到涵德是一家靠谱的基金，可以放心把钱交给我们。

亚马逊创始人贝佐斯有句名言：“一切关乎长期”，欧美机构例如桥水成立将近 50 年，千禧年也有 30 多年了。客户能够长期信任一家基金是很不容易的，一定是因为公司本身立足长期布局，足够靠谱和值得信任。

问：你们在取涵德名字的时候是出于什么考虑？

涵德顾小军：这有一定的偶然因素，因为公司早期几个成员都来自清华，清华有个亭子叫涵德园，故以此取名。很多时候冥冥之中自有安排，这个名字也和公司整体的风格和理念非常契合。

问：慈不掌兵、义不理财，过于讲义气或者在客户出现亏损时有太多内疚心对于投资而言是否是一件好事？

涵德顾小军：人非神仙，总有能力边界。我们在这方面的理解是抛开内疚，多问自己是否做到了应该做的事情，是否足够勤勉尽职。

问：常常有新闻说量化来了、量化杀游资、量化割韭菜，2023 年底对于量化的批评很多，甚至掀起投资人骂战，您认为要怎样面对这些抨击？

涵德顾小军：应对这样的舆论和市场环境，不是一家公司的事情，而是整个行业的事情。

我相信适者生存，竞争的基础是良善的监管，不能将市场变成黑暗森林。黑暗森林没有规则，而市场应该有一定的规则，而且最好不要频繁变换。当然随着市场发展和认知变化，规则也应该不断演化。例如基金业协会从 0 到 1 的建立，量化开启托管等，我认为都是很好的。另一方面，市场是持续博弈的，你想赚市场的钱，别人也想赚你的钱，这时候就是谁有能力谁赚钱。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

问：普通投资人在炒股上的决策，很多时候要比人生中其他决策更加盲目。

涵德顾小军：是的。我经常强调心态，是因为金融行业与其他行业非常不同，练习不一定能够得到正反馈。它不像技能型职业，经验增加技能也会不断提升。很多时候，大家做了正确的决策但赔了钱于是抛弃正确决策，错误决策反而赚了钱，于是把错误理论当作自己的炒股原则，最后越炒越差。投资是有门槛的，而且门槛很高。

02 涵德策略 多维数据

涵德秉持“控制风险，实现稳定收益”的投资理念，涵德投资方法以量化研究为核心，并结合中国市场。研究基于股票、期货市场价格、交易量、开仓量、波动性、基本面数据、新闻数据和交易者的行为特点等，在基础理论之上通过量化建模和数据分析，利用数学模型的统计优势发现市场交易机会，全自动、系统化地交易国内股票、期货及其他金融工具。

通过多策略、多品种投资有效降低下跌风险，追求风险收益比的最优回报。

问：您曾说到，中国 A 股市场的中期波动性呈现系统性收敛和结构性发散，请问如何理解？

涵德顾小军：A 股市场现在的交易对象在不断变化，但总的来说会朝着美股的方向走，相当于整体会越来越收敛。在过程中又会有大量的仓位调整，这又是一个波动加剧的过程，包括会有一些机构被淘汰出局，形成市场冲击。

现在市场上的策略同质性比较严重，归根结底是因为客户愿意给的时间有限，如果客户愿意给五年的时间，就可以研究一些相对长期的交易策略，整个市场也会百花齐放。

问：涵德进阶方向聚焦于多维数据，目前多维主要在怎么用？大家对 WorldQuant 最熟悉的 Alpha101 体系，其实更偏向中低频量价。

涵德顾小军：对，这和 Igor Tulchinsky（WorldQuant 创始人）的经历有关。他在最初入行时，认为量价永远做不完，所以早年它只做量价数据。2007 年美国市场同质化特别严重，量价几乎全部失效，这时 Igor 的理念发生巨大变化。当时公司一位基金经理提议进行基本面测算，数据测算结果表明基本面的回撤小很多，于是 Igor 便雷厉风行立刻进行大转向，每年花费大量资金购买所有可用数据。

我们也是这样的理念，在发展过程中不断进阶。物竞天择，适者生存，很多因子都在慢慢消失。因为随着交易增多，有些因子有效性在不断下降，整个市场正在变得越来越有效。例如股票上有个方法是连涨 5 天就做空，连跌 5 天就做多，我们刚入行时这在欧美都非常有效，2017 年之前在国内也挺有效。这种方法非常简单，但 WorldQuant 放出 Alpha101 之后大家就都会了。我们最初只有量价，从日频数据不断扩充成熟的基本面数据，分析师一致预期数据等。

不同数据会带来不同观点，量价、基本面、另类数据的逻辑完全不同，而获取数据的资金成本在某种程度上设立了门槛，10 亿以下的机构真的没有太多资金购买数据。长期看，数据一定会越来越有优势。目前国内的 Wind 处于较垄断的地位，但它的精细度不如 Bloomberg，Bloomberg 会把标题以及事件对行业和公司的影响范围整理得非常清晰，而国内数据在购买后会多一个清洗难题。

目前涵德股票因子大约在 3500-5000 量级，其中量价占 60%-70%，20%左右是基本面信息，剩下部分是机器学习。目前大家都在大力发展深度学习和算力，我们在这一领域发展并不深入，一方面是这需要投入大量资金，另一方面我们对这一领域还存有困惑。人工找因子可以解释每个因子赚的是什么钱，但机器学习之后变成了黑箱。量化因子还有个特点是一段时间的有效并不代表长期真正有效。

问：公司股票策略采用机器学习时间稍晚，主要是什么原因？

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

涵德顾小军：也是在战略上判断有误，尤其是 2018 年至 2020 年四年间核心团队就四个人，我们并不觉得需要更多人，也没有更多的人做别的事情，是维佳的加入打破了这个状态。

问：目前人工智能部分应用主要集中在哪些环节？

涵德顾小军：目前两个大块的应用相对较多。一个是在开发，现在有一些模块可以直接让它帮忙实现，稍微调整一下就可以直接用，大幅提升了效率，这是这两年一个比较大的变化。入行的时候工作对于编程能力有比较高的要求，但现在要求会相应下降一些，只要能够想到好的想法让它实现就可以。另外一个是通过分析新闻资讯获取信号，这部分目前还在探索阶段，现在积累的数据还有待完善。

问：怎么看待国内头部机构疯狂囤积算力拓展 AI 的行为？

涵德顾小军：大家都很聪明，很多机构也都很用心在做，所以国内量化基金公司能百花齐放。这些决策对他们来说是最合理的决策。头部机构最担心的是没有掌握未来产生颠覆性影响的资源，虽然不确定哪个方向能赢，但是他们有足够多的资源提前布局。

因为我们资源有限无法这样做，这时候就要进行微操，看好哪个赛道就在哪个方向投入更多。正如前面提到的，对于机器学习的不可解释性有些担忧，而且如果把所有的资源都 all in，一旦失败会满盘皆输，所以这是不同公司基于自己现状做出的最佳决策。

问：我在贵公司材料中看到高管薪酬因子，这部分能够详细介绍吗？

涵德顾小军：这种单因子并不复杂，逻辑很简单，高管拿钱越少公司就越好，因为留下的钱越多。但也有边界，高管不能拿太少，否则高管对公司的投入会很少。当然这些底层因子也需要和其他数据相结合，例如影响公司稳定性的高管离职率。这种特质因子的特点是稀疏，不像量价因子每天都有很多数据，基本面每季度也有数据，这种数据一年甚至几年都不变，就需要结合其他因素考虑，背后要有强大逻辑驱动。

问：一些 FOF 认为量化私募在另类因子的探索并没有增强其收益，只是降低其回撤。

涵德顾小军：投资最终目的是为客户盈利，长期发展的大前提是能影响客户买入和离开的时间。客户很难接受大回撤，所以降低回撤是非常有价值的事情。客户很多时候并不能坐下来慢慢探讨，他有自己要面对的现实因素，驱动做出非理性决策。所以对客户来说收益相对平缓但回撤较小的产品是要比 20%收益+10%回撤的产品更好。

问：新闻谈到另类因子时，都会将其神化，例如沃尔玛楼顶停车之类，它是裸露的 Alpha，只要能获取数据就可以得到信息，但实际量化在处理另类数据时因为涉及大量数据信息，其实会涉及很多冗长、琐碎的工作。

涵德顾小军：首先，这一类故事更多是讲给外行听的。沃尔玛停车数据只和沃尔玛有关，影响因子会很低，量化更倾向于找整个行业共有的数据。类似于京东和淘宝销售额数据，代表了公司在一段时间内真实营收的数据，才是真正的另类数据，这类数据比财报快，会有一些优势。

其次，另类数据确实会有大量的清洗工作，例如新闻的核心点在于发布时间，但很多数据是没有时间标签的，这就需要花费大量时间搜集。再次另类数据是否真实，因为很多数据商不指望从数据里面挣很多钱，所以数据填错也没人管。这应了那句话，外行看热闹，内行看门道。

问：不仅是另类数据，有时候量价数据也会丢包。

涵德顾小军：对，最开始处理量价数据也会有问题。2017 年茅台价格还不像现在这样，当时定的规矩是超过 1000 全是错的数据。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

问：另类数据很多时候会缺数据，目前海外流行使用高斯混合填充数据。

涵德顾小军：各家会有自己的方法填写缺失数据。量化和主观底层逻辑是一样的，都是用信息辅助判断，首先考虑真假，其次是对错。只是主观的管理人在自己脑子里分析，量化管理人把自己的逻辑写成程序，让程序来分析。

问：从 WorldQuant 出来创业的机构已经有八九家了，大家在投研方法论上有区别吗，涵德的特色是什么？

涵德顾小军：可以笼统地这么说，量化底层逻辑或者本质理念都没有本质区别，就是怎样用历史来拟合，减少一些过拟合，差异是如何对细节进行优化，实际上大家在这套体系下在不断成长和发展。

问：所有行业如同生命树一样，它们起源于某个地方，最初累积的知识图谱高度相似，但随着时间推移，大家会走向不同路径，差异化越来越大，现在量化领域，大家是否已经走向不同方向？

涵德顾小军：量化从黄金年代到白银年代，未来会发展到欧美阶段，头部的几家各有千秋。

问：近几年宏观因素主导市场，接下来会采取什么策略应对持续突变的市场？

涵德顾小军：我们不会主动把方向限制掉，每套方案都会有最适应的环境，环境会实时变化，不能束缚自己。在这个过程中有两个比较关键的因素，资源与人才两者结合。

人才可以外部寻找，还有内部自我生长。人的潜力是无限大的，在内部给真正有能力的人提供更好的环境让他们发挥潜力，你想象不到他们能给你创造什么样的奇迹。我们在外部也会多方寻找，无论是自己招还是从猎头等等各种渠道找到靠谱的人才。

03 存亡之际 主观决策与“一键清仓”

涵德采用自主化量化投资系统。系统包含投资研究平台、交易管理平台、绩效分析系统和风险监控系统，能够实现程序化、系统化的量化交易执行，从市场数据、信号生成、交易执行的全流程应用。

公司风控主要从模型风险、交易风险、流程管理风险三个维度进行。模型风险控制是在策略研发的过程中，将风控指标内嵌于系统中，如交易制度要求、仓位要求等。若不符合风控指标，则直接淘汰策略。交易风险控制是对交易数据、交易系统、交易品种等全方位控制，在极端情况下可“一键清仓”。流程管理风险是对交易计划、账户资金、交易结果、投资决策、权限划分等流程进行严格监控。

问：这些年您在投资上有印象深刻的事件吗？

涵德顾小军：首先是细节决定成败。刚入行的时候就是给老板打工，提供 Alpha 就可以。后来有次导师问我，为什么改了参数之后结果不升反降呢？当时我的想法是策略满足标准就直接送上，具体怎么用是老板的事情，但那次对话让我触动很深，明白了研究究竟是什么，不能浮于表面，要深入细节，这对我们后来创业的影响很大。

第二个关键点是风控。2006 年我们入行，2007 年美国经历了量化崩塌，WorldQuant 老板 Igor Tulchinsky 迅速砍仓，三天内把所有仓位全部清完，两周之内又全部加回来，决策又快又准。后来我们自己交易后发现这样的决策很难，以最近几件事情为例：

例如 2022 年的镍，我们在不断涨停过程中很纠结是继续持有还是砍掉，因为当时已经涨停三连板，交易所会三板强平，后来我们选择中庸方法，砍一半，结果当天晚上开始跌停。这已经是在当时能做出的最好判断。我们虽然赚了但是收益其实可以更好，资本市场赚钱总有机会，但一次大亏损你就没有任何机会了。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

再如 2023 年下半年欧线指数五连涨，我们也陷入纠结是否要继续交易，认真思考之后选择放弃，因为当时每天只有半个小时是正常交易，剩下三个小时都是封板状态。量化策略的信号在这种情况下已经被扭曲，只能靠运气。

再如近两天（2023 年 1 月）市场也遇到类似情况，整个市场非常混乱，随着股市下跌雪球敲入会导致股指不断卖出，进而导致基差拉大，同行开始抢跑，最后都亏钱。现在遇到问题不是要谈策略，而是如何处理，经历了这些才感受到 Igor Tulchinsky 当时的决策不是一般人能做的。

问：Millennium、Citadel 这些机构都会要求风险因子敞口必须是 0，异质性因子越高越好。你们在实际操作时真可以到达因子敞口为 0 的状态吗？

涵德顾小军：因子这一层我们允许暴露一些敞口，有的为正暴露，有的为负暴露，但在最后因子交易组合中我们会要求敞口为 0，直接抵消所有暴露。

敞口可以带来收益，但是也带来波动。回到刚才的话题，为什么国内要暴露小盘，Millennium、WorldQuant 没有，就是因为暴露可以赚取更多钱，但是会有潜在的更大风险，就像我们最近遇到的行业波动一样。客户有时看不到我们为了控制风险做的努力，只能看阶段性收益来评价为什么你们只有 8%、9%的收益，有些机构可以做到 15%。

问：在量化决策的时候是否也会有一些主观色彩？

涵德顾小军：我们团队也探讨过这一部分，我们认为关键是主观在此阶段是承担风险还是控制风险。2015 年股灾时期，整个股票市场和期货市场全部跌停。那时候我们思考要不要赌一把，但我们觉得自己没有优势，随后就清空了所有期货。

问：量化还有很好的一点是，这里面是最拥抱变化的一群人。

涵德顾小军：这与人息息相关，是终身学习还是形成一套理念沿用一生。

我是 80 后，虽然小时候相对辛苦，但是整个人生经历都在上升。2018 年之后贸易战、俄乌战争、巴以冲突以及去年的 AI，整个世界正在剧烈变化，要活下去就不能再抱着过去的观念。无论是生态环境还是社会环境，都在不断演化，我们也要根据演化的进程不断调整。监管要求越来越高，如果遵循过去的方式不愿改变，反而会被它缚住手脚。

问：WorldQuant 以前对风控非常严格，例如回撤超过 5%可能就要面临被砍掉的危险。涵德是借用了这一套体系还是有不同之处？

涵德顾小军：国内外有很多不同之处，最大的差异在于投资者。虽然近两年经过市场教育，P2P 爆雷和固收收益率下降等，使得大家开始对量化、主观的宽容度增加，但在国内每个产品背后都是有着独特需求的投资者，所以更多时候是匹配投资者收益风险偏好。

问：涵德在整个市场中募资一直是一个很稳健的公司，这是否与你们的发展理念或者前段时间业绩表现有很大关联性？

涵德顾小军：我们还是从投资理念出发，秉持的底层理念是客户赚钱才能赚钱。在此大前提下，首先会对未来能否盈利做出判断，再募资。我们希望涵德经过一两轮牛熊之后，能够树立起“值得客户信任”的标签。大盘涨的时候很多人都想要进场，这是顺人性的，往前一步类似于定投无论大盘好还是差都会购买；再进一层是逆人性的，例如在低位或回撤的时候敢买涵德，这需要很长时间的坚持才能让大家认可，但我们愿意给自己定个高远的值得追求的目标。

“高山仰止，景行行止，虽不能至，心向往之。”我们勉励自己和公司的同事不断努力，希望后续可以做得更好，给投资人和合作伙伴创造持续稳健的收益。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

08 平方和：量化策略的匠心和价值

平方和：一种统计量，用于测量数据集的观测值在均值附近的变异性，是每个数据点与均值的平方差的累积总和。

量化也像这样，遵守基本理念，留点差异，形成平衡，最后成为自己的投资之道。



01 从互联网到量化一次典型转身

已经进入量化行业 15 年，虽然穿着西装，但程序员的特质仍然嵌刻在平方和投资创始人吕杰勇身体里。他言语朴素但直面重点，有着 IT 人特有的认真，他认为自己社交能力不行：“这就是能力圈的事情。”

社交天赋不足，但技术能力点满了。吕杰勇从小学习优异，在北京大学数学系取得本科学位之后，又继续攻读了计算机硕士。毕业后先后在搜狐、腾讯从事网络大数据的数据建模和挖掘工作，负责过搜狗输入法项目与腾讯智能 AI 小 Q 机器人。

在他看来认为，输入法软件是让计算机从所有可能符合拼音条件的汉字空间中寻找最优的词句，而量化投资就是让计算机从全市场四千多个股票中寻找最优的几百个股票，其实这些问题都是空间寻优问题，只是场景不同。

而且国内最早没有量化人才储备，与互联网人才重合度高，其实就是同一拨人一部分转到金融，另外一部分还是在互联网，两个行业技术上还是相通的，也不算是很大的转变，都是基于机器学习进行统计，找到规律进行预测，没什么本质不同。

差异更多体现在。例如互联网在初期强调跑马圈地，打出概念后快速扩张，实际上管理落后、做事粗糙、容错率更高，主要依靠商业模式支撑，互联网门槛主要由平台带来，例如 QQ 只能由腾讯完成。

而金融行业是精雕细琢，强调厚积薄发。行业从业者平均年龄也是大于互联网，需要多积累几年才能发力，管理水平反而更加领先，竞争更加激烈。

2009 年吕杰勇转战公募基金，正式加入量化投资行业，先后任职于博时基金和中信证券，从事量化投资研究和公募基金专户、券商自营账户管理工作。彼时中国刚刚推出 4 万亿财政补贴计划，金融市场波澜壮阔，风光正好，而其中量化领域也正处于蓬勃发展的时期，很多华尔街的量化人才来到中国工作、扎根。在积累多年经验后，2015 年吕杰勇创建平方和投资，次年规模即突破 10 亿。

