

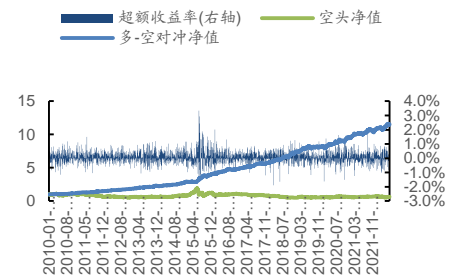
# 基于地理关联度因子研究

## 多因子 Alpha 系列报告之(四十三)

### 报告摘要:

- **因子开发迭代更新越来越重要。**近几年来,随着传统多因子模型在市场的应用逐渐广泛,因子的波动特征逐渐加大,因子拥挤等原因造成了因子的收益逐渐下降。为了能够寻找更好的 Alpha 收益来源,在多因子模型框架中,因子作为底层 Alpha 来源输入的基础,因子的开发、迭代、更新就显得越来越重要。低频相关的数据的因子开发目前难度越来越大,增量的信息越来越有限。本篇专题探讨个股基于地理关联数据在因子选股中的应用。
- **领先滞后效应与地理关联度概念。**传统的有效市场假说认为,在完全有效的金融市场上,价格能够及时、充分反映资产的所有公开信息以及私有信息。但实证研究表明,股票市场中存在着“领先滞后效应”。Parsons 和 Sabbatucci(2018)在发表论文《Geographic Lead-Lag Effects》中提出地理关联股票之间存在这一效应。这两位学者认为,总部位于相同地理区域的个股会受到同一基本面因素影响,但不同公司的股价对于新信息的反应速度存在差异,从而地理关联股票的价格变动对目标股票收益具有显著预测作用。本报告基于地理关联度研究思路,构造了地理相关系数因子及其优化因子,并研究该类因子在 A 股中的有效性。
- **地理关联度因子实证分析。**本篇专题报告共构建了六种地理相关系数类因子并在全市场范围进行了月频调仓的实证分析。实证分析结果表明, GEOG CORR、GEOG CORR P 与 GEOG CORR IP 共 3 种因子的分档效果明显。其中, GEOG CORR P 因子整体表现较好。因子整体的 IC 均值为 0.069,正 IC 占比 90%,多头相对中证 500 指数年化超额收益率为 13.98%,信息比率为 1.706。
- **相关性分析与稳健性检验。**本篇专题报告针对回测表现较好的 3 种地理相关系数类因子,进行与 BARRA 因子相关性分析与稳健性检验。实证结果表明,地理相关系数类因子能够挖掘传统因子外的增量信息,即股票之间的地理关联信息。GEOG CORR 因子、GEOG CORR P 因子在中证 1000、创业板股票池内仍具有较好表现。由于因子策略多头换手率较高,地理相关系数类因子对手续费敏感。
- **风险提示。**本专题报告所述模型用量化方法通过历史数据统计、建模和测算完成,所得结论与规律在市场政策、环境变化时可能存在失效风险;策略在市场结构及交易行为的改变时有可能存在策略失效风险。

图: GEOG CORR P 因子全市场表现



数据来源: Wind, 广发证券发展研究中心

分析师: 陈原文



SAC 执证号: S0260517080003



0755-82797057



chenyuanwen@gf.com.cn

分析师: 罗军



SAC 执证号: S0260511010004



020-66335128



luojun@gf.com.cn

分析师: 安宁宁



SAC 执证号: S0260512020003



SFC CE No. BNW179



0755-23948352



anningning@gf.com.cn

请注意, 陈原文, 罗军并非香港证券及期货事务监察委员会的注册持牌人, 不可在香港从事受监管活动。

### 相关研究:

行业聚类方法探讨:——重构 2021-11-17

行业轮动框架之十

海量技术指标掘金 Alpha 因 2021-07-30

子:多因子 Alpha 系列报告之

(四十二)

## 目录索引

一、因子挖掘思考.....	5
(一) 高频信息.....	5
(二) 低频信息.....	6
二、关联度因子研究进展.....	6
(一) 关联度因子的理论研究进展.....	6
(二) 地理关联度与 A 股市场的关系.....	7
(三) 地理关联度初步实证结果与改进.....	8
三、地理关联度因子构造方法与策略框架.....	10
(一) 因子构造方法.....	10
(二) 因子特征分析.....	11
(三) 因子策略构建.....	13
四、实证分析.....	15
(一) 数据说明.....	15
(二) 因子分档表现.....	15
(三) 因子实证结果.....	16
五、地理关联度因子与 BARRA 因子相关性分析.....	25
(一) BARRA 因子说明.....	25
(二) 地理关联度因子与 BARRA 相关性分析.....	26
六、地理关联度因子敏感性分析.....	26
(一) 敏感性分析-选股范围.....	26
(二) 敏感性分析-手续费费率.....	27
七、总结.....	29
八、风险提示.....	29
九、参考文献.....	30

## 图表索引

图 1: TOP5 省市上市公司数量与占比分年度情况	8
图 2: TOP5 省市上市公司数量增长率分年度情况	8
图 3: TOP5 省份上市公司月均收益分年度情况	8
图 4: 地理关联度因子-IC 值与 IC 累计值走势	9
图 5: 地理相关系数因子的省份分布特征	12
图 6: 地理相关系数因子的申万一级行业分布特征	12
图 7: 地理相关系数因子的市值分布特征	13
图 8: 某个股月度地理相关系数因子值与股票收益走势图	13
图 9: 全市场月度地理相关系数因子均值与中证全指月度收益走势图	14
图 10: 策略构建框架	15
图 11: 全市场 GEOGCORR 因子十档-月度调仓	15
图 12: 全市场 GEOGCORRCHG 因子十档-月度调仓	15
图 13: 全市场 GEOGCORRP 因子十档-月度调仓	16
图 14: 全市场 GEOGCORRN 因子十档-月度调仓	16
图 15: 全市场 GEOGCORRIP 因子十档-月度调仓	16
图 16: 全市场 GEOGCORRJP 因子十档-月度调仓	16
图 17: GEOGCORR 因子-IC 值与 IC 累计值走势	18
图 18: GEOGCORR 因子-多空策略净值走势	19
图 19: GEOGCORR 因子-多-中证 500 策略净值走势	19
图 20: GEOGCORRP 因子-IC 值与 IC 累计值走势	20
图 21: GEOGCORRP 因子-多空策略净值走势	21
图 22: GEOGCORRP 因子-多-中证 500 策略净值走势	21
图 23: GEOGCORRIP 因子-IC 值与 IC 累计值走势	23
图 24: GEOGCORRIP 因子-多空策略净值走势	24
图 25: GEOGCORRIP 因子-多-中证 500 策略净值走势	24
图 26: 全市场 GEOGCORR 因子	28
图 27: 全市场 GEOGCORRP 因子	28
图 28: 全市场 GEOGCORRIP 因子	28
表 1: 广发金工高频数据因子挖掘系列报告一览	6
表 2: 地理关联度因子-多空对冲策略与多头相对中证 500 指数策略-整体与分年度表现-月度调仓	9
表 3: 某个股月度地理相关系数因子值与股票收益序列相关性	14
表 4: 全市场股票地理相关系数因子均值与中证全指收益序列相关性	14
表 5: 地理相关系数类因子整体绩效表现	17
表 6: GEOGCORR 因子-整体与分年度 IC 表现	17
表 7: GEOGCORR 因子-多空对冲策略与多头相对中证 500 指数策略-整体与分年度表现-月度调仓	18

表 8: GEOG CORR 因子-分年度换手率 .....	19
表 9: GEOG CORR 因子-整体与分年度 IC 表现 .....	20
表 10: GEOG CORR 因子-多空对冲策略与多头相对中证 500 指数策略-整体与分年度表现 .....	21
表 11: GEOG CORR 因子-分年度换手率 .....	22
表 12: GEOG CORR 因子-整体与分年度 IC 表现 .....	22
表 13: GEOG CORR 因子-多空对冲策略与多头相对中证 500 指数策略-整体与分年度表现 .....	23
表 14: GEOG CORR 因子-分年度换手率 .....	24
表 15: BARRA 因子说明 .....	25
表 16: 地理相关系数类因子与 BARRA 因子相关系数 .....	26
表 17: 地理相关系数类因子在不同选股范围的整体绩效表现 .....	27

## 一、因子挖掘思考

### （一）高频信息

近年来，A股市场机构化趋势明显，量化私募机构的管理规模也迅速扩大，产生了一批管理规模超过百亿的量化私募机构。与此同时，传统的风格因子波动增大，从市场获取超额收益的难度在增加。

因子拥挤是因子收益下降的原因之一。因子代表着市场某方面的非有效性、或者是一段时期内的定价失效。当某类因子收益高的时候，会吸引更多的资金进入，从而出现因子拥挤，降低因子的预期收益。一旦新的因子被公开，套利资金的介入会使得错误定价收窄，因子收益也会跟着下降。因此，在多因子选股模型中，因子的开发和更新迭代变得越来越重要。

与低频因子相比，高频数据在用于量化投资中存在一定优势。

首先，高频价量数据的体量明显大于低频数据。以分钟行情为例，用压缩效果较好的mat格式存储2020年全市场股票的分钟行情数据（包括分钟频的开高低收价格数据、买卖盘挂单数据等），约为12GB。如果是快照行情（目前上交所和深交所都是3秒一笔）或者level 2行情，数据量要大很多。因此，高频数据因子挖掘对信息处理能力和处理效率的要求较高。而且，日内数据，尤其是level 2数据，一般要额外付费，甚至需要自行下载存储实时行情，在此基础上构建的因子拥挤度较低。

其次，高频价量数据一般是多维的时间序列数据，数据中噪声比例较高，而且与ROE、PE这类低频指标本身就具有选股能力不同的是，原始的高频行情数据一般不能直接用作选股因子，而要通过信号变换、时间序列分析、机器学习等方法从高频数据中构建特征，才能作为选股因子。此类因子与低频信号的相关性较低，而且由于因子开发流程相对复杂，不同投资者构建的因子更具有多样性。

此外，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。独立样本的增多有助于检验高频因子的有效性。

高频数据挖掘因子的难点在于数据维度大、噪声高。凭借专业投资者的经验或者是参阅已发表的文献，可以从高频数据中提炼出一部分有选股能力的特征。此外，机器学习方法擅长从数据中寻找规律和特征，是高频数据因子挖掘的有力工具。



表 1：广发金工高频数据因子挖掘系列报告一览

高频价量数据的因子化方法-多因子Alpha系列报告之（四十一）

深度学习框架下高频数据因子挖掘-深度学习研究报告之七

基于个股羊群效应的选股因子研究-高频数据因子研究系列三

基于日内高频数据的短周期选股因子研究-高频数据因子研究系列二

基于日内高频数据的短周期选股因子研究-高频数据因子研究系列一

信息不对称理论下的高频因子挖掘

再谈信息不对称理论下的因子挖掘

日内价量数据因子化研究

数据来源：Wind，广发证券发展研究中心

## （二）低频信息

以传统日频价量和更低频财务数据为基础的因子开发是一种研究途径。由于基础因子广为人知，在此基础上进行因子挖掘的收益提升空间相对有限。而且日频数据由于本身的数据量和信息量有限，过度挖掘会增大过拟合的风险。

对于低频信息的挖掘，从最近几年的进展上看，低频里的增量信息成果越来越少。从数据维度上看，低频的因子建模更多是从一些另类数据或者是新的方法、理论成果中出发构建相关的因子。如另类数据角度，从互联网中的股吧、新闻、关注度等角度，或者是专利数据、供应链相关数据等。新的理论成果如从图网络等角度出发构建相关的因子。

本篇专题报告基于个股的“关联度”角度出发，研究个股所在区域关联度角度构建因子。

## 二、关联度因子研究进展

### （一）关联度因子的理论研究进展

传统的有效市场假说认为，在完全有效的金融市场上，价格能够及时、充分反映资产的所有公开信息以及私有信息。但是，Kalok等(2005)<sup>[3]</sup>、刘菁哲(2010)<sup>[12]</sup>等众多学者通过实证研究发现，股票市场中存在着“领先滞后效应”，即不同公司对相同基本面信息的反应速度存在差异，一些公司能够迅速对新信息做出反应，另一些公司对于新信息的反应存在时滞。

