



多视角理解基于相关性的因子优化

多因子量化选股系列专题研究 | 2022.2.17

中信证券研究部

核心观点



汪洋
量化策略分析师
S1010521090005



王兆宇
首席量化策略分析师
S1010514080008



赵文荣
首席量化与配置分析师
S1010512070002



赵乃乐
量化策略分析师
S1010521050001

本报告从多个视角深入探讨基于相关性的因子优化的背后逻辑。理论视角下，从量化策略全局最优出发，得出拟合估计收益率和实际收益率在可行集上投影的夹角是正确的选择，而优化 IC 可证明与其等价；可视化视角下，相关性作为优化目标更适合区分不同因子的重要性的方向性；模型行为视角下，模型通过学习因子间的对冲关系来实现更稳定的超额收益。经组合优化后，2012 年以来回测相对中证 500 指数年化超额收益 13.39%，信息比率 2.63，较直接选取前 100 只股票有进一步提升。

■ **前言：**在报告《多因子量化选股系列专题研究—破局低信噪比：基于深度学习的因子优化研究》（2021-11-23）中，我们提出了基于相关性的目标函数。为了提升对于其背后的逻辑和模型行为的理解，本报告从三个视角——理论视角、可视化视角和模型行为视角继续深入探讨，并且补充了经组合优化后的策略测试结果。

■ **理论视角：优化 IC 等价于优化投影向量夹角。**1) 最优组合是关于估计收益率的函数：从量化策略全局最优出发，将实现的量化策略的与最优量化策略的收益之差作为损失函数，则损失取决于最优组合，而最优组合受估计收益率影响；2) 最优组合只与估计收益率的方向有关，而均方误差对于实际收益率的模也进行了拟合，导致信噪比降低；3) 根据夹角余弦值的两种计算方式可证明优化 IC 等价于优化投影向量夹角。

■ **可视化视角：相关性更关注向量方向。**假定因子与收益之间符合线性模型，采用模拟数据观察。采用均方误差作为损失函数时，函数曲面的等高线是类似的同心椭圆形；而采用相关性时，函数曲面的等高线是过原点的射线，表明相关性更适合区分不同因子的重要性的方向性。

■ **模型行为视角：学习因子间的对冲关系。**1) 因子数据及其处理流程。采用了基本面、估值等共计 21 个因子。因子和下期收益均转化为行业市值中性化的排序分位数；2) 通过求网络输出值对因子的一阶偏导可观察因子对输出值的影响。EP5 年分位数、北向持有占比以及净资产收益率（TTM）对于网络输出值的影响最大，并且模型还学习到在 EP5 年分位数与净资产收益率（TTM）之间构建对冲关系来实现更稳定的超额收益。

■ **组合优化下的模型回测。**1) 5 种基线系统。构造了等权加权、历史 IC 加权、最大化历史 IC、最大化 IC_IR、和 Loss1 神经网络 5 个基线系统用于比较。2) 模型测试：采用与上篇报告同样的股票打分进行组合优化来测试深度相关性模型（DCM），控制市值和行业的风险暴露在 0.1% 以内，个股偏离限定在 1% 以内。在 2012 年 1 月 1 日至 2022 年 1 月 19 日的测试区间上，DCM 策略实现 13.39% 的年化超额收益，信息率 2.63，超额最大回撤 5.02%。并且信息率较选取前 100 股票实现的 2.31 有进一步提升。

■ **结论与投资建议：**对基于相关性的因子合成提出了理论基础并进行了实证验证。文中提出的可视化方法和求偏导观察因子影响的方法也可作为投资者分析模型的重要手段。

■ **风险因素：**训练的随机性风险；因子效果衰减风险；历史业绩不代表未来表现。

目录

前言.....	1
理论视角：优化 IC 等价于优化投影向量夹角	1
最优组合是关于估计收益率的函数	1
最优组合只与估计收益率的方向有关	2
优化 IC 等价于优化投影向量夹角.....	3
可视化视角：相关性更关注向量方向.....	5
模型行为视角：学习因子间的对冲关系	6
因子数据及其处理流程.....	6
通过因子间的对冲关系进行因子优化	7
组合优化下的模型回测.....	10
5 种基线系统	10
策略历史表现	11
结论与投资建议.....	12
风险因素	13

插图目录

图 1：量化选股策略运行流程图	1
图 2：组合优化案例三维示意图	2
图 3：组合优化案例二维示意图	2
图 4：收益率向量及其在可行集上投影的示意图.....	3
图 5：均方误差函数（MSE）曲面.....	5
图 6：相关性函数（Correlation）曲面.....	5
图 7：相关性函数（Correlation）曲面 3D	6
图 8：各因子的平均（绝对值）梯度.....	7
图 9：三因子的平均梯度走势.....	8
图 10：四个因子的平均 IC	9
图 11：四个因子的 IC 大于 0 的次数占比	9
图 12：四个因子的 IC 序列	9
图 13：策略的净值走势	11
图 14：各模型相对中证 500 的相对收益	11

表格目录

表 1：模型涉及的全部指标	6
表 2：策略的历史表现	11
表 3：各模型的历史表现.....	12

■ 前言

在报告《多因子量化选股系列专题研究—破局低信噪比：基于深度学习的因子优化研究》（2021-11-23）中，我们基于对数据信噪比的启发式思考，提出了在深度模型中使用基于相关性的目标函数，并在策略表现上取得了显著提升，从实证维度验证了模型的有效性。

截至上一篇报告发布，我们实际上还存在诸多疑问，比如：1. 为什么在估计收益率和真实收益率之间采用相关性作为距离度量要好于欧式距离？2. 相关性的函数曲面是如何的，优化过程是否会陷入局部最优？3. 深度学习训练得到的映射关系是否可以理解或分析？

对于上述的这些问题，本报告将从三个视角——理论视角、可视化视角和模型行为视角——继续深入探讨基于相关性的因子优化的背后逻辑，并且补充了经组合优化后的策略测试，结果显示信息率能有进一步提升。

