

撰写日期：2021 年 12 月 13 日

证券研究报告--金融工程专题报告

基于谱聚类的 CTA 私募基金分类方法研究

金融工程专题报告

分析师：张青

执业证书编号：S0890516100001

电话：021-20321154

邮箱：zhangqing@cnhbstock.com

研究助理：程秉哲

邮箱：chengbingzhe@cnhbstock.com

销售服务电话：

021-20515355

相关研究报告

- 1 《金融工程：上证 50 的配置时机来了吗？-量化择时与资产配置月报》 2021-12-03
- 2 《金融工程：若春季攻势来临，该关注哪些板块？-量化策略跟踪评价月报》 2021-12-01
- 3 《金融工程：多因子选股：大类行业的选股因子及组合构建-金融工程专题报告》 2021-11-11

◎ 投资要点：

◆CTA 私募基金产品策略众多，业绩表现参差不齐，缺乏统一的分类方法：CTA 私募基金，即管理期货策略基金，在期货市场捕捉期货价格变动从而获利，由于期货合约多样性及多空双向开仓等特点，能够适应多种市场环境。然而由于私募基金信息披露约束性低，缺乏统一标准，第三方数据商提供的分类的参考价值也有待商榷，这对 CTA 私募基金产品的资产配置和微观遴选构成了挑战。

◆谱聚类天然契合对资产按收益率相关性划分的思想：谱聚类是从图论中演化出来的算法，后来在聚类中得到了广泛的应用。它的主要思想是把所有的数据看做空间中的点，这些点之间可以用边连接起来。距离较远的两个点之间的边权重值较低，距离较近的两个点之间的边权重值较高，通过对所有数据点组成的图进行切图，让切图后不同的子图间边权重和尽可能低，而子图内的边权重和尽可能高，从而达到聚类的目的。把不同的资产当做空间中的点，各点之间的边权重值就是收益率的相关性，通过谱聚类可以把资产分成彼此收益率弱相关、内部收益率强相关的集合，从而为风险分散化的资产配置打通了道路。

◆本文通过谱聚类方法构造了表现优良的 CTA 私募基金分类标准：本文所采用的 CTA 私募基金净值数据没有固定披露日期，通过筛选出考察时间段信息密度较高的产品并计算固定时点的平均净值，本文构造了不同产品之间可比的收益率。按照时间衰减滚动叠加产品之间的收益率相关性矩阵作为邻接矩阵，即可对 CTA 私募基金进行谱聚类分析。本文采取截面虚拟变量回归和轮廓系数两种方式对谱聚类结果进行评估，并且通过合成产品净值，发现谱聚类在解释力度和收益表现上优于市场第三方提供的分类标准，对于 CTA 私募基金的分散化配置和分类遴选具有一定指导意义。

◆风险提示：本报告主要采用数量化研究方法，可能存在模型设定偏差。

内容目录

1. CTA 私募基金介绍.....	4
1.1. 配置意义.....	4
1.2. 数据描述.....	4
2. 谱聚类介绍.....	5
2.1. 策略理论说明.....	5
2.2. 算法流程.....	7
3. 谱聚类 CTA 私募基金分类策略.....	7
3.1. 数据清洗.....	7
3.2. 评估方法.....	8
3.3. 策略回测.....	8
4. 结论.....	14

图表目录

图 1: CTA 私募基金产品第三方分类统计.....	5
图 2: CTA 私募基金产品每日净值披露数量.....	5
图 3: 图论切图示例.....	6
图 4: 标准净值计算方式示意图.....	8
图 5: 衰减速度示意图.....	9
图 6: 模型 $q = 5, p = 12$ 谱聚类子类净值.....	11
图 7: 三方分类子类净值.....	11
图 8: 模型 $q = 5, p = 12$ 等权合成净值.....	11
图 9: 模型 $q = 5, p = 12$ 波动率倒数合成净值.....	11
图 10: 模型 $q = 5, p = 10$ 等权合成净值.....	12
图 11: 模型 $q = 5, p = 10$ 波动率倒数合成净值.....	12
图 12: 模型 $q = 5, p = 8$ 等权合成净值.....	12
图 13: 模型 $q = 5, p = 8$ 波动率倒数合成净值.....	12
图 14: 模型 $q = 4, p = 12$ 等权合成净值.....	12
图 15: 模型 $q = 4, p = 12$ 波动率倒数合成净值.....	12
图 16: 模型 $q = 4, p = 10$ 等权合成净值.....	13
图 17: 模型 $q = 4, p = 10$ 波动率倒数合成净值.....	13
图 18: 模型 $q = 4, p = 8$ 等权合成净值.....	13
图 19: 模型 $q = 4, p = 8$ 波动率倒数合成净值.....	13
图 20: 模型 $q = 3, p = 12$ 等权合成净值.....	13
图 21: 模型 $q = 3, p = 12$ 波动率倒数合成净值.....	13
图 22: 模型 $q = 3, p = 10$ 等权合成净值.....	14
图 23: 模型 $q = 3, p = 10$ 波动率倒数合成净值.....	14
图 24: 模型 $q = 3, p = 8$ 等权合成净值.....	14
图 25: 模型 $q = 3, p = 8$ 波动率倒数合成净值.....	14

