

2022年05月12日

证券研究报告·金融工程专题报告

因子选股系列

西南证券
SOUTHWEST SECURITIES

基于相似股票历史收益的选股因子研究

摘要

在美股市场中叱咤风云的各种动量因子在我国 A 股市场中往往收效甚微, 是什么导致了同一个因子在不同的市场中经常表现出截然不同的实证结果? 近年来, 伴随着行为金融学的高速发展, 行为金融学从多个角度解释了动量效应背后的成因。本文从投资者的**投资心理**和**投资行为**出发, 探讨相似股票的历史收益是否可以**为因子选股提供新的信息**。

我们从股票价格、规模、风险、盈利能力、投资水平五个维度定义了**股票距离**, 用来衡量股票与股票之间的相似程度, 进而构建了**相似股票动量因子**, 在我国 A 股市场对其进行了有效性检验。**相似动量因子同股票的下期收益正相关**, 相似股票动量因子 **IC 均值为 0.0305**, **IR 为 0.2689**。多空组合平均月收益率为 **1.01%**, 组合年化收益率为 **12.89%**, 最大回撤仅为 **8.39%**。

在此基础上, 本文对该因子进行了相关系数分析与 Fama-Macbeth 回归分析, 结果表明, 在**排除了异质波动、流动性、动量与规模等影响因素的情况下**, **相似股票动量因子依旧显著有效, 可以为选股提供新的信息**。

我们还对相似动量因子的参数进行了**敏感性分析**。结果表明, 采用股票过去一个月的平均收益率衡量股票历史收益, 选择**距离最近的 80 只股票**或选择**股票距离小于等于 1 的股票**构建相似股票样本可以保证相似动量因子稳定有效。

最后, 我们对相似股票动量因子进行了改进。我们结合股票的历史收益, 构建了**相似预期差因子**。相似预期差因子选股卓有成效, 控制股票历史超额收益后, 信息系数 **IC 均值高达 0.0910**, 多空组合平均月收益率为 **1.44%**, 组合年化收益率为 **18.72%**, 最大回撤仅为 **8.60%**。样本期内, 使用相似预期差因子构建多头组合收益率远超市场组合, **年复合增长率高达 25.23%**。

风险提示: 本研究基于 2016-2021 年中国 A 股市场历史数据, 因子的历史收益率不代表未来收益率。若未来市场环境发生较大变化, 因子的实际表现可能与本文的结论有所差异。

西南证券研究发展中心

分析师: 邓璎函

执业证号: S1250517080005

电话: 023-67507084

邮箱: dyh@swsc.com.cn

相关研究

1. 基于吸收比率规避市场下行风险的资产配置策略 (2022-02-09)
2. 基于“集中度”识别拥挤交易的行业轮动策略 (2022-01-26)
3. 基于价格动量和基本面动量的双动量选股策略 (2021-11-19)
4. 基金管理的主动性指标构建与业绩预测 (2021-10-26)
5. BARRA 中国市场模型 (CNE6) 解读——因子选股系列 (2019-02-26)
6. 波动率因子的改进: 异质波动率——因子选股系列 (2018-08-29)
7. 基于方向波动率的选股因子研究 (2017-12-21)
8. 风险度量方法探讨——混合风险指标 (2017-09-18)



目录

1 因子构建	1
1.1 股票距离与相似股票	1
1.2 相似股票动量因子构建	1
2 因子有效性检验	2
2.1 分组检验	2
2.2 信息系数分析	4
2.3 相关分析与回归分析	4
3 敏感性分析	7
3.1 相似股票样本构建方式的敏感性分析	7
3.2 历史收益指标的敏感性分析	9
4 因子的改进与优化：引入历史收益优化选股策略	10
5 总结与展望	14
6 风险提示	15
参考文献	16



图 目 录

图 1: 相似股票动量因子分组净值图.....	3
图 2: 多空组合累计净值变化图.....	3
图 3: 相似股票动量因子累计信息系数变化示意图.....	4
图 4: 样本容量与 IC 均值的变化示意图.....	7
图 5: 西南证券 (600369.SH) 股票距离频数分布直方图.....	8
图 6: IC 均值与股票距离阈值的变化情况示意图.....	8
图 7: 历史收益的度量与 IC 均值变化示意图.....	9
图 8: 相似预期差因子分组净值图.....	11
图 9: 相似预期差因子分组净值图.....	12
图 10: 相似预期差因子累计信息系数.....	13
图 11: 相似预期差多头组合与市场指数净值变化图.....	14

表 目 录

表 1: 相似股票动量因子分组检验结果.....	2
表 2: 相似股票动量因子信息系数分析结果.....	4
表 3: 因子相关系数矩阵.....	5
表 4: 单因子回归结果.....	5
表 5: 多因子 Fama-Macbeth 回归结果.....	6
表 6: 相似预期差因子分组检验结果.....	12
表 7: 相似预期差因子信息系数分析结果.....	13



股票的历史收益是股票未来收益的重要参考, 相似股票的历史数据也是选股的重要依据。那么相似股票与历史收益的双管齐下能否相得益彰? 本文将从月度低频数据出发, 研究相似股票的历史收益对选股的指导意义。

1 因子构建

1.1 股票距离与相似股票

为了衡量股票与股票之间的相似性, 借鉴 He, Wang 和 Yu (2021) 的思路与方法, 本文将从股票价格、规模、风险、盈利能力、投资水平五个维度来刻画股票间的相似程度。我们将*i*股与*j*股在*t*时刻的距离 D_{ij}^t 定义为二者在股价、市值、账面市值比、净资产收益率、资产增长率之间的欧氏距离:

$$D_{ij}^t = \sqrt{(P_i^t - P_j^t)^2 + (Sz_i^t - Sz_j^t)^2 + (BM_i^t - BM_j^t)^2 + (RE_i^t - RE_j^t)^2 + (Iv_i^t - Iv_j^t)^2}$$

其中, P_i^t 为*i*股在*t*月末的收盘价, Sz_i^t 为*i*股去年十二月底或今年六月底总市值的自然对数值, BM_i^t 为*i*股去年十二月底资产净值与总市值之比, RE_i^t 为*i*股去年十二月底的净资产收益率, Iv_i^t 为*i*股去年十二月底的资产增长率。

为了降低各个指标不同的数量级和量纲对股票距离测算产生的偏误, 计算距离之前, 以上指标均需要在截面上进行标准化处理。并且, 为了避免动量因子对实证结果产生干扰, 我们在构建股票距离指标时并没有引入股票的历史收益率。

两只股票的距离越小, 两只股票的相似程度就越高, 当两只股票距离比较小时, 我们就可以称这两只股票为相似股票。基于以上对股票距离与相似股票的定义, 接下来我们可以构建相似股票动量因子。

1.2 相似股票动量因子构建

某只股票相似股票的历史收益是否会影响该股票未来的收益呢? Tversky 和 Kahneman (1974) 研究发现, 当人们试图对某件事物进行判断或分类时, 往往会采用**代表性启发法 (representativeness heuristic)**。当投资者判断某只股票的未来收益时, 往往会采用与其相似的某只股票的历史收益来推断其未来表现。倘若某只股票的相似股票历史收益率较高, 那么投资者对这只股票的收益预期往往也比较高。