■ 理论视角：优化 IC 等价于优化投影向量夹角

最优组合是关于估计收益率的函数

常规的量化选股策略一般按照下图所示的流程运行。首先，进行因子计算，然后采用模型对各个因子进行组合来估计各个股票的未来收益率¹，最后根据估计收益率和约束条件进行组合优化，以实现投资效用函数的最大化。由于量化策略由多个子流程构成，局部的最优不一定代表全局最优。因此，将下游任务的需求纳入考量对于实现全局最优具有重要意义。

图 1：量化选股策略运行流程图



资料来源：中信证券研究部

因子优化和组合优化实际是求解两个优化问题：

1. 因子优化：给定数据 $(\mathbf{x}_1, c_1), \dots, (\mathbf{x}_n, c_n)$ ， \mathbf{x}_i 表示第 i 个样本的因子向量， c_i 表示第 i 个样本的收益，求解最优映射 $f^*(\cdot)$ 对 \mathbf{c} 进行估计得到 $\hat{\mathbf{c}} = f^*(\mathbf{x})$ ，

2. 组合优化：求解组合优化得到最优组合 \mathbf{w}^* ，优化问题的形式为：

¹ 在均方误差优化目标下模型输出值的物理含义是收益率，改变优化目标后，其物理意义可能发生变化，分布也可能不再符合收益率的特征，但后文为了便于理解和叙述，仍然延续这一说法，将模型输出值称为估计收益率。

$$\begin{cases} \max & \mathbf{c}^T \mathbf{w} \\ \text{s.t.} & \mathbf{w} \in S \end{cases}$$

其中 S 表示可行集。 S 只取决于因子暴露和行业属性，是确定的。因此 \mathbf{w}^* 可以看做是 \mathbf{c} 的一个函数，表示为 $\mathbf{w}^*(\mathbf{c})$ 。

在这一框架下，存在最优量化策略，其实现的收益是 $\mathbf{c}^T \mathbf{w}^*(\mathbf{c})$ ，而我们构建的策略实现的收益是 $\mathbf{c}^T \mathbf{w}^*(\hat{\mathbf{c}})$ 。从全局的角度来思考，量化策略的真正优化目标应该是其业绩表现。为了简化问题，我们暂且只考虑收益，则真实的损失定义为实现的量化策略的收益与最优量化策略的收益之差，即：

$$\text{loss}_{real} = \mathbf{c}^T \mathbf{w}^*(\hat{\mathbf{c}}) - \mathbf{c}^T \mathbf{w}^*(\mathbf{c}) = \mathbf{c}^T (\mathbf{w}^*(\hat{\mathbf{c}}) - \mathbf{w}^*(\mathbf{c}))。$$

根据该式可知， $\hat{\mathbf{c}}$ 的变动带来 $\mathbf{w}^*(\hat{\mathbf{c}})$ 的改变，从而影响 loss_{real} 。后文将从组合优化的角度入手分析 $\hat{\mathbf{c}}$ 的变动如何影响 $\mathbf{w}^*(\hat{\mathbf{c}})$ 。

最优组合只与估计收益率的方向有关

组合优化中使用的优化模型一般为线性规划，即目标函数和可行集 S 均为线性函数，表达式为：

$$\begin{aligned} & \max_{\mathbf{w}} \mathbf{c}^T \mathbf{w} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{w} = \mathbf{b} \\ & \mathbf{G}\mathbf{w} \leq \mathbf{h} \end{aligned}$$

其中， $\mathbf{w} \in \mathbb{R}^{K \times 1}$ 表示每个股票的权重，是需要求解的变量。 $\mathbf{c} \in \mathbb{R}^{K \times 1}$ 表示未来收益率。 $\mathbf{A} \in \mathbb{R}^{m \times K}$ ， $\mathbf{b} \in \mathbb{R}^{m \times 1}$ ， $\mathbf{G} \in \mathbb{R}^{p \times K}$ ， $\mathbf{h} \in \mathbb{R}^{p \times 1}$ 都是约束条件中的系数矩阵和向量。其中的等式条件一般约束的是总权重和为1，这一约束条件首先将可行集确定为一个高维空间中的超平面，其它的不等式约束可视为对这一超平面又进行了分割和限制。

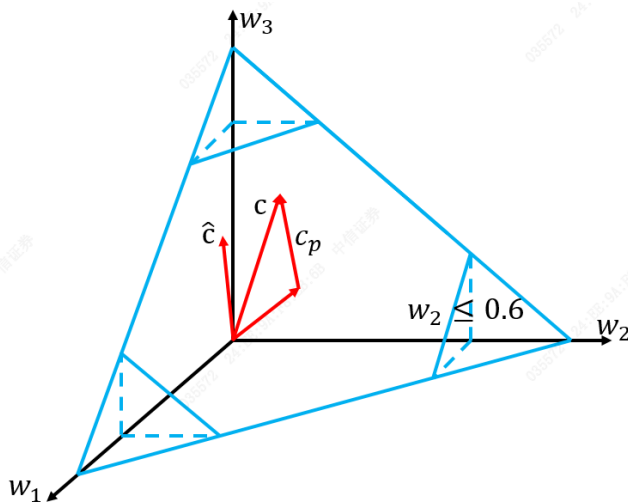
求解最优解 $\mathbf{w}^*(\mathbf{c})$ 的过程可以描述为：

1. 将 \mathbf{c} 投影到可行集的超平面上得到 \mathbf{c}_p ；
2. 在可行集上沿着 \mathbf{c}_p 的方向平移，与超平面相切的位置便是最优解。