表 1: 9 种参数的谱聚类模型回测效果比较.....	9
-----------------------------	---

表 2: 最优模型 $q = 5, p = 12$ 回测结果.....	9
-------------------------------------	---

表 3: 模型 $q = 5, p = 12$ 聚类结果与第三方分类结果交叉表.....	10
--	----

表 4: 模型 $q = 5, p = 12$ 绩效统计	11
------------------------------------	----

CTA,即管理期货策略基金,在期货市场捕捉期货价格变动从而获利,由于期货合约多样性及多空双向开仓等特点,能够适应多种市场环境。CTA 策略在多资产投资中具有重要意义,因为 CTA 与传统股票、债券资产的相关性低,具有危机中的 alpha 属性。加入 CTA 可以有效改善投资组合的有效边界,丰富投资者达到预期目标的可选路径和方案,优化组合的收益风险比。

CTA 策略配置最终要落地到具体的标的上,也即需要经历 CTA 私募基金的遴选步骤,而遴选的核心是同类产品的比较,因此遴选策略是否有效,很大程度取决于对 CTA 的分类是否合理有效。因为 CTA 策略在运作中,有些以趋势策略为主,有些以套利策略为主,且操作频率不同产品也大相径庭,而这又导致了虽然同属于 CTA 大类,但收益风险特征迥异,倘若只是笼统的在 CTA 大类中进行产品比较,显然是有失偏颇的,且从私募 FOF 的自上而下配置角度看,不对 CTA 进行更细致的分类,也不利于在 CTA 产品内部做更加细致的配置,无法确保底层资产的低相关性,从而降低了组合投资的功效。

可见,对 CTA 进而更细致的分类意义显著。然而,由于私募基金信息披露约束性低,并没有一个官方的统一分类标准,而第三方数据商提供的分类的参考价值也有待商榷。鉴于此,本文尝试采用数据挖掘的方法,基于聚类思想,从数量统计角度研究如何对 CTA 策略进行分类,以此作为后续我们展开对 CTA 细分策略配置和遴选研究的一项基础。

1. CTA 私募基金介绍

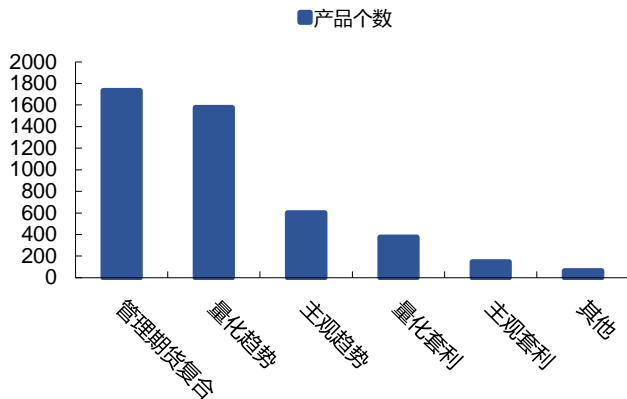
1.1. 配置意义

CTA (Commodity Trading Advisors) 即管理期货基金,从投资标的看,CTA 基金主要交易金融期货合约和商品期货合约,因为是在期货市场上交易,所以 CTA 基金可以选择做多和做空两个方向。CTA 涉及的股指期货、国债期货、商品期货等标的与其他资产或策略类型具备较低的相关性,同时期货市场的多空双向交易特征,赋予了 CTA 策略在市场上涨抑或是下跌行情中均能博取收益,进而能够适应多种市场环境,穿越牛熊。CTA 策略的加入可以改善投资组合的有效边界,丰富投资者达到预期目标的可选路径和方案,使投资者实现在相同风险条件下,提升组合的潜在获利能力,或者在相同收益条件下,降低组合的潜在风险水平。

1.2. 数据描述

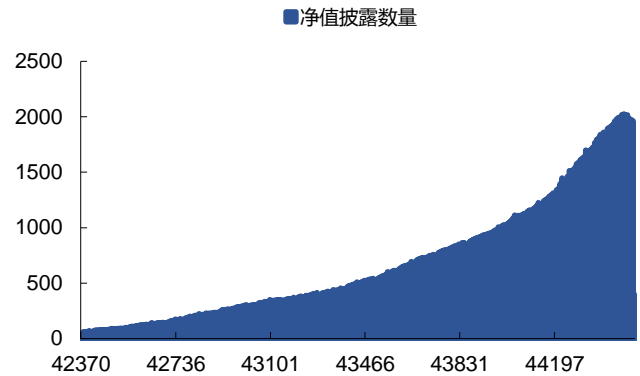
选取 4491 个 CTA 私募基金从 2016 年 1 月 1 日至 2021 年 11 月 5 日的净值数据,每日披露的净值数量逐渐增多,从第三方提供的分类标准来看产品子策略主要是管理期货复合和量化趋势。