不仅如此, 根据行为金融学中的**后悔理论 (Bell, 1982)**, 如果一只股票过去产生了较高收益, 但投资者偏偏与这支股票擦肩而过, 投资者会因此痛苦、自责后悔不已。为此, 投资者往往会寻找相似但还未上涨的股票进行投资。这也就意味着, 和历史表现优秀的股票相类似的股票的投资需求会增加。

更进一步, 如果投资者在对某只股票的投资中收获颇丰, 他的投资思维容易产生路径依赖, 之后的投资依然会偏爱相似的股票, 这就是投资者预期偏差中的**锚定效应 (Tversky and Kahneman, 1974)**。



为了检验某只股票相似股票的历史收益对该股未来收益的影响，基于上文对股票距离与相似股票的定义，给定“半径”参数 r ，我们将相似股票动量因子 $SIM_i^{(t)}$ 定义为同股票 i 在时刻 t 距离最近的 r 只股票过去一个月历史收益率的市值加权平均值：

$$SIM_i^{(t)} = \sum_{k=1}^r W_k R_k^{(t-1)} \quad W_k = Sz_k / \sum_{j=1}^r Sz_j \quad k \in \left\{ k \mid \min_k \sum_{k=1}^r D_{i,k}^{(t)} \right\}$$

其中， W_k 代表 k 股以市值加权的权重， $R_k^{(t-1)}$ 代表 k 股 $t-1$ 月的收益率。股票 i 的相似股票动量因子 $SIM_i^{(t)}$ 取值越小，意味着股票 i 的相似股票历史表现较差；相似股票动量因子 $SIM_i^{(t)}$ 取值越大，意味着相似股票历史收益率越高。

2 因子有效性检验

因子有效性检验的样本期为 2016 年 12 月到 2021 年 12 月，为了保证找到每只股票最相似 r 只股票，我们对上市交易的全部 A 股进行回溯。为了防止股票自身的动量因子对回溯测试的结果产生干扰，我们在测算股票距离时将每只股票同自身的距离设定为 $+\infty$ 。

2.1 分组检验

为了检验相似股票动量因子的有效性，给定参数 $r=80$ ，对相似股票动量因子进行分组检验。后续我们也将就因子参数的敏感性进行实证分析。

首先，根据股票距离的定义，我们计算每只股票同 A 股市场中的其他股票的距离从而确定此刻同该股最相似的 80 只股票构成相似股票样本；随后，本文用相似股票过去一个月的历史收益数据计算因子值；最后按照当期因子值大小将股票平均分为 5 组，计算下一个组内平均收益率。为了避免市值对组合收益的测算产生偏误，我们选用算数平均计算组合收益，在计算每一期的因子值和下一期收益率时，为了减少误差，剔除相关数据缺失的股票。

表 1 对应相似动量因子分组检验结果。在分组检验的基础上，我们计算了多空组合的 t 统计量，判断了多空组合月度平均收益的显著性。与此同时，根据每组因子选股的结果建立了相应的投资组合，测算了风险收益指标，结果如下表所示。

表 1：相似股票动量因子分组检验结果

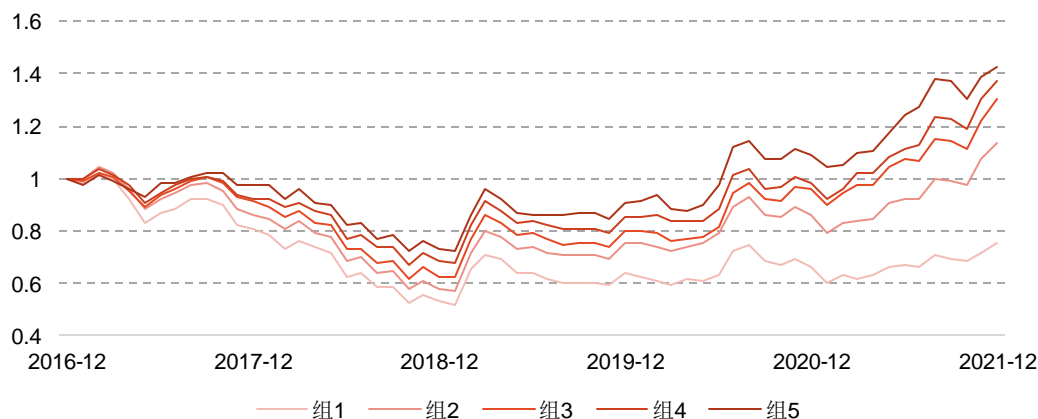
	组 1	组 2	组 3	组 4	组 5	组 5-组 1	(4+5) - (1+2)
因子均值	-0.0315	-0.0132	-0.0017	0.0105	0.0359	-	-
平均月收益率	-0.29%	0.39%	0.61%	0.68%	0.72%	1.01%***	1.30%***
t 统计量	-	-	-	-	-	2.71	2.49
年化收益率	-3.54%	4.79%	7.59%	8.50%	9.02%	12.89%	16.88%
Sharp 比率	-0.26	0.13	0.27	0.33	0.39	1.08	1.06
最大回撤率	49.05%	45.00%	40.01%	35.61%	29.15%	8.39%	11.18%
月胜率	41.67%	48.33%	45.00%	53.33%	48.33%	66.67%	66.67%

注：“***”、“**”、“*” 分别代表参数在 1%、5%、10% 的显著性水平下显著，下同，不再说明。数据来源：西南证券



根据表 1 数据, 自组 1 到组 5, 伴随着相似股票动量因子值的逐步增大, 各组平均月收益率单调递增。基于相似动量因子的分组结果分别构建投资组合在样本期进行回测, 每月的最后一个交易日进行调仓, 相似股票动量因子值较大的组合不仅在年化收益率、Sharp 比率、最大回撤率、月胜率等方面显著优于因子值较小的组合, 而且各指标均显示出显著的单调性。多空组合“组 5 - 组 1”平均月收益率为 1.01%, 在 99% 的置信水平下显著非零, 其年化收益率为 12.89%, 组合的最大回撤率仅为 8.39%, 表现稳定。