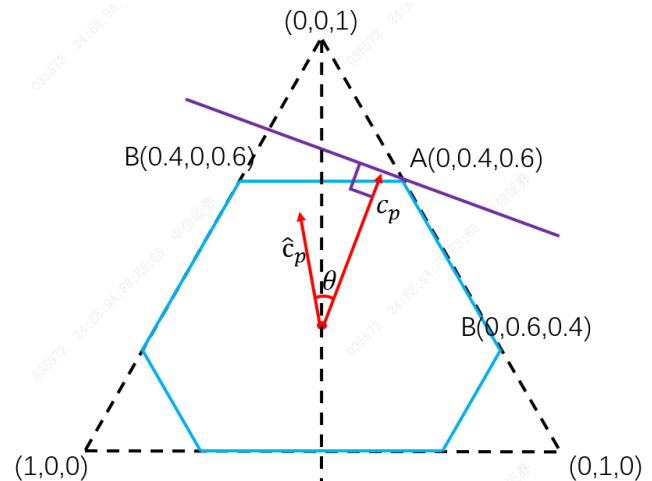
例：考虑3只股票的情形，假定股票真实的收益 $\mathbf{c} = [1\%, 3\%, 4\%]^T$ ，约束条件为权重和为1且单只股票的权重不超过0.6，则 $\mathbf{A} = [[1, 1, 1]]$ ， $\mathbf{b} = [1]^T$ ， $\mathbf{G} = [[1, 0, 0], [0, 1, 0], [0, 0, 1]]$ ， $\mathbf{h} = [0.6, 0.6, 0.6]^T$ 。对于该优化问题，其可行解的集合如下图所示。左图是三维空间的展示，权重和为1的可行集是蓝色的等边三角形，单只股票的权重不超过0.6相当于增加了三个半平面的限制，最终的可行集为一个六边形。以六边形的法向量为视角进行观察，看到的是一个二维平面上的六边形。

图 2：组合优化案例三维示意图

图 3：组合优化案例二维示意图



资料来源：中信证券研究部



资料来源：中信证券研究部

如右上图所示，沿着 c_p 平移最后与六边形相切于A点，即 $w^*(c) = (0,0.4,0.6)$ ，此时最优策略实现的收益为 3.6%。根据线性规划的相关理论和对几何示意图的观察，我们有以下重要结论：

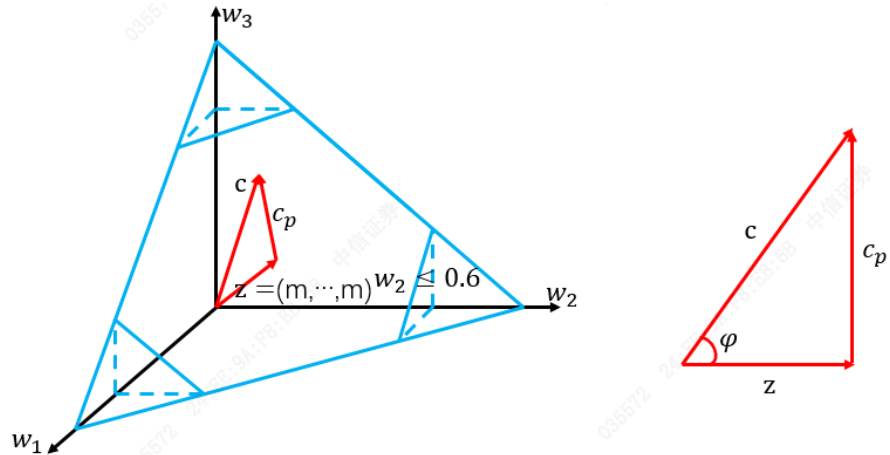
1. 在线性规划问题中，优化目标的最优解都是在极点处取得；
2. $w^*(c)$ 不是一个连续函数，当 c_p 转动很微小的夹角时， w^* 可能不变，也意味着 loss_{real} 不变，所以它也是一个非连续函数；
3. 对于给定的可行集 S ，最优解 w^* 只与 c 的方向有关，而与 c 的长度无关。那么在因子优化的步骤中，最小化 c 和 \hat{c} 的夹角是一个对 loss_{real} 很好的近似，即最大化 $\cos(c_p, \hat{c}_p)$ ；
4. 对于均方误差的优化目标，最小化的是 $\|c - \hat{c}\|$ ，这与 $\cos(c_p, \hat{c}_p)$ 的差异较大。均方误差之所以信噪比更低正是因为对向量长度也进行了优化，而这正是无效的。

下一节将通过证明 $\text{corr}(c, \hat{c})$ 与 $\cos(c_p, \hat{c}_p)$ 等价来说明优化 IC 实际就是在最大化 $\cos(c_p, \hat{c}_p)$ 。

优化 IC 等价于优化投影向量夹角

如下图所示，可行集的法向量为 $(1, \dots, 1)$ ，每个元素均为 1。 $z = (m, \dots, m)$ 是同方向的向量， $m > 0$ ，满足投影 $c_p = c - z$ ，这里不妨设 $\varphi < \frac{\pi}{2}$ 。

图 4：收益率向量及其在可行集上投影的示意图



资料来源：中信证券研究部

求解 m： 夹角的余弦值 $\cos(\varphi)$ 有两种计算方式。第一种采用向量直接计算：

$$\cos(\varphi) = \frac{\mathbf{c} \cdot \mathbf{z}}{||\mathbf{c}|| \times ||\mathbf{z}||} = \frac{m \sum c_i}{||\mathbf{c}|| \times \sqrt{nm^2}} = \frac{\sum c_i}{\sqrt{n} ||\mathbf{c}||}$$

第二种按照直角三角形直角边与斜边比值进行计算：

$$\cos(\varphi) = \frac{||\mathbf{z}||}{||\mathbf{c}||} = \frac{m\sqrt{n}}{||\mathbf{c}||}$$

两式相等得：

$$\frac{m\sqrt{n}}{||\mathbf{c}||} = \frac{\sum c_i}{\sqrt{n} ||\mathbf{c}||}$$

整理得：

$$m = \frac{\sum c_i}{n} = \bar{c}$$

证明 $\text{corr}(\mathbf{c}, \hat{\mathbf{c}}) = \cos(c_p, \hat{c}_p)$ ：

将 $m = \bar{c}$ 代入 $c_p = \mathbf{c} - \mathbf{z}$ 得：

$$c_p = \mathbf{c} - (\bar{c}, \dots, \bar{c}) = \mathbf{c} - \bar{\mathbf{c}}$$

则：

$$\cos(c_p, \hat{c}_p) = \frac{(\mathbf{c} - \bar{\mathbf{c}}) \cdot (\hat{\mathbf{c}} - \bar{\hat{\mathbf{c}}})}{||\mathbf{c} - \bar{\mathbf{c}}|| \times ||\hat{\mathbf{c}} - \bar{\hat{\mathbf{c}}}} = \text{corr}(\mathbf{c}, \hat{\mathbf{c}})$$