本报告对国内外学者基于行业关联、科技关联、供应链关联、地理关联信息的“领先滞后效应”研究成果进行了简单梳理。对于行业关联信息，Cohen和Lou(2012)<sup>[5]</sup>实证检验了，面对影响全行业的信息事件，单一经营部门公司的股价能够更迅速的反映新信息，同时对于多经营部门公司未来股票收益存在显著预测能力。胡聪慧等(2015)<sup>[10]</sup>采用A股上市公司数据验证了这一结论，并证实了集团公司股价变动的滞后性主要在于投资者关注度与处理能力有限性，以及行业估值的复杂性。向诚等(2018)<sup>[13]</sup>实证说明了行业内受关注度最高的30%公司组合的收益率，显著引领受关注度最低30%公司组合的未来收益率。段丙蕾等(2022)<sup>[9]</sup>认为行业关联回报率仅

在月度层面显著，在周度层面不显著。同时，Parsons和Sabbatucci(2018)<sup>[1]</sup>对于行业关联公司的收益预测能力的有效性提出质疑。他们认为，随着证券分析师覆盖率不断提升，股票价格的有效性增强；随着个股证券分析师重复率上升，股票价格反映的行业一致预期信息越多，因此基于行业关联构建的股票投资策略效果可能衰减。

对于科技关联信息，Lee等(2019)<sup>[6]</sup>构建科技关联指标并进行实证分析，研究结论表明科技关联企业的收益对研究企业的收益具有很强的预测能力。国内学者借鉴Lee等(2019)<sup>[6]</sup>的科技关联指标构建方法，研究该指标在我国股票市场的适用性。李绪泉等(2020)<sup>[11]</sup>的实证分析结果说明，A股市场存在科技溢出效应。段丙蕾等(2022)<sup>[9]</sup>进一步证明了科技关联因子仅在周度上具有显著收益预测能力，认为造成这一结果的原因在于A股市场中存在较多博彩倾向的散户投资者，该类投资者追涨杀跌的交易行为缩短了科技关联信息融入股价所需的时间。

对于供应链关联信息，Cohen和Frazzini(2008)<sup>[4]</sup>、Menzly和Ozbas(2010)<sup>[7]</sup>验证了公司客户信息能够有效预测公司未来股票收益。国内学者对于供应链关联相关研究相对较少，现有的研究成果也未提供在控制变量基础上，供应链关联能够有效预测股票收益的证据(段丙蕾等, 2022)<sup>[9]</sup>。

对于地理关联信息，Peng和Lin在其发表论文《Investor Attention, Overconfidence, and Category Learning》(Journal of Financial Economics, 2006)<sup>[8]</sup>中提出，总部位于同一地理区位的公司，会受到相同基本面因素的影响，从而这些公司股价都会对新信息作出反应。基于这一研究思路，Parsons和Sabbatucci在其发表论文《Geographic Lead-Lag Effects》(The Review of Financial Studies, 2018)<sup>[1]</sup>中提出地理关联公司的概念，具体指与研究个股处于相同地理区位不同行业的所有上市公司。认为地理关联公司股票与目标股票的价格变动存在领先滞后关系（本文将此关系简称为地理关联度），前者对后者未来收益具有预测能力。并且采用面板数据回归方法，实证检验了这一结论。研究结果表明：（1）在控制行业影响基础上，地理关联公司的基本面因素（EPS、销售收入、雇员数量等）变动对目标股票的基本面变动具有显著的解释能力。（2）地理关联公司股票的平均收益对目标股票未来收益具有显著的预测能力，地理关联公司股票的平均收益越高，目标股票未来收益越高。（3）由于证券分析师通常是基于行业而非省份分类的，因此，共同分析师覆盖率提升并不会导致地理关联度的领先滞后关系减弱甚至消失。

综上所述，一方面，与个股自身的动量效应或反转效应相比，地理关联度能够从股票之间的地理关联中挖掘增量信息。另一方面，与其他股票经济关联特征相比，地理关联信息对股票收益的预测能力虽然已在海外市场得到验证，但尚未应用于A股市场。这为本文通过新的经济关联特征挖掘因子提供了新思路。

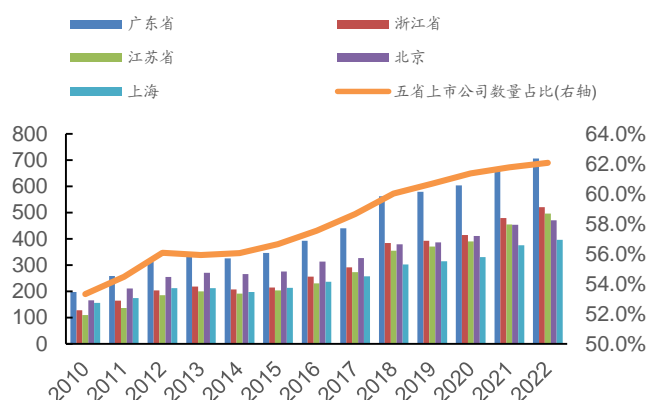
## （二）地理关联度与 A 股市场的关系

为探索地理关联度是否适合在A股市场进行实证研究，本报告首先对31个省及自治区行政区内的上市公司数量、数量增长率以及股票月均收益情况进行统计。受篇幅限制，仅对上市公司数量排名前5的省及直辖市（广东省、浙江省、江苏省、北京、上海）统计结果进行展示。

图1和图2表明，2010年至今，TOP5省份上市公司数量占全国上市公司数量的比重不断上升，从2010年占比53.31%升至2022年6月30日的62.06%。此外，不同省份的上市公司数量与数量增长率表现不一。图3表明，从不同省份的月度平均收益

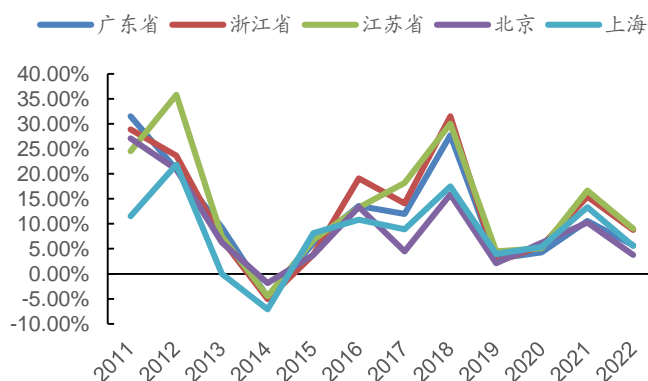
来看，虽然收益整体走势与市场行情保持一致，但是省份之间的月均收益存在显著差异。不同省份的上市公司整体表现存在地理区域上的不同。

图 1：TOP5省市上市公司数量与占比分年度情况



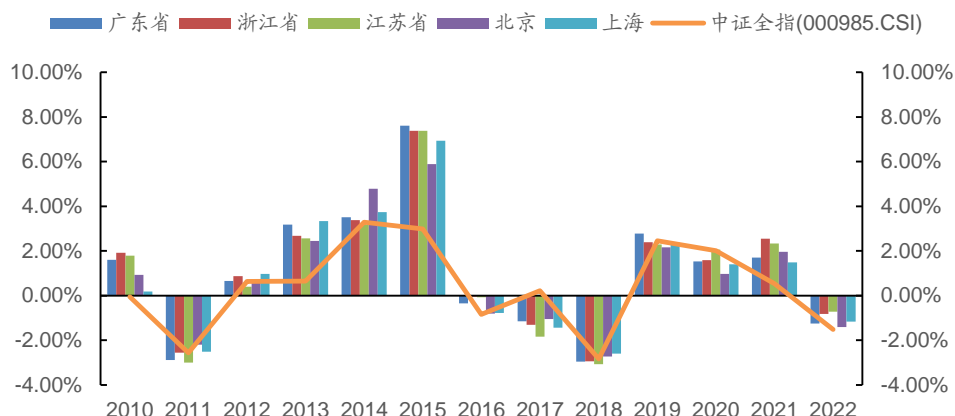
数据来源：Wind，广发证券发展研究中心

图 2：TOP5省市上市公司数量增长率分年度情况



数据来源：Wind，广发证券发展研究中心

图 3：TOP5省份上市公司月均收益分年度情况



数据来源：Wind，广发证券发展研究中心

### （三）地理关联度初步实证结果与改进

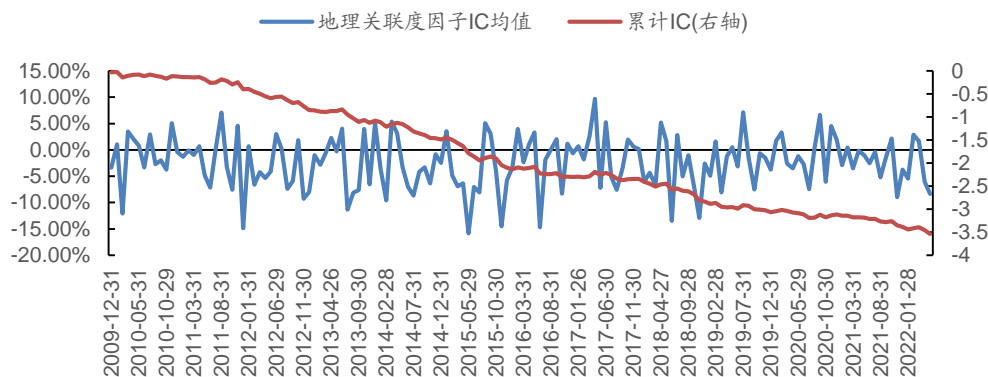
本报告借鉴Parsons和Sabbatucci在《Geographic Lead-Lag Effects》(The Review of Financial Studies, 2018)<sup>[1]</sup>中的指标构建方法，对于某支股票而言，将与之属于相同省份不同行业的所有股票的月度收益均值作为这支股票的月度地理关联度因子。参考上述论文中地理关联度分析思路：当某支股票的地理关联度因子越大时，反映出这支股票的地理关联公司股票收益平均水平越高。如果这支股票对新信息的反应速度相对滞后，那么这支股票很可能在未来一段时间内会由于基本面改善信息而表现为股价上涨。因此，地理关联度因子的投资策略为：在每个换仓日中，根据地理关联度因子值大小，在调仓日买进地理关联度最大的股票，同时卖出地理关联度最小的股票。

本报告在全市场选股范围内，对20100101-20220630区间的地理关联度因子进行月度调仓的回溯测试。实证结果表明，地理关联度因子在A股市场的选股逻辑与上



述论文中阐述的逻辑相反：因子值越小，股票未来收益越高。因子的整体IC均值为-0.024，正IC占比34.67%。地理关联度因子在多头相对中证500指数策略的表现较差，策略整体年化收益率7.52%，信息比率0.832，最大回撤17.65%。

图 4：地理关联度因子-IC值与IC累计值走势



数据来源：Wind，广发证券发展研究中心

表 2：地理关联度因子-多空对冲策略与多头相对中证 500 指数策略-整体与分年度表现-月度调仓

年份	纯多头		多空对冲策略				多头-中证 500 策略				
	累计 收益率	累计 收益率	年化 收益率	年化 波动率	信息 比率	最大 回撤	累计 收益率	年化 收益率	年化 波动率	信息 比率	最大 回撤
2010	16.76%	0.53%	0.53%	8.01%	0.066	4.37%	6.14%	6.14%	5.70%	1.077	2.57%
2011	-30.90%	0.21%	0.21%	5.49%	0.038	2.69%	4.34%	4.34%	2.70%	1.611	0.88%
2012	4.36%	5.27%	5.27%	4.22%	1.247	2.48%	4.08%	4.08%	3.26%	1.251	0.89%
2013	23.16%	9.64%	9.64%	7.33%	1.315	2.70%	6.27%	6.27%	6.79%	0.924	2.72%
2014	54.05%	13.55%	13.55%	5.90%	2.297	1.19%	11.34%	11.34%	5.90%	1.923	3.51%
2015	92.40%	22.98%	22.98%	14.64%	1.569	3.36%	36.98%	36.98%	15.46%	2.392	6.11%
2016	-5.05%	13.28%	13.28%	7.01%	1.895	1.34%	15.93%	15.93%	4.20%	3.790	0.72%
2017	-16.88%	-3.54%	-3.54%	2.96%	-1.198	3.43%	-16.46%	-16.46%	4.40%	-3.737	15.23%
2018	-24.66%	15.05%	15.05%	6.71%	2.243	1.85%	13.10%	13.10%	7.48%	1.753	1.15%
2019	22.52%	-2.23%	-2.23%	5.60%	-0.399	7.86%	-2.89%	-2.89%	5.50%	-0.525	4.91%
2020	18.88%	1.70%	1.70%	7.62%	0.223	4.35%	-2.39%	-2.39%	13.30%	-0.179	5.76%
2021	37.33%	14.14%	14.14%	8.21%	1.722	4.23%	19.01%	19.01%	14.51%	1.310	3.69%
2022	-6.52%	1.83%	3.69%	11.36%	0.325	7.01%	7.27%	15.07%	12.23%	1.232	5.29%
ALL	241.74%	135.88%	7.11%	7.67%	0.926	7.86%	147.65%	7.52%	9.04%	0.832	17.65%