图 1: 相似股票动量因子分组净值图

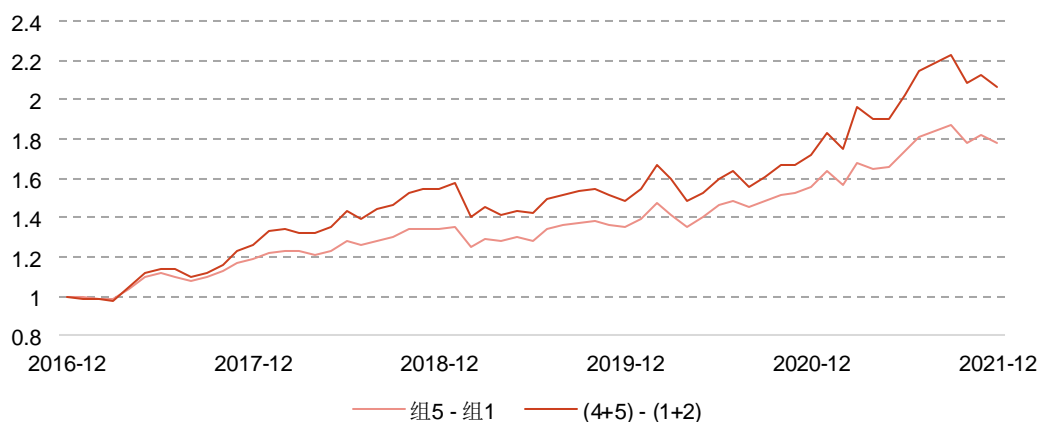


数据来源: 西南证券

根据回溯测试的结果, 我们在图 1 中绘制了样本期内相似股票动量因子的分组累计净值图。如图可见, 在整个样本期内, 不论是在 2018 年的熊市, 还是在 2019 年的牛市, 各组合的历史净值存在显著差异且具备非常显著的单调性。因此, 我们判定相似股票动量因子有效, 且当期相似动量因子值同股票下期收益正相关。

在分组净值分析的基础上, 为了进一步考察因子在样本期内的实际市场表现, 我们在图 2 中分别绘制了多空组合“组 5 - 组 1”与“(4+5) - (1+2)”的历史净值变化情况。由图所示, 在样本期内, 无论是在 2018 年的熊市还是在 2019 年后的牛市, 相似股票动量因子多空组合净值始终保持稳步增长, 并且从未出现过大幅回撤, 表现优良。

图 2: 多空组合累计净值变化图



数据来源: 西南证券



2.2 信息系数分析

为了进一步验证相似股票动量因子的有效性，我们对相似动量因子的信息系数 IC 进行了实证分析。信息系数是指当期因子值与下期股票收益率之间的秩相关系数，IC 越大，相似动量因子对股票下期收益率的预测作用也就越强。因子信息系数分析的结果见下表。

表 2：相似股票动量因子信息系数分析结果

	IC 均值	IC 标准差	t 统计量	IC 与均值同向的比例	IR
相似股票动量因子	0.0305**	0.1134	2.0832	61.67%	0.2689

数据来源：西南证券

根据表 2 数据，在样本期内，相似股票动量因子的 IC 均值为 **0.0305**，在置信水平 95% 上强烈拒绝 IC 均值为 0 的原假设，IC 与均值同向的比例高达 61.67%，IR 比率（IC 均值除以标准差）为 0.2689。相似股票动量因子具备较强的有效性和稳定性。为了进一步考察相似股票动量因子的有效性，我们在图 3 中绘制了在样本期内相似动量因子的累计信息系数变化情况。

图 3：相似股票动量因子累计信息系数变化示意图



数据来源：西南证券

如图所示，从因子的累计信息系数来看，尽管存在一定的波动，相似股票动量因子的累计 IC 在总体上以一个相对固定的速率保持平稳上升，表明因子在样本期内的表现比较稳定，相似股票动量因子同股票下期收益呈现显著的正相关关系。

综上所述，通过分组检验和信息系数分析，可以得出以下结论：**在样本期内，相似股票动量因子有效且与股票的下期收益正相关。**

2.3 相关分析与回归分析

上文我们分别通过分组检验及信息系数分析论证了相似股票动量因子的有效性，但是，该因子能否在选股中发挥作用取决于它是否可以提供目前尚未被市场定价的新信息。倘若相似动量因子提供的信息已经被市场定价或与其它已知因子提供的信息高度重合，那么即便该因子显著有效也难堪大用。因此，在论证因子有效性的基础上，接下来我们将通过相关性分析与回归分析来证明相似动量因子揭示了有关股价变化的新信息。



我们选用异质波动率（Ang et al., 2006）、Amihud 非流动性指标（Amihud, 2002）、股票前一月的历史收益率、公司的流通市值来分别刻画股票在异质波动、流动性、动量、规模四个维度的信息，以此为基础，证明相似股票动量因子为选股提供了新的信息。以上指标均选用月度数据，其中我们选用 Fama-French 五因子模型（Fama & French, 2015）作为测算异质波动率的基础定价模型。

为了考察相似股票动量因子同其它因子的内在联系，我们首先计算了相似动量因子与其它因子的次序相关系数。

表 3：因子相关系数矩阵

	相似动量	异质波动率	非流动性	动量	流通市值
相似动量	1.000	—	—	—	—
异质波动率	0.063***	1.000	—	—	—
非流动性	-0.141***	-0.121***	1.000	—	—
动量	0.543***	0.209***	-0.118***	1.000	—
流通市值	0.046***	0.002	-0.761***	0.097***	1.000

数据来源：西南证券

根据相关系数矩阵，相似股票动量因子同异质波动率、流通市值的相关性较弱，信息重合度较低，而非流动性、动量相关性较强，可能存在较大的信息重合。为了进一步验证相似股票动量因子能否为选股提供新信息，我们采用 Fama-Macbeth 方法（Fama & Macbeth, 1973）对以上因子和股票收益率进行回归分析。

首先，我们分别对以上因子进行单因子回归，检验每个因子与股票下期收益的显著性，回归方程为：

$$r_{i,t+1} = \alpha + \beta f_{i,t} + \varepsilon_{i,t}$$

其中， $r_{i,t+1}$ 代表股票 i 的下一期的收益率， β 代表因子收益率， $f_{i,t}$ 为股票 i 在 t 期的因子载荷，单因子回归结果如下表所示。

表 4：单因子回归结果

	相似动量	异质波动率	非流动性	动量	流通市值
回归系数	0.1544***	-0.5588***	0.0161***	-0.0174	0.000153
T 统计量	2.76	-3.54	2.81	-1.28	0.87
可决系数	1.01%	1.51%	7.16%	1.23%	0.39%

数据来源：西南证券

根据单因子的回归结果，动量和流通市值的因子收益率并不显著，而其它因子均在 99% 的置信水平下显著，表明相似动量、异质波动率、非流动性均对股票的下期收益率具有明显的预测作用。

因子之间可能会存在信息重叠，为了检验相似动量因子是否可以提供其它因子中未包含的信息，我们将其它因子作为控制变量，进行多因子回归分析，回归方程设定如下：

$$r_{i,t+1} = \alpha + \sum_k \beta_k f_{i,t}^k + \varepsilon_{i,t}$$



其中， β_k 代表因子 k 的因子收益率， $f_{i,t}^k$ 为股票 i 在 t 期的因子敞口。多因子的 Fama-Macbeth 回归结果如表 5 所示。