证毕。

该结论的证明表明优化 IC 实际上是优化估计收益率和真实收益率在可行集上投影的夹角，能更好的满足组合优化需求。

■ 可视化视角：相关性更关注向量方向

对目标函数进行可视化有助于我们对其进行理解。这里采用模拟数据来展示均方误差和基于相关性的目标函数。

假定有 n 只股票，股票收益受两种因子影响，且股票收益和因子是线性映射关系，即：

$$\mathbf{c} = \boldsymbol{\alpha}^T \mathbf{x} + \lambda \boldsymbol{\varepsilon}$$

其中 $\mathbf{x} \in \mathbb{R}^{n \times 2}$ 表示 n 个样本的因子， $\boldsymbol{\varepsilon} \in \mathbb{R}^{n \times 1}$ 表示噪声， λ 表示噪声系数， $\mathbf{c} \in \mathbb{R}^{n \times 1}$ 是股票收益率。首先采用标准正态高斯分布生成 100 组 2 维的 \mathbf{x} 和噪声 $\boldsymbol{\varepsilon}$ ，取 $\boldsymbol{\alpha}^* = [1, 2]^T$ ， $\lambda = 0.5$ 。根据公式能够计算得到 \mathbf{c} 。

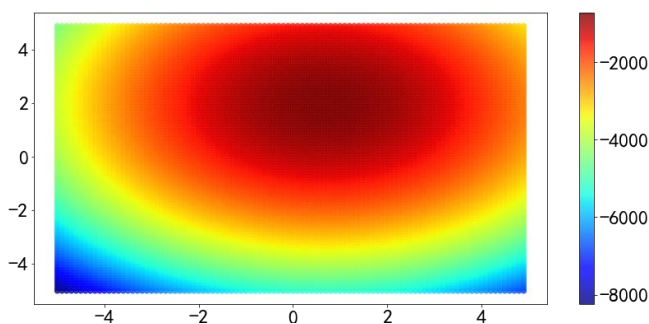
求解 $\boldsymbol{\alpha}^*$ 是迭代 $\boldsymbol{\alpha}$ 使 $\hat{\mathbf{c}}$ 和 \mathbf{c} 得更加接近的过程，距离的度量取决于我们定义的目标函数。对于不同的 $\boldsymbol{\alpha}$ (α_1 和 α_2 在 -5 到 5 变化)，都能够计算出 $\hat{\mathbf{c}}$ ，然后分别采用均方误差和相关性进行度量，得到如下的函数曲面图。横坐标表示 α_1 ，纵坐标表示 α_2 ，颜色表示 $\hat{\mathbf{c}}$ 和 \mathbf{c} 之间的距离，越接近深红，越靠近最优解。

采用均方误差度量时，函数曲面的等高线是类似的同心椭圆形，优化算法收敛结果 $\hat{\boldsymbol{\alpha}}$ 等概率落在 $\boldsymbol{\alpha}^*$ 的四周；而采用相关性度量时，函数曲面的等高线是过原点的射线，算法的收敛点等概率地落在 $\boldsymbol{\alpha}^*$ 同方向的射线附近。这表明，线性模型下虽然目标函数度量的是 $\hat{\mathbf{c}}$ 和 \mathbf{c} 的距离，优化过程会将其转变为 $\hat{\boldsymbol{\alpha}}$ 和 $\boldsymbol{\alpha}^*$ 之间的距离，并且距离的性质基本不变，即均方误差下仍然为欧氏距离，相关性为夹角距离。

为因子确定权重的目的实际是区分不同的因子的重要性和方向性，也即各个因子权重的相对大小关系。从可视化的函数曲面看相关性作为距离度量时更符合这一目的。

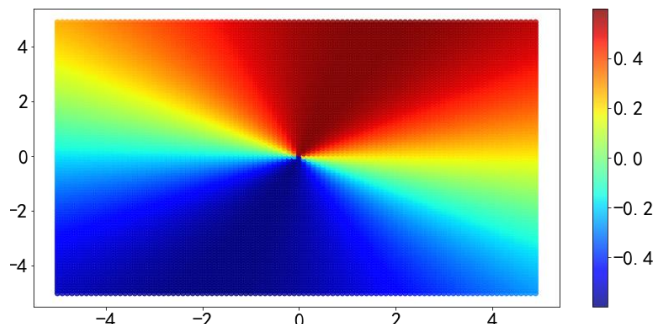
三维空间中的相关性函数的曲面有助于直观地理解优化过程，该函数虽然不是一个凸函数，但仍然能够通过梯度上升的方法收敛到最优解。