从 10 亿到如今 70 亿，不只是一个数字的变化，其中所经历的风浪，如人饮水。

2017 年后量化市场很快进入“地狱模式”，漂亮 50 行情打破了小盘自带 alpha 的天然认知，市场上以中性策略为主，Alpha 必须扛过负基差才有正收益，负基差如果特别大，或者 Alpha 不够高，就很有可能会亏钱。这个阶段里，市场对量化的预期达到了冰点，觉得量化整体要受到冲击、退出中国。与此同时，市场上出现了“真 Alpha”的呼声，所谓真 Alpha 就是必须控制住大小盘、行业的风险，并且获得一个稳健的高水平超额收益。 2017 年后，市场上对量化策略“谈风格色变”，一旦有人说某些量化策略暴露了小票或者某种风格，大家就会感到非常恐惧。这种恐惧持续了很多年才逐渐消退。

“市场不断变化，不同阶段对同样一个动作或事件的理解也会发生变化。在这种情况下，需要不断地迭代模：一是因为市场变化可能会导致因子失效；二是赛道越来越拥挤，同行之间的激烈竞争会倒逼量化私募提升模型的能力，为了保持竞争力必须不断地更新和优化模型。”平方和迭代的具体方向包括数据、系统、策略和算力四个方面。数据方面，不断更新和拓展数据广度和深度，以更好地反映市场的变化；系统方面，追求更高效、更

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

易用的特点，以提高策略开发和管理的效率；策略方面，不断探索和创新，以提供更有效、更多样的策略选择；算力方面，保证算力足够强大，但同时也需要做到相应的匹配，算力太多也可能带来成本浪费。”

02 行业思辨：量化策略的三个问题以及解决之道

2023 年，吕杰勇以《量化策略的匠心和价值》分享了当前对于量化行业的思辨。

■图表 量化策略的三个问题

策略问题	技术提升	产品设计
超额下降	超额下降倒逼技术进步	个股多空策略能显著增加收益
量化策略同质化现象明显	同涨同跌问题似是而非	打新红利仍然存在
规模容量的影响	策略容量魔咒可以破解	增强策略的配置

数据来源：平方和投资

在他看来，量化行业现在面临三个问题：第一，长期来看，整个行业的超额收益持续下跌，A 股市场的超额水平在 2015 年达到高峰之后持续衰减。第二，量化策略之间的同质化现象明显。例如 2022 年大量指数增强产品呈现出负收益，中性产品在 2021 年四季度集体回撤。第三，量化策略存在规模容量问题。随着市场规模突破万亿，很多策略的容量上限被撑破，规模上升可能对策略造成负面影响，策略的降频和投资组合的业绩下滑，导致了大规模的赎回。

长期来看，超额呈现下降趋势，但是变化缓慢且不会消失。技术进步会延缓超额下降，同时让量化投资机构技术集中。超额的下降倒逼了技术的进步，而技术进步可以延缓超额下降，对一些专注技术的量化机构，能够获得更多的先机，其他实力相对较弱的团队，则可能面临一些挑战。

对同质性进行分析。首先，把产品本身收益进行拆分比较。对一个指数增强产品来说，要先把 Beta 因素拿掉，来考察超额部分；对中性产品来说，要把基差这样的主要干扰因素剔除，来讨论策略之间的差异性。比较超额可以发现，不同频段、不同策略的 Alpha 会有相对较低的相关性，高频和中低频的 Alpha 策略，其涨跌路径是不一样的，依赖的条件也不一样。通常来说，当市场比较低迷的时候，做波动率的策略会更加艰难；当市场出现垃圾股行情的时候，基本面策略会比较困难。

市场平均水平与技术领先的管理人会有较大差距。策略的容量、稳定性、收益率这三个因素互相制约、此消彼涨，单个机构或者单个策略的容量有限，规模增加，管理难度增加。这意味着，如果一个策略的容量超过了限制，收益就会降低，而收益降低又会导致规模减小，因此需要合理预期、长期投资。

“量化策略的未来一方面需要不断优化策略本身；另一方面也需要注重产品的设计”。对于未来，吕杰勇认为，2017 年量化私募破千亿，在证券类私募中占比达 4.8%；2021 年量化私募破万亿，占比达到 20%，这是投资人用脚投票的结果。但相比美国 70%的占比，中国市场规模仍有较大提升空间。

中国量化策略仍有较长的红利周期。放眼十年以上的长周期，鲜有资产可以在收益率和战胜通胀方面取代权益类资产；中国地产投资性价比下降，非标资产规模压降，增量资金大部分将流入权益资产；A 股市场无效波动较多，基于 A 股市场的量化策略将具有长期投资价值。指数增强策略是指数投资的利器，投资人在国家指数慢牛这一逻辑的指引下，选择指数增强类的策略工具，同时获取 A 股比较显著的超额收益和长期价值。量化策略整体换手率较高，带来了收益率的补偿。比较分散的持仓，也起到了降低风险和波动的效果。

中性产品也有很强的配置价值。截至 2022 年底，市场上的量化产品中，中性产品占总数的 25%。股指期货容量稳步增长。早期沪深 300 股指期货的出现为中国量化对冲产品奠定了基础，到 2015 年中证 500 股指期货的出现进一步提升了量化中性产品的收益质量。2022 年，中证 1000 的股指期货的推出为投资者提供了更多的选择。在股灾后，股指期货持仓量一直在增加。在股指期货中，主要的参与者包括套保策略和 Beta+策略。在负基差

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

情况下，买股指期货相当于获得免费的超额收益，免费的 Alpha。而 Beta+策略的杠杆优势也使得其规模大增。此外，近年来盛行的“雪球”等策略也让股指期货市场越来越活跃，负基差越来越小，越来越适合对冲。

■图表 股指期货容量



数据来源：平方和投资

03 不自囿于方寸 关键是 10 年内坚持观点不动摇

平方和特点是全面，不自囿于任何一种方法派。

“作为老牌机构，我们不仅资历老，而且积累时间较长且全面，基本面和量价都非常充分，没有任何想法或者想要走向哪个极端。”吕杰勇明白投资是分散再分散，营销则是集中再集中。从营销角度来看，机构风格需要越极端越好，这样别人才能辨识。但从技术角度来看，反而越分散越好，所以平方和虽然没有任何标签，机器学习、高频、低频以及基本面价量成分都非常大。

公司成立之初即布局基本面因子研究，目前因子库有 3000 多个因子，实盘因子在 1000 个左右。其中量价因子占比约 50%，基本面类占比约 40%，另类因子占比约 10%。各类因子的比例会随市场行情的变化而调整。因子挖掘方面，以人工挖掘为主，占比 70%以上。公司不卷高频，不过多依赖量价因子，重视模型稳定性。平均持仓 800-1000 只，不含 T0 的情况下，双边换手年化 60-80 倍。基本面的提早布局使得平方和的策略包容性与稳定性都更强，业绩方面也能保持一定优势，与其他管理人的策略相关性较低。

目前公司为不同风险收益偏好的投资人提供四大产品体系，但平方和实际以中性策略起家也以中性见长。而指增其实作为扩大规模的大杀器，能更迅速地完成跑马圈地，吕杰勇自认为平方还没上百亿的原因之一，就是在关键时期发展期做得不好，“之前我们是不太好沟通的机构，也没有别的原因，就是不会、不擅长。”

回归到资产配置的本质，找合适的投资者、合适的资金以及合适的时机。Alpha 做得很一般但是规模很大的机构很多，这类公司主要做 2C，要求服务好，产品不能太差，类似麦当劳口感并不好，但是是成功商业项目。但 2B 是要求产品要好。指增带来“跑马圈地”是时代造就的，投资者对指增有需求，部分机构抓住了趋势并且满足了需求，从技术角度来看，大家有不同投资理念，不能说哪一种产品更值得借鉴或者更先进。量化的本分是把超额做出来，保持稳定，容量要大，商业上的成功是另外的维度。

从资产配置的角度来看，量化中性产品以其“低回撤+稳健回报”的特性，通常被称为“类固收”产品，一直是防守性品种的重要代表，也是平方和的主要赛道，即便 2022-2023 年中性产品跑赢市场，但市场上量化中性产品并不多见，归结原因，无外乎有两点：一是主观上不想做了，二是客观上可能也做不好了。管理人发行量化中性产品，后端其实只能获得 Alpha 减去贴水的分成收入，这跟指数增强产品可以同时获得 Alpha+Beta 的分成收入相比显然偏低。

与此同时，部分量化管理人的策略天然不能很好跟踪指数或者因为规模扩张等因素导致跟踪指数效果变差，那么这部分量化管理人确实也比较难在量化中性产品上有好的收益。自然，在付出更多反而获得更少、策略挑战更大的权衡下，这类量化管理人会最终不得不放弃中性产品，其实也算是理性选择，背后逻辑可以理解。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

■图表 量化中性产品的优势

收益稳定	与市场涨跌的相关性非常低，追求长期稳定的绝对回报。	风险较低	做空机制使风险程度可控，尤其是下行风险较小； 因子风控模型把控其他维度的风险。
可持续	获利主要取决于量化选股能力，无需依靠市场择时。	少犯错	通过量化投资方式，克服人性贪婪与恐惧的弱点。

数据资料来源：平方和投资，中金公司财富服务中心

2023 年讨论超额稳定性必然涉及敞口，尤其是小市值的暴露，在稳定性和超额高之间是否存在博弈或者取舍?

吕杰勇认为这个描述不准确，将超额和敞口等价直接抹杀 Alpha 上差异。这种论断的前提是 Alpha 无差异，收益都来源于敞口。通常市场带来趋势性机会类似于物竞天择，策略表现良好企业才能够生存，有些策略会逐渐消亡，有些策略则继续活跃。趋势来临时有的人能够抓住，短期内会受到赞誉。但在趋势反转后经过一个周期还能活着的人少之又少，经历了上行周期，而向下周期则通常无法承受。环境发生变化，策略适应性也会发生变化，海外投资者通常关注多年策略表现，而国内投资者则通过观察三个月或者近一年表现便作判断，这并不合理但又有其特殊合理性。另外，大部分量化私募都是因为策略容量不够，从高频逐渐增加基本面因子，降频至中低频。至今，大部分高频起家的量化私募，基本面因子的主要作用是扩大容量，而非贡献收益。

平方和的风控自评只能叫做的正常，主要控制 Barra 风险因子以及行业暴露。“但关键是能够在 10 年内坚持观点不动摇，是否长期坚持，是否坚决坚持，尤其在极端环境下，这非常难得。”在吕杰勇看来，在风控这条路上如果没有坚持住就会出现问题，公司不尝试、不采纳冒进方案，例如 Beta 择时，无论是考虑到投资者收益还是自己的稳定性，最终都没有做。

在量化不友好的时间点、艰难时刻，依然坚持风控，是一种结果，但本质更来源于一种理念。吕杰勇以 2020 年为例，白酒和消费一路大涨，后来以春节为分界线，急转直下。吕杰勇个人也想拥抱白酒，但也没有额外暴露，核心的理念在于考虑公司是想长期发展还是短期赚快钱。其次在于实力决定，策略不赚钱时没办法考虑风控只能赌，并没有选择余地，而吕杰勇对公司 Alpha 还是有一定信心的。

海外 Barra 并非必需品，但在国内目前成为通用模型，是管理人和投资人的心理安慰剂，还是能切实做好风控? 平方和吕杰勇认为许多风险因子已经失辨识度，但影响不大，能够解释风险的就是几个主要的因子，Size 风格因子是主要的风险来源，能把它解决好大部分的问题也就解决了，其次风控并非越严格越好，海外市场波动更大，有更多、更精细化对冲手段，国内市场只有少量股指期货，无法进行精细对冲，同时 Alpha 本身有效造就了客观情况。

有的要改革这个行业，把机器学习吹得神乎其神，每个概念都包装得完美，但是本质上没什么变化，平方和在机器学习方面下的成本也很多，但是并没有把机器学习当作一个噱头。

Alpha2Fund 是平方和投资的英文名，在创始之初选择这个名字，寄托着对公司投资团队寻求更多更好 Alpha 的希望，与利用机器学习不断自我提升的发展愿景。众所周知，机器学习能有效提升空间寻优搜索的效率，甚至它的非线性学习能力可以成为人类思维的补充或提升。

“平方和并非旧事物的守护者，热烈拥抱新事物。从技术角度来看，技术不分优劣，只分有效与无效。机器学习本身并不是一个革命性的事物，更多是锦上添花，与原有基础结合应用效果更好。”吕杰勇说，站在投资角度，而非营销角度来考虑，机器学习是很好的工具，它在平方和所占比重很大，目前深度神经网络主要运用在因子组合上，能够增加预测精度，在 2022 年的产品业绩上有所贡献，并且未来比重会加大。

平方和现在同时使用机器学习算法与传统算法，这并非汰弱留强的问题，而是会选择叠加，以求达到策略更好、更高、更稳定的结果。因此这个时候就谈不上谁要革谁的命，而是互补关系。目前平方和主要将人工智能应用于预测，因子挖掘和风控则会更为慎重，因为黑盒风险较大，风控方面有所探索但作用不大。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

04 特殊时期的基差管理 以不变应万变

量化中性产品通过数量化模型选择一篮子股票构成现货多头，并通过股指期货等工具持有与现货头寸相反方向的空头，实现较好的对冲效果，以获取与市场波动无关的较为稳健的超额收益。尤其是在弱势市场环境中，空头可以起到有效的下跌保护作用，同时有机会获取一定的正收益。但在对冲的过程中，需要承担基差或券息。

1 月市场关注中性产品增多，一方面市场在竞争 Alpha，另一方面基差因为雪球受到较多挫伤。作为一家中性产品为主量化管理人，平方和吕杰勇追求一种稳。

吕杰勇复盘了 1 月的市场及应对：雪球杀跌会带来市场急跌，紧接着国家队进场救市，主要聚焦沪深 300 之类的大票 ETF，此时市场风格发生变化，主要表现为小盘股下跌。历史上每次出现救市的情况，量化策略表现都一般。短期内会面临 Beta 下跌和 Alpha 下跌两个风险，但 Beta 方面量化中性产品有对冲，因此市场下跌的影响不大。Alpha 在风格变化之后也会有所回撤，历史上来看平方和基本可以控制在 1% 左右的合理范围之内。因此在这种情况下，就是等待市场恢复正常，不用采取特殊动作。

而今年的特殊之处为以前基差收敛是从负基差收敛为 0 或者正基差，这次是负基差扩大，与以前方向是相反的。基差是量化软肋，它不能通过量化分析出来，只能被动接受。如果基差快速收敛，已有中性产品会承担短期损失。这时候，平方和会第一时间与各个券商融券，等基差回来后再进行管理。但利刃反面在于对新发产品会更有利，如果基差扩大，之前的产品能赚钱但也要承担其他的成本。

基差的紊乱场景并非历史上第一次出现，往年第四季度到春节期间基差也会拉大，会有量化管理人赎回中性产品或者通过抢跑让基差收敛，雪球与中性产品在基差控制上存在对手方关系，平方和认为这些数据可以分析出一部分对手方，但并非通过量化手段或常规选股逻辑，参与者对未来市场预期持乐观或者悲观态度都会导致基差变动，但它终归不是系统化长期现象，不容易事前判断。例如雪球爆仓，雪球通常提供期货多头，一平仓基差就会放大，但当理清楚关系时基差已经放大了。基差还有一个约束是期货交易，期货的流动性不是很好，假如要平掉几十亿的交易，期货市场是承受不住这么大的交易量的。

05 当真正能管理核心资产时 量化就是发展到健康程度了

访谈前两天刚好有朋友问吕杰勇关于量化吊打主观的问题，“这让我想起了当初我们被主观吊打的时候。所以市场里没有人能永远吊打谁，大家永远都在困难中前进。”

吕杰勇谈到，量化害怕黑天鹅，2015 年市场把对冲定义为恶意做空，直接停牌不让交易，量化没法模拟这种情况。整个行业都很痛苦。2017 年行业结构性变化冲击也很大，小盘股切换到大盘股行情，历史数据上都没有出现过大盘持续上涨的行情，从来没有训练过的行情出现了，对于量化也是一个很大的考验。2021 年，行业破万亿，监管政策的影响减弱，反而行业变得疯狂了。有一些小容量策略开始降频。量化行业接下来会遇到什么样的黑天鹅或者灰犀牛，平方和吕杰勇难以预判，在他看来，遇到市场突变只能见招拆招，但把视角放入更大范围的竞争，就不能只关注量化管理者之间的竞争，跨界卷也需要注意。例如量化中性很长时间是和信托合作，对手就不只是量化私募，所以行业需要内部友好，外部再卷。

“其次要看整个私募行业是否发展健康，是否能够承担更多社会责任，如果某天国家把压箱底的钱也能让私募行业来管，私募行业合规性和各方面水平就已经达到一定程度。现在面对出现的各种乱象，确实量化私募没有资格管理国家核心资产。”

穿越历史，2023 年市场对量化污名化早在 2015 年就已经出现过。如今，市场对量化行业各类评论甚嚣尘上，平方和吕杰勇心态反而更为平稳。量化作为一个新兴事物，要占据主流地位没有挑战者是不行的，后来者从客人变成主人，必然会有挑战者。对公司、对行业都是客观规律，无需过于介意，趋势无法扭转。量化天性追涨杀跌，但除了做趋势，还可以做 Alpha，市场并不理解这些概念，而且追涨杀跌按照

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

市场交易规则盈利本身没有错误，不应该以技术和道德为标准，技术属于中性。但现在（2024 年 1 月）的结构也存在风险，例如现在跌破 2800，在趋势上追涨杀跌时突然打到 2500 再拉一下，市场总归不舒服。

海外有 Beta，国内有 Alpha，而海外自己抓 Beta 的风险还是很大，国内的 Alpha 市场足够大，平方和需要专注做好国内市场。吕杰勇坦诚，A 股由于交易深度或者投资品种不支持展现策略丰富性，整个市场量化策略会单一，导致国内量化千军万马过独木桥的意思，“但我对独木桥从不反感，大家就慢慢挤”。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