图 1: CTA 私募基金产品第三方分类统计



资料来源：华宝证券研究创新部

图 2: CTA 私募基金产品每日净值披露数量



资料来源：华宝证券研究创新部

2. 谱聚类介绍

2.1. 策略理论说明

对于无向图 G ，我们一般用点的集合 V 和边的集合 E 来描述。对于 V 中的任意两个点，定义权重 w_{ij} 为点 v_i 和点 v_j 之间的边权重。对于图中的任意一个点 v_i ，它的度 d_i 定义为和它相连的所有边的权重之和，即 $d_i = \sum_{j=1}^n w_{ij}$ 。利用每个点度的定义，我们可以得到一个 $n \times n$ 的度矩阵 D ，

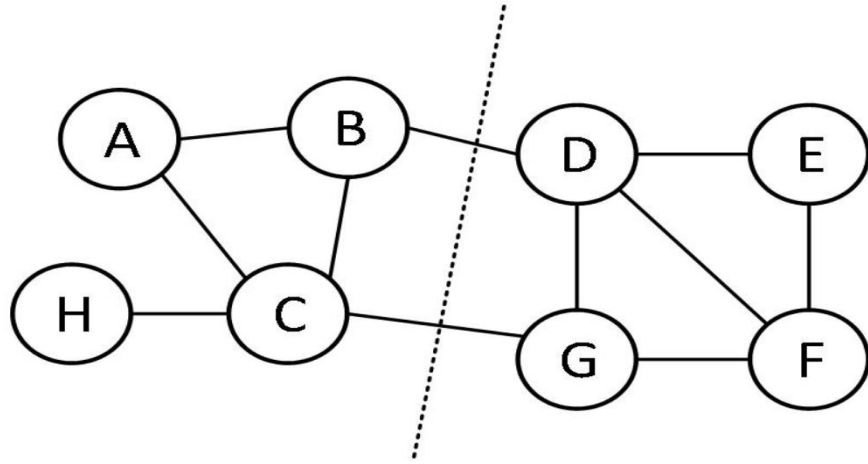
$$D = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{bmatrix}$$

利用所有点之间的边权重值，可以得到图的邻接矩阵 W ，这也是一个 $n \times n$ 的矩阵，第 i 行的第 j 个值对应权重 w_{ij} 。拉普拉斯矩阵定义为 $L = D - W$ ，对于任意向量 f ，有

$$\begin{aligned} f^T L f &= f^T D f - f^T W f = \sum_{i=1}^n d_i f_i^2 - \sum_{i,j=1}^n w_{ij} f_i f_j \\ &= \frac{1}{2} \left(\sum_{i=1}^n d_i f_i^2 - 2 \sum_{i,j=1}^n w_{ij} f_i f_j + \sum_{j=1}^n d_j f_j^2 \right) = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2 \end{aligned}$$

对于无向图的切图，我们的目标是将图 G 切成相互没有连接的 k 个子图： $A_1, A_2 \dots A_k$ ，它们互不相交且并集为 V 。切图有 $Ncut$ 切图和 $RatioCut$ 切图两种方式，这里主要介绍 $Ncut$ 切图。

图 3：图论切图示例



资料来源：华宝证券研究创新部

定义 $vol(A) = \sum_{i \in A} d_i$, $W(A, B) = \sum_{i \in A, j \in B} w_{ij}$, 则 $NCut$ 切图为

$$NCut(A_1, A_2, \dots, A_k) = \frac{1}{2} \sum_{i=1}^k \frac{W(A_i, \bar{A}_i)}{vol(A_i)}$$

引入指示向量

$$h_{ij} = \begin{cases} 0 & v_i \notin A_j \\ \frac{1}{\sqrt{vol(A_j)}} & v_i \in A_j \end{cases}$$

那么有

$$\begin{aligned} h_i^T L h_i &= \frac{1}{2} \sum_{m=1} \sum_{n=1} w_{mn} (h_{im} - h_{in})^2 \\ &= \frac{1}{2} \left(\sum_{m \in A_i, n \notin A_i} w_{mn} \left(\frac{1}{\sqrt{vol(A_i)}} - 0 \right)^2 + \sum_{m \notin A_i, n \in A_i} w_{mn} \left(0 - \frac{1}{\sqrt{vol(A_i)}} \right)^2 \right) \\ &= \frac{1}{2} \left(\sum_{m \in A_i, n \notin A_i} w_{mn} \frac{1}{vol(A_i)} + \sum_{m \notin A_i, n \in A_i} w_{mn} \frac{1}{vol(A_i)} \right) \\ &= \frac{1}{2} \left(cut(A_i, \bar{A}_i) \frac{1}{vol(A_i)} + cut(\bar{A}_i, A_i) \frac{1}{vol(A_i)} \right) \\ &= \frac{cut(A_i, \bar{A}_i)}{vol(A_i)} \end{aligned}$$