图 5：均方误差函数（MSE）曲面



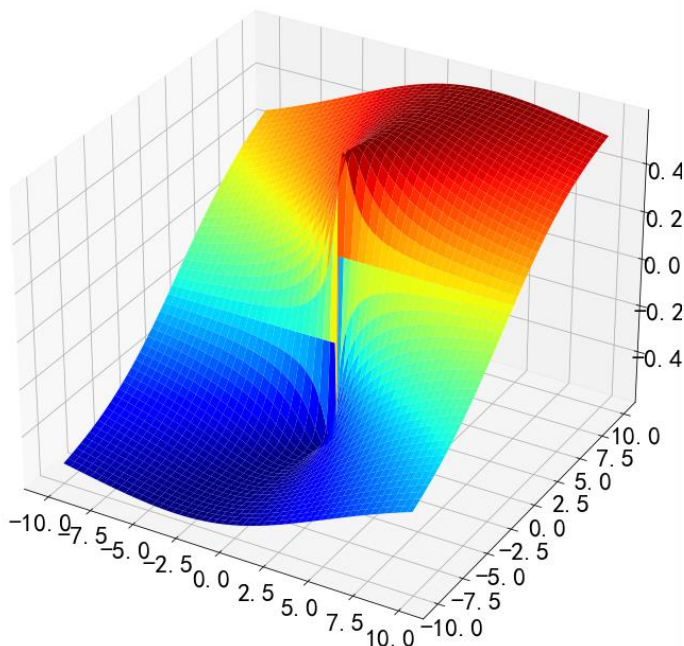
资料来源：中信证券研究部

图 6：相关性函数（Correlation）曲面



资料来源：中信证券研究部

图 7：相关性函数（Correlation）曲面 3D



资料来源：中信证券研究部

■ 模型行为视角：学习因子间的对冲关系

因子数据及其处理流程

我们采用的因子如下表所示，共计 21 个，分属 8 大类。对每一指标计算其分位数，在计算 IC 时则自动为 RankIC 以提升稳定性。具体步骤如下：

1. 对样本空间的股票按照中信一级行业分为 30 组。
2. 在每一个行业组内按市值排序分为 3 组，则截面上共计有 90 个小组。
3. 在每一个小组中对每一个因子排序，并转化为排序分位数，即：小组排序/小组股票数。
4. 对于下一期的收益，我们也采用上述同样的做法。

经过上述的处理，因子和收益都进行了行业市值中性化。

表 1：模型涉及的全部指标

指标维度	指标描述	方向
偿债能力	有息负债/总资产	-
	短期负债/货币资金	-
盈利能力	净资产收益率(TTM)	+
	销售毛利率(TTM)	+
	单季度净资产收益率	+

	单季度销售毛利率	+
	单季度扣除非经常损益后的净利润 / 净利润	+
	单季度经营活动产生的现金流量净额 / 营业收入	+
盈利波动性	净资产收益率波动性	-
	毛利率波动性	-
成长能力	单季度营业收入同比增长率	+
	单季度净利润同比增长率	+
经营能力	存货周转率	+
	应收账款周转率	+
	总资产周转率	+
分析师	一致预期营收同比	+
	一致预期净利润同比	+
	一致预期 2 年复合增长率	+
	4 周一致预期净利润变化率	+
投资者认可度	北向持有占比	+
估值分位数	EP 5 年分位数	+

资料来源：中信证券研究部

通过因子间的对冲关系进行因子优化

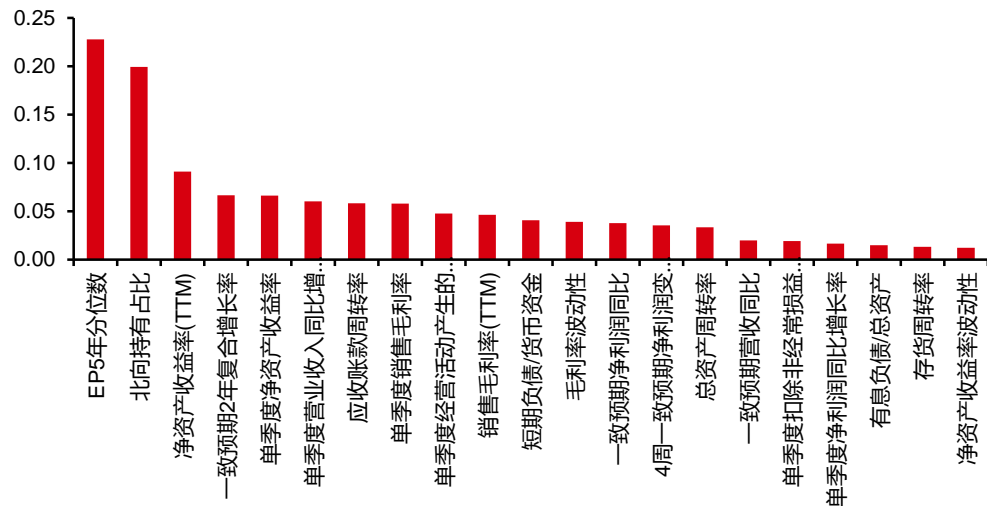
深度学习一个经常为人诟病的地方在于它的“黑盒”属性，因为经过多层非线性网络得到的函数形式通常非常复杂，且不够直观。对于这种复杂函数，我们可以根据泰勒公式能够近似原函数的理论，对 $f^*(\cdot)$ 求一阶微分来研究各个因子对模型输出值的影响。

每一次网络训练完成后，求出输出值对每一个因子的一阶偏导 f'_i ，并采用截面所有样本计算平均偏导值，即：

$$f'_i = \frac{\partial f}{\partial x_i}, \text{Mean}(f'_i) = \frac{1}{n} \sum_j f'_i(x_{1j}, \dots, x_{ij}, \dots, x_{Kj})$$

其中 x_{ij} 表示第 j 个样本的第 i 个因子值。平均偏导值可以表示该因子对网络输出值的整体影响。进一步计算时序的因子偏导绝对值的平均值可以观察各个因子的重要性，结果如下图所示。整体来看，估值因子 EP5 年分位数、北向持有占比以及财务因子净资产收益率（TTM）对于网络输出值的影响最大。