数据来源：Wind，广发证券发展研究中心

由于地理关联度因子在A股市场表现欠佳，本报告对该因子构建方法进行分析并提出因子改进思路。在地理关联度因子构建方法中，仅利用地理关联公司与研究个股归属于同一省级行政区这一特征，但未能反映地理关联公司与研究个股在股价变动上的关联程度。本报告综合考虑上述两点，在考虑股票地理特征的同时，进一步从时序角度考虑地理关联公司与研究个股在过去一段时间内价格变动的相关程度，构建地理相关系数因子(GEOGCORR)及其优化因子，并探讨这类因子在A股的有效性。

### 三、地理关联度因子构造方法与策略框架

#### (一) 因子构造方法

根据上述地理关联度研究得出的初步结论，本篇专题报告构建6种地理相关系数类因子，具体的因子定义、构造逻辑与计算方法如下。

##### 1. 地理相关系数因子

本篇报告定义地理相关系数因子(GEOGCORR)，用以度量个股与其地理关联公司股票之间的整体相关程度，具体由个股和地理关联公司股票相关系数均值表示。

以股票*i*在*t*月月末的地理相关系数因子为例，具体计算方式如下。首先，在全市场范围剔除*t*月的ST股、\*ST股、停牌股以及上市不满一年的股票；其次，筛选出与股票*i*办公地所属省份相同、申万一级行业不同的全部共*N*支股票*j*，并分别计算与股票*i*在*t*月日频收益序列的皮尔森相关系数，即 $CORR_{i,j,t}$ 。最后，对所有相关系数进行加权求和(若不做特殊说明， $w_{j,t}$ 均设置为 $1/N$ ，即等权)，得到股票*i*在*t*月月末换仓日的地理相关系数因子 $GEOGCORR_{i,t}$ 。

$$CORR_{i,j,t} = \frac{cov(R_i, R_j)}{std(R_i) * std(R_j)}$$

$$GEOGCORR_{i,t} = \sum_{j=1}^N w_{j,t} * CORR_{i,j,t}$$

$$geog_i = geog_j$$

$$industry_i \neq industry_j$$

其中， $geog_i(geog_j)$ 分别为股票*i(j)*的办公地归属省份。 $industry_i(industry_j)$ 为股票*i(j)*的申万一级行业分类。

##### 2. 地理相关系数变动因子

为研究个股与地理关联公司股票的相关程度变动，是否影响地理关联特征对个股的收益预测能力，本报告定义地理相关系数变动因子(GEOGCORRCHG)，用以度量个股与关联公司股票整体相关程度的变动情况。

以股票*i*在*t*期的因子为例，具体计算方式如下。首先，按照地理相关系数的计算方式，分别计算股票*i*与股票*j*的*t*月日频收益序列的相关系数 $CORR_{i,j,t}$ 、两者在*t*-2至*t*月共3个月的日频收益序列的相关系数 $CORR_{i,j,t-2,t}$ ；其次，对任意股票*j*，计算1月相关系数与3个月相关系数的差值，衡量股票*i*与股票*j*之间相关性变动程度；最后，对*N*个股票*j*的相关系数差值进行加权求和，得到股票*i*在*t*月月末的地理相关系数变动因子 $GEOGCORRCHG_{i,t}$ 。

$$GEOGCORRCHG_{i,t} = \sum_{j=1}^N w_{j,t} * (CORR_{i,j,t} - CORR_{i,j,t-2,t})$$

### 3. 地理相关系数拆解因子

Bollerslev等(2022)<sup>[2]</sup>在发表论文《Realized semibetas: Disentangling “good” and “bad” downside risks》(Journal of Financial Economics)中,根据市场收益与资产收益序列的符号将传统市场贝塔拆分为四个半贝塔,并实证说明了基于负市场收益与负资产收益序列协方差构建的半贝塔与资产未来收益显著正相关,基于负市场收益与正资产收益序列协方差构建的半贝塔与资产未来收益显著负相关。这一结论对本报告的启示在于:基于不同数值方向收益序列构建的相关系数,可能蕴含的信息量也存在差异。因此,本报告将股票*i*与股票*j*的收益序列进行拆分,并定义四种具体的地理相关系数拆解因子(*GEOGCORRP*、*GEOGCORRN*、*GEOGCORRIP*与*GEOGCORRJP*),用以度量个股与地理关联公司股票的调整后收益序列的相关程度。

以股票*i*在*t*月的地理相关系数拆解因子(*GEOGCORRP<sub>i,t</sub>*)为例,具体计算方式如下。首先,对于股票*i*与全部*N*个股票*j*,利用*R<sub>i</sub><sup>+</sup>*公式对其日度收益序列进行调整,也就是将负日度收益调整为0。其次,根据地理相关系数因子构造步骤,得到地理相关系数拆解因子*GEOGCORRP<sub>i,t</sub>*。其余三种地理相关系数拆解因子(*GEOGCORRN<sub>i,t</sub>*、*GEOGCORRIP<sub>i,t</sub>*与*GEOGCORRJP<sub>i,t</sub>*)构造方式同理可得。

$$R_i^+ = \max(R_i, 0) \quad R_i^- = \min(R_i, 0)$$

$$GEOGCORRP_{i,t} = \sum_{j=1}^N w_j * CORR(R_i^+, R_j^+)$$

$$GEOGCORRN_{i,t} = \sum_{j=1}^N w_j * CORR(R_i^-, R_j^-)$$

$$GEOGCORRIP_{i,t} = \sum_{j=1}^N w_j * CORR(R_i^+, R_j^-)$$

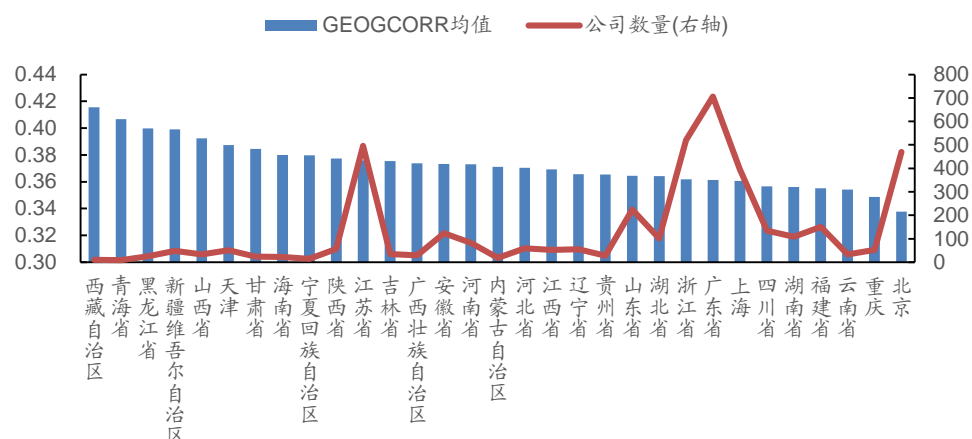
$$GEOGCORRJP_{i,t} = \sum_{j=1}^N w_j * CORR(R_i^-, R_j^+)$$

#### (二) 因子特征分析

在构建具体的因子投资策略前,本报告先对地理相关系数因子特征进行分析,以了解因子更偏向哪些省份、行业与市值特征股票,并据此判断是否需要因子层面的行业与市值中性化处理。

图5表明,地理相关系数因子在省份层面存在显著差异。具体地,*GEOGCORR*因子均值较高的前十个省份的公司数量(截至2022年6月30日)均不超过60个,江苏省上市公司数量占全市场比重超过11.89%,但仍表现出较高的*GEOGCORR*因子均值。

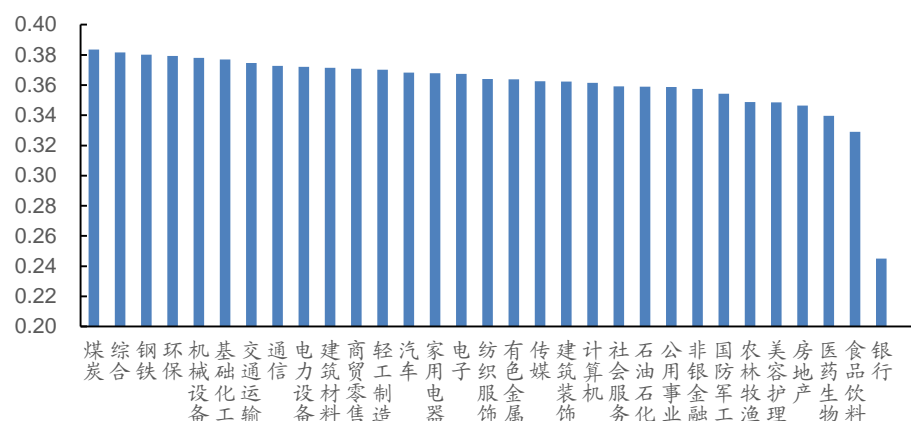
图 5：地理相关系数因子的省份分布特征



数据来源：Wind，广发证券发展研究中心

图6表明，地理相关系数因子在行业层面存在显著差异。具体地，银行业地理相关系数均值小于其他行业因子均值一半，这也与市场实际情况相吻合，即银行业股价变动与其他制造业行业相比，影响因素存在显著差别。为排除行业暴露对地理相关系数因子的影响，本报告在实证分析将对因子进行行业中性化处理。

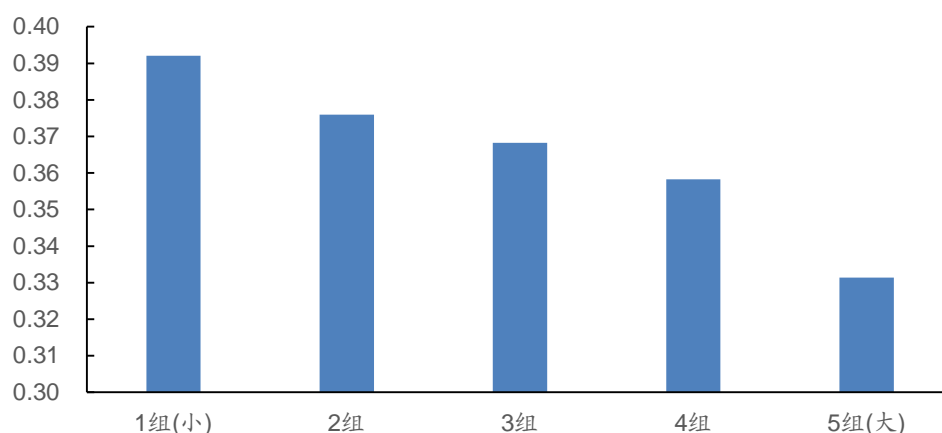
图 6：地理相关系数因子的申万一级行业分布特征



数据来源：Wind，广发证券发展研究中心

图7表明，地理相关系数因子在市值层面存在显著差异。具体表现为市值越小(大)，地理相关系数因子值越高(低)。这意味着，小市值股票更容易受到共同地理因素影响，其股价变动与地理关联公司股价变动相关程度更高。为排除市值暴露对地理相关系数因子的影响，本报告在实证分析将对因子进行市值中性化处理。