表 5：多因子 Fama-Macbeth 回归结果

回归方程	回归结果	相似动量	异质波动率	非流动性	动量	流通市值	可决系数
I	回归系数	0.1822***	-0.7399***	—	—	—	2.00%
	T 统计量	3.48	-4.04	—	—	—	
II	回归系数	0.1411**	—	-0.0170	—	—	2.58%
	T 统计量	2.62	—	-0.99	—	—	
III	回归系数	0.1786***	—	—	-0.0218	—	1.50%
	T 统计量	3.34	—	—	-1.58	—	
IV	回归系数	0.1478***	—	—	—	0.000118	1.43%
	T 统计量	2.75	—	—	—	0.90	
V	回归系数	0.1663***	-0.7965***	-0.0440*	-0.0011	0.000049	4.18%
	T 统计量	3.50	-4.35	-1.87	-0.08	0.41	

数据来源：西南证券

在回归方程 I 至 IV 中，我们分别将异质波动率、非流动性、动量及流通市值作为控制变量，检验在控制了异质波动、流动性、动量及规模信息后，相似股票动量因子是否依旧可以提供股价变化的信息。

回归方程 I 中，相似动量因子与异质波动率一同作为解释变量，因子收益率在 99% 的置信水平下均显著非零，表明相似动量因子同异质波动率的信息重合度较低，两者均对股票的下期收益率具备明显的预测作用。

在回归方程 II 中，在引入非流动性作为控制变量后，相似动量因子的显著性并没有发生改变，而非流动性因子的因子收益率变得不显著，这意味着相似股票动量因子与非流动性因子之间存在信息重叠，但在控制流动性信息后，相似动量因子依旧可以为选股提供新的信息。

在回归方程 III 与 IV 中，在引入动量和流通市值作为控制变量后，相似动量因子的因子收益率在 1% 的显著性水平下保持不变，说明在控制了动量与规模信息后，相似动量因子依旧有效。

回归方程 V 纳入了所有的解释变量，相似动量因子与异质波动率的因子收益率在 99% 的置信水平下显著，证明相似动量因子与异质波动率对股票的下期收益仍具备良好的解释能力；非流动性因子的因子收益率在 90% 的置信水平下显著，表明非流动性因子对股票的下期收益也具备一定的预测作用。

综上，在排除了异质波动、流动性、动量、规模等影响因素后，相似股票动量因子依旧可以提供目前尚未被市场定价的新信息。

3 敏感性分析

在相似股票动量因子的构建过程中，相似股票样本的选取和历史收益的衡量直接决定了相似动量因子的选股能力；不仅如此，相似股票样本的构建与历史收益的度量又具备极强的灵活性。因此，为了确保相似股票动量因子在中国 A 股市场中稳定有效，我们需要对相似股票样本的构建方式与历史收益的度量指标进行敏感性分析。

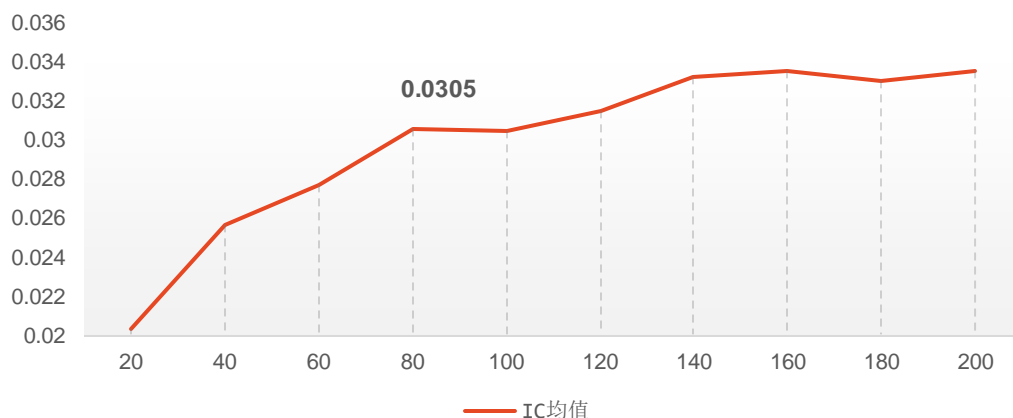
3.1 相似股票样本构建方式的敏感性分析

一共有两种构建相似股票样本的方式，分别是**固定样本容量**选取距离最近的数十只股票与**固定股票距离**选取 A 股市场中股票距离较近的所有股票。

为了保证每只股票都具有相同大小的相似股票样本，我们首先选用固定样本容量这一方式构建相似股票样本。样本容量的不同取值决定了某只股票的相似股票样本的代表性。样本容量越小，相似股票样本的特征与该股票就越接近，样本的代表性也就越强；但同时，缩小样本容量也将导致误差增大，难以归纳总结出一般性的统计规律；反之，样本容量越大，样本的同质性就越强，导致样本难以刻画不同股票的特殊性，从而降低相似股票动量因子的有效性。因此，选定恰当的样本容量构建相似动量因子至关重要。

为了选出恰当的样本容量，我们在图 4 中绘制了相似股票动量因子的信息系数 IC 均值随样本容量变化的图像。

图 4：样本容量与 IC 均值的变化示意图



数据来源：西南证券

如图所示，当样本容量小于 80 时，随着样本容量的扩大，信息系数 IC 均值显著增大，相似股票动量因子的有效性大幅提升；当样本容量大于 80 时，进一步扩大样本容量，信息系数 IC 均值的的增长出现波动，相似动量因子的有效性提升微乎其微。

因此，我们判断，当样本容量为 80 时，即选择某只股票 80 只最相似股票来构建相似动量因子就可以使相似股票动量因子具备相当的有效性及稳定性。

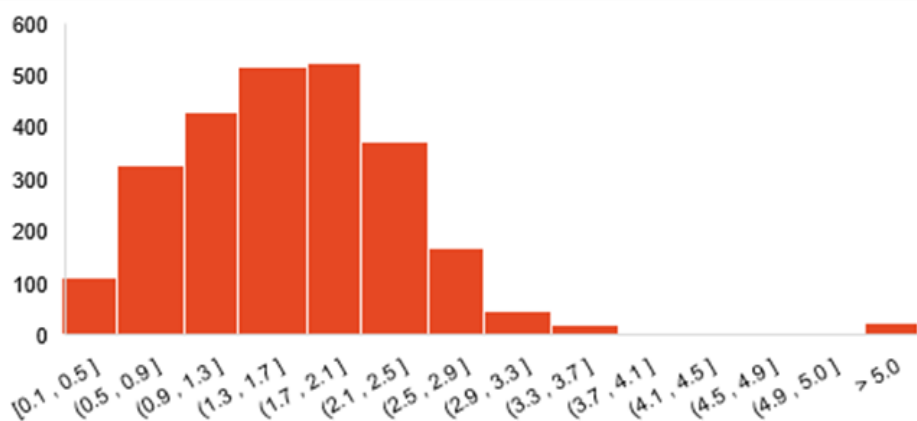
但是，固定样本容量构建相似股票样本也会带来两方面隐患：一方面，难以保证相似股票样本中所有的股票都与该股票高度相似；另一方面，难以保证相似股票样本涵盖 A 股市场中所有的相似股票。



