

## 基本面量化系列

# 挖掘基本面有效因子,探索行业视角量化投资

中信证券研究部 数据科技组 张若海

2022年7月26日

## 核心观点



从基本面量化的发展趋势看,算法模型的升级和数据指标的挖掘是核心创新方向。比较传统交易数据,基本面数据蕴含大量产业认知,是基本面量 化可持续的创新方向。基于近年来机器学习模型的发展,结合产业因子挖掘的文本分析与量化回测深度结合,产业视角的基本面量化因子挖掘逐步 可突破传统人力边界。本报告围绕核心因子的挖掘和使用,提供通用视角下,多行业基本面量化研究新思路。

- **自主构建以行业基本面为核心的有效数据矩阵,三大数据集群保证因子完整、有效、稀缺。**我们的底层数据基于三大数据 集群:标准财务+行业运营核心数据集群、另类特色专业数据集群、投研流程数据集群,保证指标专业性和完整性,同时借助另类数据提升因子的及时性和高频性。基于市场公开数据,我们采用自然语言处理技术实现周-月-季频的集群更新和指标维护,从海量非结构化数据构建另类因子,提升数据稀缺性。
- 探索深度个股基本面因子挖掘,创新专家视角沉淀有效因子。专家视角下的因子蕴含大量背景知识,以分析师财务模型为例,核心因子的敏感性体现在计算逻辑中。我们基于nlp技术并结合量化分析抽取财务模型底层逻辑的核心因子,并且自下而上将公司级因子归集到行业层面,动态看待标的和行业的变化驱动因子。进一步反向推导计算公式得到核心因子的敏感性,并对相应因子进行标准化和数字化。
- 探索行业视角量化投资,构建"政策+景气+ESG"投资新思路。基于自主构建的另类数据集+行业经营数据集,自上而下我们以政策文本+行业运营指标做一级行业赛道选择,自下而上以中信证券自主ESG评分体系为标的筛选系统。基于以上策略,在月度轮动视角,首先面对的问题是如何评价和筛选出长期稳定有效的指标。回测效果看,2018-01至2022-06,景气度+政策+esg组合:累计收益率146.5%,年化收益率29.4%,最大回撤率25.0%,年化波动率26.9%,对标沪深300,组合超额收益率为141.6%,超额胜率58.0%。
- **数据创新与算法优化双轮驱动。**基本面指标方面,聚焦自下而上专家经验抽取,从重点个股财务模型,形成前瞻视野因子抽取的系统工程。算法优化方面,聚焦政策+景气指标,进一步探索机器学习模式下,指标权重的有效反馈与自动修正,指标前瞻性的动态市场反馈。
- **风险因素:**数据计算风险,数据标注不准确风险,模型失效风险。

## CONTENTS

## 目录

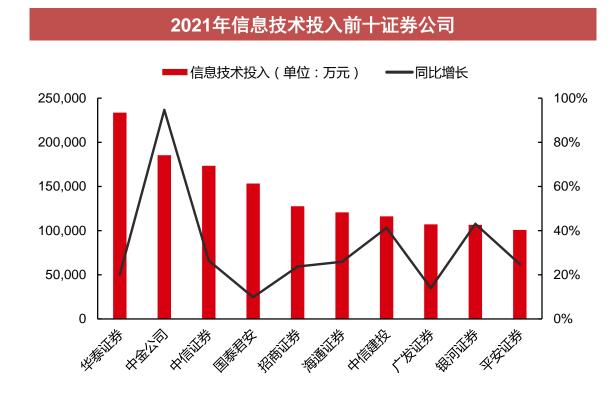
- 1. 数据爆发+技术升级,因子挖掘为基本面量化投资新趋势
- 2. 基本面量化有效因子挖掘和应用(政策力度指标+行业景气指标+ESG评分指标)
- 3. 探索"政策力度+景气程度+ESG质量"投资新思路

## 1 另类数据爆发叠加量化技术升级,数据科技驱动投资值得期待



■ 金融IT投入逐年增长,券商信息技术能力或将持续提升。证券行业对信息科技重视程度不断增强,行业信息技术投入逐年增长。2021年全行业信息技术投入金额338.20 亿元,同比增长28.7%,占2020年度营业收入的7.7%。2017年至今证券行业在信息技术领域累计投入近1200亿元,行业持续加大信息技术领域的投入为行业数字化转型和高质量发展奠定坚实基础。

#### 2017年至2021年证券公司信息技术投入 ■ 金额(单位:亿元) 一同比增长 35% 400 350 30% 300 25% 250 20% 200 15% 150 10% 100 5% 50 0 0% 2017 2018 2019 2020 2021



# 1 另类数据爆发叠加量化技术升级,数据科技驱动投资迎来新场景



- 另类数据作为新数据源,升级了过往研究方式中的信息来源。
- 机器学习方法弥补传统模型不足,在量化领域中发挥着越来越重要的作用。

	传统数据与另类数据对比					
	类型	来源	样例			
———传统数据	标准化、结构化	上市公司、交易所、指数提供商等	价格、成交量、财务报表等			
	结构化或者非结构化(例如:文本、图像、视频和HTML)	个人产生、网络平台、传感器、商业过 程等	消费者行为、电子收据、新闻、用户点 评、天气等			

资料来源:中信证券研究部

,一个大量,一个大量,一个大量,一个大量,也不是一个大量,也不是一个大量,也不是一个大量,也不是一个大量,也不是一个大量,也不是一个大量,也不是一个大量,也不是 ————————————————————————————————————					
	有监督学习算法	无监督学习算法	强化学习算法		
概念简介	是从有标签的训练数据中学习模型,以便对未知或未来的 数据做出预测	处理的是无标签或结构未知的数据	通过与环境的交互来提高其预测性能		
典型算法	KNN、回归模型、决策树、随机森林、逻辑回归、SVM等	关联规则、K-means、主成分分析等	马尔科夫决策过程、Q-learning等		
金融应用场景	股票走势预测、邮件过滤、欺诈检测	用户画像、文本聚类、文章推荐	量化交易		
有效指标挖掘应用	匹配,如用于行业有效指标检验,用于行业预期收益率的 预测	比较匹配,如行业影响因子聚类,将因 子分为上涨因子和下跌因子	暂未形成匹配和应用方案		