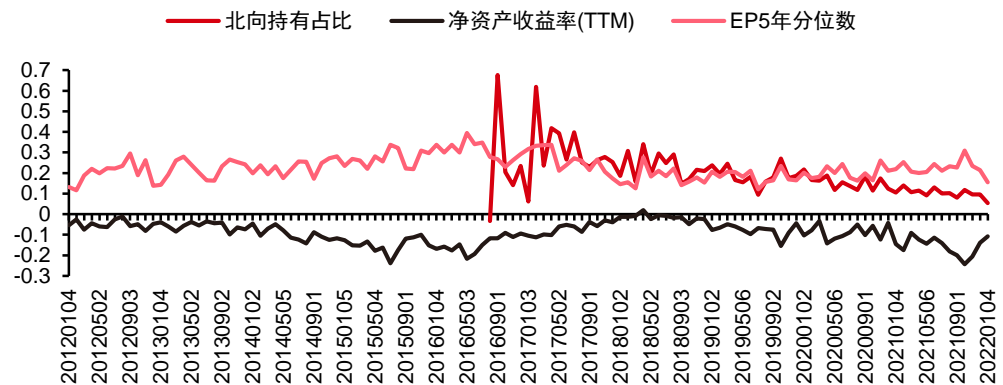
图 8：各因子的平均（绝对值）梯度



资料来源：Wind，朝阳永续，中信证券研究部测算

从时序走势来看，北向持有占比在最初的两年比较有效，但是影响随时间逐渐衰减。而净资产收益率（TTM）和 EP5 年分位数并没有衰减的趋势，仅呈现一定的趋势性。值得注意的是，逻辑上随着净资产收益率（TTM）升高，未来收益率应该也越高，即梯度为正值，而图中显示其为负梯度。除此之外，还能观察到净资产收益率（TTM）和 EP5 年分位数呈现较明显的负相关性，相关系数达到-0.61。该结果意味着负梯度可能并不是出现了错误，而是模型学习到了因子间的对冲关系。

图 9：三因子的平均梯度走势



资料来源：Wind，朝阳永续，中信证券研究部测算

将 EP5 年分位数与净资产收益率（TTM）的差作为一种新的因子，如果该因子能够实现更好的效果，便能对上述猜想进行验证。将 EP5 年分位数的梯度作为因变量，净资产收益率（TTM）的梯度作为自变量构建线性回归方程，得到系数为-0.6347，则估值 ROE 对冲因子的计算式为：

$$\text{估值 ROE 对冲因子} = \text{EP5 年分位数} - 0.6347 * \text{净资产收益率 (TTM)}$$

常规思路下，应该选取估值低且 ROE 高的股票，则估值 ROE 等权因子计算式为：

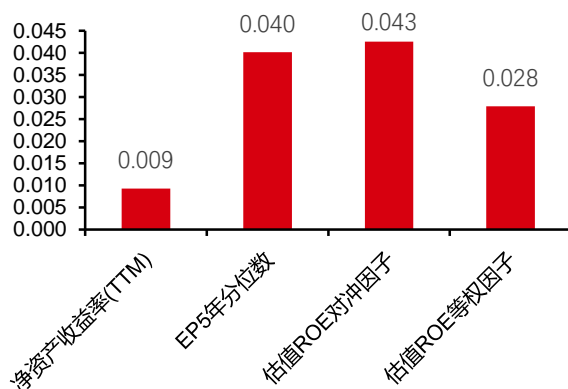
$$\text{估值 ROE 等权因子} = \text{EP5 年分位数} + \text{净资产收益率 (TTM)}$$

EP5 年分位数、净资产收益率 (TTM) 以及估值 ROE 等权因子作为对照，结果如下图所示。从平均 IC 和 IC 大于 0 的占比来看，估值 ROE 对冲因子均好于三个对照因子，而按照常规思路构建的估值 ROE 等权因子效果则是位于两个单因子之间。

从 IC 时间序列来看，EP5 年分位数和净资产收益率 (TTM) 因子有较强正相关性，并且均有较强的风格轮动性，在风格产生转换时可能产生较大回撤，如 2018 年 Q4 和 2021 年 Q1，通过等权方式构建复合因子也很难避免这一问题。而对冲方式的逻辑可理解为选择了同等 ROE 水平下估值更低的股票，实际上是 ROE 的估值溢价逻辑，不同于单因子下风格延续的逻辑，因此能够有效解决风格切换的问题。

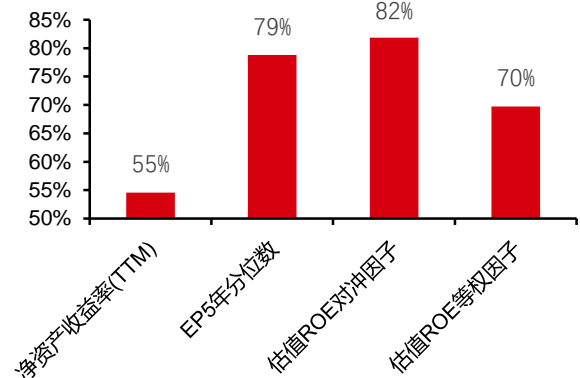
总的来说，基于相关性的深度因子优化可能能够寻找因子间的对冲关系来实现了更加稳定超额收益。

图 10：四个因子的平均 IC



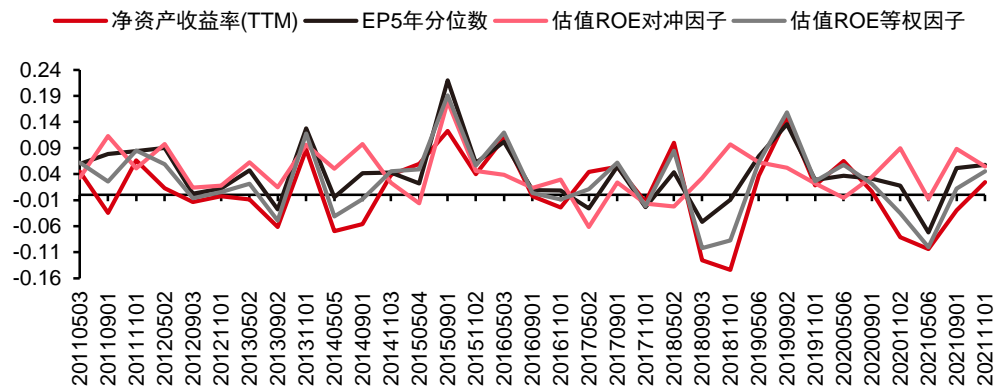
资料来源：Wind，朝阳永续，中信证券研究部测算

图 11：四个因子的 IC 大于 0 的次数占比



资料来源：Wind，朝阳永续，中信证券研究部测算

图 12：四个因子的 IC 序列



资料来源：Wind，朝阳永续，中信证券研究部测算

■ 组合优化下的模型回测

在之前发布的研究报告《多因子量化选股系列专题研究—破局低信噪比：基于深度学习的因子优化研究》（2021-11-23）中，我们直接选取分数最高的 100 只股票测试了深度相关性模型（DCM）的效果。后文将采用同样的股票打分进行组合优化，控制市值和行业的风险暴露，测试组合优化后的策略效果。由于控制了行业和市值的暴露后跟踪误差已经可以控制在较小范围，所以不需要限制其它风格因子的暴露。