图 7：地理相关系数因子的市值分布特征



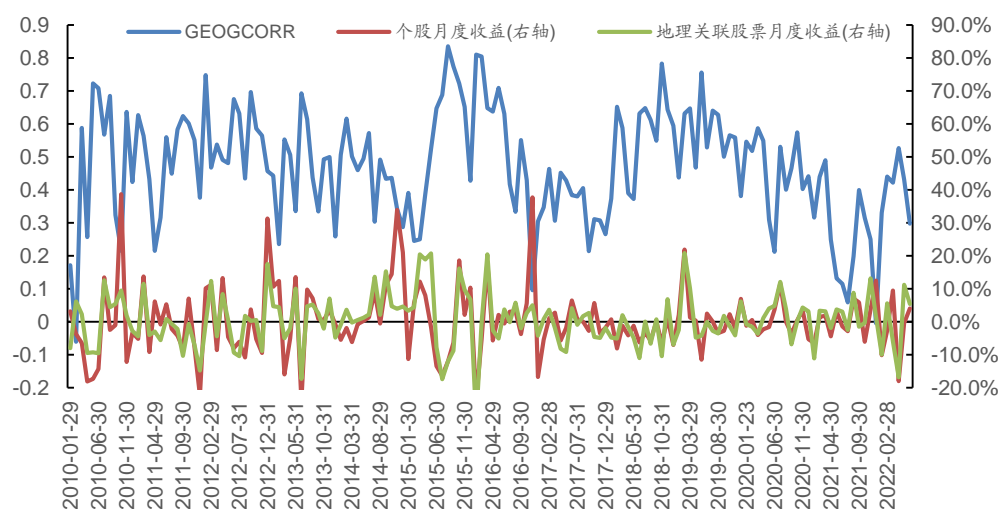
数据来源：Wind，广发证券发展研究中心

### （三）因子策略构建

以图8中某个股为例，对于一支股票而言，它与地理关联公司股票的收益序列相关性越强，隐含的信息在于这些股票受共同基本面因素影响更大，因此股票价格变动的一致性越高，这支股票在下期出现收益反转的可能性越大，那么在下月内持有该股票获得股票反转收益的概率越大。

对于整体样本区间而言，地理关联公司股票与研究个股在月度收益上表现出较高的相关性(相关系数0.693)，这一相关性越强，研究个股在下期的月度收益越高(因子值与个股未来1月收益相关系数0.034)。对于地理关联公司股票平均月度收益为负的样本期而言，上述地理相关系数因子对个股未来反转收益预测效果越显著。

图 8：某个股月度地理相关系数因子值与股票收益走势图



数据来源：Wind，广发证券发展研究中心



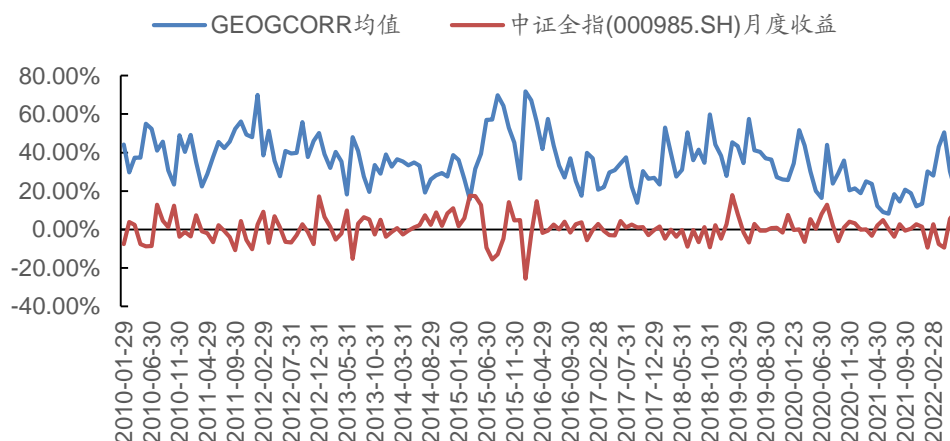
表 3: 某个股月度地理相关系数因子值与股票收益序列相关性

	因子值 均值	个股收益与关联 股票平均收益	因子值 与个股收益	因子值与个股 未来 1 月收益
总样本	0.470	0.693	-0.310	0.034
关联股票平均收益>0 样本	0.434	0.474	-0.196	-0.014
关联股票平均收益<0 样本	0.508	0.691	-0.304	0.108

数据来源: Wind, 广发证券发展研究中心

从全市场地理相关系数因子均值与中证全指指数收益走势可以看出, 整体而言, 全市场股票月度因子均值为0.358, 与中证全指指数同期收益呈负相关关系(相关系数-0.379), 与下期收益呈正相关关系(相关系数0.017)。仅考虑中证全指指数收益为负的月份时, 因子均值增长至为0.410, 因子均值与中证全指指数同期收益相关系数增强至-0.620, 与下期收益相关系数增强至0.077。该统计结果与上述个股分析保持一致, 说明: (1) 全市场股票地理相关系数因子均值与下期市场收益序列呈正向变动关系。当期的全市场因子均值越高, 下期反转效应越强, 市场收益越高; (2) 第(1)点结论在中证全指指数收益小于0时效果更显著。

图 9: 全市场月度地理相关系数因子均值与中证全指月度收益走势图



数据来源: Wind, 广发证券发展研究中心

表 4: 全市场股票地理相关系数因子均值与中证全指收益序列相关性

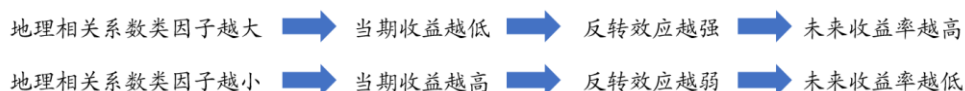
	因子值均值	因子值与 市场同期收益	因子均值与市场 未来 1 月收益
总样本	0.358	-0.379	0.017
中证全指收益>0 样本	0.314	0.202	0.021
中证全指收益<0 样本	0.410	-0.620	0.077

数据来源: Wind, 广发证券发展研究中心

根据地理相关系数因子与个股、全市场收益走势相关性分析, 当某支股票的地理相关系数因子值越大时, 该股票当期的收益率越低, 下期的反转效应越显著。

据此，本篇专题报告构造如下交易策略：根据地理相关系数类因子值，在调仓日买入地理关联程度最高的股票，同时卖出地理关联程度最低的股票。

图 10：策略构建框架



数据来源：广发证券发展研究中心

## 四、实证分析

### （一）数据说明

选股范围：全市场

股票预处理：剔除非上市、摘牌、ST/\*ST、涨跌停板、上市未满1年股票

因子预处理：MAD去极值、Z-Score标准化、行业市值中性化

回测区间：2010.01.01 – 2022.06.30

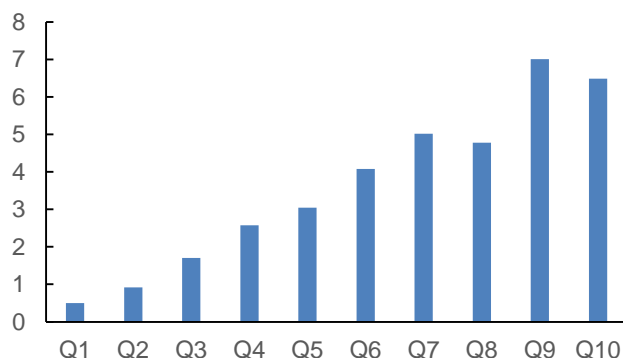
分档方式：根据当期股票的因子值，从小到大分为十档

调仓周期：每个月最后一个交易日以收盘价调仓

交易费用：千分之三（卖出时收取）

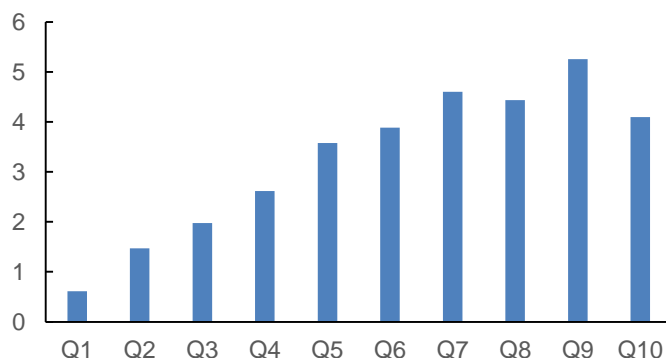
### （二）因子分档表现

图 11：全市场 GEOGCORR因子十档-月度调仓



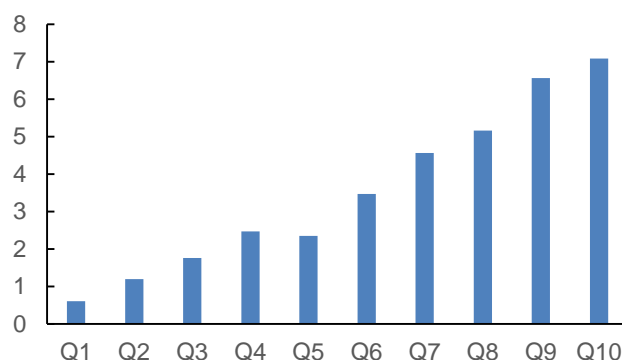
数据来源：Wind，广发证券发展研究中心

图 12：全市场 GEOGCORRCHG因子十档-月度调仓



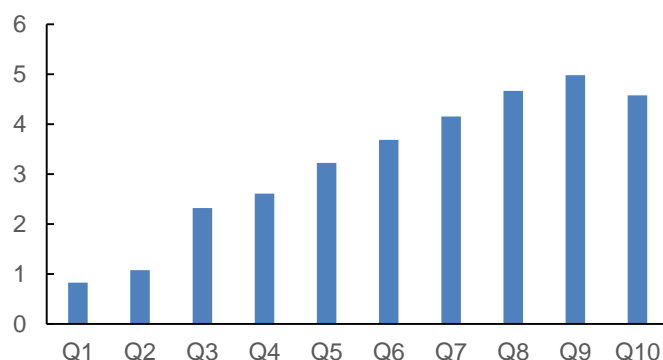
数据来源：Wind，广发证券发展研究中心

图 13: 全市场 GEOGCRRP因子十档-月度调仓



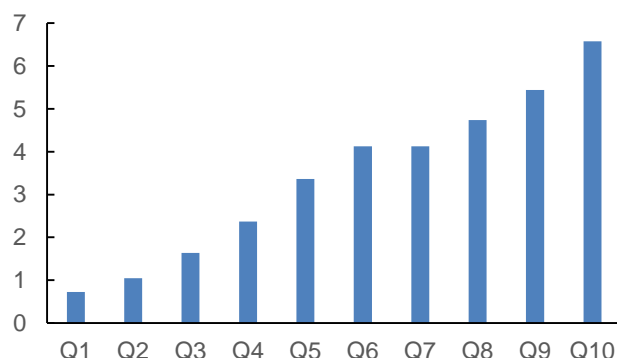
数据来源: Wind, 广发证券发展研究中心

图 14: 全市场 GEOGCRRN因子十档-月度调仓



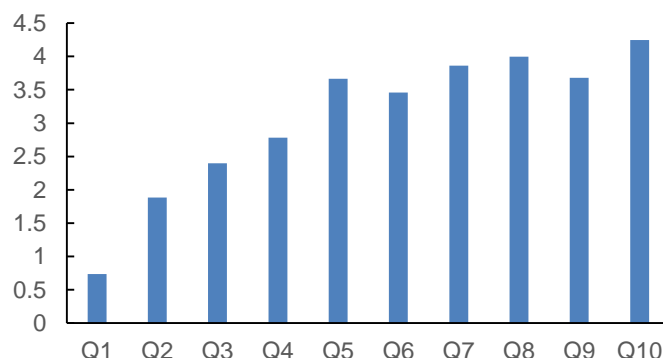
数据来源: Wind, 广发证券发展研究中心

图 15: 全市场 GEOGCRRIIP因子十档-月度调仓



数据来源: Wind, 广发证券发展研究中心

图 16: 全市场 GEOGCRRIJP因子十档-月度调仓



数据来源: Wind, 广发证券发展研究中心

在月度调仓的历史回测下, 6种地理相关系数类因子的整体分层效果表现不一。具体而言, 地理相关系数拆解类因子中, GEOGCRRP因子与GEOGCRRIIP因子分层效果显著, 分层收益区分度高。

### (三) 因子实证结果

本小节中, 首先, 将对构建的6种地理相关系数类因子在IC、多空策略、多头相对基准策略以及换手率方面的回测表现进行整体展示。其次, 通过对比各因子在回测期中的绩效, 选取GEOGCRRN因子、GEOGCRRP因子与GEOGCRRIIP因子, 对其分年度表现进行详细展示。

#### 1. 整体表现

整体来看, 6种地理相关系数类因子在选股方向上与构建的交易策略保持一致, 即个股与地理关联公司股票的整体相关性越高, 股票未来收益表现越好。各因子的IC表现、多空对冲策略表现与因子分档测试结果一致, GEOGCRRN因子具有最高的IC均值与多空年化收益, 其次为GEOGCRRP因子与GEOGCRRIIP因子, 但GEOGCRRP因子由于回测期的最大回撤相对更低, 信息比率高于其他因子。多头相对基准策略表现出现分化, 上述两种地理相关系数拆解因子表现优于地理相关系数因子。各因子的多头平均换手率在80%左右。具体来看, GEOGCRRN、GEOGCRRP