对应的 $NCut$ 切图函数表达式为

$$NCut(A_1, A_2, \dots, A_k) = \sum_{i=1}^k h_i^T L h_i = \sum_{i=1}^k (H^T L H)_{ii} = tr(H^T L H)$$

最终的优化目标变为

$$\underset{H}{\operatorname{argmin}} \operatorname{tr}(H^T L H) \quad \text{s.t. } H^T D H = I$$

令 $H = D^{-1/2} F$ ，则有 $H^T D H = F^T F = I$ ，优化目标变为

$$\underset{F}{\operatorname{argmin}} \operatorname{tr}(F^T D^{-1/2} L D^{-1/2} F) \quad \text{s.t. } F^T F = I$$

观察到 F 的子向量 f_i 为单位正交基， $D^{-1/2} L D^{-1/2}$ 为对称矩阵，若 f_i 是 $D^{-1/2} L D^{-1/2}$ 的特征向量，此时 $f_i^T D^{-1/2} L D^{-1/2} f_i$ 的最大值为 $D^{-1/2} L D^{-1/2}$ 的最大特征值，最小值为 $D^{-1/2} L D^{-1/2}$ 的最小特征值。优化目标是找到 k 个最小单位正交基，通过找到最小的 k 个特征值对应的特征向量，可以近似解决优化问题。得到 k 个特征向量组成的矩阵 H 后对矩阵 H 进行行标准化，由于我们在使用维度规约的时候损失了少量信息，导致优化后的指示向量 f_i 对应的 H 现在不能完全指示各样本的归属，因此一般在得到 $n \times k$ 维度的矩阵 H 后还需要对每一行进行一次传统的聚类，比如使用 $K - Means$ 聚类。

2.2. 算法流程

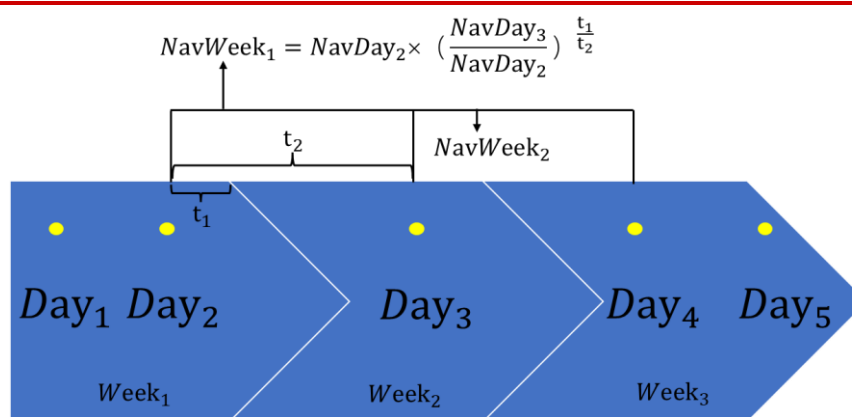
- 1) 根据集合 V 中点关系构建邻接矩阵 W ，构建度矩阵 D ，计算出拉普拉斯矩阵 L
- 2) 构建标准化后的拉普拉斯矩阵 $D^{-1/2} L D^{-1/2}$
- 3) 计算 $D^{-1/2} L D^{-1/2}$ 最小的 k 个特征值所各自对应的特征向量 f
- 4) 将各自对应的特征向量 f 组成的矩阵按行标准化，最终组成 $n \times k$ 维的特征矩阵 H
- 5) 把 H 中的每一行作为一个样本，共 n 个样本，用输入的聚类方法如 $K - Means$ 进行聚类。
- 6) 得到簇划分 $C(c_1, c_2, \dots, c_k)$ 。

3. 谱聚类 CTA 私募基金分类策略

3.1. 数据清洗

本文使用周度收益率进行计算，把某一时点过去 48 周的数据作为训练集进行谱聚类，未来 12 周数据作为测试集观察样本外分类效果，并以 12 周为单位向前滚动。由于各 CTA 私募基金产品没有固定的净值更新规则，在每时点选择过去 48 周内每周空缺值不超过 5 个的产品作为待聚类样本，然后计算产品每周最后一次更新的净值与下周第一次更新的净值在本周末的平均数作为标准净值，当有一数据为空时填充 Nan，最后用前一个非缺失值填充并计算收益率。

图 4：标准净值计算方式示意图



资料来源：华宝证券研究创新部

每次训练加测试的数据时长共 60 周，计算标准净值时去掉了最后一周，计算收益率时去掉第一周，为避免未来数据去掉了训练集和测试集交界的一周，最终周度收益率训练集 46 周，测试集 11 周。

3.2. 评估方法

采用轮廓系数和 Brown and Goetzmann (1997)¹ 中对分类结果作为虚拟变量进行截面回归的方式来评估聚类结果。 a 为某个样本与其所在簇内其他样本的平均距离， b 为某个样本与其他簇样本的平均距离，则针对某个样本的轮廓系数 s 为 $(b - a) / \max(a, b)$ ，聚类总的轮廓系数 sc 为 $1/N \sum_{i=1}^N s_i$ 。轮廓系数取值范围为 $[-1, 1]$ ，取值越接近 1 则说明聚类性能越好，相反取值越接近 -1 则说明聚类性能越差。