组合优化设置：

1. 行业和市值的暴露均限定在 0.1%以内。
2. 个股偏离限定在 1%以内。

5 种基线系统

我们构建如下 5 种模型作为对比的基线系统：

1. 等权加权。直接将所有因子的值等权加总。
2. 历史 IC 加权。计算每个因子过去 3 年的平均 IC，并以 IC 值作为权重加总所有因子。
3. 最大化 IC。按照最优解形式计算权重并以此加总所有因子。
4. 最大化 IC_IR。按照最优解形式计算权重并以此加总所有因子。
5. Loss1 神经网络，采用加权的均方误差作为优化目标，网络结构与 DCM 完全一致。

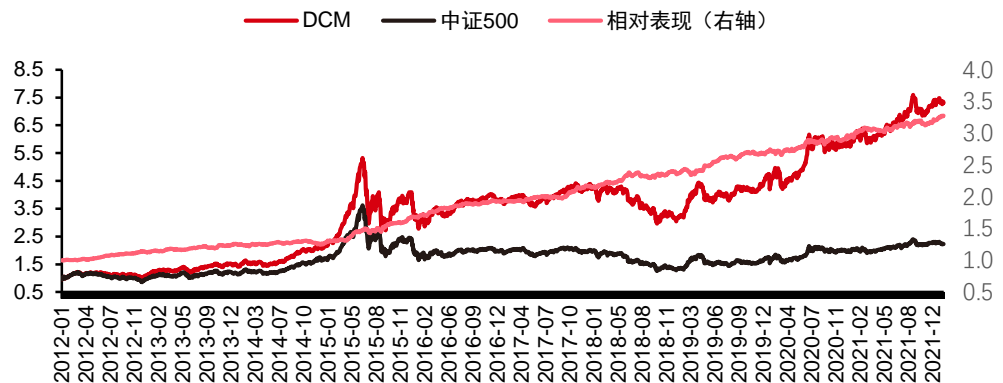
对于以上五种基线系统，以全市场非 ST 股票为样本空间，均选择分数最高的 100 只股票，行业内按自由流通市值加权，行业间按中证 500 行业权重加权；涉及最优化的方法 3、4、5 均采用过去 3 年的数据；调仓频率为月度，交易手续费设为单边 3%。DCM 模型也采用相同的设置，与基线系统唯一的差异就是各自输出的打分不同。测试区间为 2012 年

1月1日至2022年1月19日。

策略历史表现

在测试区间上，DCM 策略整体实现 13.39%的年化超额收益，信息率 2.63，超额最大回撤 5.02%。并且信息率较选取前 100 股票实现的 2.31 有进一步提升。分年度来看，除 2014 年和 2018 年偏弱外，其它年份均能实现 10%以上的超额收益。

图 13：策略的净值走势



资料来源：Wind，朝阳永续，中信证券研究部测算

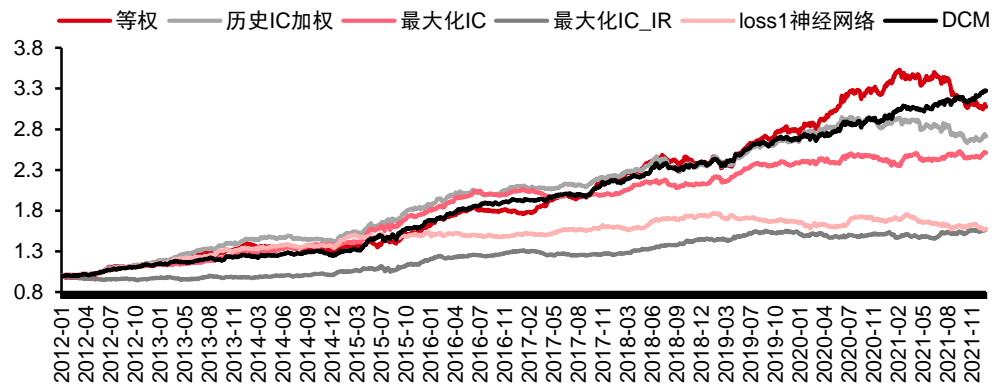
表 2：策略的历史表现

	年化收益 (%)	年化超额收 益(%)	年化波动率 (%)	年化超额波 动率(%)	夏普比率	信息率	最大回撤 (%)
全部	21.59	13.39	25.20	5.09	0.86	2.63	48.81
2022	-37.22	11.35	12.17	3.41	-3.06	3.33	2.97
2021	25.44	10.07	15.41	5.41	1.65	1.86	9.86
2020	33.48	12.90	24.85	5.98	1.35	2.16	15.28
2019	41.57	15.68	21.04	4.87	1.98	3.22	16.04
2018	-26.09	6.90	24.06	5.00	-1.08	1.38	32.62
2017	11.53	11.73	13.27	4.26	0.87	2.75	10.51
2016	-5.73	11.78	30.17	3.96	-0.19	2.97	30.78
2015	89.24	46.96	45.34	7.59	1.97	6.18	48.81
2014	38.63	0.55	18.93	4.58	2.04	0.12	11.81
2013	30.43	13.39	21.60	4.30	1.41	3.11	15.47
2012	15.63	13.04	22.28	3.54	0.70	3.69	19.65