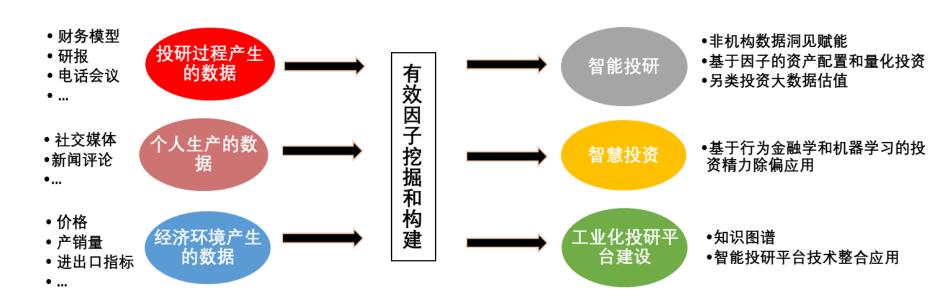
资料来源:中信证券研究部

# 1 基本面量化策略寻求破局:从财务+运营指标到核心因子挖掘 yai



- 基本面量化研究困境:
  - > 财务指标被广泛应用于基本面量化策略,但存在重要信息忽视和失真现象。
  - ▶海量数据"因子化"不足,原始数据难以直接适用于基本面量化策略。
- 有效因子挖掘目标:从经济活动、个人生产、投研过程产生的数据出发,基于nlp技术并结合量化分析挖掘面向投资应用 因子。

#### 基本面量化研究有效因子上下游



资料来源:麦肯锡,中信证券研究部



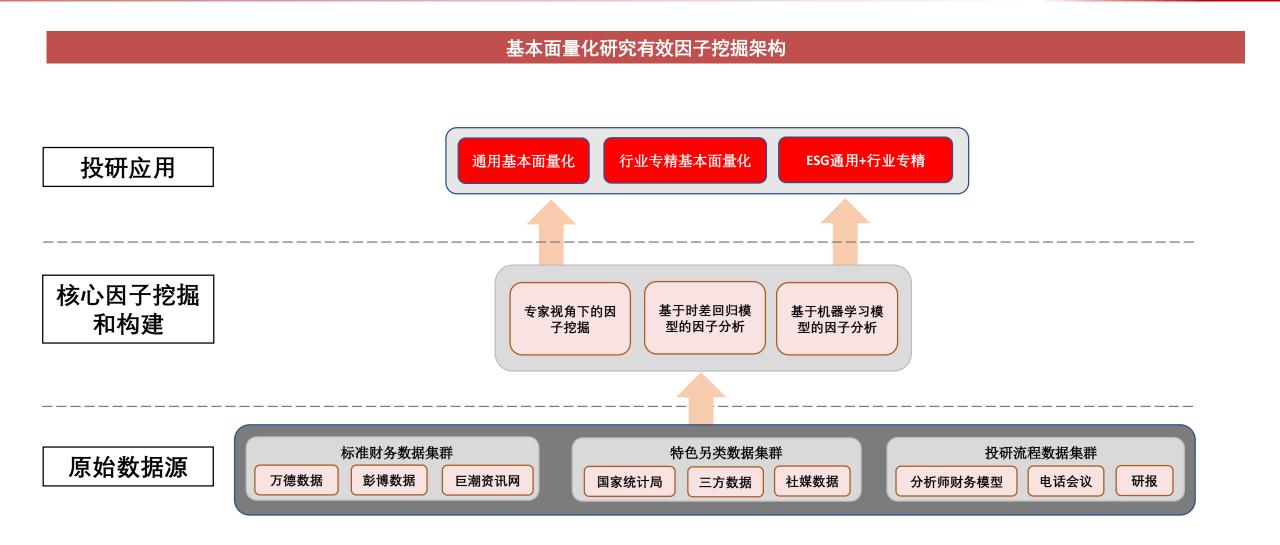
## CONTENTS

## 目录

- 1. 数据爆发+技术升级,因子挖掘为基本面量化投资新趋势
- 2. 基本面量化有效因子挖掘和应用(政策力度指标+行业景气指标+ESG评分指标)
- 3. 探索"政策力度+景气程度+ESG质量"投资新思路

### 2基本面量化研究有效因子挖掘架构





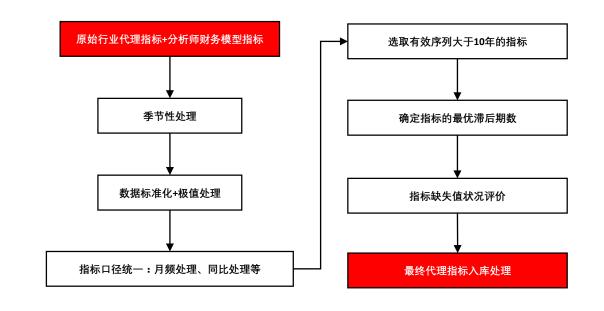
资料来源:中信证券研究部



- 构建三大数据集群,保证投研资产数字化完整、有效、稀缺。
- 设计代理指标程序化处理流程:季节性调整、数据标准化、缺失值处理以及口径统一处理。
- 设计初代理指标评价筛选流程:有效时间序列长度评价、最优滞后期数评价、缺失情况评价。

:	三大数据集群构建行业	业因子底层数据库	
数据源分类	数据库名称	数据来源	更新频率
标准财务数据集群	标准财务数据	Wind、巨潮资讯网等	季度
	经济数据	国家统计局、工信部、财政 部、商务部等	月度
特色另类数据集群	政策文本数据	中央、国务院、15个国家部 委、17个地方省市政府	周度
19 COOOCAMANIA	公司、行业扩产指标	各公司扩产能公告	月度
	社媒数据	B站、抖音、小红书等	月度
	地产、线下门店数据	贝壳、大众点评、百度地图 等	月度
	行业投融资数据	清科私募通	月度
专业数据集群	产品价格、产销量等指标	百川盈孚、卓创资讯、高工 锂电等	月度
	ESG+专利数据	轶鼎科技、专利局等	月度