与GEOGCRIP共3个因子在IC分析、多空策略绩效、多头相对基准策略绩效上总体表现较好。

表 5：地理相关系数类因子整体绩效表现

因子名称	因子代码	IC 均值	IC T 统计量	正 IC 占比	多空策略表现			多头相对基准策略			多头平均 换手率
					年化 收益	信息 比率	最大 回撤	年化 收益	信息 比率	最大 回撤	
相关系数	GEOGCR	0.074	13.207	88.00%	22.11%	2.84	6.39%	13.21%	1.55	12.14%	77.77%
相关变动	GEOGCRCHG	0.044	9.711	82.67%	16.04%	2.238	7.22%	9.26%	1.101	15.22%	83.60%
相关拆解	GEOGCRRP	0.069	15.297	90.00%	21.19%	3.234	4.57%	13.98%	1.706	10.52%	81.92%
	GEOGCRRN	0.052	9.147	77.33%	13.93%	1.718	7.50%	10.08%	1.238	14.43%	79.40%
	GEOGCRIP	0.063	15.336	88.67%	18.92%	3.099	4.22%	13.48%	1.659	10.45%	84.62%
	GEOGCRJP	0.04	7.649	76.00%	14.51%	1.713	4.99%	9.48%	1.157	14.66%	83.97%

数据来源：Wind，广发证券发展研究中心

## 1. GEOGCR因子具体表现

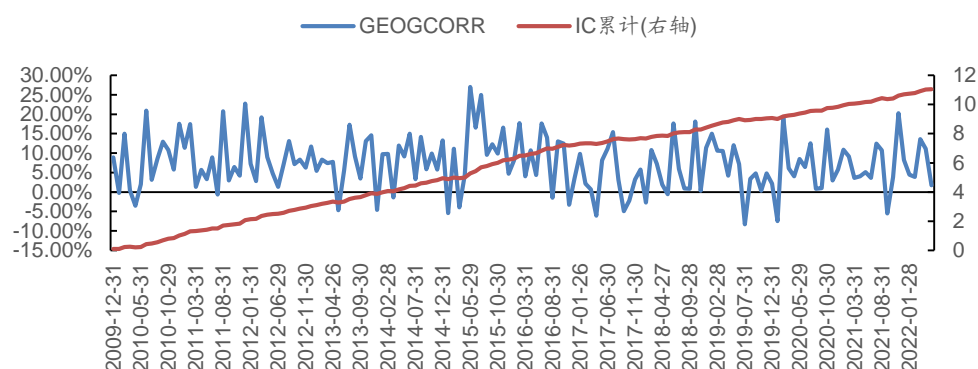
表 6：GEOGCR 因子-整体与分年度 IC 表现

范围	IC 均值	IC 标准差	IC 最大值	IC 最小值	IC T 统计量	IC 累计值	正 IC 占比
2010	0.071	0.072	0.21	-0.035	3.416	0.846	83.33%
2011	0.083	0.07	0.208	-0.006	4.077	0.993	91.67%
2012	0.091	0.063	0.227	0.013	4.968	1.09	100.00%
2013	0.083	0.058	0.173	-0.047	4.971	0.992	91.67%
2014	0.074	0.059	0.15	-0.046	4.298	0.885	83.33%
2015	0.113	0.098	0.27	-0.054	4.01	1.362	83.33%
2016	0.085	0.069	0.177	-0.033	4.252	1.021	83.33%
2017	0.037	0.065	0.154	-0.061	1.953	0.441	75.00%
2018	0.055	0.069	0.181	-0.027	2.778	0.664	83.33%
2019	0.063	0.063	0.15	-0.083	3.471	0.76	91.67%
2020	0.06	0.073	0.192	-0.074	2.866	0.724	91.67%
2021	0.07	0.063	0.203	-0.055	3.823	0.84	91.67%
2022	0.072	0.046	0.136	0.017	3.83	0.431	100.00%
ALL	0.074	0.068	0.27	-0.083	13.207	11.049	88.00%

数据来源：Wind，广发证券发展研究中心

2022 指截止至 2022 年 06 月 30 日，如下若无特别说明，该表述与此一致

图 17: GEOGCORR因子-IC值与IC累计值走势



数据来源: Wind, 广发证券发展研究中心

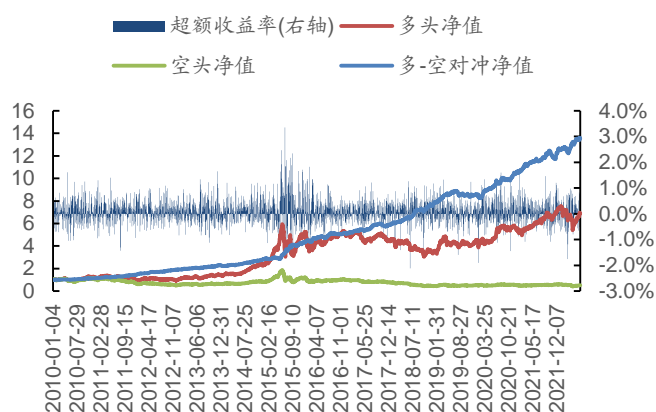
表 7: GEOGCORR 因子-多空对冲策略与多头相对中证 500 指数策略-整体与分年度表现-月度调仓

年份	纯多头		多空对冲策略				多头-中证 500 策略				
	累计 收益率	累计 收益率	年化 收益率	年化 波动率	信息 比率	最大 回撤	累计 收益率	年化 收益率	年化 波动率	信息 比率	最大 回撤
2010	19.86%	12.42%	12.42%	5.43%	2.289	1.74%	9.39%	9.39%	4.50%	2.088	1.02%
2011	-22.54%	29.15%	29.15%	5.26%	5.541	0.00%	16.07%	16.07%	4.51%	3.565	0.32%
2012	14.10%	28.07%	28.07%	6.77%	4.148	0.00%	13.82%	13.82%	4.33%	3.189	0.59%
2013	33.18%	20.24%	20.24%	4.26%	4.746	0.91%	14.93%	14.93%	4.36%	3.425	1.21%
2014	58.95%	23.78%	23.78%	6.91%	3.44	1.40%	15.21%	15.21%	8.69%	1.751	5.45%
2015	122.06%	48.13%	48.13%	15.77%	3.051	3.25%	57.02%	57.02%	13.24%	4.306	2.77%
2016	-3.91%	16.38%	16.38%	7.30%	2.244	0.89%	18.62%	18.62%	6.57%	2.836	1.01%
2017	-10.56%	14.04%	14.04%	6.04%	2.326	2.64%	-10.10%	-10.10%	4.98%	-2.029	10.14%
2018	-25.99%	26.22%	26.22%	6.25%	4.195	0.02%	11.17%	11.17%	7.18%	1.556	2.22%
2019	28.73%	12.29%	12.29%	6.84%	1.797	4.03%	1.98%	1.98%	6.14%	0.323	4.68%
2020	25.74%	21.91%	21.91%	8.52%	2.573	2.97%	3.47%	3.47%	11.51%	0.301	4.72%
2021	35.82%	19.01%	19.01%	8.44%	2.253	4.76%	17.68%	17.68%	12.21%	1.448	2.08%
2022	-6.81%	8.51%	17.74%	6.25%	2.84	0.65%	6.66%	13.76%	9.28%	1.483	3.81%
ALL	548.84%	1114.05%	22.11%	7.78%	2.84	6.39%	371.81%	13.21%	8.53%	1.55	12.14%

数据来源: Wind, 广发证券发展研究中心

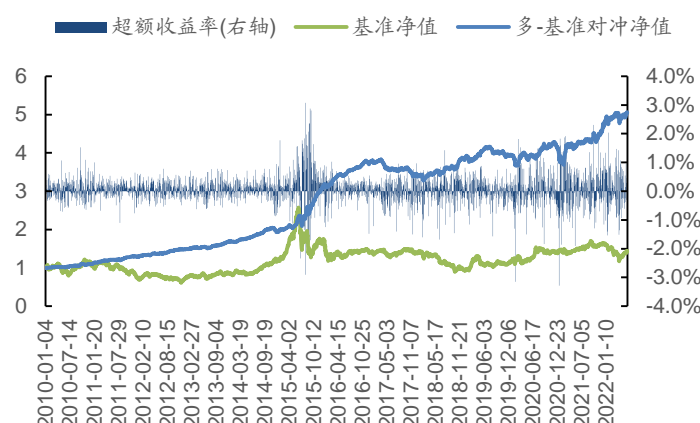


图 18: GEOGCORR因子-多空策略净值走势



数据来源: Wind, 广发证券发展研究中心

图 19: GEOGCORR因子-多-中证500策略净值走势



数据来源: Wind, 广发证券发展研究中心

表 8: GEOGCORR 因子-分年度换手率

统计区间	均值	最大值	最小值	标准差	累计值
2010	75.89%	84.06%	63.91%	0.068	834.83%
2011	79.00%	85.21%	71.34%	0.04	948.05%
2012	77.64%	87.71%	70.26%	0.047	931.64%
2013	77.16%	85.91%	64.73%	0.062	925.96%
2014	80.53%	86.85%	72.90%	0.038	966.35%
2015	80.56%	85.92%	71.83%	0.043	966.68%
2016	77.95%	84.07%	73.66%	0.034	935.45%
2017	74.53%	85.14%	64.92%	0.07	894.32%
2018	74.04%	79.08%	69.67%	0.027	888.47%
2019	74.30%	80.12%	67.87%	0.037	891.58%
2020	79.19%	83.53%	70.75%	0.04	950.29%
2021	81.76%	85.36%	72.30%	0.038	981.15%
2022	78.93%	86.34%	70.10%	0.055	473.57%
ALL	77.77%	87.71%	63.91%	0.052	11588.33%

数据来源: Wind, 广发证券发展研究中心

在全市场选股中, GEOGCORR因子的选股区分度较高, 因子IC均值为0.074, 正IC占比88.00%, 除2017年外, 分年度IC均在0.05以上。在多头相对中证500指数的回测中, 策略整体的年化超额收益率为13.21%, 信息比率为1.55。在2015年市场趋势较大时表现相对较好, 除2017年外, 其余年份均可取得超额收益。整体换手率保持在77.77%左右。

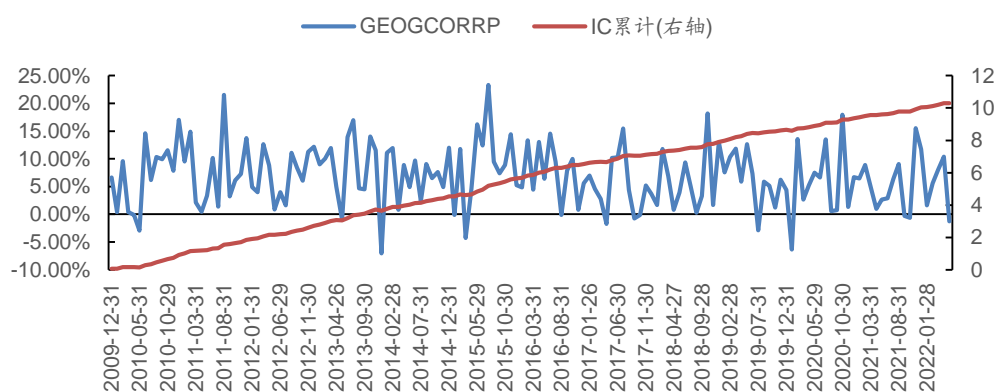
## 2. GEOGCRRP因子具体表现

表 9: GEOGCRRP 因子-整体与分年度 IC 表现

范围	IC 均值	IC 标准差	IC 最大值	IC 最小值	IC T 统计量	IC 累计值	正 IC 占比
2010	0.062	0.055	0.146	-0.029	3.892	0.743	83.33%
2011	0.081	0.068	0.215	0.004	4.142	0.971	100.00%
2012	0.073	0.043	0.137	0.009	5.823	0.874	100.00%
2013	0.094	0.052	0.17	-0.01	6.315	1.127	91.67%
2014	0.059	0.053	0.119	-0.071	3.86	0.707	91.67%
2015	0.097	0.073	0.233	-0.043	4.566	1.16	83.33%
2016	0.075	0.048	0.145	-0.001	5.44	0.899	91.67%
2017	0.052	0.05	0.155	-0.017	3.602	0.628	75.00%
2018	0.055	0.053	0.182	0.003	3.641	0.665	100.00%
2019	0.07	0.047	0.13	-0.029	5.2	0.841	91.67%
2020	0.056	0.068	0.18	-0.064	2.887	0.677	91.67%
2021	0.053	0.046	0.155	-0.006	3.98	0.636	83.33%
2022	0.06	0.051	0.117	-0.013	2.912	0.36	83.33%
ALL	0.069	0.055	0.233	-0.071	15.297	10.289	90.00%