例如, 倘若市场中与某只股票距离较近的股票只有 20 只, 如果我们选用距离最近的 80 只股票构建相似股票动量因子, 那么其余 60 只距离较远的股票就会降低相似股票样本的代表性, 进而降低相似动量因子选股的有效性; 反之, 假设市场中同某只股票高度相似的股票共有 200 只, 若我们只选用距离最近的 80 只股票, 那么其余 120 只高度相似股票就没有被纳入相似股票样本, 造成有效信息的流失。

为了解决以上两方面的弊端, 接下来我们尝试通过固定股票距离构建相似股票样本。为了选取恰当的股票距离构建相似股票样本, 我们需要根据股票距离的分布来确定纳入相似股票样本的距离阈值。为此, 我们绘制了 A 股市场中股票距离的频数分布直方图, 由于数据量较大, 为了直观展示股票距离的分布, 在此我们仅以我司西南证券 (600369.SH) 2021 年 11 月底与 A 股市场中其他股票的距离分布为例。

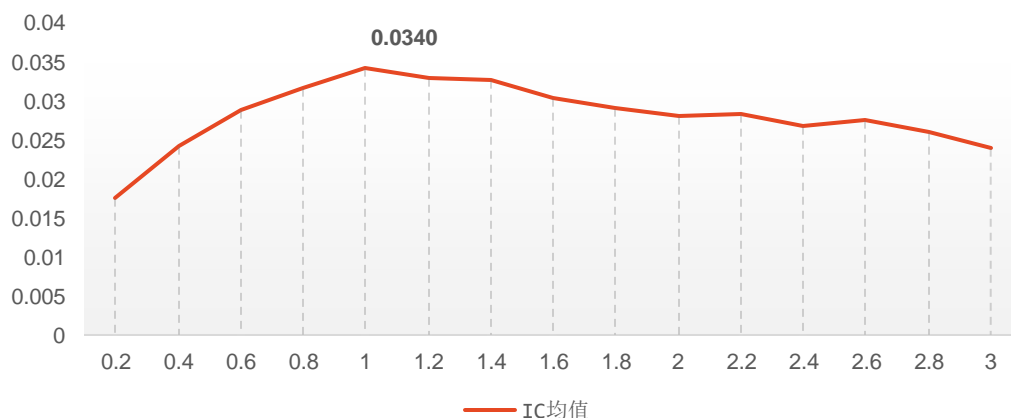
图 5: 西南证券 (600369.SH) 股票距离频数分布直方图



数据来源: 西南证券

如图所示, 我国 A 股市场中股票距离的分布主要集中在区间 $[0, 3]$ 之内, 以此为依据, 为了寻找合适的距离阈值构建相似股票样本, 我们在图 6 中绘制了相似动量因子信息系数 IC 均值与股票距离阈值的变化情况。

图 6: IC 均值与股票距离阈值的变化情况示意图



数据来源: 西南证券



如图所示, 伴随着股票距离阈值的逐步提升, 相似动量因子的 IC 均值先增大后减小, 在股票距离阈值设置为 1 处到达**最大值 0.0340**, 表明选取股票距离小于或等于 1 的股票构建的相似股票样本最具代表性。得益于相似股票样本代表性的提高, 相较于固定样本容量构建的相似股票动量因子, 信息系数 IC 均值有所提升。

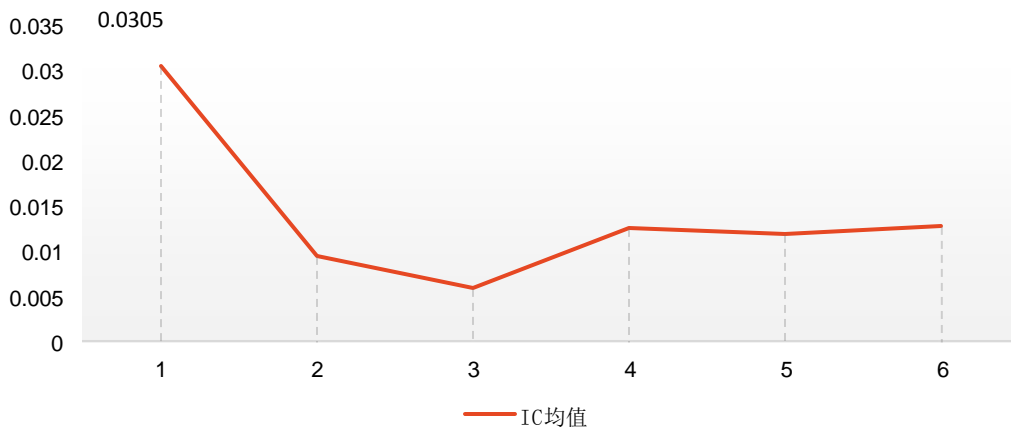
3.2 历史收益指标的敏感性分析

一方面, 相似股票的历史收益率刻画了过去一段时间内**资本市场资金的偏好**; 另一方面, 相似股票的历史收益率也是投资者**选股的重要依据**。

那么过去多长时间的股票历史收益率对未来资本市场资金偏好的预测效果最好? 过去多长时间的相似股票历史收益率对投资者的投资决策影响最大? 为了回答以上两个问题, 首先我们将历史收益率定义为过去 m 个月的股票的市值加权平均收益率。

以此为基础, 为了寻求历史收益率合适的度量, 在图 7 中我们绘制了 IC 均值随 m 不同取值的变化情况。

图 7: 历史收益的度量与 IC 均值变化示意图



数据来源: 西南证券

如图所示, 选用过去一个月的相似股票历史收益构建相似动量因子, 信息系数 IC 均值为 0.0305, 相似动量因子选股较为有效; 选用过去两个月、过去三个月、过去四个月、过去五个月、过去六个月的市值加权平均收益, 信息系数 IC 均值大幅降低, 降至 0.01 左右, 相似动量因子近乎失效。

因此, 在样本期内, 我们判断**选用过去一个月相似股票的历史收益信息构建相似动量因子可以确保因子兼具较强的稳定性与有效性**。



4 因子的改进与优化：引入历史收益优化选股策略

在上文我们构建了相似股票动量因子，并讨论了相似股票动量因子的有效性及其参数的敏感性。在此基础上，我们将尝试改进相似动量因子，以期进一步提高相似动量因子的选股能力。

通过设定股票距离的阈值来构建相似股票动量因子使因子的有效性有所提升。在此基础上，为了更好的发挥相似股票动量因子的选股优势，我们将引入股票的历史收益率，优化选股策略。

从投资者的投资心理出发，投资者在预测某只股票未来的收益率时一般会采用**代表性启发法 (Tversky and Kahneman, 1974)**。因此，投资者对某只股票及其相似股票往往持有相同的收益预期，这也就意味着，如果某只股票的相似股票近期均表现良好且已经实现了较高收益，而这只股票本身还尚未实现较大涨幅，那么投资者往往对这只股票的未来收益充满期待，倾向于买入持有该股票。