#### 行业因子程序化处理、评价筛选和入库流程



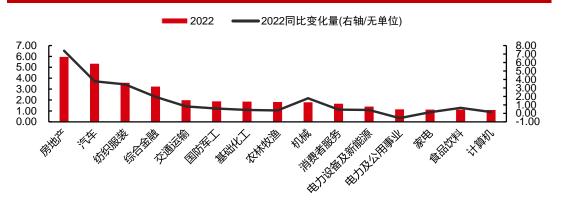
# 2 政策文本数据——聚焦中央办公厅+国务院+各部委,向产业与行业对齐 y



#### 部分中信一级行业政策情绪力度年度扫描

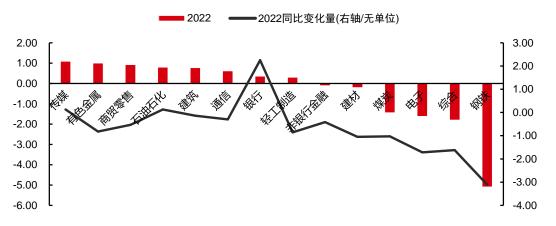
中信一级行业	2018	2019	2020	2021	2022
房地产	-0.12	0.77	-0.13	-1.40	5. 98
汽车	2. 48	1.02	1.87	1.58	5. 35
纺织服装	1.00	1.00	1.97	0.20	3. 61
综合金融	0.00	1.00	1.86	1.30	3. 25
交通运输	0.40	0.11	1.64	1. 19	2.00
国防军工	0.85	-0.43	1. 29	1.33	1. 90
基础化工	0.11	0.57	3. 17	1. 47	1.88
农林牧渔	0.68	1. 27	1.40	1.49	1.84
机械	0.91	0.17	1. 47	0.04	1.81
消费者服务	0.82	1.00	1. 77	1.24	1. 68
电力设备及新能源	1.02	1.01	1.46	1.02	1. 42
电力及公用事业	0.81	-0. 79	1.30	1.74	1. 17
家电	0.88	1.08	1. 22	1.00	1. 14
食品饮料	1.02	0.96	1.55	0.50	1. 14
计算机	0.80	1.00	0.98	0.96	1.11
传媒	0.62	1.01	1.80	0.96	1.09
有色金属	1.04	1.88	2. 09	1.82	1.00
商贸零售	0.32	0.35	1. 24	1.46	0.92
石油石化	-0.33	1.29	0.88	0.67	0.80
建筑	0.80	0.96	2. 53	0.91	0.77
通信	0.66	1.06	1.43	0.91	0.62
银行	-2.60	-2.14	-1.06	-1.90	0.36
轻工制造	0. 47	1.39	1. 97	1.16	0.30
非银行金融	-0.85	-0.33	1.02	0.32	-0.10
建材	1.00	1.69	1.50	0.85	-0.20
煤炭	0.30	-1.20	1.07	-0.41	-1.44
电子	0.94	1.05	1. 23	0.10	-1.61
综合	0.74	0.64	0.87	-0.17	-1.79
钢铁	-1.23	-1.58	-1.64	-1.98	-5.09

#### 部分中信一级行业政策情绪力度边际变化值



资料来源:中国政府网,其他政府官网,wind,中信研究;数据截止2022.05.31

#### 部分中信一级行业政策情绪力度边际变化值



资料来源:中国政府网,其他政府机构官网,wind,中信证券研究部;数据截止2022.05.31

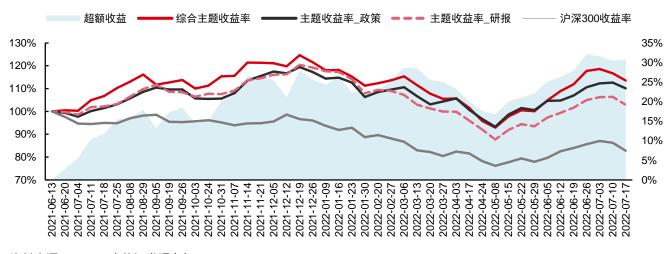
# 2 政策周报推荐主题对应指数回测,历史推荐主题超额收益30.38%



10

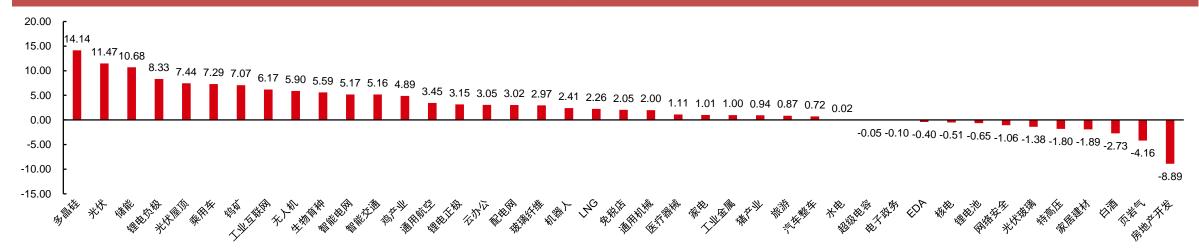
- 从2021年6月至上周的回测结果显示,历史周报的主题 推荐相对沪深300超额收益率为30.38%。
- 通过对过去10次外发周报回测结果进行统计可以看到, 收益率排名前五的主题分别为:多晶硅、光伏、储能、 锂电负极、光伏屋顶。
- 注:回测使用标的均为Wind主题指数,每周对推荐的所有主题指数等权买入,指数购买方式是按照主题指数权重股及其对应权重买入;交易时间为每周一开盘对上周的推荐结果买入并在本周五收盘时卖出;为方便计算目前没有计入交易成本。

#### 历史周报推荐主题(对每次周报推荐主题的对应指数持有一周)收益率



资料来源: Wind, 中信证券研究部

#### 过去10次外发周报推荐主题统计——分主题周度涨跌幅累加(%)



资料来源:Wind,中信证券研究部

## 2 构建多维行业经济数据矩阵,打造中观景气底层支持体系



■ 中观景气度跟踪的数据支持体系结合国家统计局、行业专业数据商等多个数据源,为28个中信**证券一级**行业构建代理指标库,纳入了超过300个行业高频指标。

底层数据库及其获取途径和更新频率						
数据源分类	数据库名			更新频率		
	行业标准财务数据	Wind	基于Wind并结合nlp提取	季		
	一————行业经济数据	国家统计局、工信部、财政部、商务部等	nlp提取	月		
行业特色数据源	行业政策文本数据	中央、国务院、15个国家部委、17个地方省市政府	nlp提取	周		
	行业相关扩产公告	Wind	基于Wind并结合nlp提取	月		
		B站、抖音、小红书等	数据商、nlp提取	日		
		贝壳、大众点评、百度地图等	数据商、nlp提取	日		
	行业投融资数据	清科私募通	数据商	月		
行业专业数据源	主营产品价格、产销量等数据	百川盈孚、卓创资讯、高工锂电等	数据商	月		
	研发专利数据	轶鼎科技、专利局等	数据商	月		