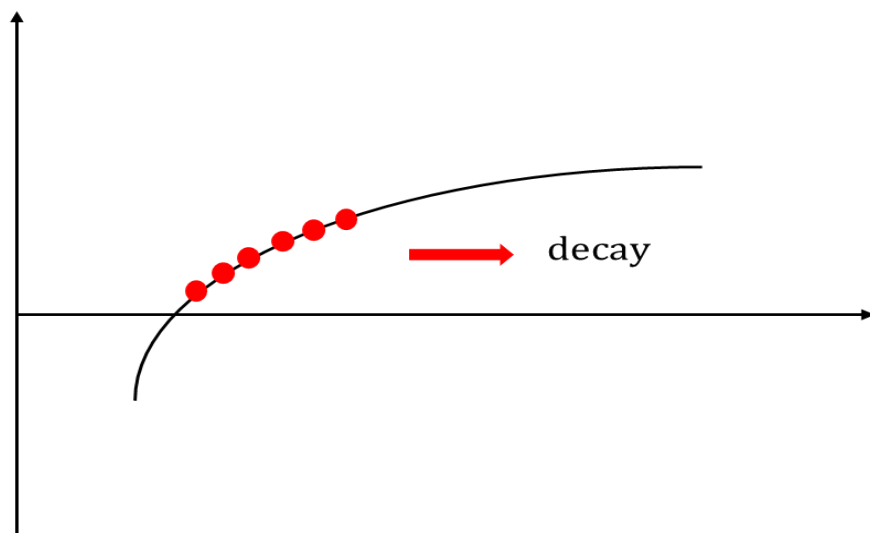
虚拟变量回归中根据每个截面 adj_R^2 的大小判断分类结果对 CTA 私募基金收益率的解释力度，并对多期 adj_R^2 进行检验查看是否显著不为 0。除谱聚类的分类结果外，本文还对第三方分类标准进行了相同处理，作为谱聚类的效果参考。

3.3. 策略回测

在训练区间内滚动叠加不同 CTA 私募基金的相关系数矩阵再取指数作为谱聚类所需的邻接矩阵，计算相关系数矩阵的时间长度为 p 周，滚动时间单位为一周，叠加矩阵的衰减速度为 $\log(1 + decay + 0.01 \times (i + 1))$, $i = 0, 1, 2 \dots$ 。通过对数函数上取一系列等间隔的点，保证了衰减速度随时间间隔增大越来越快， $decay \geq 0$ 代表衰减系数，可以控制这一系列点在对数函数上的初始位置。

¹ Brown, Stephen J., and William N. Goetzmann. "Mutual fund styles." *Journal of financial Economics* 43.3 (1997): 373-399.

图 5：衰减速度示意图



资料来源：华宝证券研究创新部

控制聚类分类数量为 $q = 3, 4, 5$ ，相关系数矩阵时长为 $p = 8, 10, 12$ ，测试共 9 种模型的效果，结果如下，可见聚类效果随分类数量增加显著提升，相关系数矩阵时长对聚类效果影响不明显。总体而言谱聚类分类效果显著优于第三方数据商提供的分类标准。

表 1：9 种参数的谱聚类模型回测效果比较

模型参数	谱聚类训练集-adjR2	谱聚类测试集-adjR2	三方分类测试集-adjR2
q=3,p=8	0.0664	0.0286	0.0123
q=3,p=10	0.0612	0.0295	0.0123
q=3,p=12	0.0562	0.0285	0.0123
q=4,p=8	0.0839	0.0373	0.0123
q=4,p=10	0.0765	0.0369	0.0123
q=4,p=12	0.0707	0.0394	0.0123
q=5,p=8	0.0914	0.0440	0.0123
q=5,p=10	0.0858	0.0454	0.0123
q=5,p=12	0.0828	0.0465	0.0123

资料来源：华宝证券研究创新部

最优模型 $q = 5, p = 12$ 具体的回测结果如下，可以看出模型的轮廓系数随着时间逐渐减小，但是测试集-adjR2 并没有显著变化。随着时间推移数据的总体质量越来越高，符合筛选条件的待聚类 CTA 产品也越来越多，用有限的特征去拟合越来越多的样本，轮廓系数的下降符合预期，但是测试集-adjR2 几乎不变说明谱聚类分类效果稳健，具有一定的样本外预测能力。