资料来源：Wind，朝阳永续，中信证券研究部测算

对比各个基准模型，DCM 在年化超额，信息率和超额最大回撤三个方面均是最优。

图 14：各模型相对中证 500 的相对收益



资料来源：Wind，朝阳永续，中信证券研究部测算

表 3：各模型的历史表现

	年化收益 (%)	年化超额收益 (%)	年化波动率 (%)	年化超额波动率 (%)	夏普比率	信息率	最大回撤 (%)	超额收益最大回撤 (%)	最大回撤天数
等权	20.83	12.63	25.21	6.07	0.83	2.08	46.93	13.60	65
历史 IC 加权	19.38	11.17	25.36	5.63	0.76	1.98	49.28	10.97	817
最大化 IC	18.47	10.26	25.06	4.69	0.74	2.19	52.94	6.39	817
最大化 IC_IR	13.10	4.90	24.61	4.56	0.53	1.07	54.18	6.71	817
loss1 神经网络	13.11	4.91	24.92	4.87	0.53	1.01	59.06	12.18	817
DCM	21.59	13.39	25.20	5.09	0.86	2.63	48.81	5.02	65

资料来源：Wind，朝阳永续，中信证券研究部测算

结论与投资建议

本报告从三个视角，包括理论视角，可视化视角和模型行为视角对基于相关性的因子优化进行了更加深入的探讨，主要有以下结论：

1. 理论视角：从量化策略全局优化的角度出发，需要将组合优化的需求纳入考虑。组合优化中最优组合只与估计收益率的向量方向有关，因此优化估计收益率和实际收益率在可行集上投影的夹角是正确的选择，我们进一步证明了优化夹角等价于优化 IC。
2. 可视化视角：欧式距离曲面的等高线是同心的椭圆，而相关性距离曲面的等高线是过原点的射线，后者更关注向量的方向性，也就更适合学习不同因子的相对权重。
3. 模型行为视角：训练完成后，模型输出值对因子求一阶偏导可以用来观察各个因子对输出值的影响，历史中影响最大的因子分别是估值因子 EP5 年分位数、北向持有占比以及财务因子净资产收益率（TTM）。模型还学习到了 EP5 年分位数和净资产收益率（TTM）之间的对冲关系从而实现更稳定的超额收益。
4. 组合优化下的模型回测：在 2012 年 1 月 1 日至 2022 年 1 月 19 日的测试区间上，DCM 策略整体实现 13.39% 的年化超额收益，信息率 2.63，超额最大回撤 5.02%。并且信息率较选取前 100 股票实现的 2.31 有进一步提升。

总的来说，从量化策略全局最优出发得出拟合夹角是正确的选择，并且证明其与优化IC等价为基于相关性的因子优化提供了理论支撑；另一方面，实证结果验证了该方法的有效性。除此之外，我们采用的可视化方法和求偏导观察因子影响的方法可作为投资者分析模型的重要手段。

■ 风险因素

训练的随机性风险；因子效果衰减风险；历史业绩不代表未来表现。

■ 相关研究

多因子量化选股系列专题研究—破局低信噪比：基于深度学习的因子优化研究

(2021-11-23)

量化策略专题研究—寻找股权激励背后的预期差

(2022-01-26)

量化策略专题研究—财务视角下的负向组合构建研究

(2021-05-20)

量化策略专题研究—寻找业绩预期未被透支的黑马股

(2021-05-20)

量化策略专题研究—量化视角下的市场风格分析

(2021-04-10)

量化策略专题研究—AH溢价因素的分析与策略应用

(2020-11-02)

量化策略专题研究—量化新闻情绪，把握投资先机

(2020-06-02)

量化策略专题研究—以史为鉴，可知兴替：模式识别视角的行业轮动策略

(2020-01-09)

量化策略专题研究—机器学习在量化投资中的应用探讨

(2020-02-05)

量化策略专题研究—行业趋势配置模型研究

(2020-03-25)

量化策略专题研究—公募持仓视角下的行业配置策略

(2020-03-25)

量化策略专题研究—顺势而为：行业趋势配置模型研究

(2020-04-09)

量化策略专题研究—寻找分析师前瞻观点中的Alpha

(2020-05-14)



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。

评级说明

投资建议的评级标准		评级	说明
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；韩国市场以科斯达克指数或韩国综合股价指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅 20%以上
		增持	相对同期相关证券市场代表性指数涨幅介于 5%~20%之间
		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
	行业评级	强于大市	相对同期相关证券市场代表性指数涨幅 10%以上
		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

其他声明

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”。

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 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.（金融服务牌照编号：350159）分发；在美国由 CLSA group of companies（CLSA Americas, LLC（下称“CLSA Americas”）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：孟买（400021）Nariman Point 的 Dalamal House 8 层；电话号码：+91-22-66505050；传真号码：+91-22-22840271；公司识别号：U67120MH1994PLC083118；印度证券交易委员会注册编号：作为证券经纪商的 INZ000001735，作为商人银行的 INM000010619，作为研究分析商的 INH000001113）；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）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：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经营范围包括证券投资咨询业务。

美国：本研究报告由中信证券制作。本研究报告在美国由 CLSA group of companies（CLSA Americas 除外）仅向符合美国《1934 年证券交易法》下 15a-6 规则定义且 CLSA Americas 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载任何观点的背书。任何从中信证券与 CLSA group of companies 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。

新加坡：本研究报告在新加坡由 CLSA Singapore Pte Ltd.（资本市场经营许可持有人及受豁免的财务顾问），仅向新加坡《证券及期货法》s.4A（1）定义下的“机构投资者、认可投资者及专业投资者”分发。根据新加坡《财务顾问法》下《财务顾问（修正）规例（2005）》中关于机构投资者、认可投资者、专业投资者及海外投资者的第 33、34 及 35 条的规定，《财务顾问法》第 25、27 及 36 条不适用于 CLSA Singapore Pte Ltd.。如对本报告存有疑问，还请联系 CLSA Singapore Pte Ltd.（电话：+65 6416 7888）。MCI (P) 024/12/2020。

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。

一般性声明

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。

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所包含的材料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、价值及收益可能会受汇率影响而波动。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所载明的内容产生的直接或间接损失承担任何责任。

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。