## 2 方向预判:准确率高达79.8%,回测期间多次全行业命中



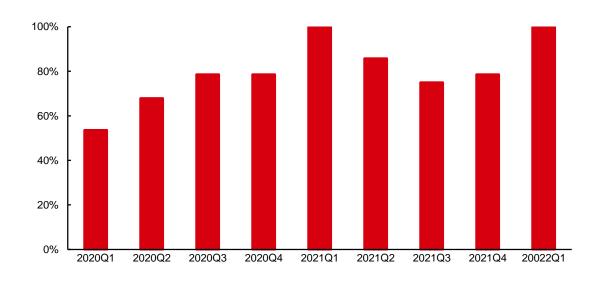
■ 由于行业营收同比是季频指标,而模型预测的景气指数仅含当月信息。为了使两者口径近似可比,以景气指数3个月均值的同比,作为模型对景气度方向的预测结果。所以景气度方向变化定义为:

$$trend_t = sign[(f_t + f_{t-1} + f_{t-2}) - (f_{t-12} + f_{t-13} + f_{t-14})]$$

■ 回测区间为2020-01至2022-03,将景气度变化方向与真实营收同比方向进行逐期对比,并计算命中率。结果显示,景气方向平均命中率高达79.8%。其中,相比于2020年,2021年模型命中率提升15.2%,表明随着数据不断积累模型预测能力有望逐步增强。

# 景气指数变化方向定义 第t-14个月至第t个月的景气指数 均値 作差 均値 模型預測方向 当前季度的真实营收同比 作差 真实方向

#### 2020年起各季度景气度方向命中率情况



资料来源: Wind, 中信证券研究部

## 2 泛周期高适用,新成长可优化

(CITIC SECURITIES

■ 从单个行业的预测效果来看,景气度模型对周期类行业(如:石油石化、基础化工等)预测较好,而对成长性行业(通信、计算机、电力设备及新能源)仍有可优化的空间。

中信证券一级行业景气指数评价(2020年起)						
中信证券行业分类	区间命中率	方向准确率	中信证券行业分类	区间命中率	方向准确率	
石油石化	89%	100%	建筑	56%	89%	
保险	89%	89%	汽车	56%	89%	
煤炭	89%	78%	交通运输	67%	78%	
钢铁	89%	78%	传媒	78%	67%	
电力及公用事业	67%	100%	纺织服装	56%	78%	
基础化工	78%	89%	房地产	67%	67%	
轻工制造	78%	89%	国防军工	44%	78%	
消费者服务	67%	100%	商贸零售	56%	67%	
家电	78%	89%	电子	44%	78%	
机械	56%	100%	计算机	56%	67%	
建材	67%	89%	证券	67%	44%	
食品饮料	67%	89%	电力设备及新能源	56%	56%	
农林牧渔	67%	89%	医药	44%	67%	

## 2 评分体系指标设计:投资导向下,结合国际标准与中国智慧



一级议题	二级议题	三级议题	指标名称
环境	污染排放	企业污染排放水平	单位收入颗粒物排放
环境	污染排放	企业污染排放水平	单位收入Nox排放
环境	污染排放	企业污染排放水平	单位收入SO2排放
环境	污染排放	企业污染排放水平	单位收入氨氮排放
环境	污染排放	企业污染排放水平	单位收入COD排放
环境	污染排放	企业环境信息披露合规性	排污许可执行报告披露比例
环境	污染排放	企业污染排放风险	重污染天气限工限产风险
环境	污染排放	企业污染排放风险	环境行政处罚
环境	企业环境管理	企业环境管理体系评价	环境管理体制认证
环境	企业环境管理	企业绿色生产水平	工信部绿色生产认证
环境	气候与能源	企业能源管理水平	能源管理体制认证
环境	气候与能源	企业对气候问题重视度	是否披露碳排放
社会责任	员工	企业安全管理体系评价	安全生产认证
社会责任	员工	企业安全管理体系评价	安全生产处罚
社会责任	员工	职工薪酬与福利	职工人均薪酬
社会责任	员工	职工薪酬与福利	职工人均社会保险费
社会责任	社会	企业业务的政策环境	企业的政策量化得分
社会责任	消费者	企业产品、服务质量	质量管理体系认证
社会责任	消费者	企业产品、服务质量	市场监管机构处罚
社会责任	消费者	客户隐私保护	信息安全管理体系
社会责任	供应链	企业供应链稳定性评价	企业供应链CR5集中度
社会责任	供应链	企业供应链关系评价	企业应付账期水平
社会责任	供应链	企业供应链关系评价	企业应收账期水平
公司治理	商业道德	企业涉及诉讼	刑事诉讼
公司治理	商业道德	反腐败	纪委监委调查
公司治理	ESG治理	ESG信息披露	单独披露CSR报告
公司治理	ESG治理	ESG信息披露	CSR报告进行第三方审计
公司治理	内部治理合规性	上市监管负面事件	交易所、证监会监管与处罚
公司治理	内部治理合规性	信息披露水平	交易所企业信息披露等级评价
公司治理	内部治理合规性	审计结果	财报审计意见结果
公司治理	内部治理合规性	审计结果	内控审计意见结果
公司治理	管理层评价	管理层激励制度	股权激励方案
公司治理	管理层评价	管理层收入评价	管理层总薪酬
公司治理	管理层评价	管理层收入评价	管理层薪酬/平均薪酬
公司治理	管理层评价	管理层收入评价	管理层薪酬/营业收入
公司治理	管理层评价	管理层稳定性	管理层平均任期
公司治理	管理层评价	管理层稳定性	管理层平均年龄
公司治理	管理层评价	管理层稳定性	管理层增减持比例
公司治理	管理层评价	管理层稳定性	管理层增减持数值