基于这一逻辑，在相似股票动量因子的基础上，我们引入股票的历史收益率，构建**相似预期差因子** $\Delta ER_i^{(t)}$ 如下所示，

$$\Delta ER_i^{(t)} = SIM_i^{(t)} - r_i^{(t)}$$

其中， $SIM_i^{(t)}$ 代表股票*i*在*t*时刻的相似股票动量因子值， $r_i^{(t)}$ 表示股票*i*在*t*时刻前一个月的月度收益率。相似预期差因子越大，股票与其相似股票的历史收益的差额就越大，投资者预期这只股票未来的涨幅空间也就越大。

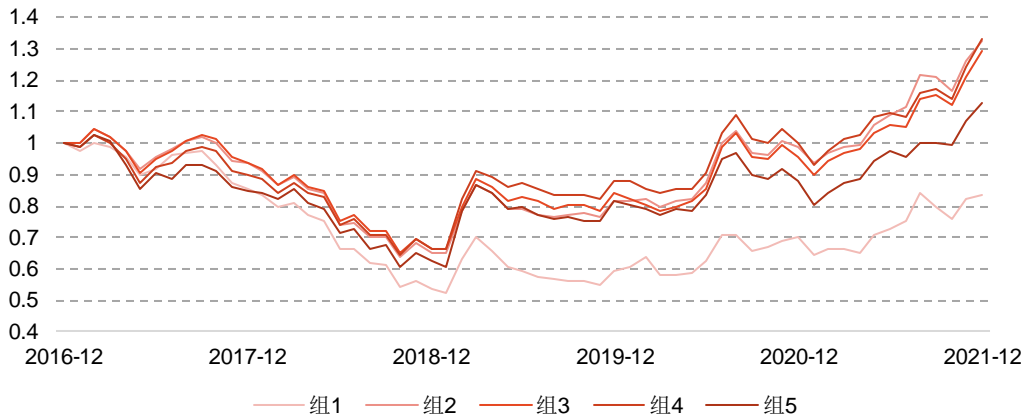
在样本期内，我们对相似预期差因子 $\Delta ER_i^{(t)}$ 进行了回溯测试。根据相似预期差因子值的大小将 A 股市场中的所有股票平均分成五组，以此为基础，分别构建投资组合，在每月的最后一个交易日进行调仓。

回溯测试的结果显示，相似预期差因子的**信息系数 IC 均值**高达 **0.0450**，相较于通过设定距离阈值构建相似股票样本的相似动量因子 $SIM_i^{(t)}$ (0.0340)，相似预期差因子的选股能力更胜一筹。

为了进一步考察相似预期差因子的有效性，根据分组检验的结果，我们在图 8 中绘制各个投资组合在样本期内累计净值的变化情况，为了避免市值因子对累计净值的变化造成干扰，我们采用简单算数平均计算各个投资组合的月度收益率。



图 8：相似预期差因子分组净值图



数据来源：西南证券

如上图所示，在样本期内，根据相似预期差因子值的大小分组构建的投资组合的累计净值并未显示出明显的单调性。尽管平均因子值最小的组 1 的累计净值始终位居低位，但是因子均值较高的组 2、组 3 以及组 4 的累计净值却未呈现出显著差异，并且因子均值最高的组 5 累计净值出现了反转，其净值显著低于组 2、组 3 以及组 4。相较于相似股票动量因子，尽管相似预期差因子的信息系数 IC 大幅提升，但净值分组情况却远不及相似股票动量因子。

为什么累计净值的分组情况会呈现如此异象？回答这一问题需要对因子均值最大的组 5 的因子值进行分拆。相似预期差因子的因子值较大根据股票本身的盈亏可以分为以下两种情况：股票本身历史收益率为正以及股票本身收益率为负。

$$\Delta ER_i^{(t)} = SIM_i^{(t)} - r_i^{(t)} \begin{cases} < SIM_i^{(t)} & r_i^{(t)} > 0 \\ > SIM_i^{(t)} & r_i^{(t)} < 0 \end{cases}$$

当股票自身历史收益率为正时，相似预期差因子值较大意味着相似股票的历史收益较高，而股票本身的涨幅尚小，投资者对该股票的未来收益预期较高，倾向于买入或持有该股；股票本身收益率为负时，相似预期差因子值较大的原因一方面源于相似股票历史收益较高，另一方面由于股票本身收益率较低，负负得正，最终导致相似预期差的因子值较大。在相似股票历史收益表现良好的情况下，这只股票股价却大幅下调，这意味着广大投资者并不看好该股，投资者倾向于卖出或做空这只股票。

由于在组 5 中根据相似预期差因子值大小选择的股票中即包含了投资者预期收益较高的优质股，也包含了投资者一致不看好的劣质股，所以相似预期差因子平均值最大的组 5 累计净值出现了反转。

为了避免股票自身的盈亏对相似预期差因子的测算造成偏误，我们首先根据股票前一个月的市场表现将股票分为强势组与弱势组。**强势组**由前一月**超额收益为正**的个股组成，**弱势组**由前一月**超额收益小于或等于零**的个股组成。超额收益指个股收益与市场收益之差，市场收益以万得全 A 指数月度收益率为基准。



结合近期时事热点，俄乌冲突致使全球股票市场情绪低迷，投资者避险情绪高涨，A股市场近期月度收益为正的个股寥寥无几，倘若我们采用绝对收益对股票进行初步筛选，筛选后的股票样本容量极小，因子选股会产生较大误差。为了提高相似预期差因子的适用性，我们选用超额收益对股票进行初步筛选。

在强势组内，我们根据相似预期差因子值的大小构建投资组合进行分组检验。为了避免股票市值对回溯测试的结果产生干扰，我们采用简单算数平均的方法计算组合收益率。回溯测试的结果如下表所示。

表 6：相似预期差因子分组检验结果

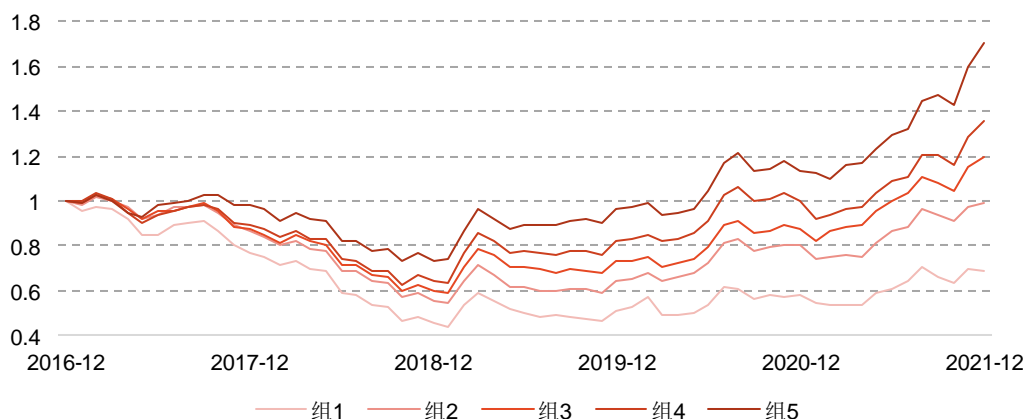
	组 1	组 2	组 3	组 4	组 5	组 5-组 1	(4+5) - (1+2)
因子均值	-0.2408	-0.1016	-0.0583	-0.0307	-0.0006	-	-
平均月收益率	-0.42%	0.14%	0.44%	0.65%	1.02%	1.44%***	1.95%***
t 统计量	-	-	-	-	-	3.39	3.37
年化收益率	-4.93%	1.69%	5.41%	8.08%	12.95%	18.72%	26.08%
Sharp 比率	-0.31	-0.02	0.18	0.32	0.61	1.47	1.52
最大回撤率	54.77%	36.17%	30.31%	27.55%	29.29%	8.60%	11.09%
月胜率	38.33%	48.33%	50.00%	55.00%	60.00%	73.33%	71.67%