表 2：最优模型 $q = 5, p = 12$ 回测结果

date	轮廓系数	谱聚类训练集-adjR2	T-test	谱聚类测试集-adjR2	T-test	三方分类测试集-adjR2	T-test
2016/11/28-2017/02/19	0.468	0.053	0.037	0.043	0.282	(0.007)	0.699
2017/02/20-2017/05/14	0.417	0.077	0.017	0.035	0.119	(0.024)	0.009
2017/05/15-2017/08/06	0.337	0.057	0.015	0.059	0.002	(0.006)	0.402
2017/08/07-2017/10/29	0.321	0.069	0.003	0.052	0.006	(0.001)	0.820
2017/10/30-2018/01/21	0.307	0.100	0.003	0.055	0.006	0.021	0.122
2018/01/22-2018/04/15	0.281	0.085	0.000	0.026	0.026	0.023	0.021
2018/04/16-2018/07/08	0.306	0.078	0.006	0.042	0.007	0.020	0.021
2018/07/09-2018/09/30	0.268	0.060	0.016	0.049	0.007	0.024	0.009

2018/10/01-2018/12/23	0.290	0.082	0.004	0.041	0.002	0.013	0.071
2018/12/24-2019/03/17	0.308	0.101	0.003	0.030	0.003	0.014	0.114
2019/03/18-2019/06/09	0.317	0.084	0.000	0.041	0.001	0.017	0.027
2019/06/10-2019/09/01	0.294	0.065	0.000	0.035	0.000	0.012	0.035
2019/09/02-2019/11/24	0.324	0.070	0.000	0.058	0.000	0.012	0.000
2019/11/25-2020/02/16	0.277	0.103	0.002	0.053	0.000	0.013	0.026
2020/02/17-2020/05/10	0.270	0.068	0.003	0.033	0.006	0.014	0.103
2020/05/11-2020/08/02	0.256	0.111	0.003	0.060	0.000	0.011	0.008
2020/08/03-2020/10/25	0.269	0.110	0.000	0.040	0.001	0.027	0.001
2020/10/26-2021/01/17	0.267	0.076	0.000	0.055	0.003	0.023	0.001
2021/01/18-2021/04/11	0.258	0.104	0.000	0.059	0.000	0.013	0.001
2021/04/12-2021/07/04	0.269	0.102	0.000	0.060	0.001	0.018	0.005
2021/07/05-2021/09/26	0.281	0.087	0.000	0.052	0.003	0.021	0.008

资料来源：华宝证券研究创新部

以最优模型 $q = 5, p = 12$ 在2021/07/05 – 2021/09/26选择的 CTA 私募基金产品为例，观察谱聚类分类与第三方分类的关系。可以看出谱聚类分类 1 和分类 4 与图 1 中第三方分类的分布有显著不同，分类 1 对量化趋势产品有很强的捕捉能力，分类 4 对主观趋势、量化套利、主观套利等第三方分类中占比很少的产品有额外的捕捉能力。

表 3：模型 $q = 5, p = 12$ 聚类结果与第三方分类结果交叉表

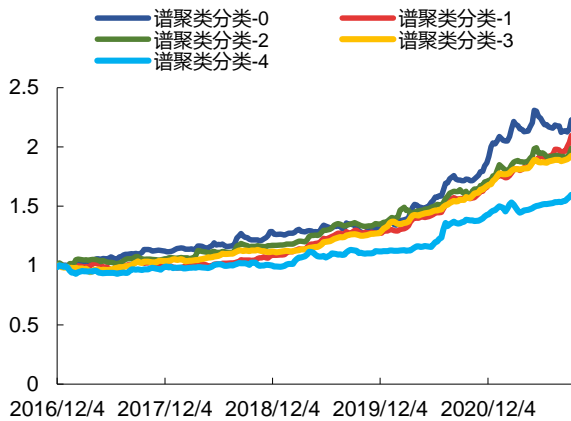
谱聚类	管理期货复合	主观趋势	量化趋势	主观套利	量化套利
0	82	11	164	0	3
1	83	25	61	11	36
2	96	21	72	10	7
3	131	15	59	2	6
4	67	34	23	33	29