投资导向下,结合国际标准与中国智慧。在中信证券 ESG指标体系构建的整体流程中,我们全面落实国际视 野与中国智慧的结合,切实探索如何在与国际ESG评价 体系接轨的同时又能结合我国的发展阶段,从而推动 ESG理念从评价研究走向实战落地。

环境

• 结合监管体系,重点关注环境风险。重点关注企业在生产经营过程中对外部环境所的影响,包括污染排放、能源消耗和碳排放等。

社会责任

• 引入中国智慧,关注国家和社会的全面发展。我们内立足中国发展阶段,对该理念进行了一定的拓展和延伸,着重突出了上市公司对国家和社会全面发展的责任,引入了企业所属行业的政策环境变量。

公司治理

• **关注治理合规性,增强指标确定性。**在公司治理中,我们一方面重点关注其治理的合规性,尤其是是否符合上市公司治理规范。另一方面,我们尝试基于投资逻辑和理解对其治理架构的有效性进行评价

## 2 评分体系指标权重:以实质性为核心,结合量化回测与基本面逻辑



行业	公司治理	环境	社会责任
传媒	48%	14%	38%
电力及公用事业	50%	17%	33%
电力设备及新能源	53%	15%	32%
电子	45%	19%	35%
房地产	55%	19%	26%
纺织服装	51%	16%	33%
非银行金融	52%	24%	24%
钢铁	49%	19%	31%
国防军工	51%	16%	33%
机械	50%	16%	34%
基础化工	49%	22%	29%
计算机	48%	16%	36%
家电	48%	19%	34%
建材	44%	20%	36%
建筑	48%	21%	31%
交通运输	52%	15%	33%
煤炭	42%	30%	27%
农林牧渔	53%	13%	34%
汽车	42%	25%	34%
轻工制造	45%	18%	37%
商贸零售	55%	12%	33%
石油石化	49%	23%	29%
食品饮料	46%	23%	31%
通信	47%	18%	35%
消费者服务	46%	13%	41%
医药	47%	17%	35%
银行	62%	14%	24%
有色金属	48%	27%	26%
综合	46%	20%	34%
综合金融	63%	2%	36%
平均值	49%	18%	32%

- 指标权重设计思路:以实质性为核心,结合量化回测与基本面逻辑
  - ▶ 核心假设: 我们将指标权重视为该指标的实质性体现,因此权重差异性的本质是不同行业中ESG指标的实质性不同,在此前版本中,我们完全依据主观逻辑判断为指标赋权,为更科学量化地设计指标权重,我们假设指标在历史时期对股票收益的分层效果能一定程度反映该指标在特定行业内的实质性
  - ▶ **量化回测**:采用了ESG指标在行业中多空组合的累计收益率作为其实质性判断的基础。依据该行业累计多空组合收益排名将指标分为0~4共计5挡,其中0指在该行业中无数据或无区分度,最终将各个指标分档得分加权,即为该行业中企业的指标权重。
  - ▶ 逻辑判断: 同时我们认为指标历史收益表现并不能完全表征该指标的实质性,且评分体系中也应当突出企业ESG表现,例如部分事件型指标对某个行业不敏感可能是由于其发生频率较低或行业内区分度较低,例如银行业的社会责任报告披露率已达100%,我们结合主观认知对其部分指标的分档结果进行调整。
- 最终,我们会结合指标覆盖率,将行业内覆盖率低于20%的指标权重设置为0。整体来看,各行业中公司治理、环境和社会责任权重平均为49%、18%、32%,行业间差距较大,综合金融的环境平均权重仅为2%。

资料来源: 中信证券研究部测算

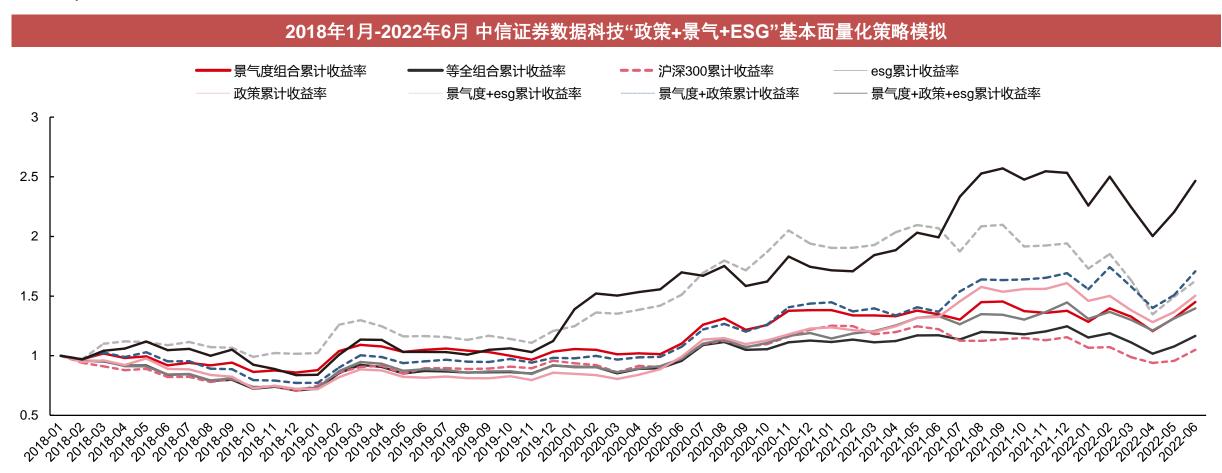
## CONTENTS

## 目录

- 1. 数据爆发+技术升级,因子挖掘为基本面量化投资新趋势
- 2. 基本面量化有效因子挖掘和应用(政策力度指标+行业景气指标+ESG评分指标)
- 3. 探索"政策力度+景气程度+ESG质量"投资新思路

## 3 政策力度+景气程度+ESG质量投资策略结果

■ 2018-01至2022-06景气度+政策+ESG组合:累计收益率146.5%,年化收益率29.4%,最大回撤率25.0%,年化波动率26.9%。沪深300:累计收益率4.9%,年化收益率1.4%,最大回撤率22.2%,年化波动率18.1%,组合超额收益率为141.6%,超额胜率58.0%。