注：“***”、“**”、“*” 分别代表参数在 1%、5%、10% 的显著性水平下显著，下同，不再说明。数据来源：西南证券

根据分组检验的结果，伴随着相似预期差因子均值的逐步增大，各组平均月收益率单调递增，Sharp 比率、月胜率也呈现出显著的单调性。多空组合“(4+5)-(1+2)”平均月收益率为 1.95%，在 99% 的置信水平下显著非零，其年化收益率高达 26.08%，最大回撤率为 11.09%。相较于相似股票动量因子，相似预期差因子在平均月收益率、组合年化收益率、Sharp 比率、月胜率方面全面占优。

根据分组检验的分组结果，分别构建投资组合，每月的最后一个交易日进行调仓。在样本期内，投资组合的累计净值变化情况如下图所示。

图 9：相似预期差因子分组净值图



数据来源：西南证券



如图所示，在控制了股票本身历史收益的情况下，在整个样本期内，各组合的历史净值存在显著差异且具备非常显著的单调性。相较于控制股票自身超额收益之前，累计净值的分组情况得到极大改善，说明在控制了股票自身超额收益后，相似预期差因子的选股能力大幅提高。

为了进一步检验相似预期差因子的有效性，下面我们将对相似预期差因子进行信息系数分析，信息系数分析的结果详见表 7。

表 7：相似预期差因子信息系数分析结果

	IC 均值	IC 标准差	t 统计量	IC 与均值同向的比例	IR
相似预期差因子	0.0910***	0.0947	7.4468	75.00%	0.9609

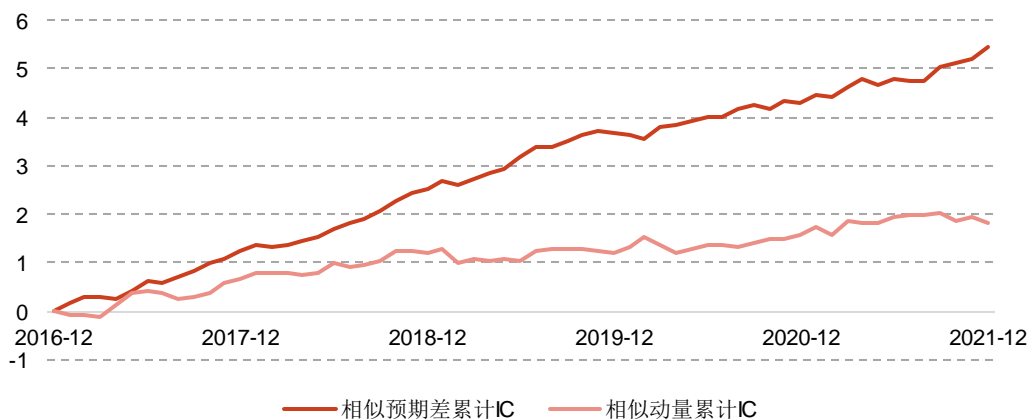
数据来源：西南证券

根据表 7 数据，在样本期内，相似预期差因子的信息系数 IC 均值高达 **0.0910**，取显著性水平 0.01，强烈拒绝 IC 均值为 0 的原假设；IC 与均值同向的比例高达 **75%**，IR 比率（IC 均值除以标准差）为 **0.9609**。信息系数分析的结果表明，相似预期差因子在样本期内具备极强的有效性和稳定性。

相较于相似股票动量因子，在样本期内，相似预期差因子在 IC 均值、IC 标准差、IC 显著性、IC 与均值同向的比例、IR 值等方面全面占优。

为了进一步比较样本期内相似预期差因子与相似股票动量因子信息系数 IC 的稳定性，在图 10 中，我们绘制了样本期内相似预期差因子与相似股票动量因子累计信息系数的变化情况。

图 10：相似预期差因子累计信息系数



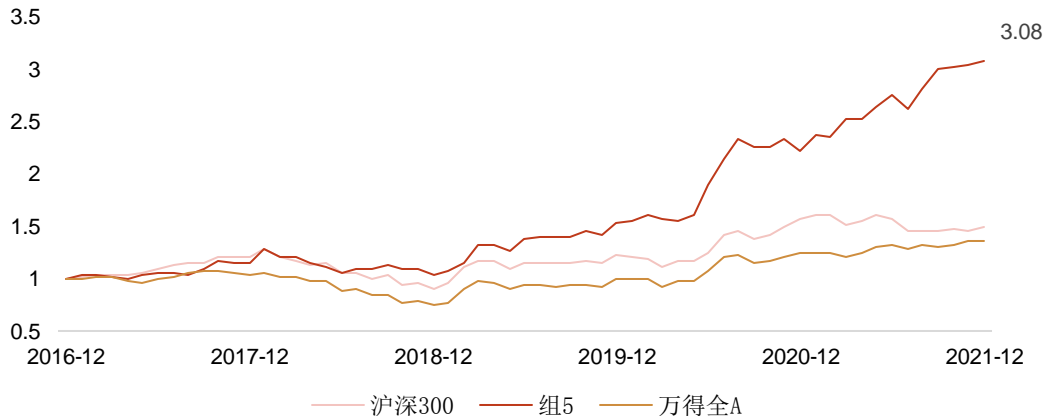
数据来源：西南证券

如图所示，在样本期内，相似预期差因子的累计 IC 一直以一个相对稳定的速度持续增长，说明相似预期差因子在样本期内表现稳定。相较于相似股票动量因子，相似预期差因子的稳定性与累计值均显著占优。

最后，为了模拟相似预期差因子在未来 A 股市场的实际表现，我们绘制了通过相似预期差因子构建的多头组合沪深 300 指数与万得全 A 指数在样本期内的净值变化图。需要特别注意的是，为了保证累计净值的可比性，我们采用市值加权的方式来测算投资组合的月度收益率，计算方法同沪深 300 指数与万得全 A 指数保持一致。