08 天算量化：人算不如天算

谋士以身入局，举棋胜天半子，是一种天算。

人算不如天算，则是另外一种天算。

对天算量化来说，人算不如天算是一种技术信仰。“我们相信技术对于算数的作用，是超越人类的。”

但细看天算的发展，胜天半子何尝不是对投研的倔强。

01 人算不如天算 以身入局

天算的成立是人算。

2011 年从清华大学毕业后，何天鹰加入中金公司，从事海外量化对冲基金业务，彼时国内量化行业刚刚起步，而在海外市场，量化投资早已成为主流投资策略。

此时的何天鹰更多是基金研究，尽调了许多海外量化机构，认识到量化在中国市场未来的发展潜力，并最终决定前往麻省理工学院深造，专攻金融工程领域。

此时的刘骏洲正在波士顿东北大学攻读金融学，室友刚巧和何天鹰是高中同学，接机时因此结识，发现彼此都对量化投资抱有相似的兴趣和想法，一见如故。

刘骏洲父亲从事粮食贸易，涉及从美国、巴西和阿根廷进口大豆，并且也从事期货大宗商品交易。所以刘骏洲自小天然地对金融敏感，虽然尚未接触到量化投资的概念，但学生时期数学成绩一直很好，“数学只要逻辑正确，写出来的答案一定是正确的，是逻辑很清晰的一门学科，而基于数学模型的决策可能有很大的潜力。”在刘骏洲看来，主观更像艺术，而自己更喜欢科学。

2014 年何天鹰回国，9 月份募集了第一笔资金。恰逢中国股市经历牛熊转换，何天鹰采用的股指期货策略在一年的时间实现了 206.9%的收益率，最大回撤仅为 10.2%，这就为公司赢得了第一笔起始资金。

2014 年创业，2015 年就经历转折。正如当时的所有量化机构一样，在 2015 年的股灾中股指期货的交易受到了限制，行业面临重大挑战。天算也面临选择：是继续交易股票还是转向商品期货？

最终天算选择了股票市场，主要基于两个原因：1.股票市场的规模更大；2.何天鹰在股票研究方面有更深入的专长。于是公司积极招募人才，积极布局指数增强、市场中性等策略。

02 人无我有 人工智能的天坑与绕行

2017 年，对于公司来说又是一次跃迁。这一年，天算关注到了 Alpha Go 与李世石的围棋比赛，何天鹰萌生了将人工智能和机器学习技术应用于股票市场的想法。成为国内最早用人工智能做非线性模型的管理人之一。

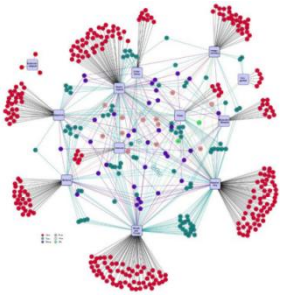
当时由于公司刚创办，资金有限，于是通过财务股东和自营资金的支持，购买了大量的计算资源，是国内首批建立起人工智能投研体系的管理人之一。“要想在一个行业里取得领先，一定是敢于创新、敢于尝试新事物的团队。我觉得这点是很重要的，是任何一个行业都是，但量化更卷。”

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

天算把所有的应用分为三个阶段：人无我有、人有我优、人优我先。在初期公司作为先行者也踩过很多的坑，“但不能因噎废食，要有创新对应的接受能力。最开始用人工智能的时候，我们更多的都是乐观评价它的好处，并不看重风险，但是经历过 2021 年的回撤之后就要考虑怎样兼顾风险和收益。在踩过坑之后，会开发出自己的手段更好地使用这些工具。”

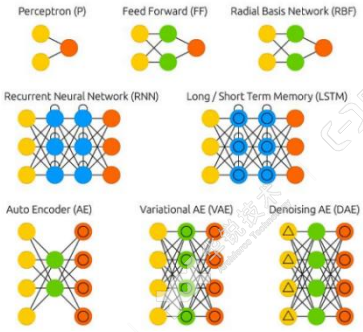
目前天算有因子组、风控组、技术组、算法组、交易组 5 个组，因子方面，因子库有核心库、备选库和试错库。月度更新，备选库周度更新，试错库日度更新。核心库 5000 个，实盘在用 2000-3000 个区间，80%人工挖掘，20%机器挖掘。其中量价因子占比 70%左右，基本面因子占比 20%左右，舆情、分析师、另类因子占比 10%左右。组合方面，非线性模型占比 80%左右，传统线性模型 20%左右。

■图表 因子分类图谱示例



因子类别	量价 50%基本面 20%舆情 10%分析师 10%另类 10%
因子数量	核心库 5000+；备选库 3 万+；试错库 100 万+
挖掘机制	逻辑意义类：人工挖掘；统计意义类；机器挖掘
筛选机制	评价多维度；筛选多标准；深度与广度并重
更新频率	核心库月频；备选库周频；试错库日频

■图表 天算因子组合



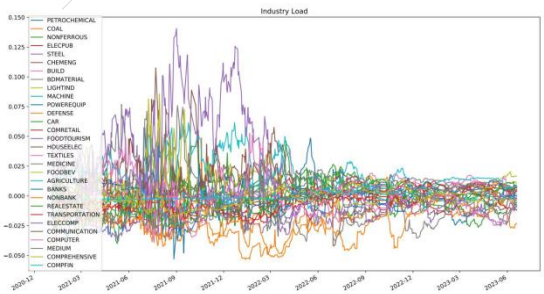
算法类别	人工智能非线性模型 80%；传统线性模型 20%
算法数量	核心库 100+；备选库 1000+；试错库 100 万+
研发逻辑	预测目标（输出）->因子 选择（输入）->算法设计
筛选机制	定量标准+定性标准+优胜劣汰
更新频率	核心库双周频；备选库周频；试错库日频

数据资料来源：天算量化

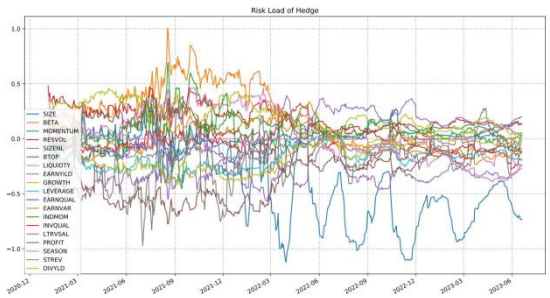
天算合伙人刘骏洲谈到，2017 年作为第一批开始用人工智能的机构，得益于市场拥挤度低，天算超额做到了 35%～40%。2018 年下半年随着市场变得拥挤，超额变得比较平，如何突破成了关键问题。在此之前天算的交易逻辑是每天收盘之后给未来的股票数据做预测进行打分，得到一个持仓列表，这种方式优势在于静态数据，易于收集且对处理时长要求不高，在此之后公司做出的突破是将日间交易改为日内交易，难度在于如何通过技术输出流动数据并给出预测。从日间预测改为日内之后，天算超额又上来并在 2019 年获得了金牛奖。

天算的方法论在于将每天的 Alpha 进行拆解，分为 pure Alpha 与 risk Alpha。risk Alpha 就是行业和风格暴露带来的收益。因为深度学习模型会选择超跌行业，天算没有主动对行业和风格做出约束，在半导体芯片行业和中证 500 指数在行业的配比相比超配 15%，导致 2021 年天算第四季度回撤 10%，规模也有一定程度下降。而在这一波回撤中，天算发现其回撤主要来自 Risk Alpha。于是从 2022 年 2 月份开始，天算加入了优化器，把行业敞口控制在 2%以内，开始用主动工具严格约束，风格因子除市值之外都控在 0.4 以内，本质上改变了 Risk Alpha 的走势。超额及回撤都取得了不错的效果。

■图表 策略换代前后：行业敞口对比



■图表 策略换代前后：风格敞口对比



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

03 风控升级、预测升级和组合升级

天算 2022 年做了三大升级：风控升级，组合升级（将 5 位 PM 的子信号进行加权线性融合），以及最重要的预测升级。因为优化器这一主动工具虽然可以把整体的行业敞口控制住，敞口变小后稳定性会提高，但超额收益也会受损。

预测升级上，除了传统 Alpha 模型对未来股票收益进行打分之外，天算额外对行业收益率以及市值强弱打分，最终把这三个打分表进行结合，前者权重占比 80%，后两者权重占比 15%～20%。也可以理解为市值轮动和行业轮动。

“市值因子其实很有意思，它有一定延续性，在 A 股一般来说大小市值进行切换后会延续一段时间。”这部分的测算是偏长周期的预测。具体实现是把过去 10 年数据按大小市值分组，可以预测出到底几天作为周期的延续概率大，模型一旦捕捉到这样的趋势就用来投资，以最快速度实现大小票趋势策略切换。这也是天算在市值的处理上做的创新，从而它跟百亿量化间的平均相关性低于 0.5。

行业轮动同时也是基于行业收益率目标给全市场来打分，包含行业间资金流向因子、行业龙头股强弱因子、行业内各股票相关性因子等。例如，当预测未来银行股更强，就会在所有银行股股票上做更多权重。

国内目前行业因子涉及很多基本面研究，图神经网络也在此有相当一部分应用。天算的行业因子同样涵盖较多基本面和量价，“行业间相关性是重要因子，最终目标是把无序的行业波动变得相对有序。这一部分我们除量价之外基本面也会应用，从换手率就可以体现出来，天算升级之前换手在年化 120~150 倍，升级之后换手只有 40 倍。”

最终天算模型现在每天都会进行收益拆解归因，例如会用滤波器把行业划为两个点，在回撤里看每个行业因轮动赚钱的情况，最终加权成一条线，体现所有因为行业暴露带来的收益。

目前天算会在控制市值风格的同时做轮动模型，把风格因子控在 0.4，市值因子控在 0.5 以内。但并不总是稳定小市值，当大市值强的时候也会暴露大市值。“全市场 5000 多股票，市值配平是挺难的一件事儿，”刘骏洲谈到，目前整体行业敞口是在 ± 2%左右，大部分小于 3%。而最早用 Barra 模型来做风控的一家管理人，用自建的风险数据库去做风险模型，风格因子暴露之前是 ± 1%左右标准差，现在严格控制到 ± 0.4%左右。对天算整个 2023 年的收益归因的话，80%左右收益来自 Pure Alpha，20%左右来自 Risk Alpha。

2023 年行业主题较突出，例如上半年 AI 主导的行情。刘骏洲也坦诚，由于量化持仓很分散即便超配了 AI，但在这种权重股领涨行情下确实不利于量化发挥。“行业轮动毕竟也是难的事情，我们并不是把它当成收益主要来源，只是个辅助功能，哪怕预测够准也只会超配 2%。”

04 创新的人才体系与窘境

不管是 AI 的运用、还是行业与市值的轮动，都根植在天算的 DNA 中。与之相伴的是相匹配的人才培养机制和产品商业体系。

天算的投研体系自定义为积木式投研体系，公司采用流水线和基金经理结合的方式，同时以内部培养为主，外部招聘为辅。目前 5 个基金经理有 4 个都是自己培养，最初从因子挖掘开始搭建全流程的培养体系，因子环节能很好地验证个人数据处理能力、编程能力。

因子挖掘之后天算一般会让其着手做风控，通过 Barra 体系和 A 股的特点做风险约束模型。最终再让其进行因子组合的尝试，一整个流程下来我们会很清晰地知道大家的能力点在哪里，有的人适合短周期，有人适合长周期，有的人优化器做得

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

好，其他同事就不必在此领域发力更多，每个基金经理小组可以单独挖也可以用公司的因子，可以自己做优化器也可以用公司的优化器，众人拾柴火焰高，“我们称其为积木式投研体系，通过这样的体系来进行相关性低的配置分流。”

刘骏洲认为，同事之间就需要两样东西，情绪价值与物质价值。通过这种感情基础与融洽的氛围，以及合理的分配机制使得大家的向心力更强。目前天算基本上会把业绩报酬 50%分给管理团队，核心员工都有股份，是非常偏平台化的机构。

但创新总是面临很多窘境，虽千万人吾往矣是一种豪迈，也是当前量化面临的现实。

在天算看来，一方面量化需要提升自身透明度，另外一方面确实也需要公众的支持。

天算呼吁量化能够透明，因为现在量化模型已经较为复杂，很难通过估值表的信息倒推出模型，例如天算现在是 5 个基金经理小组，估值表最终体现大概 30 个线性策略加权结果，并且进行独立风控。所以如果一个管理人不敢提供他的估值表，可能并不是担心策略暴露，而是因为言行不一。目前天算向合格投资者提供投前的月度估值表和投后的周度估值表。“现在有太多路演、太多故事，我希望量化不要讲太多故事，而是用数据说话，有多少超额、是在什么样的风格下暴露出来的都可以从数据观测到。投资人可以看到量化持仓对于量化未来发展是很核心的，既然公募可以、主观配合，那量化一样可以提供。”

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

09 图灵基金：90 后掌门人的七十二般变化

历史能不能代表未来？

算法是不是越复杂越好？

鲁棒性检验该怎么做？

在图灵基金的 PPT 最后一部分风险控制上，他提了如上三个问题，并推荐了一本书《失灵》“尽量在概率的适用范围做事情，同时请为不可预测的小概率事件做好预案。”

王亚民，一个 90 后量化创始人，经过近 10 年的量化市场周期，对于这个市场仿佛有一种无限的求知欲。

01 90 后量化掌舵人 从高频跨越 Alpha

2014 年，90 后王亚民从伦敦大学学院（UCL）金融数学硕士毕业后，先后在摩根士丹利、瑞士信贷，从事外汇高频和交易行为的大数据分析。“期货老一派交易员虽然没有采用严肃的量化方法，但当时确实也非常赚钱，有些期货公司的朋友询问我是否可以回国尝试一下。”

王亚民拿到数据后简单算法跑了一下，发现在回测的角度确实可以赚钱。经过三个月的策略开发之后，从 2015 年 3 月开始实盘，此次小试牛刀让王亚民看到了国内量化的潜力。2015 年底，图灵成立。

从 2016-2018 年，图灵开始时主要以高频商品期货为主，后端提成很高，而且没有营销成本，量化工作很是惬意。但 2018 年市场格局开始发生变化，四大天王百亿私募开始出现。“回过头想，那时候刚毕业没几年，自己还是太年轻了，只有二十六七岁，有些缺乏远见，只想着眼前的交易策略，没有考虑长远的大趋势。”

王亚民不避讳彼时的青涩，以及对量化前辈的赞叹。2016 年左右，量化选股的商业思路都是如何进行量化对冲，但就有前辈提出指数增强会成为量化规模最大主菜单，并进行早期的赛道布局和开拓。有早期机构用最原始自有资金自购，5 年不计提业绩报酬，拿着自己业绩报酬与客户共同承担市场波动，在市场上换得信任。这以后 5 年量化行业快速发展，到 2021 年，500 指数增强赛道沉淀几千亿级别资金，布局比较早的机构许多都成为了后来的头部管理人。这个过程，体现了机构在长远商业趋势上的前瞻性。一些头部机构引领技术革新，愿意开拓进取，在人工智能等领域做非常多前瞻性投入，这都是非常值得全行业骄傲的。

量化行业快速发展，王亚民有着自己的节奏，大概年轻，他并不一味追求商业利益，而是更关注研究边界的拓展和进化。在此期间，图灵开始从高频跨越到股票研发。“这其实是有很多痛苦的，我们以为在期货市场做得不错的策略或许可以直接迁移到股票这边，但事实并非如此。”

王亚民解释道，高频更多的是“点突破”，不需要太大的数据量，只要掌握一种交易技巧或交易思路就可以尝试去做，而在选股策略中，股票的底仓换仓风险是没办法用高频高胜率的方式解决，因为当天只能换一次。在期货市场中可以采用短时间内大量重复的方式来做好风控。“一个胜率 70%的模型，每天预测 1000 次，那么这 1000 次样本总和为正的概率也差不多是 100%。但股票并非真正意义上的 T+0 产品，这是两者之间的本质性差别。在当下预测接下来一天的股票收益率是一件很麻烦的事情，存在太多信息不透明导致不确定性因素。但是在截面上看，这些风险因子都被抵消掉了。在长周期上，资产间的 Beta 会互相抵消，噪音会被过滤，相对收益的可预测性要比绝对收益可预测性更高。因而在预测周期拉长后，截面模型的重要性尤为凸显。”

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

“行业内很少有人能够同时兼顾高频时序和截面 Alpha 两套研究框架，所以很多做高频的公司后期转型都是找到另外一个策略合伙人。”王亚民当时也是这样的想法，当时市场流行的是老少搭配，有一定江湖阅历和行业地位或者资本金积累的资深人，再寻找年轻策略合伙人，而自己作为一个 90 后要招聘合伙人会面临诸多挑战。“但好在，年轻的时候至少有机会可以把一切推倒重来。”在 2018 年下半年，王亚民及其团队开始自己做选股，从“高频时序”奔赴“中频截面”。

从 0 到 1 的过程是很痛苦的，那段时间王亚民经常半夜打车去办公室“推电闸”：“先是算力跟不上，加了服务器之后又遇到散热问题，办公楼中有很多是中央空调，不能控制单独房间的空调。于是我们又搬到了郊区中环附近，可以独立控制空调的办公楼。但当服务器加到第 20 多台，机柜加到第 3 个，虽然散热没有问题，但是会跳闸。写字楼的供电标准跟不上了，所以经常夜里要过去开电闸。”最终，经过反复试错后，图灵把大量算力都放到了专业的数据中心进行运维（IDC）。

团队一起走来，一起经历过所有困难。“2021 年 500 指增超额回撤 3.5%，当时我们内心非常痛苦，刚拿到资管牌照募集了 10 亿左右，大多数初始资金是在高点直接回撤，没有任何安全垫。虽然客户没有埋怨，2021 年 10 月份，一家券商自营电话联系我，不要压力太大。另一家反映他们回撤已经到 5%、7%，而我们这边在 10 月的回撤只有 2%+，已经很好，但是实际上客户确实也亏了钱，我们内心压力非常大，觉得非常对不起客户。”至此，公司全体投研更加注重风控在各个环节及各个产品线上的渗透，“我们内部很多想法和理念都非常统一，也有充分的沉淀，投研人员极度稳定，因而早前的市场经验与教训，一直被继承下来。”