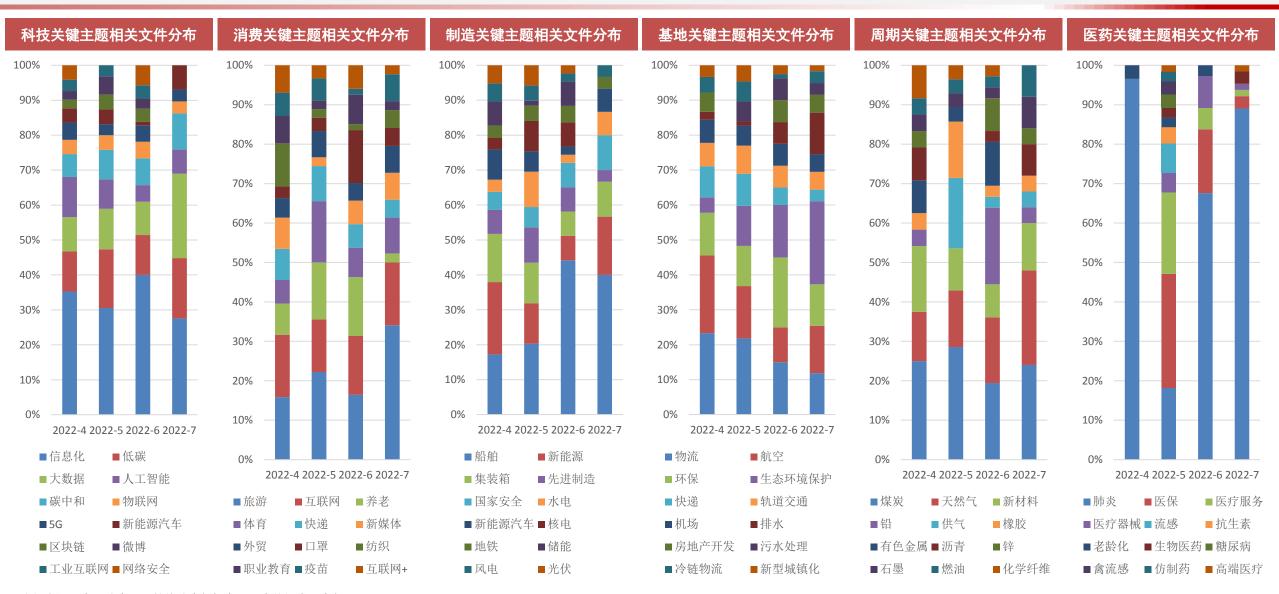
## 3 政策力度+景气程度+ESG质量具体应用说明



- 政策配置算法说明:
  - ▶ 行业力度计算方式: 行业政策力度=当月政策热度\*0.5+前一个月政策热度\*0.3+前两个月政策热度\*0.2;
  - > 行业配置方式: 月度调仓, 进行行业间比较, 每次将政策力度前三名行业进行等权买入。
- 中观景气算法说明:
  - ▶ 行业景气程度方式: 行业景气程度=当月景气度绝对位置\*0.5+当月景气变化\*0.5。其中,景气度绝对位置为近八年行业景 气度分位数,景气变化为景气度的同比、环比等权求和值;
  - ▶ 行业配置方式: 月度调仓,进行行业间比较,每次将景气程度前三名行业进行等权买入。
- ESG配置算法说明:
  - ▶ ESG评分方式: 自建ESG评级体系, 从环境、社会责任以及公司治理三大维度进行企业精准评价;
  - ▶ 行业配置方式: 月度调仓,对于选中的行业,选出ESG评分前20%的个股进行等权买入。
- 模拟调整时间说明:
  - ▶回测区间为2018-01至2022-06,每月最后一个日历日发出调仓信号,次月第一个交易日调仓。暂不考虑手续费。

# 3 未来应用升级与探索——通用因子指向更精准,行业因子挖掘更前瞻 <sup>y</sup>



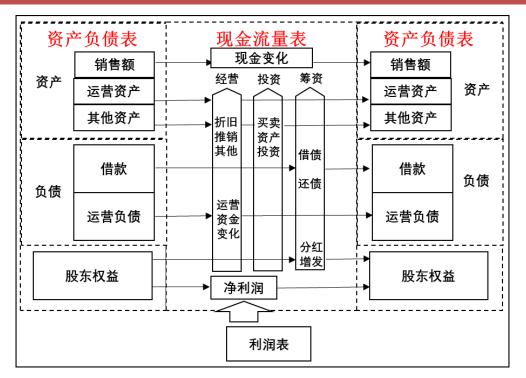


资料来源:中国政府网,其他政府部门官网,中信证券研究部注:数据统计截至2022年07月15日

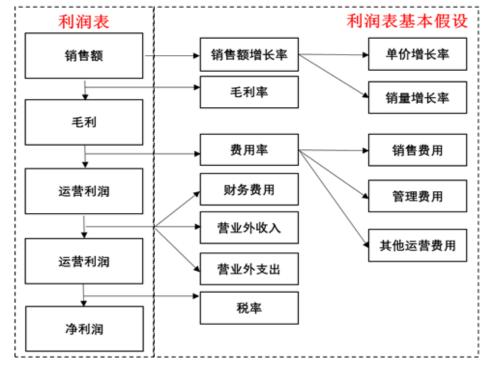
## 3未来应用升级与探索——通用因子指向更精准,行业因子挖掘更前瞻。(CITIC SECURITIES

- 公司财报反映创造利润的过程,利润表披露历史核心影响因子。
- 基于分析师财务模型数据天然优势,自下而上,关注经营假设以及财务关键因子。
- 基于分析师逻辑和预期,开展公司级核心因子抽取与行业归集,以及核心因子对财务指标的敏感性统计。

#### 财务模型的三张表



#### 分析师进行利润表预测的过程



资料来源:中信证券研究部

资料来源:中信证券研究部

## 3 未来应用升级与探索——通用因子指向更精准,行业因子挖掘更前瞻。



- 基于分析师逻辑和预期,开展公司级核心因子抽取与行业归集,以及核心因子对财务指标的敏感性统计。
  - ▶ 基于计算机技术抽取财务模型底层逻辑的有效因子,借助计算逻辑反向推导其敏感性;
  - ▶ 自下而上将公司级因子归集到行业层面,更多考虑的是共线因子的选择,动态看待标的和行业的变化驱动因子。