数据来源: Wind, 广发证券发展研究中心

图 20: GEOGCRRP因子-IC值与IC累计值走势



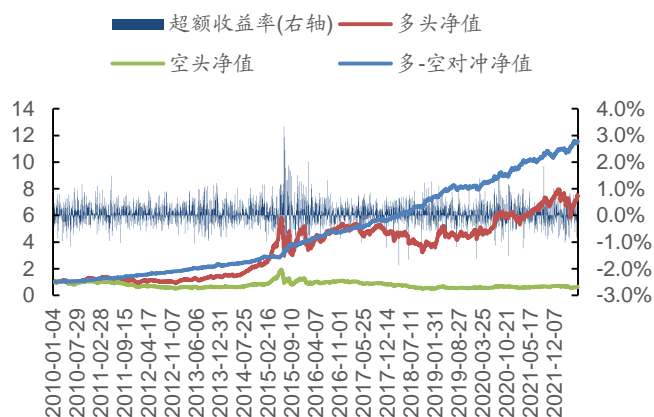
数据来源: Wind, 广发证券发展研究中心

表 10: GEOGCRRP 因子-多空对冲策略与多头相对中证 500 指数策略-整体与分年度表现

	纯多头		多空对冲策略				多头-中证 500 策略				
	累计	累计	年化	年化	信息	最大	累计	年化	年化	信息	最大
年份	收益率	收益率	收益率	波动率	比率	回撤	收益率	收益率	波动率	比率	回撤
2010	21.24%	18.80%	18.80%	4.34%	4.332	0.58%	10.74%	10.74%	3.84%	2.795	1.47%
2011	-21.35%	25.57%	25.57%	6.17%	4.142	0.42%	17.60%	17.60%	5.06%	3.481	0.28%
2012	10.70%	22.25%	22.25%	3.67%	6.058	0.00%	10.48%	10.48%	3.43%	3.053	0.00%
2013	33.86%	26.34%	26.34%	4.77%	5.516	0.32%	15.27%	15.27%	4.47%	3.419	0.41%
2014	58.68%	23.22%	23.22%	6.51%	3.565	0.00%	14.92%	14.92%	6.49%	2.299	3.17%
2015	119.61%	37.55%	37.55%	13.82%	2.717	4.57%	55.88%	55.88%	11.44%	4.884	2.31%
2016	-1.04%	20.03%	20.03%	6.19%	3.237	2.17%	22.52%	22.52%	7.24%	3.112	0.17%
2017	-8.00%	15.88%	15.88%	4.20%	3.781	0.86%	-7.56%	-7.56%	4.73%	-1.598	7.97%
2018	-23.14%	23.46%	23.46%	5.07%	4.627	0.00%	15.38%	15.38%	8.33%	1.845	1.26%
2019	28.66%	15.14%	15.14%	5.10%	2.966	1.17%	1.81%	1.81%	5.51%	0.328	4.59%
2020	25.86%	17.11%	17.11%	7.44%	2.298	1.77%	3.40%	3.40%	11.99%	0.283	4.65%
2021	33.03%	13.78%	13.78%	6.50%	2.119	3.19%	15.22%	15.22%	11.98%	1.271	2.50%
2022	-4.54%	7.36%	15.25%	6.34%	2.405	1.09%	8.98%	18.77%	9.01%	2.083	2.85%
ALL	608.48%	1004.48%	21.19%	6.55%	3.234	4.57%	413.10%	13.98%	8.19%	1.706	10.52%

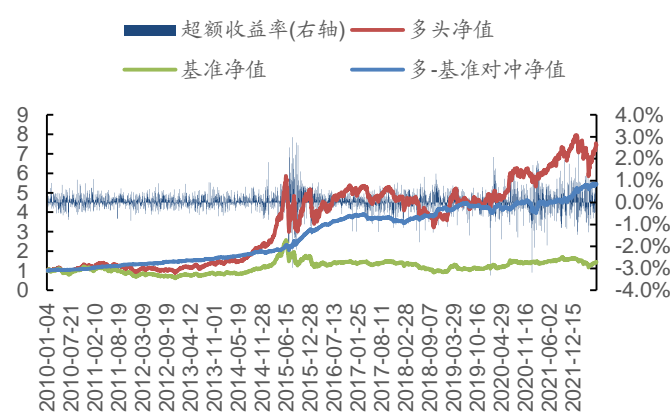
数据来源: Wind, 广发证券发展研究中心

图 21: GEOGCRRP 因子-多空策略净值走势



数据来源: Wind, 广发证券发展研究中心

图 22: GEOGCRRP 因子-多-中证 500 策略净值走势



数据来源: Wind, 广发证券发展研究中心

表 11: GEOGCORRP 因子-分年度换手率

统计区间	均值	最大值	最小值	标准差	累计值
2010	80.21%	88.41%	72.93%	0.05	882.36%
2011	84.24%	89.61%	79.29%	0.033	1010.93%
2012	81.20%	87.28%	75.37%	0.034	974.46%
2013	82.05%	92.63%	74.88%	0.059	984.60%
2014	81.58%	86.24%	73.76%	0.038	978.93%
2015	83.53%	89.19%	78.67%	0.029	1002.40%
2016	80.48%	83.56%	76.71%	0.021	965.72%
2017	81.81%	87.04%	78.00%	0.027	981.76%
2018	81.24%	87.19%	78.03%	0.027	974.93%
2019	79.44%	83.38%	75.08%	0.031	953.23%
2020	81.45%	89.35%	74.78%	0.04	977.39%
2021	85.02%	89.78%	80.84%	0.03	1020.27%
2022	83.13%	87.69%	79.18%	0.034	498.77%
ALL	81.92%	92.63%	72.93%	0.038	12205.73%

数据来源: Wind, 广发证券发展研究中心

在全市场选股中, GEOGCORRP 因子表现出较好的选股区分度, 因子IC均值为 0.069, 正IC占比90.00%, 分年度IC均值最小出现在2017年, 为0.052。在多头相对中证500指数的回测中, 策略整体的年化收益率为13.98%, 信息比率为1.706。

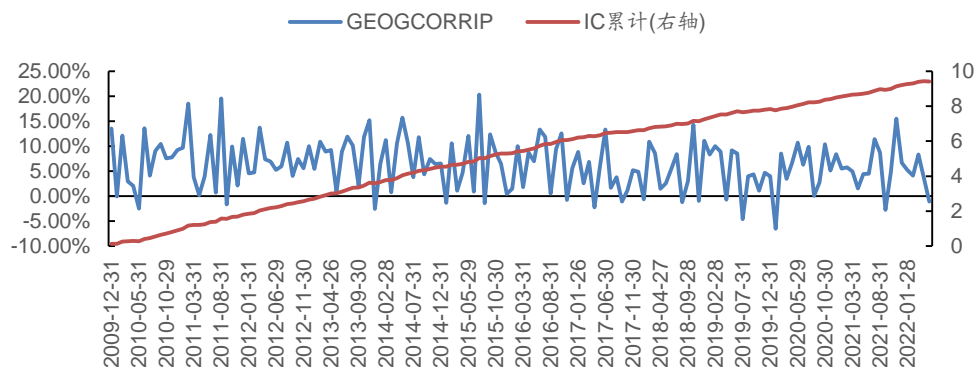
### 3. GEOGCORRIP因子具体表现

表 12: GEOGCORRIP 因子-整体与分年度 IC 表现

范围	IC 均值	IC 标准差	IC 最大值	IC 最小值	IC T 统计量	IC 累计值	正 IC 占比
2010	0.067	0.054	0.136	-0.025	4.332	0.805	83.33%
2011	0.074	0.07	0.195	-0.017	3.634	0.883	91.67%
2012	0.073	0.031	0.137	0.041	8.264	0.878	100.00%
2013	0.088	0.041	0.152	0.008	7.414	1.055	100.00%
2014	0.072	0.051	0.157	-0.026	4.87	0.866	91.67%
2015	0.068	0.065	0.203	-0.014	3.6	0.811	83.33%
2016	0.064	0.053	0.134	-0.008	4.157	0.764	91.67%
2017	0.043	0.043	0.133	-0.022	3.457	0.517	83.33%
2018	0.047	0.05	0.143	-0.012	3.299	0.569	75.00%
2019	0.054	0.048	0.111	-0.046	3.878	0.649	83.33%
2020	0.051	0.049	0.107	-0.065	3.596	0.613	91.67%
2021	0.061	0.046	0.155	-0.027	4.541	0.729	91.67%
2022	0.045	0.032	0.084	-0.011	3.406	0.27	83.33%
ALL	0.063	0.05	0.203	-0.065	15.336	9.409	88.67%

数据来源: Wind, 广发证券发展研究中心

图 23: GEOGCRIP因子-IC值与IC累计值走势



数据来源: Wind, 广发证券发展研究中心

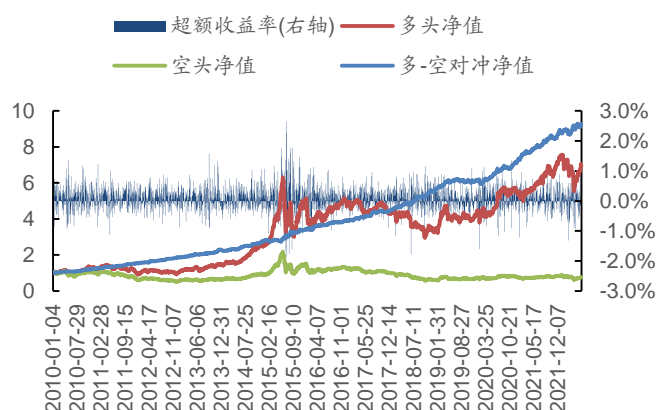
表 13: GEOGCRIP 因子-多空对冲策略与多头相对中证 500 指数策略-整体与分年度表现

年份	纯多头		多空对冲策略				多头-中证 500 策略				
	累计收益率	累计收益率	年化收益率	年化波动率	信息比率	最大回撤	累计收益率	年化收益率	年化波动率	信息比率	最大回撤
2010	24.29%	23.39%	23.39%	5.79%	4.036	0.81%	13.57%	13.57%	5.36%	2.53	2.17%
2011	-22.61%	23.74%	23.74%	5.67%	4.187	0.26%	16.37%	16.37%	4.70%	3.481	1.09%
2012	11.26%	20.83%	20.83%	4.80%	4.338	0.36%	10.81%	10.81%	4.90%	2.206	0.58%
2013	36.22%	23.71%	23.71%	4.86%	4.881	0.22%	17.18%	17.18%	5.85%	2.937	0.79%
2014	68.90%	21.43%	21.43%	7.62%	2.812	0.35%	22.45%	22.45%	8.01%	2.804	2.76%
2015	97.37%	18.70%	18.70%	11.31%	1.653	4.20%	41.80%	41.80%	13.76%	3.038	5.69%
2016	-5.31%	13.80%	13.80%	5.77%	2.39	0.65%	16.96%	16.96%	7.02%	2.418	0.45%
2017	-9.56%	14.09%	14.09%	4.94%	2.856	1.36%	-9.20%	-9.20%	3.95%	-2.331	9.20%
2018	-25.50%	21.26%	21.26%	5.21%	4.085	0.06%	11.80%	11.80%	6.94%	1.702	1.91%
2019	29.63%	12.08%	12.08%	5.20%	2.322	2.48%	2.78%	2.78%	5.42%	0.512	3.98%
2020	27.46%	17.86%	17.86%	6.86%	2.604	3.90%	4.95%	4.95%	10.84%	0.457	3.91%
2021	38.18%	21.27%	21.27%	4.03%	5.273	0.02%	19.75%	19.75%	10.91%	1.81	2.11%
2022	-7.07%	5.13%	10.52%	6.02%	1.746	1.58%	6.34%	13.09%	8.37%	1.563	3.68%
ALL	557.80%	772.19%	18.92%	6.10%	3.099	4.22%	385.96%	13.48%	8.13%	1.659	10.45%