02 一年新增 50 亿 客制化精细赛道

2023 年图灵规模新增 50 亿规模，到达 70 亿总量，但公司并非只关注主流 500 指增赛道，而是客制化程度非常高的细分赛道，甚至包括如科创 50 指增、中证红利指增等。

做指增并非挑选出一揽子的 Beta，而是一种精细的雕刻。

王亚民认为，500 的问题在于其中位数市值大约为 260 亿，为中盘。而市值这个风险因子的收益率，在全部 A 股的各个区域的变化非常大，并不能被经典的线性（Size）或者非线性（Nonlinear Size）所很好地刻画，从 2022 年到 2023 年跌幅来看，沪深 300、中证 500 以及中证 1000 的跌幅只有几个点的差异，即市值 100 亿左右的公司和市值 1000 亿以上的公司相比收益率没有大的差别，但从 80 亿以下开始，小市值风格的收益率开始疯狂“上翘”。

“你应该用什么样的思路做东西？应该用什么样的选股域？”实际上很难决定。如果全 A 选股做指增，选择持有一部分市值 30-50 亿的上市公司，在这几年肯定是占到极大好处。但这与 500 主要中枢是否一致？王亚民曾经进行过研究，没有任何选股技巧的情况下，可以直接以每隔 20 亿左右市值范围计算区间做等权指数，类似万得小市值指数的编制算法。实际上在 60 亿以下市值范围，例如 40-60 亿等区间，其等权指数已经强于中证 500 至少年化率有 20 个百分点。“我们比较排斥‘模糊’，无论对内做研究还是对外做客户沟通，确定赛道的真实标准很重要。例如：实际选股域、实际基准、风控参数，这些都是很重要的。这样可以与客户进行更加透明的沟通，对内也能取得真实的研究进展。对外和对内，都要真实。”

中证 500 已经是一个拥挤的赛道，有 170 多家管理人，在这样的市场背景下，图灵着重全市场指增和沪深 300 指增。沪深 300 和全市场指增的思路比较明确，沪深 300 采取成分内选股的极端方式，能够做到年化超额 9-10%，在别家机构出现单周 1%回撤时，图灵的回撤只有 0.1%；全市场指数的市值本就是 50-60 亿左右，有一半以上都是 60 亿以下的，自然纳入了一部分 60 亿左右的股票，并可以获得不错的弹性。

“量化工具是一种实现商业价值的技术手段。并非盲目地追求谁算力最强，谁超额最高，而是尊重客户需求。这并非在

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

制作奥赛题，没有标准答案。真正唯一的答案只是客户是否满意。在当前社会中，安全性、稳健性和理性预期会逐步变得越来越重要。在这种情况下，经营思路和理念确实应该发生改变。脱离原来销售理念，现在我们更多的是追求实打实地把资产做好。”在王亚民看来，当前环境下，可以预测到未来资产和资管将发生巨大颠覆性改变。未来带给资管的是全新问题，即如何真正意义上创造不是刚兑、定制化资产，甚至是全天候资产。这些目前对资管还有很多挑战，机构需要在新挑战上实现自己的价值，这也是图灵的重点。“短期规模永远会有波动，也许今年增加几十亿，也许下降几十亿，但是公司安全边际也较高，因为公司人均创收、人均管理规模较高，在这种情况下，我们愿意花费很多耐心研究新产品、新问题、新边界，短期规模并不重要，这是对未来展望和经营理念，希望在时代中能够真正解决痛点。在这段时间中，能够为行业扎实地创造价值。”

图灵的机构资金占比高达 90%，这些资金更加追求稳定性及信息沟通的透明性。按照客户的需求行事非常重要，客户自己选择的模式可以更好地接受其收益并配置组合风险，机构要做的是保持充分透明的沟通。为了提高沟通效率，图灵此前甚至会把各个赛道、所有产品的中位数持仓市值直接公布给机构客户。

沪深 300 选择全成分股的方式也是基于对客户负责的考虑，若允许 20-30%的部分在沪深 300 以外选股，理论上的长期年化收益会更高，但是目前银行、保险、券商等机构资金存在的痛点是久期不够长，因此长期的年化 2%-3%的提升，如果意味着短期最大回撤会因此放大同样的幅度，那么意义并不大。“10 年小票跑赢 8 年，拉长后没有问题。问题在于如果今年考核无法通过，可能投资人就不会持有到我们想要的时间。你与他谈论 5 年或者 10 年没有意义。”

为了能够更好匹配客户需求，图灵会与客户进行充分沟通，包括历史收益及回撤、预期收益及回撤、资金久期等，并依此在自研的流水线因子组合及风控层面进行调整，制定高度定制化的产品。

以科创 50 指增为例，图灵曾接收到了一笔 1 亿的券商自营定制单，在持有期内，年化收益 8%左右、最大超额回撤 1%左右。因为基准指数只有 50 只股票，科创 50 指增很难进行风控，“这是因为 Barra 在设计的时候没有考虑特异性股池，进行回归时，100 个样本点或者 200 个样本点与 5000 个样本点，最后回归出来的风险因子的收益率肯定不一样，有时候甚至因子收益率的方向会相反。”

■图表 图灵各指增线风控标准

	全市场	300	1000	小市值
业绩基准	全市场等权指数*	沪深 300	中证 1000	万得小市值指数
成分股占比	100%	100%	40-50%	90%
持股数量	800-1200 只	250-300 只	800-1200 只	800-1200 只
换手倍数	80-100 倍	80-100 倍	80-100 倍	80-100 倍
Barra 风险因子	0.3 标准差	0.2 标准差	0.3 标准差	0.3 标准差
行业偏离	≤3%	≤2%	≤3%	≤3%
个股偏离	≤1%	≤0.5%	≤1%	≤1%
运行起始时间	2022/9	2023/5	2023/5	2023/5

03 七十二般变化 图灵的因子构建方法论

图灵的优势在什么地方？

王亚民回答非常干脆直接，“量价的基本功也是非常扎实，因子数量及因子挖掘的流程、效率在行业内都做得不错。”

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

最早做高频的时候，王亚民很少与同行见面的，因为高频更多的是技巧性的东西，很难共享。而中低频与资管了解市场需求和未来发展方向，以调整研发思路，对于多因子的研究，许多机构讳莫如深，但王亚民的分享充满着细节。

他说最近公司在研究一个新品标的，花费两个月时间生产了大概 3700 多个因子。“图灵因子库的因子的构造方式有许多种，例如常规高频 k 线统计算法、偏度、峰度等涉及许多统计算法，不同构造方式当中又存在许多可变化的算法，因而可以用一个比较相似的逻辑构造出上百种不同形式。”

例如高频 K 线的选择可以是 1 分钟/3 分钟/5 分钟，偏度选择上行/下行/全盘，统计时段选择全天/半天/开盘/收盘，平滑函数选择 average/min/max，平滑周期选择 5 天/10 天/20 天，包括大单，有些机构按全年分位数计算，有些机构则按绝对单量计算，……存在许多变化的可能性，不能说这是机器的因子，因为大单逻辑或者 K 线偏度逻辑本身具有金融意义，不同单量则有不同信息含量，代表某些具体的交易行为，因此每一个选择都自有其内在的逻辑意义，具体选择哪个指标要看测试结果，不能轻易认为 2 万元订单为大订单，实际上 20 万元或者 200 万元订单是最显著。一个大的逻辑之下会产生许多低相关的因子，这便是考验算力的时候。

当前图灵量价因子占比 70%-90%，在王亚民看来，信息最终都会在量化当中最终呈现。图灵目前使用的模型有很多，包括线性模型、决策树模型、神经网络模型等，不同模型的策略类型存在差异，但总体更加倾向于高频。对于日频以下的频位段会使用机器学习的方法，在整体偏线性的大模型之下会包括权重不那么高的非线性子模型。

04 量化小市值的风险聚集与情景预演

“历史能不能代表未来？

算法是不是越复杂越好？

鲁棒性检验该怎么做？”

过拟合，量化机构似乎已经极少再去思考这个问题，但王亚民还在思考。

目前整个 A 股在使用机器学习算法时，要求这些金融数据的“独立同分布”属性。然而目前在实际数据上存在很大的结构化差异。例如，市场长期风格、交易制度、各类投资者占比，尤其是量化交易自身在某些股票区域的占比，都存在很大差异。例如 2017 年前几乎没有大量化公司，现在有接近 1 万亿资金每天在交易，即使平均单边换手 10%、双边换手 20%，也即年化 50 倍换手率，那么每天也相当于有 2000 亿量化在交易，而这段时间 A 股正常交易额为 6000 亿或者 7000 亿，其中量化占比就能达到接近 1/3，尤其是在中小盘区域，这个占比可能会达到 40%-50%。对于选股流动性冲击远比直观能看到的要大。

当下做量价的会更喜欢中小票，因为中小票的对手方更多的是个人投资者，相对而言博弈的确定性会高很多，而沪深 300、中证 500 所代表的大票越来越难做，在做宽基指增时，若不把选股往后拉，很难获取 Alpha，但整个投资结构在不断分化。而这些信号大多数聚焦在中小票上，因此对真实选股流动性冲击不止如此。在这种情况下，数据与以前结构会发生很大改变。

“这个形势终有一天会发生逆转，我认为这对国内量化会是一个很大考验。对于个体单位来说，所有做算法交易人都有这种经验：如果股票已经占据全市场股票日均交易量 5%甚至 10%，算法交易的绩效就很难跑赢 VWAP。”

王亚民做了一番推演：目前 A 股流动性第 2000 到第 4000 这部分区域（接近中证 2000）股票平均交易金额正常情况下约为 5000 万元，这已经不算太小票，至少不属于微盘股。如果计算尽量不超过 5%交易额占比，实际上为 250 万元。如果某量化机构持仓这个区域的 1000 支票，250 万 × 1000 支票是 25 亿。也就是说，要不影响算法交易的绩效，该区域的股票一共只能承载 25 亿。如果该机构偏小市值的赛道中，有一半的股票属于该区域，则反算赛道最佳容量只有 50 亿。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

“所以我们面临如何平仓？”这样的拥挤问题在量价上面并未很快显现，拥挤度也并不会立即给出反馈结果，某只股票不断上涨，拥挤度逐渐升高，不断购买时反而会不断推高股价，需要等这个正反馈一直到达极限，再一次性反馈。

所以一方面图灵在利用海外资产集中度比较高的样本数据做情景预演，同时更加强调因子的平衡性。在极端区间表现极端的因子都不会去采用，虽然它可能在近段时间表现特别好，但是会对它在某些测试区间里表现单独制定指标进行排除。如果某个因子在某一年表现非常好，其他年份则是报废的情况，图灵就不会采用这个因子。“图灵更为看重因子的稳定性，此外在组合方法上做好分散。有差异也会有共性，历史上总会有一些风险在不断复现。”

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

10 稳博投资：模糊的正确与量化的精确



思想的碰撞。

投资的直觉。

技术的专注。

——这是稳博投资对于策略诞生的总结。

01 思想的碰撞 模糊的正确与思维的演绎

同寝室创业，郑耀和殷陶是量化行业的佼佼者。2014 年，两人毕业正好赶上 A 股小牛市，听说炒股挺赚钱的，于是写点程序、注册个公司，有了稳博。

2015 年从 0 自创高频交易、统计套利；2018 年叠加深度学习算法；2020 年融合 Alpha 横截面方法论，2021 年持续创新迭代，成为今日稳博。

上下求索，今日方成。在稳博看来，投资策略诞生首先源于思想的碰撞。

碰撞来源于不同，今日的稳博在招募各类不同背景的人：海外基金经理、公募基金经理甚至包括经济学家……除此之外，公司每年还会组织多期招募组织训练营进行交流。“招募不同背景和方法讨论出来人，认知才会有所不同，即使是海外大厂也会有截然不同的研究体系，有的讲究因子数量和差异性，有的就一两百个因子讲究研究深度，发挥极致。市场变化很快，我们需要寻找新研究方法，不仅是研究某个领域，而是如何进行研究，如何组织到一起工作，这也是公司保持竞争力的核心点。”公司创始人郑耀解释道。

这种碰撞也牵引着稳博的组织变革。

2019 年之前郑耀和殷陶两人配合写策略，手把手带出第一批人才，稳博就变成了多 PM 架构。但两人随后发现，团队往往朝着同一个赚钱的方向研究，很难横向拓展到其他高潜方向，而对于机构来说，这是投研储备工作的必要。

因此稳博重组结构，将 Alpha 部分称为核心组，采用流水线制，由殷陶、郑耀带队，涵盖国内数据投资先行者与多位人工智能专家。高频部分则设为多 PM 制，由国际知名量化对冲基金经理、公募明星基金经理、经济和统计专家共同组成，专注显著的超额收益实力和策略研究创新。

公司策略方面则讲究攻守兼备：“在长周期上讲逻辑，在短周期上讲求机器学习、算法预测。”当前，稳博把因子分为统计性因子和逻辑性因子，其中统计性因子入库标准为：低频因子统计时长不低于 10 年，T0 因子统计时长不低于 6 个月，以此兼顾长期历史有效和当下市场环境有效。

逻辑性因子则会检测每次逻辑成立时，对应的胜率以及极端情况下胜率和表现，因子提交后，在预设时间内通过样本外数据验证，统计学上显著有效，最终被采纳，因子通过检验时长 1 个月-1 年不等。

郑耀把投资方法分为归纳法和演绎法，量化机构里少见的，2021 年稳博开始招聘经济学方面人才。“传统的归纳法有一个弊端，市场上常常发生历史上未曾出现过的事件，这时候难免遇到策略失效或产生巨大风险点，通过演绎法可以一定程度上缓解该问题，从更加宏观的视角去看待正在发生的或即将发生的事情。”

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

落足到投研的实际，这批经济学人才主要为公司提供宏观因子，同时也会为投研以及风控寻找方向。例如是挖掘中证 1800 以内的 Alpha，还是挖掘价量 Alpha。偏基本面还是更偏向公司治理，偏微观还是宏观，判断市场资金偏成长还是价值，判断下一步应该研究哪个方向更容易成功。

也有机构招聘主观类研究员或基金经理，但主要方式为自下而上挖票，稳博更多是从上至下的研判。郑耀通过大量例子来说明这个问题：今年 1 月份市场看到一致预期很好，实际上每年 1 月都是如此，为什么？上市公司 1 月 31 日发布一年预报，预期类因子就会很强，但去年一度分析师因子失效，使得量化 Alpha 整体非常难受，衰减非常严重，其中很大问题是没有真正做到宏观或者微观，这并非方法论本身失效。再如近两年北向资金流出，一般而言，北向购买原因与分析师看好并推票的逻辑非常相似，北向资金的进出也会带来分析师类 Alpha 发生变化，所以需要先进行宏观风格中性化，不然就会变成负向 Alpha。再例如宏观方面，中美宏观，从美股到 A 股如何传导，或者 2023 年的 AI 行情，从英伟达到国内产业链如何传导，服务器厂商中哪些环节最受益？需要结合产业链数据进行分析，哪些节点非常细致又不可或缺，都有哪些上市公司在参与。“甚至包括更看好 300 还是微盘，这都是灵魂拷问，需要从更宏观的角度看待，例如从资金角度看存量投资人进入还是离开等。”

演绎法也被稳博用于风控，具体落足在市场宏观风险预判以及追寻策略失效的原因。

■图表 稳博在 Alpha 风险优化的实践



在稳博看来，从演绎法角度策略失效一般可分为本质、内因和外因：包括新的市场环境和历史规律有较大变化；规模超出容量，因子拥挤；大体量资金投资偏好转变。所以风控方式强调核心逻辑，不干预策略运行，争取屏蔽确定性高的回撤，以此规避风险。

同时 Alpha 还会仔细推演、反复验证之后渐进迭代。Alpha 投研团队每 2-3 个月出一组新的组合结果进行追踪，对于每一个上线追踪的策略，周度对前述所有环节进行复盘，针对已上线的 Alpha 策略和交易执行策略，进行周度打分和做出是否调仓决策。

如果一个 Alpha 因子的表现与以往不同，且存在无法解释的风险，那么可以按照先将可解释的风险敞口关闭。在量化投资中，分散投资通常是更受欢迎的方法，而不是集中在个别权重上，这在一些特殊的 2/8 行情下可能会导致显著的负超额。要解决这种情况，可以使用演绎法的方法论，把个股权重放开，或者去约束市值的敞口，甚至在特殊事件场景中把市值敞口暴露偏大。

当然绝对的演绎可能更像主观的推演，量化仍然是稳博的底色，“大部分时候对我们自己的 Alpha 是没有太明确的观点

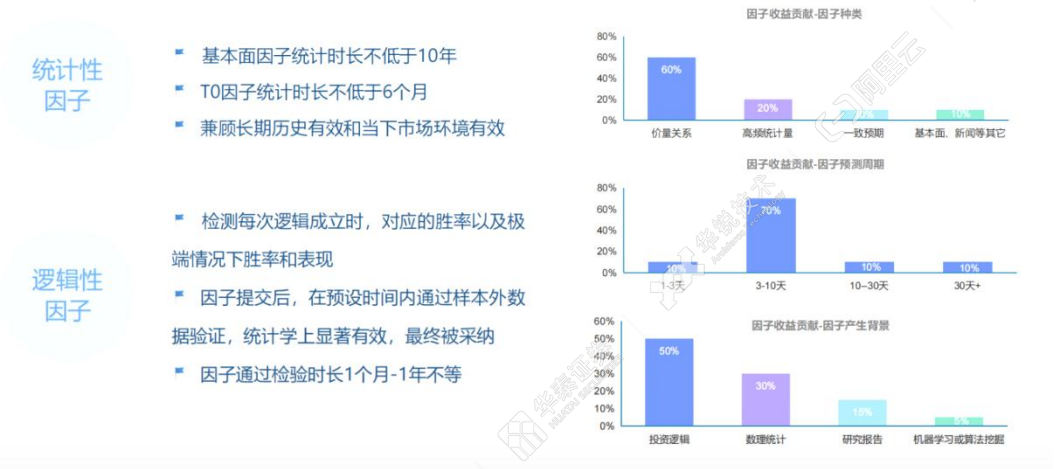
官方下载及活动链接—https://BigQuant.com/quantpaper/2023