	某企业分析师财务模型主营产品列表以及因子敏感度							
时期	所属中信一级行业	所属中信二级行业	所属中信三级行业	主营产品	指标名称	营收敏感度	毛利敏感度	
2022E	电力及新能源设备	电源设备	太阳能	半导体器件	成本	0	-0.06	
2022E	电力及新能源设备	电源设备	太阳能	半导体器件	收入	0.009	0.068	
2022E	电力及新能源设备	电源设备	太阳能	半导体硅片	成本	0	-0.032	
2022E	电力及新能源设备	电源设备	太阳能	半导体硅片	收入	0.008	0.06	
2022E	电力及新能源设备	电源设备	太阳能	光伏硅片	成本	0	-1.601	
2022E	电力及新能源设备	电源设备	太阳能	光伏硅片	收入	0.233	1.673	
2022E	电力及新能源设备	电源设备	太阳能	组件	收入	0	-0.291	
2022E	电力及新能源设备	电源设备	太阳能	组件	收入	0.048	0.348	

资料来源:中信证券研究部

## 3 未来应用升级与探索——通用因子指向更精准,行业因子挖掘更前瞻。

- 对于行业因子进行有效性检验时,经常采用量化分析的手段从回归系数、IC均值等维度来判断因子的优劣。除此之外,机 器学习够分辨出非线性的因子优势,分辨出这个因子的因果性而非相关性。
  - ▶ 传统线性回归模型进行因子相关性计算;
  - ▶ 非线性模型XGBoost进行因子重要性计算。

#### 构建机器学习系统的路线图 特征选择 特征抽取和比例 降低维度 取样 标签 训练数据 学习算法 最终模型 训练数据 标签 测试数据 原始数据 标签 预处理 学习 评估 预测 模型选择 交叉验证 度量新能 超参数调优

#### XGBoost模型训练示意图 训练 原始数据集 弱学习器1 计算残差 训练 加成 残差 $E_1$ 弱学习器2 计算残差 强学习器 残差 $E_2$ 弱学习器3 训练 残差 $E_m$ 弱学习器m

22

## 风险因素



- 数据计算风险;
- 数据标注可能不准确的风险;
- 模型失效风险。



# 感谢您的信任与支持!

THANK YOU

张若海(数据科技首席分析师)

执业证书编号: S1010516090001



证券研究报告 2022年7月26日

#### 分析师声明

主要负责撰写本研究报告全部或部分内容的分析师在此声明:(i)本研究报告所表述的任何观点均精准地反映了上述每位分析师个人对标的证券和发行人的看法;(ii)该分析师所得报酬的任何组成部分无论是在过去、现在及将来均不会直接或间接地与 研究报告所表述的具体建议或观点相联系。

#### 一般性声明

本研究报告由中信证券股份有限公司或其附属机构制作。中信证券股份有限公司及其全球的附属机构、分支机构及联营机构(仅就本研究报告免责条款而言,不含CLSA group of companies),统称为"中信证券"。

本研究报告对于收件人而言属高度机密,只有收件人才能使用。本研究报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。本研究报告仅为参考之用,在任何地区均不应被视为买卖任何证券、金融工具的要约 或要约邀请。中信证券并不因收件人收到本报告而视其为中信证券的客户。本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要,不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具, 本报告的收件人须保持自身的独立判断并自行承担投资风险。

本报告所载资料的来源被认为是可靠的,但中信证券不保证其准确性或完整性。中信证券并不对使用本报告或其所包含的内容产生的任何直接或间接损失或与此有关的其他损失承担任何责任。本报告提及的任何证券或金融工具均可能含有重大的风险, 可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可跌可升。过往的业绩并不能代表未来的表现。

本报告所载的资料、观点及预测均反映了中信证券在最初发布该报告日期当日分析师的判断,可以在不发出通知的情况下做出更改,亦可因使用不同假设和标准、采用不同观点和分析方法而与中信证券其它业务部门、单位或附属机构在制作类似的其他 材料时所给出的意见不同或者相反。中信证券并不承担提示本报告的收件人注意该等材料的责任。中信证券通过信息隔离墙控制中信证券内部一个或多个领域的信息向中信证券其他领域、单位、集团及其他附属机构的流动。负责撰写本报告的分析师的薪酬 由研究部门管理层和中信证券高级管理层全权决定。分析师的薪酬不是基于中信证券投资银行收入而定,但是,分析师的薪酬可能与投行整体收入有关,其中包括投资银行、销售与交易业务。

若中信证券以外的金融机构发送本报告,则由该金融机构为此发送行为承担全部责任。该机构的客户应联系该机构以交易本报告中提及的证券或要求获悉更详细信息。本报告不构成中信证券向发送本报告金融机构之客户提供的投资建议,中信证券以及 中信证券的各个高级职员、董事和员工亦不为(前述金融机构之客户)因使用本报告或报告载明的内容产生的直接或间接损失承担任何责任。

#### 评级说明

投资建议的评级标准		评级	说明
	股票评级 业	买入	相对同期相关证券市场代表性指数涨幅20%以上
报告中投资建议所涉及的评级分为股票评级和行业评级(另有说明的除外)。评级标准为报告发		増持	相对同期相关证券市场代表性指数涨幅介于5%~20%之间
布日后6到12个月内的相对市场表现,也即:以报告发布日后的6到12个月内的公司股价(或行业指数)相对同期相关证券市场代表性指数的涨跌幅作为基准。其中:A股市场以沪深300指数为基准,新三板市场以三板成指(针对协议转让标的)或三板做市指数(针对做市转让标的)为基准;香港市场以摩根士丹利中国指数为基准;美国市场以纳斯达克综合指数或标普500指数为基准;韩		持有	相对同期相关证券市场代表性指数涨幅介于-10%~5%之间
		卖出	相对同期相关证券市场代表性指数跌幅10%以上
		强于大市	相对同期相关证券市场代表性指数涨幅10%以上
国市场以科斯达克指数或韩国综合股价指数为基准。		中性	相对同期相关证券市场代表性指数涨幅介于-10%~10%之间
		弱于大市	相对同期相关证券市场代表性指数跌幅10%以上