图 11: 相似预期差多头组合与市场指数净值变化图



数据来源: 西南证券

如图所示, 在样本期内, 根据相似预期差因子构建的多头组合表现全面优于万得全 A 指数; 在绝大部分时间里, 多头组合表现显著优于沪深 300 指数。通过相似预期差因子选股构建投资组合可以获得显著的超额收益, 截止样本期末, 多头组合最终累计净值高达 **3.08**, 实现增长率 **208%**。

5 总结与展望

本文在定义股票距离的基础上, 构建了**相似股票动量因子**, 通过分组检验与信息系数分析检验了该因子的有效性, 结果表明, 相似动量因子与股票的下期收益具有非常显著的**正相关关系**。在此基础上, 本文通过 Fama-Macbeth 回归分析, 验证了在控制了异质波动、流动性、动量与规模信息的情况下, **相似动量因子依旧能够提供股价变化的新信息**, 具备较强的选股能力。

我们还对因子的参数进行了敏感性分析。根据敏感性分析的结果, 采用**股票过去一个月的平均收益率衡量股票历史收益**, 选择某只股票距离最近的 **80 只股票**或选择同某只股票**距离小于等于 1 的股票**构建相似股票样本可以保证相似动量因子稳定有效。

在相似股票动量因子有效的基础上, 为了充分发挥相似动量因子的选股能力, 我们同时使用股票的历史收益与相似股票动量因子, 构建了**相似预期差因子**。相较于相似动量因子, 在控制了股票过去一个月的市场表现的情况下, 相似预期差因子的选股能力出众, **信息系数 IC 均值高达 0.0910**。在样本期内, 使用相似预期差因子构建多头组合可以获得远超市场收益的超额收益。

本研究选用股票价格、规模、风险、盈利能力、投资水平五个维度的欧式距离来度量股票与股票间的相似性, 未来可以选用其他指标与不同形式的距离来刻画股票间的相似性, 进而让相似股票样本更具代表性, 从而提高相似股票动量因子与相似预期差因子的有效性。

另一方面, 若想进一步提高相似股票动量因子与相似预期差因子的有效性, 未来还可以尝试引入行业变量, 或在行业层面上使用相似股票动量因子或相似预期差因子构造投资组合及轮动策略。



6 风险提示

本研究基于 2016-2021 年中国 A 股市场历史数据，因子的历史收益率不代表未来收益率。若未来市场环境发生较大变化，因子的实际表现可能与本文的结论有所差异。



参考文献

- AMIHUD, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5, 31-56.
- ANG, A., HODRICK, R. J., XING, Y. & ZHANG, X. 2006. The cross - section of volatility and expected returns. *The journal of finance*, 61, 259-299.
- BELL, D. E. 1982. Regret in decision making under uncertainty. *Operations research*, 30, 961-981.
- FAMA, E. F. & FRENCH, K. R. 2015. A five-factor asset pricing model. *Journal of financial economics*, 116, 1-22.
- FAMA, E. F. & MACBETH, J. D. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81, 607-636.
- HE, W., WANG, Y. & YU, J. 2021. Similar Stocks. *Available at SSRN 3815595*.
- TVERSKY, A. & KAHNEMAN, D. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185, 1124-1131.



分析师承诺

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，报告所采用的数据均来自合法合规渠道，分析逻辑基于分析师的职业理解，通过合理判断得出结论，独立、客观地出具本报告。分析师承诺不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接获取任何形式的补偿。

投资评级说明

公司评级

买入：未来 6 个月内，个股相对沪深 300 指数涨幅在 20% 以上
持有：未来 6 个月内，个股相对沪深 300 指数涨幅介于 10% 与 20% 之间
中性：未来 6 个月内，个股相对沪深 300 指数涨幅介于-10% 与 10% 之间
回避：未来 6 个月内，个股相对沪深 300 指数涨幅介于-20% 与 -10% 之间
卖出：未来 6 个月内，个股相对沪深 300 指数涨幅在-20% 以下

行业评级

强于大市：未来 6 个月内，行业整体回报高于沪深 300 指数 5% 以上
跟随大市：未来 6 个月内，行业整体回报介于沪深 300 指数-5% 与 5% 之间
弱于大市：未来 6 个月内，行业整体回报低于沪深 300 指数-5% 以下

重要声明

西南证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会核准的证券投资咨询业务资格。

本公司与作者在自身所知知情范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

《证券期货投资者适当性管理办法》于 2017 年 7 月 1 日起正式实施，本报告仅供本公司客户中的专业投资者使用，若您并非本公司客户中的专业投资者，为控制投资风险，请取消接收、订阅或使用本报告中的任何信息。本公司也不会因接收人收到、阅读或关注自媒体推送本报告中的内容而视其为客户。本公司或关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行或财务顾问服务。

本报告中的信息均来源于公开资料，本公司对这些信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告，本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，本公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

本报告及附录版权为西南证券所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用须注明出处为“西南证券”，且不得对本报告及附录进行有悖原意的引用、删节和修改。未经授权刊载或者转发本报告及附录的，本公司将保留向其追究法律责任的权利。



西南证券研究发展中心

上海

地址：上海市浦东新区陆家嘴东路 166 号中国保险大厦 20 楼

邮编：200120

北京

地址：北京市西城区金融大街 35 号国际企业大厦 A 座 8 楼

邮编：100033

深圳

地址：深圳市福田区深南大道 6023 号创建大厦 4 楼

邮编：518040

重庆

地址：重庆市江北区金沙门路 32 号西南证券总部大楼

邮编：400025

西南证券机构销售团队

区域	姓名	职务	座机	手机	邮箱
上海	蒋诗烽	总经理助理 销售总监	021-68415309	18621310081	jsf@swsc.com.cn
	崔露文	高级销售经理	15642960315	15642960315	clw@swsc.com.cn
	黄滢	高级销售经理	18818215593	18818215593	hying@swsc.com.cn
	王昕宇	高级销售经理	17751018376	17751018376	wangxy@swsc.com.cn
	陈燕	高级销售经理	18616232050	18616232050	chenyanyf@swsc.com.cn
	陈慧琳	销售经理	18523487775	18523487775	chhl@swsc.com.cn
	薛世宇	销售经理	18502146429	18502146429	xsy@swsc.com.cn
北京	李杨	销售总监	18601139362	18601139362	yfly@swsc.com.cn
	张岚	销售副总监	18601241803	18601241803	zhanglan@swsc.com.cn
	王兴	销售经理	13167383522	13167383522	wxing@swsc.com.cn
	来趣儿	销售经理	15609289380	15609289380	lqe@swsc.com.cn
	王一菲	销售经理	18040060359	18040060359	wyf@swsc.com.cn
	王宇飞	销售经理	18500981866	18500981866	wangyuf@swsc.com
广深	郑龔	广州销售负责人 销售经理	18825189744	18825189744	zhengyan@swsc.com.cn
	陈慧玲	销售经理	18500709330	18500709330	chl@swsc.com.cn
	杨新意	销售经理	17628609919	17628609919	yxy@swsc.com.cn
	张文锋	销售经理	13642639789	13642639789	zwf@swsc.com.cn
	龚之涵	销售经理	15808001926	15808001926	gongzh@swsc.com.cn