的，只会觉着有一些模糊的正确。”郑耀表示，在这种模糊的正确的前提下再去判断风险大概有多大，当风险有点大的时候就会收紧风险。

既然是模糊的正确，那也自然会有错误，2023 年中特估行情起势，稳博早前判在 4 月中旬年报季加强基本面配置，但结果在 3 月最后一周基本面就开始走得非常强，同行已经开始卷入，Alpha 有效性已经逐渐变弱。但无法预判，很多时候必须在时间维度上到达某一个点，视线才会明朗起来，因为它与归纳法不一样，IF A 就有 B。在演绎法中条件太多，无法归纳出市场究竟会发生什么，只能在特定场景下进行推理，才能推理出 ABCD 等概念。

2023 年的 AI 和中特估行情令郑耀印象深刻，带来的影响是：稳博开始推动关系型数据和关系型 Alpha 的研发。

■图表 Alpha（因子形成）



02 技术的专注 风险度量与技术的专注

微观的博弈，量化的 Alpha 在细节之中产生。

“如果量化投资能解释全市场 10%的交易行为，已经非常了不起了，但在这之外还有 90%的未知情况，也会给量化投资带来冲击和风险。”在 2023 年 4 月的中国量化投资白皮书论坛上，郑耀发表主题演讲《探与思：量化投资风险优化》，讨论到了马柯维茨模型、Barra 模型的局限性以及当前公司在拥挤度，与上下行波动率方向的探索。

郑耀解释道：马柯维茨模型对 5000 只股票做最优搜索，往往会陷入股票数量非常多、数据量过大，在求最优解的过程中计算复杂度高等问题中。除此之外，求解得到的是数学意义上的最优解，不一定是真实的最优解。

而 Barra 模型相较于均值方差投资组合，其求解过程大幅简化可以显著缩短运算时间，不再局限于全局最优解目标，而是将优化目标函数转换为处理 Barra 暴露等因素，而每一个 Barra 暴露因子也具有一定的实际意义，这既使得更容易求解，也避免了陷入精确错误逻辑的问题。同时，随着优化算法的不断进展，Barra 优化器方面在求解计算性能上也有很好的优化。

许多量化私募在 Barra 模型上表现得非常严格，但同时依旧存在很大的超额波动，包括前期大幅上涨和后期的快速回撤。在投资过程中出现因子过度拥挤的现象，这种现象也可以被定义为一种风险。例如去年，一些量化管理人的上行波动率和下行波动率有较大差异，回撤较小而温和，上涨强劲而凶猛。

“在构建投资组合时，是否可以不以指数为基准，并将上行波动率和下行波动率分开考虑？”郑耀解释道：一些以绝对收益为目标的 Alpha 策略，买入了一些低估值的优质股票，其下行波动率较低，最多横盘，而在上涨时可能会有比较大的涨幅，

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

这样的配置会造成不同的上下行波动率。然而，如果要以指数为基准构建投资组合，根据 Barra 模型的解释与指数挂钩，要将上下行波动率分开考虑的难度就更大了。因此稳博认为，可以通过期权工具把指数增强产品的上下行波动率调整为有偏的，在指数下跌的时候还能跑出正超额，在指数上涨的时候也仅需要付出一些成本较低的期权费。在量化选股或绝对收益类产品中采用这种方法更符合投资者对市场的直观认知，因为投资者的想法很简单，希望在下跌时不会亏损太多，在上涨时可以真正赚钱，并且长期持有后会获得收益，这与之前的 Barra 控制风格的方法是有所不同的。

郑耀解释，上下行波动率分离想法与市场结构化行情有关。当市场出现奇怪热点时，采用这种方法可以让持仓更加集中在有爆发力股票上。这也是 Alpha 一种手段。在进行 Alpha 研究时，获取收益来源并非均匀，而是非常尾部收益。平时收益不会超额，甚至偏负，但是尾部非常强，突然进行一大堆股票组成。Alpha 与传统方法有所不同。

优化模型方程其目标是优化组合的预测收益，在目标收益的基础上，减去组合的风险扣分以及交易成本。这里面存在一个非常有趣的现象：风险因素的衡量不一定是单一周期的优化，而有可能是多周期的，随着管理规模的变化，优化方程也会发生很大的变化。例如如果管理规模非常大，对某只短期内看空，长期看好上涨的股票，也需要较早布局购买，否则在未来可能无法购买足够的数量；但是当管理规模较小时则可以不急于操作。

一般的风险约束是控制单一因子、行业、个股最大持仓的暴露边界，如将单只股票权重控制在千分之五左右。个股最大换手规模与管理规模息息相关，对超额波动率优化也有深远影响。当管理规模足够大时，风险控制变得更加困难，有可能出现调整速度跟不上市场风险切换的情况，导致风险暴露更加突出。

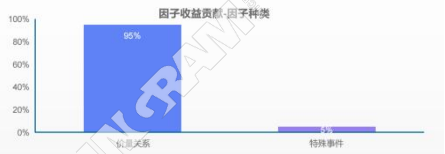
在风险优化方面，除了使用因子模型进行预测，稳博还会经常使用机器学习方法，其中强化学习是一个经典的方法，以收益或惩罚的方式，用于决定调整风格的配置和确定增减风险敞口。“确实可以产生效果，但可解释性变得更弱。”在中低频的 Alpha 上，稳博目前还没有采用强化学习的方法来调整风险暴露。

03 投资的直觉 市场教会了我们什么

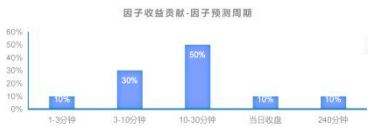
和郑耀交流，充满着一种详细颗粒度所带来的满足感，量化数据的精确，萃取了数字之后的洞察。

在稳博的眼中，想法的诞生来自对手盘分析、博弈性的思考、人工智能算法的应用。从 0 自创高频交易、统计套利，稳博对于对手盘、博弈的思考体现在公司文化之中。

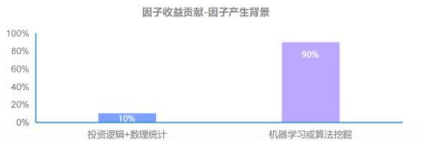
■图表 因子收益贡献



■图表 因子预测周期



■图表 因子产生背景



交易单最能够看到对手方：郑耀认为很难区分交易单为量化还是其他，但市场的每一个交易对手都有他自己的逻辑和交易喜好，从每一笔成交的蛛丝马迹中，可以总结出一些和你成交的交易对手的统计性的一些交易偏好。需要注意的是，每一笔成交都是买卖双方的分歧创造的，创造这一笔成交的双方谁是更“聪明”的一方，需要持续的思考和总结。

往往微观上交易的巨大分歧并不发生在某一只股票上，而是在有共性的一篮子股票当中，比如 Alpha 机构开始大量买入银行板块等。当风格发生切换时就会在优化器中加入某一大类的惩罚，引导持仓向某一方向调整。“我们微观是非常喜欢做这种事情的，但因为解释力度有限，还是需要结合更多维度的信息来把确定性提高一些。”郑耀表示。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

在郑耀眼中，主观的投资方式、市场逻辑近年发生了一些变化，之前市场驱动的逻辑和 Alpha 来源很多都已经消失，“行业已经不再是选股聚合的维度了，取而代之的是概念。”

过去大家会在某一行业内聚集，现在更多的是围绕某一热点甚至奇怪的概念点聚集，牵涉到的行业更多了。比如 ChatGPT 的概念跨了电子、计算机、传媒、通信四个行业，可以受益的股票五花八门但实际上其本质都有一个概念——ChatGPT，在这样的情境下怎么做关系图，怎么抓这一概念的动量，“这些问题反倒成了重要的课题。”

“所以讨论不是在左侧，而是如何在右侧第一时间抓住它，即如何将其聚集起来观察，而不是单一地观察，因为单一观察永远是看反转，对于个股涨多永远是看跌。”郑耀认为，对于行业或者板块而言，市场并非如此，而是体现出动量效应。一个事件型的热点背后会有 5 只或 50 只股票，从熊市或平市转为牛市一定是先有一些点拎起来，但此时往往比较危险，容易被虚晃一枪。

这也刺激了公司关系型数据的展开以及大模型的应用，“所以未来在做 Alpha 时会重点储备，将股票联动做到一起，尽量能够在一个热点刚冒出来的时候就去抓住，即做好关联性的 Alpha。”郑耀介绍道，实体识别是大模型非常擅长的领域，与图神经网络串起来的。要研究两家上市公司之间是否存在关系时，要去看两家公司的产业链、产品链、股权链以及是否在同一时间被新闻提及等因素，GPT 可以帮忙识别出许多包括公司、行业、产业、产品在内的实体，再根据彼此之间图神经网络，大语言模型便可以实体被提及的频次等为依据激活其中部分实体的未来行情，进而做更加全面的信息挖掘。”图神经网络各大机构已基本研究透彻，但利用大语言模型做实体识别仍是比较新颖的方式，这部分未来会在公司量化模型中占多大权重，郑耀认为很难讲得清楚。

量化本身也是对手方。从 2021 年底稳博开始研究因子拥挤度问题，以应对新能源回撤等市场现象提出的话题，以 2021 年 9 月的市场行情为例，此前量化投资在市场中的占有率大约只有 5%到 10%。但是在 2021 年的一波大幅下跌之后，市场占有率迅速攀升到了 20%到 30%的区间。由于量化投资人的微观交易策略存在一定程度的同质化，历史规律随之变化，当时行业的底层 Alpha 来源比较贫瘠，例如近年来人工智能成为一个热门话题，但有些事件对于传统的量化方法来说可能是负面的 Alpha，很多热点 AI 股按照传统归纳法并不是量化眼里的“好”股票，甚至根据未来三年的财报预期，这些公司的成长可能并不会有很大幅度的提升，但是它们的股价仍然可以翻倍。这样的情况偏离了传统量化方法基于基本面的定义，往往无法利用这些基本面数据，只能采用一些简单的行业动量逻辑来获取 Alpha，但这种行业动量能够带来多少超额收益仍然是一个大问题，可能甚至是反转因子。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

11 玄信资产：我们正在经历的风暴

关于玄信：8 年国内量化投资经验。管理规模约 15 亿， 其中自有资金 5 亿，为高频期货策略，另外十亿资管资金是量化股票策略。公司目前拥有一个庞大完整的团队，团队人数 61 人，80%以上都是投研人员。

策略架构方面，玄信一直沿用流水线的投研体系，期望收益与策略的可解释性能够达到平衡，底层 5000 个因子通过非线性的模型组合成 100 个+特征，再使用线性进行组合形成策略。

关于吴星玮：创始人兼投资总监，清华大学工业工程学士（年级第一），加州大学伯克利分校统计学硕士、运筹学博士，5 年美国市场量化研究经验，先后在巴克莱、安盛等知名机构工作，8 年中国市场量化研究经验。

在 2024 年初市场经历剧烈波动的背景下，玄信资产管理公司的负责人吴星玮接受了我们的专访，详细解释了公司在面临极端市场情况时的策略调整和未来规划，以下为对话的精华摘要：

问：这周（2 月 7 日）极端行情下，我们如何处理以及如何看待当前面临的巨大挑战？

玄信：截至上周，我们的表现非常优异，中性产品收益仍然为正，超额回撤 0.9%。但本周发生的事情已经超过了所有模型能够的控制。行业偏小票引发的超级亏损，原本可以通过平仓事先释放，但无论是 DMA 还是中性产品，由于交易限制都无法平仓，不带杠杆的资金普通私募产品至少可以调仓或者掉头，带杠杆的继续承压，所以大家只能切换到 1800 成分股，所有资金涌向成分股之际，形成比平常更严重的踩踏。即便再稳定稳健的管理人，都面临非常大的回撤。

我们即便没有风格暴露，但目前也已经转变为防御模式。

问：策略切换到防御模式，相当于人在干预它，遇到危机时刻还是人的决策？

玄信：首先，这个危机在历史数据中从未出现过，我们所看到的 2008、2015 年股灾也没有出现过类似的情况。因此量化历史数据中如果出现过这种情况，模型才能够学习。因此这次的决定一定是主观的。我们刚才主观分析的逻辑是目前整个量化在抱团指数成分，主要基金也在购买 ETF，因此成分股非常强势。截至上周，小盘对我们的影响还有限。抱团行为并非量化无法模型化的行为，此前我们也会提前关注征兆。但这种判断并非归纳法逻辑，只能运用经验或者演绎法逻辑进行判断。

本轮量化踩踏肯定是系统性的问题，接下来无论从管理人、投资人、券商还是监管人、都需要反思，这都是非常重要的一课。

问：非数理性或者非统计信息发生政策改变导致的突发时刻．量化管理人员应该如何切换或者采取相应的环境措施？

玄信：我认为所有决定都非常困难，无论是空仓、指数复制还是继续承担亏损都是艰难的决定。我们主要进行量化和统计，现在没有统计意义，所以做任何决定可能都是运气。

问：现在进入防御模式，出现什么信号会切换到正常模式？

玄信：这是个好问题。首先我们需要从市场角度预估 DMA 是否已经企稳以及整个小盘股的超额回撤是否已经完成。

例如明天小盘股与大盘的差距开始缩小，我们会进行实时模拟观察平仓资金和赎回已经开始企稳。

问：本轮行情与美国 2007 年行情有什么特别之处？

玄信：第一速度快：2007 年美国量化崩盘，经过 3 天、4 天的跌幅达到 30%-40%，随后迅速恢复。美国的发展速度非常快。反弹完毕当时文艺复兴回撤了 40%，但随后一两周恢复。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

其次聚焦小市值：当时美国集中的并非小票，且主观并未遭遇踩踏，拥有更好的流动性，另外美国拥有更多的做市商、高频交易商。整个市场微盘流动性非常少，每天几百万就能打个涨跌停，而本轮国内量化超发规模集中在中小市值上，风险释放并不会如此快速出清。

第三拥挤度：由于管理规模过大，同类型策略的拥挤度无法评估数据，只能从主观感觉美国至少没有这么拥挤。**问：除了踩踏更快之外，还有哪些特征值得我们做另一个样本学习？**

玄信：我很好奇他们与投资者的沟通方式，投资者可以忍受回撤并与管理人共同承担风险。如果那轮投资者不能忍受全部止损，很多基金都会清盘。**问：关于市值暴露，管理人都声称控制市值 0.3 倍标准差，玄信则是 0.6 倍标准差，市值控制相对严格，然而仍然存在小盘或者微盘暴露，行业里大家对此如何看待？**

玄信：我们市值暴露 0.6 是上限，实际持仓没到 0.3 左右。市值暴露只是风控的一方面，还包括非线性市值和市值中位数。另外我们不会碰 25 亿以下市值的上市公司。上周纯粹小盘风格会亏钱，但其他估值和 Beta 因子则相反，所以整个风格并没有形成亏损。如果有机构纯粹市值暴露大或者 Pure Alpha 不好，整个策略回撤就会比较大。

问：刚才您提到不接触 25 亿以下的上市公司，之前您接受另一次访谈时提到 40 亿，风控是否经历过变化？

玄信：我们并非按照绝对数量计算，而是按照市场中位数计算，市场全部排名末端 10%。去年在 40 亿左右，今年就是 25 亿。

问：光控市值因子并没有用处？

玄信：我们已经设置了程序，预计基差何时开始平衡。

问：本轮风格转换是否与之前有所不同？

玄信：任何事情都不可能一直持续，从去年 12 月份已经有一些迹象，微小盘不可能一直上涨，总有到顶的时候，其估值基本面已经无法承担到达临界点。其次年底没有新增资金，从今年 1 月公募放开可以卖出，整个指数首先开始下跌，同时去年开始有些投资者获利了结，开始赎回微小盘品种，引起风格变化。之后国家队救市过程呈连锁效应逐步发展，而非某个事件立即触发。逐渐累积下来的，到本周累积到最后一步，所有资金开始买成分股。

问：从去年 12 月份开始，市场都在讨论小市值的风格转向问题。管理人有的认为变化不是一蹴而就的，量化来得及响应，而且换手率很高，如果频率足够高就可以追赶上这样的风格变化。您如何看待这几种答案？这波行情过后，大家对于预设的态度是否会发生变化？

玄信：大家都忽视了这个问题，包括我们自己都没有预测到这么快。因为如果偏小市值自己就可以平仓。现在因为交易受限大家不平仓，全部转到成分股，这样所有策略都失灵了。无论做什么，除非做 t0 的不会亏损。在本周所有选股情况下，无法有任何策略能够抵挡。

问：在调整成分股时，是否直接将选股域定在成分股以内，例如从全 A 切过去还是硬切，后续是否涉及精细化处理因子？

玄信：各家的情况不同，例如我们在，在成分股里进行微调，超额权重的上限正常情况下设置为 0.3%左右，现在到了 0.1%。

问：春节之后您认为仍然维持防御状态，还是有其他观点量化？

玄信：我们仍然需要维持防御状态，我认为大盘 Beta 下行空间有限。

问：投资人端是否会要求赎回或切中性部分？

玄信：我们中性、指增的情况较好，只要是行业平均水平，另外以高频起家，除了选股之外，还有持续上涨的策略。赎回压力基本可控，最近股价上涨，中性客户要求平仓赎回，我们会根据监管要求处理。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

目前市场高压仍然聚集在 DMA。

问：您在会议开始时提到这次大型回撤将引发一波行业洗牌。对于一些风格稳健的机构而言，这是否是一些新的机会？接下来行业层面会发生什么变化？

玄信：我们希望在行业能够平稳度过这波风波，参考国外发达国家每次出现问题后，我们都会学习最近的行为，监管也会对整个过程进行评估，以便更好地监管极端风险。

量化过去几年的稳定收益，过程中确实存在风险，相对于其他固收产品，量化一周回撤 5%很快就会回去，并非像真正的行业爆雷。

问：接下来策略升级方向？

玄信：今年我们首先学习从 1 月份到现在交易行为以及基差。从 1 月市场急剧下跌，雪球触发敲入基差扩大，量化平仓和 DMA 又把基差打回去，雪球又平，到 2 月份之后雪球基本出清。基差波动非常大，基差一旦增强直接买股指其实就可以了。昨天（2 月 7 日）ICIM 的 2409 贴水 17%，然后再跌周一基差一点就触发了全平，并且需要考虑一些特殊交易限制，因为这部分已经超过了所有策略研究的意义。