#### 免责声明



#### 特别声明

在法律许可的情况下,中信证券可能(1)与本研究报告所提到的公司建立或保持顾问、投资银行或证券服务关系,(2)参与或投资本报告所提到的公司的金融交易,及/或持有其证券或其衍生品或进行证券或其衍生品交易。本研究报告涉及具体公司的 披露信息,请访问https://research.citicsinfo.com/disclosure。

#### 法律主体声明

本研究报告在中华人民共和国(香港、澳门、台湾除外)由中信证券股份有限公司(受中国证券监督管理委员会监管,经营证券业务许可证编号:Z20374000)分发。本研究报告由下列机构代表中信证券在相应地区分发:在中国香港由CLSA Limited (于中国香港注册成立的有限公司)分发;在中国台湾由CL Securities Taiwan Co., Ltd.分发;在澳大利亚由CLSA Australia Pty Ltd.(商业编号:53 139 992 331/金融服务牌照编号:350159)分发;在美国由CLSA(CLSA Americas, LLC除外)分发;在新加坡由CLSA Singapore Pte Ltd.(公司注册编号:198703750W)分发;在欧洲经济区由CLSA Europe BV分发;在英国由CLSA(UK)分发;在印度由CLSA India Private Limited分发(地址:8/F, Dalamal House, Nariman Point, Mumbai 400021;电话:+91-22-66505050;传真:+91-22-22840271;公司识别号:U67120MH1994PLC083118);在印度尼西亚由PT CLSA Sekuritas Indonesia分发;在日本由CLSA Securities Japan Co., Ltd.分发;在韩国由CLSA Securities Korea Ltd.分发;在马来西亚由CLSA Securities Malaysia Sdn Bhd分发;在菲律宾由CLSA Philippines Inc.(菲律宾证券交易所及证券投资者保护基金会员)分发;在泰国由CLSA Securities (Thailand) Limited分发。

#### 针对不同司法管辖区的声明

中国大陆:根据中国证券监督管理委员会核发的经营证券业务许可,中信证券股份有限公司的经营范围包括证券投资咨询业务。

中国香港:本研究报告由CLSA Limited分发。 本研究报告在香港仅分发给专业投资者(《证券及期货条例》(香港法例第571 章)及其下颁布的任何规则界定的),不得分发给零售投资者。就分析或报告引起的或与分析或报告有关的任何事宜,CLSA客户应联系CLSA Limited的罗鼎,电话:+852 2600 7233。

**美国:** 本研究报告由中信证券制作。本研究报告在美国由CLSA(CLSA Americas, LLC除外)仅向符合美国《1934年证券交易法》下15a-6规则界定且CLSA Americas, LLC提供服务的"主要美国机构投资者"分发。对身在美国的任何人士发送本研究报告将不被视为对本报告中所评论的证券进行交易的建议或对本报告中所述任何观点的背书。任何从中信证券与CLSA获得本研究报告的接收者如果希望在美国交易本报告中提及的任何证券应当联系CLSA Americas, LLC(在美国证券交易委员会注册的经纪交易商),以及 CLSA 的附属公司。

**新加坡:** 本研究报告在新加坡由CLSA Singapore Pte Ltd.,仅向(新加坡《财务顾问规例》界定的)"机构投资者、认可投资者及专业投资者"分发。就分析或报告引起的或与分析或报告有关的任何事宜,新加坡的报告收件人应联系CLSA Singapore Pte Ltd.,也让:80 Raffles Place, #18-01, UOB Plaza 1, Singapore 048624,电话:+65 6416 7888。因您作为机构投资者、认可投资者或专业投资者的身份,就CLSA Singapore Pte Ltd.可能向您提供的任何财务顾问服务,CLSA Singapore Pte Ltd豁免遵守《财务顾问法》(第110章)、《财务顾问规例》以及其下的相关通知和指引(CLSA业务条款的新加坡附件中证券交易服务C部分所披露)的某些要求。MCI(P)085/11/2021。

**加拿大:**本研究报告由中信证券制作。对身在加拿大的任何人士发送本研究报告将不被视为对本报告中所评论的证券进行交易的建议或对本报告中所载任何观点的背书。

**英国:**本研究报告归属于营销文件,其不是按照旨在提升研究报告独立性的法律要件而撰写,亦不受任何禁止在投资研究报告发布前进行交易的限制。本研究报告在英国由CLSA (UK)分发,且针对由相应本地监管规定所界定的在投资方面具有专业经验的 人士。涉及到的任何投资活动仅针对此类人士。若您不具备投资的专业经验,请勿依赖本研究报告。

欧洲经济区:本研究报告由荷兰金融市场管理局授权并管理的CLSA Europe BV 分发。

**澳大利亚:**CLSA Australia Pty Ltd("CAPL")(商业编号:53 139 992 331/金融服务牌照编号:350159) 受澳大利亚证券与投资委员会监管,且为澳大利亚证券交易所及CHI-X的市场参与主体。本研究报告在澳大利亚由CAPL仅向"批发客户"发布及分发。 本研究报告未考虑收件人的具体投资目标、财务状况或特定需求。未经CAPL事先书面同意,本研究报告的收件人不得将其分发给任何第三方。本段所称的"批发客户"适用于《公司法(2001)》第761G条的规定。CAPL研究覆盖范围包括研究部门管理层 不时认为与投资者相关的ASX All Ordinaries 指数成分股、离岸市场上市证券、未上市发行人及投资产品。CAPL寻求覆盖各个行业中与其国内及国际投资者相关的公司。

**印度:**CLSA India Private Limited,成立于 1994 年 11 月,为全球机构投资者、养老基金和企业提供股票经纪服务(印度证券交易委员会注册编号:INZ000001735)、研究服务(印度证券交易委员会注册编号:INH00001113)和商人银行服务(印度证券交易委员会注册编号:INM000010619)。CLSA 及其关联方可能持有标的公司的债务。此外,CLSA及其关联方在过去 12 个月内可能已从标的公司收取了非投资银行服务和/或非证券相关服务的报酬。 如需了解CLSA India "关联方"的更多详情,请联系 Compliance-India@clsa.com。

未经中信证券事先书面授权,任何人不得以任何目的复制、发送或销售本报告。

中信证券2022版权所有。保留一切权利。