资料来源：华宝证券研究创新部

注：剔除了在三方分类中属于“未知”的产品，下同

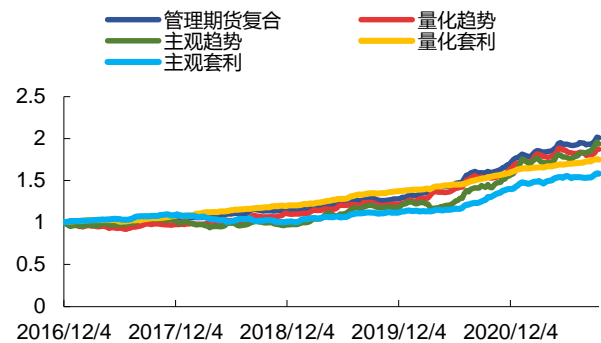
不同期谱聚类实验的标签虽然数值相同但是含义不同，本文尝试根据谱聚类结果与第三方分类的关系将不同期的聚类结果联系起来。将 $t(t = 1, 2, 3 \dots T)$ 期谱聚类 $i(i = 1, 2, 3 \dots I)$ 分类 SC_{ti} 下所有产品管理期货复合、主观趋势、量化趋势等的占比减去图 1 中第三方分类的分布比例，得到该谱聚类分类 SC_{ti} 的特征，然后对共 $T \times I$ 个样本计算相关系数再进行一次谱聚类。为了尽可能使同一期 $SC_{ti}(t = \hat{t}, i = 1, 2, 3 \dots I)$ 聚类后处于不同的分类中，改写邻接矩阵使得彼此的边权重 $w_{ij} = -T \times I$ ，这里使用了负的权重值因此不能使用 $Ncut$ 切图改用 $RatioCut$ ， $RatioCut$ 切图将 $Ncut$ 指示向量中的 $vol(A_j)$ 改为 $|A_j|$ 并且不需要计算 $D^{-1/2}$ ， $|A_j|$ 表示集合 A 中点的个数。设聚类之后某一期内重复的分类个数为 $sd_t(sd_t = 0, 1, 2 \dots I - 1)$ ，计算 $SD = (1/T) \sum_{t=1}^T sd_t$ 评估将同一期 SC_{ti} 划分到不同分类的效果，模型 $q = 5, p = 12$ 二次聚类的 SD 值为 0.215 左右，作为对比计算得到随机分类的 SD 值为 0.32 左右。对谱聚类之后仍重叠的时段予以二次调整，确保同一期 $SC_{ti}(t = \hat{t}, i = 1, 2, 3 \dots I)$ 聚类后处于不同的分类中。以模型 $q = 5, p = 12$ 为例，分类后的子类净值走势如下，可以发现谱聚类的子类 0 和 4 差异较大，较三方分类的子类区分度明显。

图 6：模型 $q = 5, p = 12$ 谱聚类子类净值



资料来源：华宝证券研究创新部

图 7：三方分类子类净值



资料来源：华宝证券研究创新部

出于分散投资的目的，并不需要了解不同期聚类结果之间的衔接关系，只需将每一期的分类结果用一定方式组合。对 $q = 3, 4, 5$ $p = 8, 10, 12$ 共 9 种模型的子类净值用等权法和波动率倒数权重两种方法进一步合成，其中波动率从训练期数据得到。

以模型 $q = 5, p = 12$ 为例统计策略绩效，结果发现聚类合成净值在收益率上表现优于三方合成净值，但是考虑风险后在波动率、最大回撤等指标上表现较差。画出 9 种模型的合成净值曲线进行直观比较，发现各模型中谱聚类的期末净值均高于第三方，尤其是 2020 年下半年开始谱聚类合成净值逐渐扩大与三方合成净值的差距，说明本文的谱聚类方法对 CTA 私募基金配置具有一定指导意义。

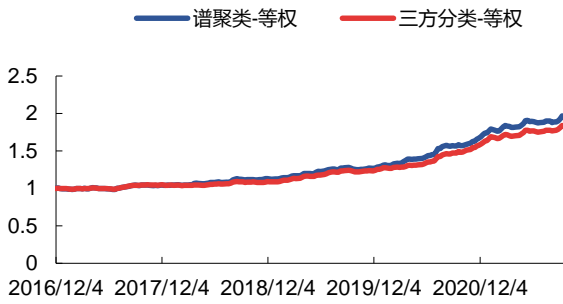
表 4：模型 $q = 5, p = 12$ 绩效统计

统计指标	谱聚类-等权	三方分类-等权	谱聚类-波动率倒数加权	三方分类-波动率倒数加权
累计收益率	95.53%	83.16%	93.89%	81.95%
最大回撤	-2.42%	-1.70%	-2.43%	-1.48%
年化收益率	14.82%	13.29%	14.62%	13.13%
年化波动率	4.92%	4.15%	4.33%	3.30%
滚动 1 年平均最大回撤	-1.79%	-1.59%	-1.33%	-1.07%
滚动 1 年平均波动率	4.66%	3.84%	4.12%	3.13%
Calmar 比率	8.28	8.36	10.99	12.25
滚动 1 年平均收益率	16.45%	14.53%	16.11%	13.87%
滚动 1 年最大收益率	38.11%	34.92%	35.52%	29.09%
滚动 1 年最小收益率	3.50%	3.28%	3.93%	5.15%
滚动 1 年胜率	100.00%	100.00%	100.00%	100.00%
夏普比率	2.71	2.84	3.03	3.52