问：关于投研层面，您之前提到 5000 个因子会导致底层较为臃肿。因此您重新将 5000 个因子通过非线性模型的方式制作成 100 个左右的特征，然后再制定策略。请问这样做的目的是什么？中间的流程是什么样子，究竟有什么优势？

玄信：我们希望平衡解释性和复杂度。

我们底层有一半是人定义的因子，另外 45%是机器挖掘的因子。由于因子层面较多，我们想将整个特征提取到信号特征层面。主要使用深度学习树模型，从数据中寻找规律将其合成为信号特征，共 100 多个。囊括不同预测周期，如几分钟、几小时、几天、几个小时，再在上层，我们通过线性方式将其组合进行交易。

所以我们有三层操作信号，从因子到信号特征再到最后组合模型。早前从因子到信号主要用统计方法继续学习，解释性不多，只能看到原始输入。现在从信号到整个模型，我们使用线性方法可以有更强解释性，每个信号特征会给它做标签，包括基本面、日内量价、资金量价、另类。到最后模型，我们会对信号特征的表现进行跟踪也就拥有更强解释性，然后再从因子到合成。

问：回顾 2023 年，您认为整个市场策略层面发生了哪些变化？

玄信：市值之所以被称为风险因子，是因为存在风险。我们在 2023 年一开始就预测到了风险，1000 指增的市值风险一直控制到-0.1，完全没有赚到去年微小盘的钱。去年我们实现了 1000 指增的超额 11%，500 超额达到 14%，所以这波回撤我们并未受到影响。另外，我们最初想制定稳定长期策略以实现持续发展。所有模型中包含 30%的基本面因子，同时因为我们最初进行高频交易，所以日内执行采用了所有高频相关措施。

2023 年，我们在机器学习方面从 100 多个信号增加到 130 多个信号特征，同时加大了深度学习非线性模型的投入，2023 年增加了十几个投研，都是从深度挖掘日内量价和基本面因子。虽然 2023 年基本面因子失效，但是我们始终保留这部分。如果只使用量价，那么容易偏小市值。所以我们可以利用这部分资金抵消偏小市值的问题。

时间上来看，2023 年市场首先炒作 AIGC，当时行业集聚在传媒通信和 Ai 方面。因为机器学习的预测性不够强，很难，所以我们后续也加强这方面的模型，基本在 3 月之后情况就相对稳定。

问：公司因子优势包括深度、高效和快捷，这三个方面如何理解？

玄信：大家所做的事情都与考试相同，关键在于谁做得更深入、更细致，我们在数据方面做得更加细致，包括逐笔和海

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

量数据。在南京建有独自的机房，拥有 500 多张显卡，在算力原始数据上的投入巨大。股票方面的因子特征研究员都来自清北，然而无法应对当前行情。

问：如何追求模型在不同市场风格下的平稳度，你们会对各种风格市场进行划分吗？

玄信：我们主要从数据层面进行讨论。主要关注风险因子收益，包括交易量刻画和重要监管政策发布日期。策略上的回撤不仅关注数据的每一段，而是分时段关注表现。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

12 致诚卓远：《破局》量化人的宏观书



“你们这么低调，为什么会接受我们的访谈。”

访谈结束，致诚卓远创始人史帆穿着一双毛拖鞋，把编写组送到了电梯口。“因为我有一本新书要发了。”说完他自己笑了起来，“这本书不是讲量化的，只是名字有点俗气，叫《破局》。”

在文章中，他如此写道，“累牍知著，拙出素心。愿本书能为改造我们的世界做出一些贡献。”

01 量化方法论：物理与量化中的推导

致诚卓远，即便在 Geek 频出的量化行业，也是有特殊之处：总部在昆明，纯本土方法派，早期量化四大天王，但保持低调，在 2021 年量化大爆发的大年，却保持着自己的募资节奏。创始人 2023 年写了本书，却是研究宏观经济学的。仿佛在持续表达一种态度，或是打破一种偏见：“量化不只是代码、技术、我们也在关心着这个世界。”

史帆，北京大学物理学院 00 本 06 硕，毕业后加入建行云南省分行。

物理学严格遵守：现象→归纳→假设→建立数学模型解释现象→求证→修正模型→再求证的逻辑，对于规律归纳以及建模，在史帆看来物理主要通过观察现象，并且试图用一两个公式整合现象，然后将其进行外推。数学做出假设用数学方法推导整个世界，观察假设下的世界是否与真实世界相同。两者本质是找到世界过去发生的事物，在过去发生的事物未来会以某种方式重现，与量化投资的逻辑非常接近且一脉相承。

在银行，史帆从事的是金融大数据分析的活，例如银行体系内对于违约概率的预测方法，与量化一致，本质都是根据已知的信号去研究未来的收益率。“采集到一些 X，然后我去预测一个我目前是看不到的，或者是我未来看到或者是我永远看不到的一些 y，然后这些 y 可能有一些意义。股票也是如此，因子是看到的信号，研究未来收益率以及两者的相关关系，因此我们将其较好地迁移到股票上。只不过银行数据做拟合，R 方可以达到 0.9，但股票因为噪音特别多，R 方可能 0.1 都达不到”。

于是史帆就开始了“一个人瞎鼓捣，鼓捣了七年。”2014 年创立致诚卓远量化投资私募，任总经理及基金经理。几乎靠自己一个人便完成了整套量化模型的构建。历经 2014 年金融股爆发、2015 年股灾、2016 年贴水成本上升及英国脱欧公投事件、2017 年小市值因子失效、2018 年动荡下跌、2019 年量化交易拥堵等行情，

一开始史帆基于日线做策略，后面逐渐扩展到分钟线，体系越来越庞大。单枪匹马了好几年，史帆发现很难顾及策略广度的延伸，便选择在上海设立办公点，形成公司的因子研发中心，而史帆则从事最后的组合优化。

现在史帆的工作量加大了，要看所有的因子、理解并选择，但他并不认为这是一种负担与效率的浪费。“这样其实可以弥补层级带来的许多制度性障碍，反而能够提高效率，减少许多重复性的工作。一个团队有 4、5 个人，也就有一些工作被重复了 4、5 遍，但在这里这些内容都成了一个整体。”

02 策略的克制：统计套利短周期细分赛道

自己鼓捣的一点好处就是整套研究体系与主流研究体系差异比较大。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

致诚卓远投资理念为统计套利，核心在于股票市场未来的运行方式与过去存在某些特点上的相似性，可以用市场过去场景的条件频率近似地替代未来市场的条件概率。

致诚卓远的策略以量价因子为主，布局在短周期、偏中高频的量化细分赛道。其收益来源更多是来自纯 Alpha，即股票在其交易中枢之间的交易性机会。这类的机会不依赖于常见风格或热点追踪，而更多关注的是个股的价格偏离，也就是当股票本身存在较大波动和分化的时候，策略能够更好地获利。目前模型已将 Alpha 策略和程序化 T0 有机结合，全行情历史回测，每日复盘，3 到 6 个月定期更新迭代。

这样的策略模式就决定了其策略具有较强的稳定性。因为策略的收益来源是股票本身的价格偏离，而并不是寻找可以获利的行业或风格，只要股票本身存在交易和波动，策略就能够有获利来源，而这种获利来源并不会轻易地消失。

一条赛道上可以承载的容量是有限的，从路径选择到收益率，致诚卓远和别家机构的相关性都很低，比较典型就是 2021 年量化机构发生大回撤的时候，致诚卓远收益还涨得很快，实际上就是因为选股逻辑不同，在市场减仓的时候加仓。“我这套东西实际上是外界是不了解的，所以这个地方存在一个单向透明。”史帆说道，去同质性使得策略衰减程度较弱，在一定时间内，市场容量不易受到同行业的冲击。

“我喜欢跟踪业内新鲜的研究报告，如果出现颠覆性的方法论，我们会投入大量精力学习。观察它们对我们解决的问题有无帮助。我认为最大的变革可能是完全新的东西，如果出现，旧方法将彻底被打败。”史帆说。

致诚卓远采用了协同化的开发模式，投研体系上，由量化研究员负责另类数据研究及多样化数据源的收集和整理，同时部分研究员从事因子挖掘及结合学术前沿理论对机器学习/深度学习等方法论进行优化；由交易人员负责交易算法的提升和保证日常交易的执行；再由 IT 系统架构工程师分阶段完成和各家券商系统反采流程，提升内部硬件、机房及相关配置性能，在多方面保证投资策略与交易策略的最优化执行。

从研究和交易层面来看，相关人员每周定期向基金经理输出研究结果，由基金经理统一在策略层面定期更新迭代。基金经理史帆同时在纵向继续精耕原有以中高频量价 Alpha、程序化 T0 为核心的模型体系。目前，投研团队已搭建市场成熟化流程作业投研体系，以老带新，逐步进入投产阶段。团队时刻保持对于量化投资的热情，将充分的精力投入在模型上，不断更新迭代策略，扩充收益来源以及因子数量。

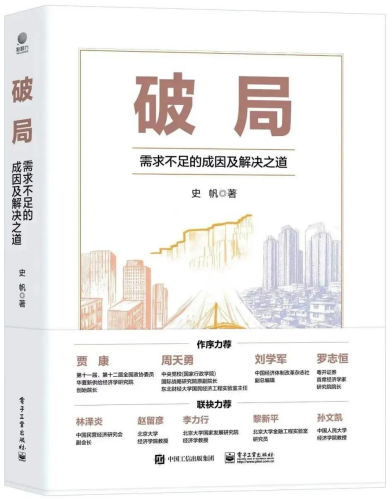
2014 年的 11 月 19 日，在公司在成立第二天，市场暴涨遇到暴涨时，大盘指数超了六七个点。市场通用做法是做空 IF 沪深 300 股指期货，许多机构在后续市场反转之时落入陷阱甚至湮灭倒闭，“如果我们选择了诱惑，那么成立的第一个月就会倒闭。”初创时遇到的极端行情，其风控理念可能就会镌刻进 DNA 整个发展长河之中。“我们不去做短期看没有风险，但本质是有风险的东西。”致诚卓远坚持不做人为的风格和行业偏离。对于需要严格对标的量化对冲策略，保持无 Beta 风险敞口，完全对冲。此外，致诚卓远自建了事前、事中、事后完整的风险管理系统，均实现了程序化系统监控。

当然选择也需要定力：“当别人都在扎堆买股票并且都获得了不错的收益时，要不要跟风，要不要冲？”史帆在遇到这种情况时也很纠结，2020 年由于在纠结之后选择不去追市场风格，业绩跟同行相比有一定差距，大家都很煎熬，但随后凭借策略优势，2021 年 9 月公司业绩又排到了很靠前的位置。

控制规模反而不需要定力，史帆一直认为规模过大并无太大意义。如果规模达到 100 亿元或者 500 亿元，收入将翻 5 倍，但生活质量改变几乎为 0。并无太大意义。只是近两年，市场上不上百亿，好像就不好意思出来打招呼，致诚卓远才又把规模上控一点维持在 100-200 亿之间，保证规模能够与收益、风险相匹配。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

“我就是喜欢做研究。”或许与史帆的兴趣有关，他喜欢做研究并且希望快速得到验证反馈，而量化正好契合这个需要，也正因此，因为这个纯粹的起点，在面对规模增长利益增长的时候他选择了克制。



03 《破局》：量化之外的世界

回到《破局》，一个量化之外的世界。

高盛发布了最新关于《2024 年中国经济展望》的报告，预测 2024 年中国实际 GDP 增长率约为 4.8%，相对于 2023 年的 5.3%有所下降。放眼全球经济，世界各国都不同程度地陷入了经济困境。而我国在疫情后经济并没有如约报复性的增长，有人说中国 2023 年经济情况像是日本 90 年代后期低速发展的状态。究其原因是我国在出口、投资下滑，消费一直未能支撑起经济发展的情况下，陷入总需求不足的困境。

史帆认为：辜朝明关于资产负债表危机的观点，从现象上描述了危机过后日本企业不愿意放贷，将利润用于持续归还贷款的现象。但其对原因的解释归结为不同条件下微观主体偏好会改变，这很难让人信服。

针对为何会发生类似的危机，在《破局：需求不足的成因及解决之道》中，史帆提出了一些有别于前人的观点。其逻辑概括为：拿到资本收益的部分人群，在特定情况下，可以进行的投资是有上限的；同时由于其财富经过 1980 年以后不管是西方偏自由主义还是中国进行改革开放 40 年的积累，总量已经较高，因此其边际消费倾向也是低的。于是这部分人群注定无法将其获得的收入全部用于投资或者消费，最终只能形成某种形式的债务金融资产。

而债务金融资产的产生，必然离不开社会上其他成员债务金融负债的成对产生。这在 2008 年前的美国表现为部分完全没有偿还能力的个人的负债，并最终导致次贷危机；而在此之后表现为几乎世界上所有国家国债永无止境的飙升。一旦试图控制财政赤字不再增加，政府负债不再飙升，那么这部分无法投资或者消费掉的钱，就将导致总需求出现缺口。

史帆认为，解决之道在推进“国股民营”。

坚持稳中求进，以进促稳，先立后破。社会要能稳定持续增长，所有问题的核心在于，各种带有比例属性的值（收入占比、资本回报率等）不能大幅度变动。而资本收益者由于承担了经营风险，稳态下其财富增长速度一定会高于社会平均。这二者间形成了不可调和的矛盾。

用这套体系看待类似日本失去三十年，或者大萧条，其结果和马克思所描述的资本主义固有危机是接近的。只是从风险的角度会比从剥削的角度更加能够定位资本收益率始终超越社会平均增速这一现象的原因，也更加能明白为何过去四十年全球经济高速增长时这个问题不严重，而又为何在此时这个问题才展露出来。核心都是因为之前四十年，较高的资本回报率起点使得企业获得的收入能以资本回报率下降的代价超越资本增量上限进行投资；而资本回报率下降到 4%以后再也无法持续这种状态。

同时从风险的角度出发，也就将不再存在阶级对立，因为不管是资本收益者，还是劳动者，都是这个经济稳态里面的一个螺丝钉，任何个体并没有能力违背这个经济规律。如果使用阶级斗争的方式使得企业的资本回报率低于风险溢价，那么企业自己的经营也难以为继，将会选择缩小生产规模以使得资本回报率提升，而这是一个多输局面。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

20 世纪已经无数人论证过公有制是克服资本主义固有危机的唯一选择，只不过随后在实践中，受限于当时的历史局限性，选择了计划经济这样一条路径。而后续苏联及中国的实践中证明计划经济并不能很好地激发劳动者的积极性以及完成资源配置的任务。在我们认清所谓资本主义固有危机，本质是风险溢价使得现实的资本回报率均衡点偏离了自由市场主义者理论上所需要的 r=利率这个均衡点后，就可以抛开有关阶级的讨论，仅仅把这个问题作为一个小的技术障碍来，想办法找到一种方法来修正它。

在这里，史帆提出了一种新的所有制形式——“国股民营”。使得公有制以及完全的市场竞争，以及市场配置一切资源可以同时达成。

“国股民营”的实行，离不开两个前提。一是资本收益者和企业经营管理者的分离；二是资本收益者和投资决策者的分离。这在我国可能尚处于发展阶段，但在美国这样一个发展到金融资本主义的国家，这两个前提都已经是存在的。

目前美国市场上管理层持有公司股权超过 10%的上市公司仅占 3%，可以说在美国这样一个成熟的资本市场上，已经完全实现了经营者和所有者的分离。美国股市另外一个显著特征就是机构投资者已经占据绝对主流。大量的财富并不会由其所有人自行交易股票，而是通过至少一层以上的委托代理关系交由其他专业的投资机构进行打理。资本收益者、投资决策者、企业经营管理者这三个身份，通常已经分别由购买各种基金的购买者、基金管理人、职业经理人三个不同的人来完成。

在这个基础上，只需要由公有资金来替代基金购买者，而维持基金管理人和职业经理人角色不变，就可实现完全的“公有制+自由市场”。在这样的体系下，政府仅仅需要通过不同的资管计划，以一个网络的形式多层间接持有最终全市场大部分公司的股权，同时并不干涉企业运行，即可实现在将所有企业利润收归国有的基础上仍然完全保持市场化运行。

实际上如果财政将资金分散至各省管理，各省仅用其投资收益支撑自身财政运行。由于任何一家公司实际上追溯其所有者可能是所有省份都会持有其股票，此时也没有任何一个官员或者政府组织可以对其经营直接进行干涉。除非这些企业涉及非常关键的国计民生，那么中央政府可以特殊地将该企业的股份划转至中央管理。

“国股民营”的方法如果希望在中国尝试，距离落地还有一段路要走，表现在中国上市公司中，管理层持有的股票份额仍较高，以及投资者仍然以散户为主，还需要进一步的针对性改造以实现金融资本和产业资本的分离。如果未来，能够得以实施，恰好也印证了马克思主义的观点，社会主义是资本主义发展的高级阶段。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

13 中欧基金：曲径通幽 公募基金的基本面量化之路

择高处立，向宽处行。在私募基金为主要玩家的量化领域，中欧基金的量化团队，以丰富的基本面和另类数据因子，和多元化的人才布局，找到了一条通向高 Alpha 的公募之路。

曲径

复旦大学数学系学士，卡内基梅隆大学计算金融硕士。

2007 年起任千禧年（Millennium Partners）高级研究员、投资经理，中信证券另类投资部投资经理。

2015 年加入中欧基金，任量化投资部总监。

01 量化踩踏启示录 量化人无法避免的课程

2007 年 8 月，曲径刚刚加入纽约的千禧年基金总部，在华尔街工作第一年，就经历了 2007 年 8 月的 Quants Meltdown。“早期的从业经历，奠定了我职业生涯的底色，从第一年起我就学习到，做模型要时刻避免与其他人发生拥挤，尽量往人少的地方去挖 Alpha。”