数据来源: Wind, 广发证券发展研究中心

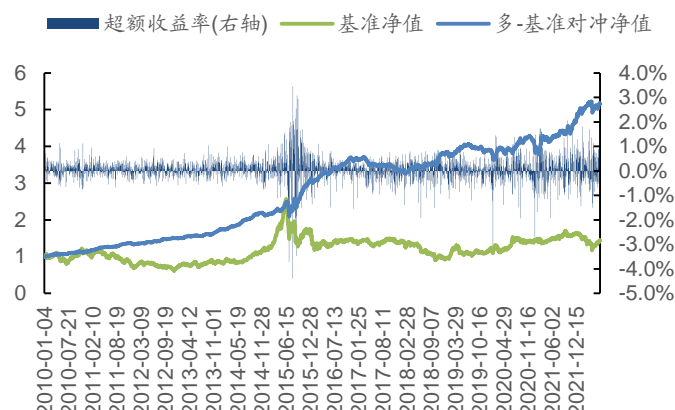


图 24: GEOGCRIP因子-多空策略净值走势



数据来源: Wind, 广发证券发展研究中心

图 25: GEOGCRIP因子-多-中证500策略净值走势



数据来源: Wind, 广发证券发展研究中心

表 14: GEOGCRIP 因子-分年度换手率

统计区间	均值	最大值	最小值	标准差	累计值
2010	85.47%	90.30%	80.30%	0.033	940.13%
2011	85.16%	91.76%	77.92%	0.041	1021.89%
2012	86.63%	91.76%	77.89%	0.039	1039.60%
2013	85.33%	91.87%	78.34%	0.039	1024.02%
2014	84.19%	88.26%	80.54%	0.022	1010.28%
2015	85.38%	90.10%	74.59%	0.048	1024.61%
2016	84.75%	89.69%	79.84%	0.034	1016.98%
2017	81.42%	86.69%	73.60%	0.036	977.08%
2018	83.34%	87.32%	77.41%	0.031	1000.05%
2019	83.77%	88.55%	77.18%	0.033	1005.28%
2020	84.82%	88.46%	79.71%	0.027	1017.84%
2021	85.48%	89.50%	81.27%	0.022	1025.81%
2022	84.10%	85.57%	81.80%	0.015	504.60%
ALL	84.62%	91.87%	73.60%	0.035	12608.17%

数据来源: Wind, 广发证券发展研究中心

在全市场选股中, GEOGCRIP因子表现出较好的选股区分度, 因子IC均值为0.063, 正IC占比88.67%。在多头相对中证500指数的回测中, 策略整体的年化超额收益率为13.48%, 信息比率为1.659。整体换手率保持在84.62%左右。

综合上述3个因子的实证分析结果, GEOGCRIP因子整体表现优于GEOGCRIP因子与GEOGCRIP因子。由于该因子对地理相关系数进行拆解, 仅考虑了研究个股与地理关联股票在日度收益不为负情况下的相关关系, 因此, 更能精确地度量多头策略所需要的基于地理共性挖掘的正向共同基本面信息。因此该因子在因子分档测试、多头对冲策略方面都具有更好的表现。

## 五、地理关联度因子与 BARRA 因子相关性分析

CNE6版本的BARRA因子模型将股票因子划分为市场、价值、成长、盈利、波动等大类因子。传统BARRA因子已经在市场中得到验证，能够较有效地刻画股票特征。本节将对GEOG CORR因子、GEOG CORRP因子、GEOG CORRIP因子与BARRA因子进行相关性分析。

### （一）BARRA 因子说明

在CNE6版本的BARRA因子模型中，共详细阐述了数大类因子。本报告仅在各因子类型中选取其中一个因子进行相关性说明。首先，对因子类型及计算公式进行如下说明：

表 15: BARRA 因子说明

因子类型	因子名称	因子计算公式/说明
Beta	HBETA	对个股在过去 252 个交易日的日频收益率序列与同期市场指数收益率序列进行回归，得到市场指数的回归系数 $\beta$ 即为 HBETA 因子 $r_t^i - r_t^f = \alpha + \beta r_t^m + e_t$
book to price	BTOP	最新报告期的股东权益与当前流通市值的比值
earnings quality	ATURNOVER	最新财报的营业收入与当期平均资产总值的比值
earnings variability	VSAL	过去五期财报营业收入的标准差/过去五期财报营业收入的均值
earnings yield	CETOP	个股历史 12 个月滚动的经营现金流净额与当前流通市值的比值
growth	EGRO	用个股过去五年的每股盈利对时间做回归，得到的斜率再除以五年每股盈利的平均值
investment quality	AGRO	用过去五年的总资产对时间做回归，得到的斜率再除以五年的平均值。数值乘-1
leverage	MLEV	长期资本与股东权益的比值，其中长期资本包括普通股市值、最新优先股市值以及付息债券市值
liquidity	STOM	过去 21 个交易日的换手率之和的对数值
long term reversal	RSTR	计算过去 504 个交易日的个股超额对数收益率与过去 21 个交易日的超额对数收益率的差额累计和，得到 RSTR 因子 $RSTR = \sum_L^{T+L} w_t [\ln(1 + r_t^i) - \ln(1 + r_t^f)]$
mid capitalization	MIDCAP	用截面上个股的对数流通市值的三次方对对数流通市值进行回归，对回归残差进行 MAD 去极值与中性化处理，获得 MIDCAP 因子
momentum	HALPHA	对个股在过去 252 个交易日的日频收益率序列与同期市场指数收益率序列进行回归，得到截距项 $\alpha$ 为 HALPHA 因子 $r_t^i - r_t^f = \alpha + \beta r_t^m + e_t$
profitability	ROA	个股历史 12 个月滚动净利润与最新报告的总资产的比值
residual volatility	HSIGMA	对个股在过去 252 个交易日的日频收益率序列与同期市场指数收益率序列进行回归，得到残差的标准差即为 HSIGMA 因子 $r_t^i - r_t^f = \alpha + \beta r_t^m + e_t$
size	LNCAP	个股的对数流通市值

数据来源：Wind，广发证券发展研究中心

## （二）地理关联度因子与 BARRA 相关性分析

本小节将数据预处理（MAD法去极值、Z-Score标准化、行业市值中性化）后的GEOGCORR因子、GEOGCORRP因子、GEOGCORRIP因子与BARRA因子进行Spearman秩相关性分析，相关性结果见下表16。如表所示，地理相关系数因子与HBETA、BTOP因子存在10%左右的正相关性。与STOM、HALPHA、HSIGMA因子之间存在12%-22%的负相关性。整体而言，地理相关系数类因子能够挖掘传统因子外的增量信息，具体为股票之间的地理关联信息。

表 16：地理相关系数类因子与 BARRA 因子相关系数

	GEOGCORR	GEOGCORRP	GEOGCORRIP
HBETA	10.64%	8.97%	9.43%
BTOP	14.95%	12.95%	10.44%
ATURNOVER	-1.38%	-0.46%	-0.44%
VSAL	-6.20%	-4.15%	-4.62%
CETOP	4.99%	4.95%	3.97%
EGRO	-2.97%	-1.65%	-2.00%
AGRO	4.22%	1.65%	2.26%
STOM	-21.21%	-18.50%	-15.93%
RSTR	-11.12%	-8.16%	-6.82%
HALPHA	-19.30%	-14.95%	-12.32%
ROA	-3.03%	0.27%	0.54%
HSIGMA	-14.50%	-12.91%	-11.64%
MIDCAP	1.56%	2.21%	2.55%
LNCAP	-5.22%	-3.43%	-4.00%
MLEV	8.35%	6.11%	4.42%

数据来源：Wind，广发证券发展研究中心

## 六、地理关联度因子敏感性分析

### （一）敏感性分析-选股范围

本报告在实证分析部分，仅在全市场选股范围内，详细展示了GEOGCORR因子、GEOGCORRP因子、GEOGCORRIP因子的选股策略表现。但是因子特征分析部分也表明，地理相关系数类因子与股票的市值大小存在显著关系，小市值股票往往具有更高的地理相关系数因子。因此，本节将重点分析上述3种因子对于选股范围的敏感性。具体而言，测试了沪深300、中证500、中证800、中证1000、创业板选股范围内，各因子的多空对冲策略、多头相对基准指数策略与换手率的整体表现。

表 17：地理相关系数类因子在不同选股范围的整体绩效表现

因子代码	选股范围	多空策略表现			多头相对基准策略			多头平均 换手率
		年化 收益	信息 比率	最大 回撤	年化 收益	信息 比率	最大 回撤	
GEOGCROR	全市场	22.11%	2.84	6.39%	13.21%	1.55	12.14%	77.77%
	沪深 300	8.34%	0.761	11.03%	2.26%	0.193	24.12%	76.63%
	中证 1000	24.03%	2.527	5.04%	13.89%	1.651	6.67%	78.38%
	中证 500	12.68%	1.375	7.63%	5.28%	0.738	13.76%	78.58%
	中证 800	11.12%	1.355	7.87%	5.23%	0.443	29.47%	77.18%
	创业板	19.76%	1.755	14.66%	8.14%	0.461	42.32%	80.28%
GEOGCRORP	全市场	21.19%	3.234	4.57%	13.98%	1.706	10.52%	81.92%
	沪深 300	7.70%	0.747	30.48%	4.35%	0.411	29.87%	80.96%
	中证 1000	22.62%	2.608	5.48%	14.82%	2.021	7.34%	82.91%
	中证 500	11.11%	1.221	19.94%	7.04%	0.976	12.34%	82.37%
	中证 800	10.27%	1.34	12.72%	6.71%	0.621	24.42%	81.61%
	创业板	20.00%	1.826	19.25%	10.80%	0.626	33.33%	84.14%
GEOGCRORIP	全市场	18.92%	3.099	4.22%	13.48%	1.659	10.45%	84.62%
	沪深 300	4.89%	0.458	14.59%	4.32%	0.409	19.51%	85.09%
	中证 1000	14.03%	1.98	6.44%	9.54%	1.342	5.06%	85.41%
	中证 500	10.47%	1.407	10.71%	6.43%	1.015	6.32%	85.19%
	中证 800	8.00%	1.233	9.29%	5.17%	0.507	27.18%	84.94%
	创业板	16.75%	1.509	23.57%	7.84%	0.467	33.69%	85.74%

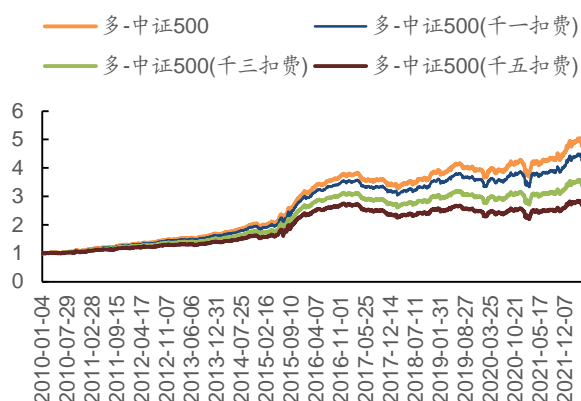
数据来源：Wind，广发证券发展研究中心

实证结果显示，多空策略方面，3种因子在中证1000、创业板选股范围中存在较好表现，GEOGCROR因子与GEOGCRORP因子年化收益率均高于全市场选股范围同策略表现。多头相对基准指数策略方面，GEOGCROR因子与GEOGCRORP因子在中证1000选股范围内的年化收益率、信息比率均优于全市场选股范围同策略表现。

## （二）敏感性分析-手续费费率

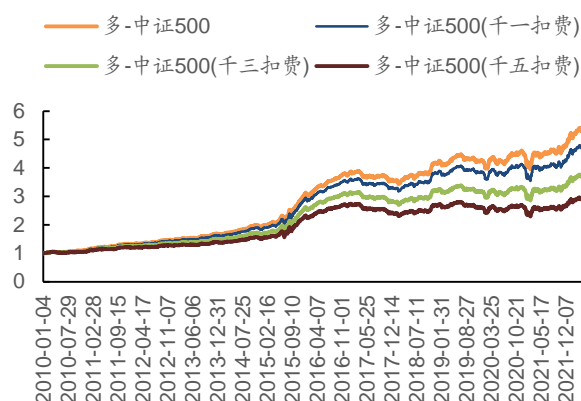
尽管上述分析表明，基于GEOGCROR因子、GEOGCRORP因子、GEOGCRORIP因子的选股策略能够在全市场表现相对较好，但是在整体与分年度换手率分析中，两类因子都呈现出高换手率的特征，这将使得策略对于手续费的设置较为敏感。因此，本节将重点分析3种因子对于手续费的敏感性。具体而言，测试了全市场选股范围内，因子在千一、千三、千五手续费费率设置下的多头基准(中证500)净值走势。