太阳底下无新鲜事，所以当时钟拨到 2024 年 1 月的时候，曲径并不惊讶。

2024 年 1 月，A 股市场尤其是微盘股的剧烈调整，引发了一波量化私募的信任危机。“和 2007 年的美股一样。MIT 的 Andrew Lo 教授有一篇论文，分析了 2007 年 8 月美国量化基金集体回撤的本质。分析认为，是由部分基金大规模且快速的平仓，引发了短期价格冲击，诱导了同质化策略集体回撤，进而滚动发生赎回和策略踩踏。A 股这次的量化危机，基本是同样复现。”

关于本次深幅调整，我们究竟应该学到什么？关于策略拥挤、关于行业变迁、关于监管变化。

“易经讲，亢龙有悔，当策略曲线‘完美’得超过常识，往往蕴含着后来的危机。”曲径看来，“公开资料显示，2006 年时，美国头部量化机构持仓重合度有 60%，交易非常密集，而踩踏也是冲高见顶之后的解体和分化。同样，A 股量化策略呈现了较高的重合度，我们可以预见到拥挤踩踏迟早会发生。”

因此，在曲径的带领下，中欧基金在自己的 500 指数增强、空气指增类产品中，同质化策略较少，微盘股更是严格控制。

02 以人才为核心 以研究为基础

在私募基金为主要玩家的市场，中欧量化以完善的投研流水线为基础，吸引了一批优秀人才。

“中欧已经搭建了一套健全的多层次投研框架，投研工作基本以流水线方式进行，包括数据清洗和处理，另类数据的引入，因子的编制和测算，多类组合方式。以及前端和事中的风控模块、产品管理和归因分析等。未来的重点是对策略进一步的迭代、筛选、评价和执行。同时引入更多先进的技术以提升当前框架，包括但不限于深度学习模型、更丰富多元的另类数据分析等等。”曲径说，“我们会以投研框架的重要节点为抓手，持续引入适配人才。”

当前中欧基金的投研队伍，有多位从百亿私募加盟的投资经理，从千亿跨境投资机构加入的另类数据科学家，以及自主培养的一批核心投研人员。

“人才是投资收益的基石，外部经验人员的招聘，我更看中人才的适配性。而初级研究员和应届生的选拔，更重要的是

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

研究潜力。通过科学培养，职位适配，让每个人都能发挥出自己最擅长的部分，把智慧转变成研究成果，并被合理运用到投资中。同时在组内会有适度的换岗、调整，这样做的原因是希望研究员包括投资经理找到自己最擅长的部分，同时也更具有全局观，以激发更多的创意和思路。”

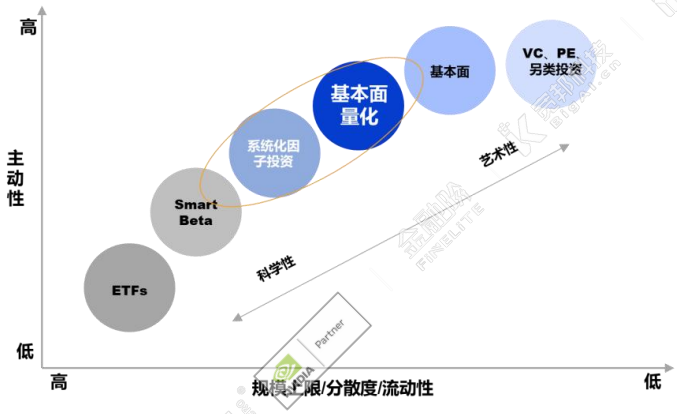
截至 2024 年一季度，中欧量化投资部在职基金经理的管理规模（包含季报所披露的基金经理兼任的私募资产管理计划）已经突破 200 亿，而中欧基金的量化投研团队并没有急剧扩张，对此曲径也有更具体的规划。（数据来源：基金定期报告，截至 2024/3/31）

“2023 年量化策略明显进入了拥挤期。如果把风险因子剥离，仅从 Pure Alpha 上观察，2023 年四季度，普通价量因子的 Pure Alpha 已经难创新高，同时，新价量因子的 Alpha 衰退速度也在加快，说明量化策略逐步进入了资金过度拥挤的状态。”

“与此同时，我们观察到，从业人员在 2023 年大批量扩容，这也是一个过度拥挤的前兆。因此我在 2023 年放慢了招聘速度，我认为不妨等市场慢下来，再补充合适的投研人员。行业发展的长期趋势一定是渐进的，而短期看具有周期性。量化投资本身是一个系统化的工作，只有在每个研究节点都做出更好的成果，让策略逐步精细化，才能在激烈的竞争中胜出，业绩才能经得起 Alpha 周期波动的风险。”

“不追求盲目的扩张，而是以策略的重点研究方向为基础，希望每一个投研人员的加入，都能为当前体系再增厚 Alpha。”曲径说，“中欧的量化投研是一个敏捷的队伍：不赛马、不内耗、能打胜仗。”

03 丰富的另类因子 让投资向宽处行



中欧量化在行业早期，就充分重视基本面和另类因子，在数据上和人才上都较早地进行了布局。2022 年和 2023 年，基本面策略和另类数据的 Alpha 并不突出，似乎不是市场主流。同时，量化投资发生同质化倾向：广泛使用相同的数据源，包括但不限于日内价量、会计报表、一致预期数据等；量化模型随着行业人才的流动，也逐步趋同。而中欧基金一直在基本面和另类数据等相关领域持续投入，潜心专注在自己擅长的领域，通过基本面模型，相信优秀的企业的利润增长，终会体现在股价上，创造长期合理的超额回报。对此，曲径有自己见解。

“任何策略，不可避免都是周期性的，相较于基本面和另类因子在前几年的不错表现，2022 年和 2023 年以来，A 股市场持续在消化高估值，市场的主导权转移到交易类策略这边。但我仍然认为，对于规模天花板的打开，对于 Alpha 的多元化和稳定性，基本面和另类因子或是更好的解决方案”。

另类数据和基本面研究的门槛较高，对行业基本面的理解是使用另类数据的重要前提，而量化团队在基本面研究上又比较薄弱，如何突破基本面研究的壁垒，成为量化团队亟需解决的核心问题。2024 年初，中欧基金斩获过金基金奖、金牛奖双料大奖的主动基金经理王健，加入了量化部，成为行业中主动基金经理加入量化投资的第一案例。这次“转会”受到了同业很大关注。（注：王健在管产品中欧新动力混合斩获七年期开放式混合型持续优胜金牛基金奖（中证报，2022）金基金・偏股混合型基金十年期奖（上证报，2022））

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

“另类数据和基本面研究，是有明显的研究壁垒的，和运用各类模型深挖交易数据因子不同，基本面研究的时序往往更短，逻辑性强于相关性。基本面基金经理加入我们部门是一种双向赋能，一方面主动基金经理很需要另类数据产生的价值：通过量化数据激发投资灵感、通过另类数据跟踪个股景气。同时，也为量化团队带了不可多得的基本面经验，揭示了很多基本面研究中的核心变量。抓住核心变量，同时应用逻辑上正确的另类数据，对量化研究员进一步了解资本市场的底层运作规律是一种极大的助力，并且能够帮助他们通过新的角度去增厚组合的 Alpha。基本面研究和量化系统，在我们的投研框架下，形成了共赢关系。这种丰富的投研土壤，有利于研究团队创造出更稳定、慢衰退的量化策略。”

04 创业文化 以奋斗为本

“中欧是一个创业团队！”

曲径在 2015 年加入中欧基金时，正赶上了中欧的高速发展期。办公室的空间，匹配不上投研人员的扩充。某个周一来上班时，董事长办公室的墙消失了，拆除的位置，变成了几排研究员的工位。

“那时候大家都深度投入在工作中，研究、投资、协作，整个公司从上到下都保持了创业状态。而这种企业文化，持续到了现在。”

如今，随着中欧基金的发展，量化部也逐步扩充为以主动投资方法论为基础的基本面量化组（Quantamental Investment），以多元量化模型为核心的系统化投资组（Systematic Investment），以及公募指数组（Index Funds），三者相互协作，成为了上下游的关系。

当然，中欧量化也经历过压力时刻，2022 年-2023 年市场对于部分公司的估值定价持续下行，基本面为主的量化策略整体没有产生足够 Alpha。

“另类数据在国内并不是量化投研的主要挖掘方向，但是以成熟市场经验来看，基本面因子和另类因子所带来的 Alpha 将会占比越来越大，过去几年我们的资源在这两者上投入较多。但是，当前 A 股相对于成熟市场，个股收益率的区分度（Return Dispersion）还是非常大的，以交易特征为基础的价量因子，依然有足够厚的 Alpha。因此在接下来的战略方向上，我们需要在研究上加强对交易特征和价量因子的重视程度。从 2023 年开始，一面继续拓宽另类数据，一面逐步增加了对价量因子的重视，通过从百亿私募引入人才，补齐了技术指标的相对短板，在坚持基本面逻辑选股的原则之上，让中短期策略的锐度得到了提升。”

目前团队基本面量化组和系统化投资组的人数基本相同，除了注重数据及基本面因子储备的传统优势，数据处理也更为精细化，充裕的算力投入也为新一代量化模型的开发提供了基础。两个团队协同工作，成为数据和逻辑的上下游关系，极大地提升了整个投研的工作效率，形成彼此驱动。

接下来，曲径还将为团队持续引入人才，“我们持续在寻找深度学习等领域的人才，希望整个团队在技术储备上，能迎接量化投资在新环境下的挑战。”中欧基金非常珍惜人才培养，在新人职业生涯的初期，给予充分的空间和试错机会，在更长期的时间维度，使团队保持足够的竞争力。“中欧基金崇尚平等共赢的工作氛围，希望团队成员能 5 年、10 年，以致更久地一起工作、成长，共同打造一只优秀的公募基金量化团队。”

基金有风险，投资需谨慎。以上内容仅供参考，不预示未来表现，也不作为任何投资建议。其中的观点和预测仅代表当时观点，今后可能发生改变。未经同意请勿引用或转载。基金管理人承诺以诚实信用、勤勉尽责的原则管理和运用基金资产，但

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

不保证基金一定盈利，也不保证最低收益。基金的过往业绩并不预示其未来表现，基金管理人管理的其他基金的业绩并不构成基金业绩表现的保证。您在做出投资决策之前，请仔细阅读基金合同、基金招募说明书和基金产品资料概要等产品法律文件和风险揭示书，充分认识基金的风险收益特征和产品特性，认真考虑基金存在的各项风险因素，并根据自身的投资目的、投资期限、投资经验、资产状况等因素充分考虑自身的风险承受能力，在了解产品情况及销售适当性意见的基础上，理性判断并谨慎做出投资决策。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

14 上海卓胜：“老派”人工智能量化的反其道

卓胜是一家新公司，2022 年 9 月获得牌照。直接从人工智能切入量化投资，但创始人马万里却认为这是：“正统且老派的做法。”

01 “老派”人工智能量化的反其道 投资是实用主义

马万里北京大学本硕毕业之后，先在新加坡进行了两年的金融工作，随后回国先后加入两家国内头部私募，从事多因子研究，并在 2022 年创立卓胜私募。

新兴的人工智能量化是卓胜的典型标签，公司每日 7x24h 机器挖掘特征，存有 100000+特征，但在马万里看来，公司并未跳出量化原有体系：“数据、模型、算法是投研体系中最核心的三部分，我们目前使用的从数据到交易整个链条底层，行业已经非常成熟，我们只是在此基础上使用组合方法和因子挖掘方法进行创新。”

传统量化机构的路径一般是从多因子过渡到人工智能量化，传统多因子部分以财务表现、市场行为、金融环境等为基本逻辑，利用多种估值统计手段，从多角度深刻挖掘大量的低相关性因子，利用优化算法进行因子合并，使其对未来股价具有高度的预测能力，选取强势股票组合获取收益。

卓胜从 2021 年初开始，即站在行业肩膀上，使用以神经网络为主的非线性模型制作 Alpha 信号，并进行交易。底层基本以高频量价数据为主，学习过往市场的微观结构和行为模式，特征形成非线性模型，最后到线性组合时，使其逐步演变为可预测的线性因子。其中量价占比 80-90%，基本面占比 10-20%，采用 100 倍或者 150 倍的换手策略。

但到 2022 年，卓胜又在反其道而行之，往线性模型发展，开始测试上实盘。

“非线性是我们最早制定的策略，经过这些年的发展，已经形成了一个成熟的策略体系，并不新颖。只是我们在实盘时使用这个东西较早且有效，在这两年的市场行情中表现不错，市场上很多人都知道我们的方向，我们因此原来主要的策略体系也是非线性模型。实际上在这之前，我更早研究的是多因子线性组合，只是后来在非线性模型上做得更好。”马万里解释道；目前，卓胜策略体系包含传统的多因子和非线性模型，两者结合生成组合信号。

线性模型会给策略体系带来优势，基本面占比提升，换手率不高，容量可以扩大，另一方面也能带来低相关性的补充以及提升稳定性，两者实现叠加效果。

到目前为止，卓胜非线性组合也有机构认为人工智能因为学习过快，所以很多时候陷入 Smart Beta 的陷阱，但在马万里看来，无论是线性还是非线性，都可以学习并刻画风格等各种特征。非线性模型的主要特点是适应速度较快，能够更快适应市场风格变化和行业转换，这与换手也有关，另一部分是它的底层以交易数据为主；第三点是模型学习能力较强。

非线性是否是 Smart Beta，需要从不同角度分析。从 2020 年开始，正是小盘和微盘的三年大行情，除了中间一段时间出现反转外，正好遇上非线性模型兴起，并且大量被大家使用。两者自然组合选择小票模型是属于片面判断。

“资方现在非常担心 AI 模型表现良好，在市值风格切换后到 2024 年能否适应？”编写组问道。

“更大的问题难道不应该是，线性模型在过去三年的大行情中一直无法选择表现好的票？”马万里笑道。对于量化机构而言，市场确实需要差异化竞争，然而对客户而言，最终关注的就是收益率。收益率来自何处并不重要。作为私募基金，不能为了标榜自己要做什么，而应是给客户提供更好的投资体验，这是不行的，投资肯定是实用主义，而非理想主义。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

02 数据探索 从中频出发 向低频和高频方向扩展

大力出奇迹，“上亿的参数才会接触到 90%的市场规律。”马万里曾经如此形容，这是否意味着参数越多才能够更完整地描绘市场？

马万里解释，参数越多，其他方面性状也不会逊色。模型越大，对模型的预测力和学习能力肯定有很重要的帮助。但在量化投资中，人工智能是一种可以帮助进行搜索研究的辅助工具。但想使用它直接制定策略，不太容易实现。“我们是一家正统且按照自己量化梳理逻辑进行研究的公司，非常强调在整个量化投资过程中尽量减少犯错，并且尽量减少人为错误。谁犯错误少，谁就发展得更快更好。”

所以卓胜采用的也不是端到端的做法。“端到端对原始数据输入后可以产生的内容完全没有把握。在短期内能学到一些东西，表现不错，然而我们对他没有长期的信心。”所以卓胜的路径是向模型输入衍生数据，所谓衍生数据，即是从原始交易所出来的交易数据中进行二次开发挖掘，包括多频段时点切片、隔日波段等等，根据自己的交易逻辑提取出的一些因子。这个过程中会做到高频数据降频处理优化，从而提高输入质量，包含更多信息含量，这些数据本身就带有 Alpha 信息，放入特征工程中进行原始输入，生成的特征与直接丢入交易所的原始数据产生的特征不同。

随着策略的完善，卓胜除了增加策略类型，也在扩充各类数据、补充交易频段。公司原本以高频价量为主。结合日频价量和较少的基本面数据，另外包括分析师数据和新闻等。

2023 年，卓胜的另外一个工作是从中频出发，同时向低频和高频方向扩展。从 2023 年底，卓胜开始测试高频类的日内 Alpha 短周期信号，以及算法上的更短周期预测。在加入高频和低频领域后，做叠加滚动优化，补足了原本体系中缺失部分，希望能为投资人提供更全面的投资体验，确保策略更加健全。

03 行业的未来：普及量化任重道远

一年的实践，卓胜规模已经到达 12 亿，随着规模的扩张，个人的精力也会发生变化。马万里将 80%-90%的时间用于策略管理和投研团队管理。“我们需要思考如何提高团队建设效率，在底层基础数据和 Alpha 因子特征这两个层面上完全平衡，大家共同研究。”现在卓胜还设有策略研讨会，使得每个人都清楚自己研究的方向，数据方面也使用相同的数据，避免重复赛道竞争。

量化投资的核心一直是人。“从长期角度来看，只要团队表现良好、人数和基础工程能力充足，大部分团队都会朝一致方向发展。市场数据并非内幕，人员和服务器都可以购买，因此最后的决策与策略负责人有很大关系。”

管理半径扩大，不可避免会丧失微观视角，但马万里坚定地谈到，“我对每个人研究的因子非常清楚，因此调配可以较为容易实现。”2021 年，马万里一个人即可完成所有数据和交易，如果自己不犯错就不会出错。随着规模、人员扩大，也在关键流程上逐渐实现自动化，叠加外部监控，团队管理效率也在提高。“如果 30 个人同时犯错的概率就会降低很多。”

机构在避免自己犯错，向公众普及量化是一件任重道远的事情。“每次路演都在向客户讲解量化投研的思维逻辑体系和交易逻辑，但大部分客户对量化仍然较为陌生，专业背景知识也较少。机构投资者主要从尽调指标出发，对公司的了解不会过多猜测。但个人投资者不太了解，就会有猜测，就会有一些想象空间。”马万里认为，不断推动量化投资教育是全行业的事情，中国投资者教育处于起步阶段，在投资者金融教育基础不牢的情况下，要实现教育量化投资更加困难。在许多客户还未理解私募概念的情况下，与其讨论量化私募也是非常困难。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

“今年发生的事情让我感到震惊，竟然有不托管的私募。我们第一次听说以前不知道有不托管的私募。希望这些事情最后得到规范处理，这样对于行业健康发展，大家的印象观感会更好，更有人愿意倾听。”

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

CHAPTER

第七章

附录



量化课程及活动链接—https://BigQuant.com/quantpaper/2023

01 数据挖掘

人们针对自己的问题，通过一定的数据分析过程和方法在海量的、杂乱无章的、不清晰的并且随机性很大的数据中找到其中蕴含的有规律并且有价值和能够理解应用的知识过程。在量化投资中，数据挖掘的主要技术包括关联分析、分类/预测、聚类分析等。

02 机器学习

MachineLearning，计算机通过模拟或实现人类的学习行为，能够处理大量非结构化数据，以及文本、音频、视频和图像。以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。

运用到量化投资，机器学习的方法，就是金融计量方法的延伸，可以从线性模型到非线性模型，低维度到高维度，运算速度非常快。比如可以通过对成交量，波动率，价格，涨跌幅等等特征的学习，也可以通过对新闻文章、分析师笔记、分析师通话等这些文本情感特征进行学习，而通过对这些特征的学习，就可以生成对股票不同周期的预测，再把这些策略做优化组合，就形成了一套完整的策略。

03 深度学习

DeepLearning，深度学习是机器学习的一个子集，是机器学习中一种基于对数据进行特征学习的算法。深度学习也是神经网络技术的一种。它的特点是，只要有足够的学习数据，神经网络自身就可以将数据群的特征自动提取出来。在此之前的图像和数据的解析，需根据各个数据和问题进行析取算法的操作。但是，深度学习则不需要人为操作，而是自动提取特征。稍微粗略地说，就是只要向神经网络注入数据，就可以任意提取特征。

04 行业中性

与基准指数的行业权重保持一致，没有任何偏差，这也是指数增强基金的通常做法。

不同的行业的波动率是不同的，譬如银行股、公用事业、必需消费品行业等，波动率就很低。那么不进行行业中性的调整完全按照波动率排序选股的话，就会导致选出来的股票行业权重和宽基行业权重偏离非常大，这会造成行业偏离的风险。

05 对冲成本

国内常见的对冲方式包括期货对冲、融券对冲、收益互换空头对冲以及期权对冲。国内主流的中性策略，是基于量化模型构建指数增强股票组合，同时做空对应指数、相等规模的股指期货。中性产品的收益构成，即为多头端收益+空头端收益-对冲成本，也就是超额收益α+做多指数β-做空指数β-对冲成本。由于做多与做空指数的收益相互抵消，所以最终产品收益=超额收益α-对冲成本。

国内主流中性策略通常利用股指期货对冲市场风险，成本主要源于股指期货价格和现货价格之间的差值，也就是基差。当期货价格＞现货价格时，基差为正（升水）；当期货价格＜现货价格时，基差为负（贴水）。对冲成本对产品收益的贡献并不一定为负。比如，在升水行情下建仓，基差为正，对冲后的收益则为超额α+正基差。

值得注意的是，对冲成本在建仓时就已确定。短期基差的收敛和扩张会导致产品净值的浮动，但只要不平仓，短期基差的波动并不会影响最终期货交割时中性产品的收益。

05 过度拟合

过拟合（over-fitting）是所建的机器学习模型或者是深度学习模型在训练样本中表现得过于优越，导致在验证数据集以及测试数据集中表现不佳。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

07 多因子

在金融领域，因子通常被定义为对于从某种角度和方式、对于当前或是未来市场的总结或预测。对于当前市场的总结和刻画的因子，我们通常称为风险因子；对于未来市场的预测的因子，我们通常称之为 Alpha 因子。另外，在传统数据科学和机器学习领域，其更普遍的名字叫“特征”（Feature），“因子挖掘”类似于机器学习中的“特征工程”。

■图表 10 种常见风格因子的定义

因子名称	因子指标构建	因子方向
市值（Size）	总市值	小市值-大市值
估值（Value）	市盈率、市净率、市现率	低估值-高估值
成长（Growth）	净利润同比增长、营业收入同比增长、净资产收益率 TTM 等	高成长-低成长
红利（Dividend）	股息、每股股利、股利支付率等	高红利-低红利
流动性（Liquidity）	当月以来换手率、三个月换手率、六个月换手率等	低流动性-高流动性
动量（Momentum）	月涨跌幅、六个月涨跌幅、十二个月最高价等	低动量-高动量
波动率（Volatility）	波动率（一年）、十二个月振幅、本月振幅等	低波动率-高波动率
杠杆（Leverage）	流动比率、资产负债率、长期负债占比等	低杠杆-高杠杆估值
质量（Quality）	每股收益、每股息税折旧前利润、销售净利率等	高质量-低质量

08 因子组合

因子整合是指在量化投资模型中，将多个因子进行组合和优化，以构建更有效的投资策略的过程。不同的因子能够反映出企业基本面、价量趋势等不同维度的信息。但是，单个因子的力量远远不能满足需求。在挖掘出了优秀的因子后，还要用适合的方法将这些因子搭配和组合，使得模型能取“百家之长”，这个过程即为因子整合。因子整合的目的是为了提高模型的预测能力和稳定性。采用科学的因子整合方法可以有效减少模型从预测到实现过程中的损耗。通过组合多个因子，可以充分利用不同因子的信息，从而降低模型对某个因子的过度依赖。此外，因子整合有助于消除模型中的冗余因子，即那些对股票收益解释力较弱的因子，进而提高模型的效率和精度。

09 事件驱动

对于一般的多因子策略，它会对于每只股票在每个采样的时刻做出打分和预测，而事件驱动策略会预先定义“事件”或“场景”，寻找当“事件”或“场景”发生时收益最大化的资产，或者降低当“事件”或“场景”发生时预测未来收益的难度。对于“事件”或“场景”的选择上，一般有两个特点：

稀疏性：“事件”的特殊性导致其出现的频率必然不会太高，覆盖的股票范围往往也只是全市场的一个子集。因此，在研究的时候对其收益的评估和归因、风险控制等会和传统的多因子策略有很大的不同；

可解释性：由于要求“事件”本身的特殊性，“事件”本身的选择一般会具有较强的经济含义和可解释性。常见的事件通常会偏基本面或是跟公司财务和运营相关的事务相关，但也有一些特殊的事件是由“量价”的信息驱动，比如一些 T0 或择时策略中会用到“突破”、“反转”或“联动”的概念。

10 指数增强

指数增强策略是利用量化投资方法进行主动管理，以股票市场指数（如沪深 300、中证 500、中证 1000）作为业绩基准，在全市场范围内选股，并保持满仓运作，以同时获取指数收益（Beta）以及超越指数的超额收益（Alpha）的一种主动型量化

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

投资策略。

指数增强策略的目标是在控制与对标指数的跟踪误差的前提下，尽可能获取更高的超额收益。大部分量化管理人通常使用多因子模型开发指数增强策略。

11 多因子流程

标准的多因子模型通常包含因子研发、因子组合、组合优化、交易执行 4 个环节。因子研发的过程，指的是从海量的市场数据中，寻找对股票收益率有一定预测或解释效果的规律。因子通常以对全市场股票的打分或排序的形式体现。因子可以按照使用的数据类型，分为量价、基本面、分析、另类等类别，也可以按照预测周期，分为日内、隔日、长周期等类别。除了关注单个因子的预测效果外，量化研究员还需要尽可能地寻找更多的低相关且有效的因子。

因子组合可以采用线性组合与非线性组合，线性组合一般适用于因子数量较少，因子逻辑清晰的情形，可归因、可解释性较强。非线性组合一般可以采用机器学习模型，适用于因子数量较多的情形，最终的结果很难归因到某一个因子上。

经过因子组合得到对于全市场股票的综合预测或排序后，对于指数增强策略，并不能简单地选取预期收益率最高的一组股票作为最终的投资组合，同时还要兼顾所对标指数的成分股、行业、风格等因素，在控制投资组合与对标指数跟踪误差的前提下，选取预期收益率更高的一篮子股票。除此之外，还要考虑股票的流动性与交易量之间的矛盾，当策略规模逐渐增大时，需要降低策略的换手，或者选择流动性更高的股票，以降低交易成本。综合考虑这些因素，并得到最终投资组合的过程称之为组合优化。

在最终的交易执行层面，量化策略一般采用算法交易的方式，并对于不用换仓的底仓，还可能会叠加一些 T0 交易以增厚收益。算法交易通常会采用拆单的方式，以 TWAP 或 VWAP 的方式交易，以尽可能地减少对市场的冲击。

11 中性策略

量化中性策略，是在指数增强的基础上，同时等市值做空所对应指数的股指期货，以对冲 Beta 风险，同时承担一定的对冲成本，获取纯 Alpha 收益的策略。由于股指期货采用保证金交易，通常需要占用 15%-20%的资金，因此量化中性策略的多头仓位通常只有指数增强策略的八成左右。

除了使用股指期货对冲的方式外，还可以采用指数 ETF 融券对冲、一篮子成分股融券对冲等方式。融券对冲的优势在于，对冲成本事先确定，无需承受期货基差的波动，并且采用两融信用交易，无需占用保证金，多头头寸可以 100%满仓运行，资金利用率更高；但同时会牺牲产品的流动性，例如融券一般有到期日期，提前还券需要承担罚息等。

12 其他（空气指增、灵活对冲）

指数增强策略的收益可以拆分为 Beta 收益与 Alpha 收益，而这两部分收益背后的风险特性大相径庭。Beta 收益的年化波动约为 20%，而 Alpha 收益的波动（跟踪误差）一般为年化 5%-8%。站在风险平价的角度，两者的直接叠加，并不是最优的解决方案。因此，市场上出现了空气指增和灵活对冲策略，二者分别相对于指数增强和量化中性策略，有着更高的收益弹性。

空气指增策略提高了 Alpha 的波动，不再对标某个特定指数，因此不必考虑成分股、行业、风格的限制，目标是博取更高的相对于全市场股票的超额收益。具体操作层面，可以在前述“组合优化”环节，去除或者大幅放宽与对标指数跟踪误差的限制，直接选取预期收益率最高的一组股票。空气指增追求绝对收益，长期来看收益高于指数增强，但短期会面临跑输所有宽基指数的风险。

灵活对冲策略使用部分对冲的方式，降低了 Beta 的波动，同时节约了部分对冲成本，风险特性介于指数增强与量化中性策略之间。除此之外，管理人还会基于对 Beta 的预测，灵活调整对冲敞口。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

13CTA

CTA（CommodityTradingAdvisor）是运用客户委托的资金投资于期货、期权等市场，利用标的价格变动而获利的一种投资策略。对量化 CTA 而言，基于信号构建的方法论，主要可以分为时序与截面两类。时序方法追求绝对收益，关注标的指标在时间序列上的变化；截面方法借鉴股票多因子思想，一般通过对多个品种的相对强弱排序进行预测，做多预测表现较好的标的，做空预测表现较差的标的。

14 趋势跟踪策略

趋势跟踪策略是时序方法中具有典型代表性的一类策略。它基于历史数据构建信号与规则，判断趋势行情的始末，当识别出行情处于上涨趋势时，就看多买入；当识别出行情处于下跌趋势时，就看空卖出。

趋势跟踪策略具有典型的右侧交易属性，因为趋势跟踪信号只具有对于过去趋势的跟随性，并且是滞后的，同时不具有对未来趋势的预测性。也就是说，趋势跟踪策略并不能预测趋势什么时候开始、什么时候结束，因此无法精准地在市场底部和顶部进出场，而是主要捕捉趋势的中间部分，所以往往在幅度较大、持续时间较长的流畅趋势行情中可以取得较多的收益。如 2020 年、2021 年 5 月、2022 年一季度都是趋势行情持续时间较长、幅度较大，伴随市场波动率上升的阶段，趋势跟踪策略在这些时段的都取得了较为亮眼的收益。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

鸣谢

在撰写本书的过程中，引用了大量来自量化私募基金的PPT和公开资料。这些资料为本书提供了宝贵的参考和启示，使得本书能够更深入地探讨量化投资策略和实践。

在此，谨向所有以上机构表示由衷的感谢。

此外，还要感谢所有在互联网上公开分享研究成果和市场分析的个人和机构。你们的慷慨分享不仅丰富了本书的研究内容，也推动了整个量化投资领域的知识传播和交流。

尽管本书在引用这些资料时已尽力注明出处，但如有疏漏之处，敬请谅解，并欢迎各位读者和相关机构指正。

再次感谢所有为本书提供帮助和支持的机构和个人!



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

关于我们
About Us

宽邦科技

宽邦科技是一家人工智能平台科技公司，成立于2016年，核心团队主要来自微软亚洲研究院等一线AI企业和金融机构。在AI平台、投资算法、量化引擎、新型投资大数据等技术上持续前沿探索，研发了企业级全栈AI平台BigAI和低门槛、端到端的AI投资平台BigQuant等金融科技基础平台，为投资者和投机构提供投资管理的从数据分析、因子挖掘、策略研究、AI建模、组合构建、回测模拟到实盘交易的全周期AI赋能，已服务数十万量化投资者和银行、证券、保险、基金等多家头部金融机构。

华泰证券

华泰证券股份有限公司（“华泰证券”或“公司”）是一家领先的科技驱动型综合证券集团。自1991年成立以来，华泰证券积极把握中国资本市场改革开放的历史机遇，在业内率先以金融科技助力转型，用全业务链服务体系为个人和机构客户提供专业、多元的证券金融服务，综合实力和品牌影响力位居国内证券业第一方阵，步入国际化发展的全新阶段。秉承“高效、诚信、稳健、创新”的核心价值观，华泰证券将全面实施数字化赋能下的财富管理和机构服务“双轮驱动”发展战略，致力成为兼具本土优势和全球影响力的一流投资银行。我们的优势：科技驱动的中国证券行业转型开拓者；开放的数字化财富管理平台为更多客户提供高效的专业服务；充分把握新经济崛起机遇的一流投资银行业务；兼具规模优势与创新优势的综合性资产管理平台；全球布局创造发展新机遇；兼具专业和技术优势的全面风险管理；一流的人才团队和多元的股东基础。

金融阶

金融阶，金融科技人才猎聘专家，为职帮人力机构旗下细分品牌。从2009年成立的职帮人力金融事业部至今，在金融行业的人才猎聘领域已有15年的持续专注与积累。

金融阶作为国内最早专注量化金融领域的猎聘机构之一，凭借在量化投资、对冲基金、金融科技、Ai、区块链等领域人才猎聘方向的资源积累、专业理解、服务效率等多方面领先优势，已获得了业内量化机构和Quant精英的青睐。



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

关于我们
About Us

华锐技术

华锐技术是中国领先的分布式基础软件公司和证券资管行业核心业务平台提供商，公司的使命是提升中国金融基础设施自主能力。公司2017年首创了证券行业第一个分布式技术实验室，研发了世界级分布式低时延消息中间件AMI，拥有业界领先的分布式高性能计算平台和全栈信创解决方案。公司打造了新一代分布式证券、资管核心业务系统，为证券、基金、银行、保险等金融机构提供极具竞争力的核心交易、实时清算、实时风控、高速行情、投资管理、理财营销、注册登记等系统和解决方案，帮助客户实现核心业务系统分布式转型，迎接数字化时代的变革。

阿里云

阿里云成立于2009年，是全球云计算服务重要引领者，基础设施遍及30个地理区域，89个可用区。阿里云专注服务私募量化行业，解决方案覆盖行情接入，策略研发，策略回测，实盘交易等核心场景，涉及计算、存储、网络、数据库、机器学习等服务。通过强化全球基础设施，助力私募量化机构应对瞬息万变的金融市场，多年来落地诸多私募机构客户案例。Gartner2024年报告，阿里云在亚太地区市场份额上升0.8%，继续保持第一。

AMD

AMD 是高性能与自适应计算领域的领先企业，致力于提供优质的产品和服务，助力客户解决各种重大的挑战。我们的技术推动着数据中心、嵌入式系统、游戏和 PC 市场迈向未来。

AMD 于1969 年在硅谷创立，最初只有几十名员工，从那时起 AMD 便踏上创新之路。致力于引领半导体产品领域的最前沿。如今，AMD 已经成长为一家现代化的全球性企业，凭借先进技术和诸多突破性行业创新，树立现代计算新标杆。



官方下载及活动链接—https://BigQuant.com/quantpaper/2023

如您希望就相关问题进一步交流，请联系：



林晓明
华泰证券研究所
金融工程首席研究员
17611535731
linxiaoming@htsc.com
S0570516010001/BPY421



梁举
宽邦科技创始人
18111650803
jliang@bigai.cn



何子修
金融阶创始人
13923705098
ivan.ho@finelite.cn



何康
华泰证券研究所金工
量化投资首席研究员
15210597767
hekang@htsc.com
S0570520080004/BRB318



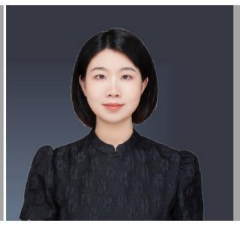
黄俊
金融阶
合伙人/资深猎头顾问
13129561205
junco.huang@finelite.cn



历彦启
阿里云
18310658114
licheng.lyq
@alibaba-inc.com



陈旭
宽邦科技
13540394180
xchen@bigai.cn



高媛
华锐技术
18516230203
gaoyuan
@archforce.cn



张雯倩
AMD
泛金融行业销售总监
18721694881
kitty.zhang@amd.com



徐颖博
13910872271
cathy.xu@
ingrammicro.com

2023年研讨会

2022年研讨会

2021年研讨会

纸质版购买

华泰研究·上海
上海市浦东新区东方路 18 号保利广场C座 1 楼

华泰研究·深圳
深圳市福田区益田路 5999 号基金大厦 10 楼

华泰研究·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

宽邦科技·成都
成都市高新区交子大道中海国际D座 1701

金融阶·深圳
广东省深圳市宝安区新安街道海纳科技大厦7028

华锐技术·上海
浦东新区杨高南路1998号6、9号楼

官方下载及活动链接—https://BigQuant.com/quantpaper/2023

《2023年度中国量化投资白皮书》编委会

主编：梁 举　林晓明　何子修
撰稿：黄 俊　陈　旭　何　康　高　媛　王晨宇　卓伟鹏　张雯倩　庄鑫博　历彦启 ChatGPT
美编：陈 旭　Midjourney

本报告(含电子版)版权归属联合出品方宽邦科技、华泰证券、金融阶所有。未经正式授权，任何转发、影印、盗版、引用、篡改的行为均属侵权，版权方将保留对其法律追诉之权利。

官方下载及活动链接—https://BigQuant.com/quantpaper/2023