图 26: 全市场 GEOGCORR因子



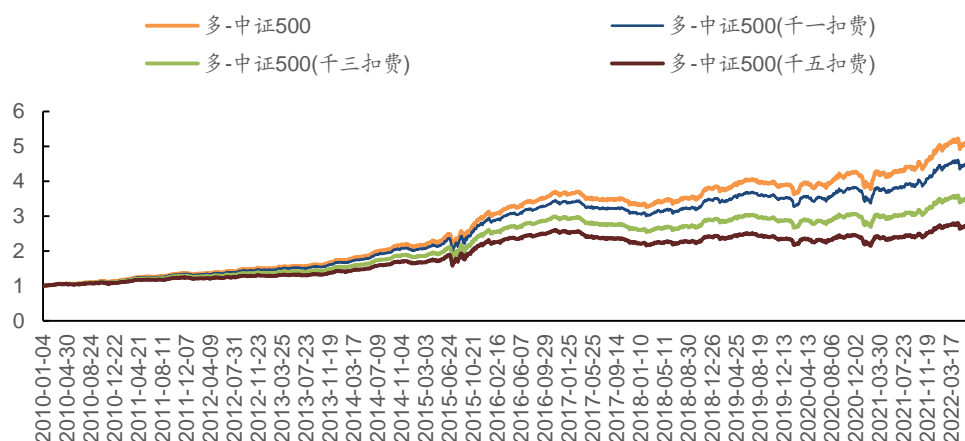
数据来源: Wind, 广发证券发展研究中心

图 27: 全市场 GEOGCORRP因子



数据来源: Wind, 广发证券发展研究中心

图 28: 全市场 GEOGCORRIP因子



数据来源: Wind, 广发证券发展研究中心

结果表明, 3种因子策略扣除千五手续费后, 仍然能够在长期获得超额收益, 但随着手续费水平上升, 策略净值表现整体下降。因此, 当使用以上地理相关系数类因子进行选股, 应当对手续费费率予以重点关注。



## 七、总结

本篇专题报告借鉴Parsons和Sabbatucci在《Geographic Lead-Lag Effects》(The Review of Financial Studies, 2018)<sup>[1]</sup>中提出的地理关联度概念，通过改进地理关联度指标，构造共6种地理相关系数因子及其优化因子(地理相关系数变动因子、地理相关系数拆解因子)，用来衡量个股与地理关联股票之间的相关程度，以期从共同基本面因素变动中获取个股反转收益。

从分档结果来看，6种地理相关系数类因子中，*GEOGCORR*因子、*GEOGCORRP*因子与*GEOGCORRIP*因子在全市场选股范围内的分档效果明显。

从整体表现来看，全市场选股范围内，月频调仓频率下，*GEOGCORR*因子、*GEOGCORRP*因子与*GEOGCORRIP*因子的IC均值在0.06以上，正IC占比均超过85%，多头相对中证500策略中，上述三种因子年化收益率均在13%以上，信息比率均超过1.5，多头换手率在80%左右。

从分年度表现来看，*GEOGCORRP*因子表现较好。因子IC均值为0.069，正IC占比90%。在多头相对中证500指数的回测中，策略整体的年化收益率为13.98%，信息比率为1.706，整体换手率保持在81.92%左右。

通过对数据预处理后的*GEOGCORR*因子、*GEOGCORRP*因子、*GEOGCORRIP*因子和BARRA因子进行相关性分析，可以发现，地理相关系数类因子能够挖掘传统因子外的增量信息。因此，可以作为新因子加入多因子模型中。

此外，本报告进一步对*GEOGCORR*因子、*GEOGCORRP*因子、*GEOGCORRIP*因子进行了选股范围与手续费率方面的敏感性测试。测试结果表明，前两种因子在中证1000、创业板选股范围内仍具有较好表现。由于因子策略多头换手率较高，地理相关系数类因子对手续费敏感。因此，当考虑利用地理相关系数类因子进行选股时，应当设定合适的选股范围并对手续费率加以考虑。

## 八、风险提示

本专题报告所述模型用量化方法通过历史数据统计、建模和测算完成，所得结论与规律在市场政策、环境变化时可能存在失效风险；

本专题策略模型在市场结构及交易行为的改变时有可能存在策略失效风险。

## 九、参考文献

- [1]. A, P.C., S. Riccardo, and T. Sheridan, Geographic Lead-Lag Effects. The Review of Financial Studies, 2020. 33(10).
- [2]. Bollerslev, T., A.J. Patton, and R. Quaedvlieg, Realized semibetas: Disentangling “good” and “bad” downside risks. Journal of Financial Economics, 2022. 144(1): p. 227-246.
- [3]. Chan, K. and A. Hameed, Stock price synchronicity and analyst coverage in emerging markets. Journal of Financial Economics, 2006. 80(1): p. 115-147.
- [4]. Cohen, L. and A. Frazzini, Economic Links and Predictable Returns. The Journal of Finance, 2008. 63(4).
- [5]. Cohen, L. and D. Lou, Complicated firms. Journal of Financial Economics, 2012. 104(2).
- [6]. Lee, C.M.C., et al., Technological links and predictable returns. Journal of Financial Economics, 2019. 132(3): p. 76-96.
- [7]. Menzly, L. and O. Ozbas, Market Segmentation and Cross-predictability of Returns. The Journal of Finance, 2010. 65(4): p. 1555-1580.
- [8]. Peng, L. and W. Xiong, Investor attention, overconfidence and category learning. Journal of Financial Economics, 2006. 80(3): p. 563-602.
- [9]. 段丙蕾, 汪荣飞, and 张然, 南橘北枳: A股市场的经济关联与股票回报. 金融研究, 2022(02): p. 171-188.
- [10]. 胡聪慧, et al., 有限注意、行业信息扩散与股票收益. 经济学(季刊), 2015. 14(03): p. 1173-1192.
- [11]. 李绪泉, 张自力, and 赵学军, 科技溢出与有限关注度对股票收益的影响研究. 价格理论与实践, 2020(06): p. 105-108.
- [12]. 刘菁哲, 股票市场领先滞后效应的研究综述. 科技和产业, 2010. 10(04): p. 93-95.
- [13]. 向诚 and 陆静, 投资者有限关注、行业信息扩散与股票定价研究. 系统工程理论与实践, 2018. 38(04): p. 817-835.

## 广发金融工程研究小组

罗 军：首席分析师，华南理工大学硕士，从业 16 年，2010 年进入广发证券发展研究中心。

安 宁：联席首席分析师，暨南大学硕士，从业 14 年，2011 年进入广发证券发展研究中心。

史 庆 盛：资深分析师，华南理工大学硕士，2011 年进入广发证券发展研究中心。

张 超：资深分析师，中山大学硕士，2012 年进入广发证券发展研究中心。

陈 原 文：资深分析师，中山大学硕士，2015 年进入广发证券发展研究中心。

樊 瑞 铎：资深分析师，南开大学硕士，2015 年进入广发证券发展研究中心。

李 豪：资深分析师，上海交通大学硕士，2016 年进入广发证券发展研究中心。

周 飞 鹏：资深分析师，伯明翰大学硕士，2021 年加入广发证券发展研究中心。

季 燕 妮：高级分析师，厦门大学硕士，2020 年进入广发证券发展研究中心。

张 钰 东：高级分析师，中山大学硕士，2020 年进入广发证券发展研究中心。

## 广发证券—行业投资评级说明

买入：预期未来 12 个月内，股价表现强于大盘 10%以上。

持有：预期未来 12 个月内，股价相对大盘的变动幅度介于-10%~+10%。

卖出：预期未来 12 个月内，股价表现弱于大盘 10%以上。

## 广发证券—公司投资评级说明

买入：预期未来 12 个月内，股价表现强于大盘 15%以上。

增持：预期未来 12 个月内，股价表现强于大盘 5%-15%。

持有：预期未来 12 个月内，股价相对大盘的变动幅度介于-5%~+5%。

卖出：预期未来 12 个月内，股价表现弱于大盘 5%以上。

## 联系我们

	广州市	深圳市	北京市	上海市	香港
地址	广州市天河区马场路 26 号广发证券大厦 35 楼	深圳市福田区益田路 6001 号太平金融大厦 31 层	北京市西城区月坛北街 2 号月坛大厦 18 层	上海市浦东新区南泉北路 429 号泰康保险大厦 37 楼	香港德辅道中 189 号李宝椿大厦 29 及 30 楼
邮政编码	510627	518026	100045	200120	-
客服邮箱	gfzqyf@gf.com.cn				

## 法律主体声明

本报告由广发证券股份有限公司或其关联机构制作，广发证券股份有限公司及其关联机构以下统称为“广发证券”。本报告的分销依据不同国家、地区的法律、法规和监管要求由广发证券于该国家或地区的具有相关合法合规经营资质的子公司/经营机构完成。

广发证券股份有限公司具备中国证监会批复的证券投资咨询业务资格，接受中国证监会监管，负责本报告于中国（港澳台地区除外）的分销。广发证券（香港）经纪有限公司具备香港证监会批复的就证券提供意见（4 号牌照）的牌照，接受香港证监会监管，负责本报告于中国香港地区的分销。

本报告署名研究人员所持中国证券业协会注册分析师资质信息和香港证监会批复的牌照信息已于署名研究人员姓名处披露。

## 重要声明

广发证券股份有限公司及其关联机构可能与本报告中提及的公司寻求或正在建立业务关系，因此，投资者应当考虑广发证券股份有限公司及其关联机构因可能存在的潜在利益冲突而对本报告的独立性产生影响。投资者不应仅依据本报告内容作出任何投资决策。投资者应自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或者口头承诺均为无效。

本报告署名研究人员、联系人（以下均简称“研究人员”）针对本报告中相关公司或证券的研究分析内容，在此声明：（1）本报告的全部分析结论、研究观点均精确反映研究人员于本报告发出当日的关于相关公司或证券的所有个人观点，并不代表广发证券的立场；（2）研究人员的部分或全部的报酬无论在过去、现在还是将来均不会与本报告所述特定分析结论、研究观点具有直接或间接的联系。

研究人员制作本报告的报酬标准依据研究质量、客户评价、工作量等多种因素确定，其影响因素亦包括广发证券的整体经营收入，该等经营收入部分来源于广发证券的投资银行类业务。

本报告仅面向经广发证券授权使用的客户/特定合作机构发送，不对外公开发布，只有接收人才可以使用，且对于接收人而言具有保密义务。广发证券并不因相关人员通过其他途径收到或阅读本报告而视其为广发证券的客户。在特定国家或地区传播或者发布本报告可能违反当地法律，广发证券并未采取任何行动以允许于该等国家或地区传播或者分销本报告。

本报告所提及证券可能不被允许在某些国家或地区内出售。请注意，投资涉及风险，证券价格可能会波动，因此投资回报可能会有所变化，过去的业绩并不保证未来的表现。本报告的内容、观点或建议并未考虑任何个别客户的具体投资目标、财务状况和特殊需求，不应被视为对特定客户关于特定证券或金融工具的投资建议。本报告发送给某客户是基于该客户被认为有能力独立评估投资风险、独立行使投资决策并独立承担相应风险。

本报告所载资料的来源及观点的出处皆被广发证券认为可靠，但广发证券不对其准确性、完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策，如有需要，应先咨询专业意见。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券的立场。广发证券的销售人员、交易员或其他专业人士可能以书面或口头形式，向其客户或自营交易部门提供与本报告观点相反的市场评论或交易策略，广发证券的自营交易部门亦可能会有与本报告观点不一致，甚至相反的投资策略。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且无需另行通告。广发证券或其证券研究报告业务的相关董事、高级职员、分析师和员工可能拥有本报告所提及证券的权益。在阅读本报告时，收件人应了解相关的权益披露（若有）。

本研究报告可能包括和/或描述/呈列期货合约价格的事实历史信息（“信息”）。请注意此信息仅供用作组成我们的研究方法/分析中的部分论点/依据/证据，以支持我们对所述相关行业/公司的观点的结论。在任何情况下，它并不（明示或暗示）与香港证监会第5类受规管活动（就期货合约提供意见）有关联或构成此活动。

## 权益披露

(1) 广发证券（香港）跟本研究报告所述公司在过去12个月内并没有任何投资银行业务的关系。

## 版权声明

未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。