资料来源：华宝证券研究创新部

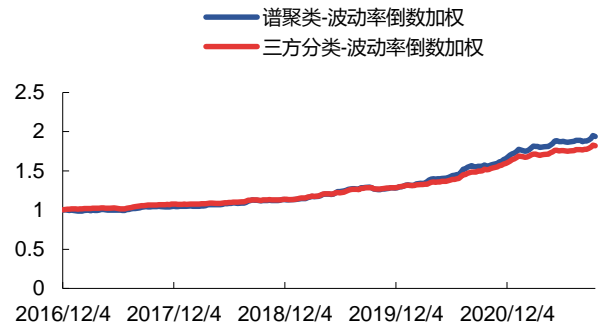
图 8：模型 $q = 5, p = 12$ 等权合成净值

图 9：模型 $q = 5, p = 12$ 波动率倒数合成净值



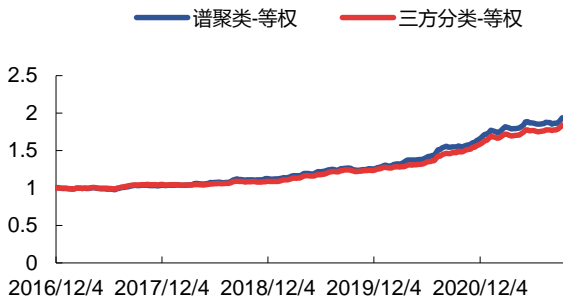
资料来源：华宝证券研究创新部

图 10：模型 $q = 5, p = 10$ 等权合成净值



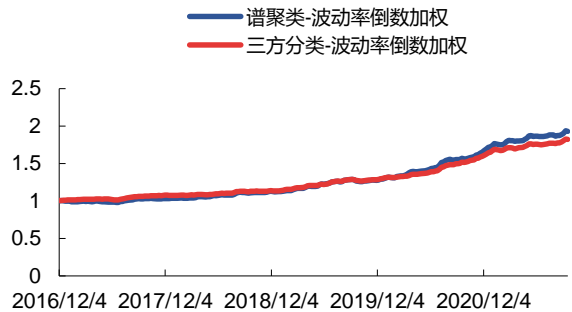
资料来源：华宝证券研究创新部

图 11：模型 $q = 5, p = 10$ 波动率倒数合成净值



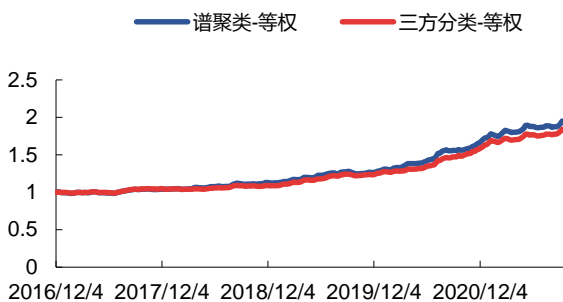
资料来源：华宝证券研究创新部

图 12：模型 $q = 5, p = 8$ 等权合成净值



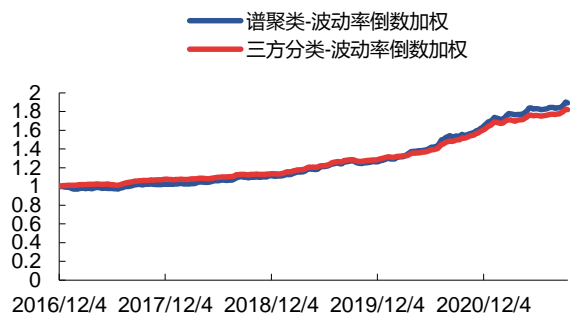
资料来源：华宝证券研究创新部

图 13：模型 $q = 5, p = 8$ 波动率倒数合成净值



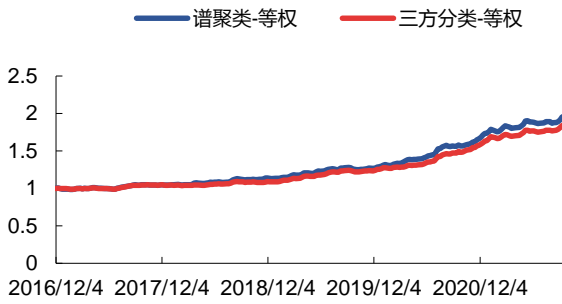
资料来源：华宝证券研究创新部

图 14：模型 $q = 4, p = 12$ 等权合成净值



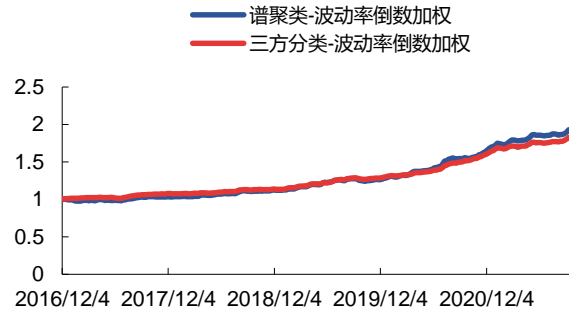
资料来源：华宝证券研究创新部

图 15：模型 $q = 4, p = 12$ 波动率倒数合成净值



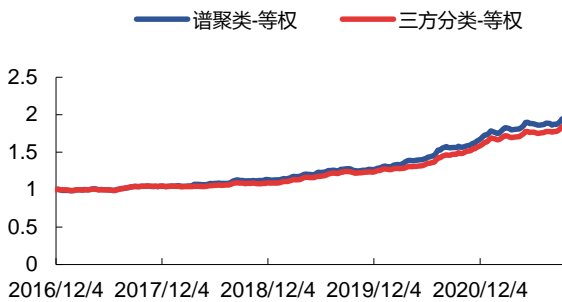
资料来源：华宝证券研究创新部

图 16：模型 $q = 4, p = 10$ 等权合成净值



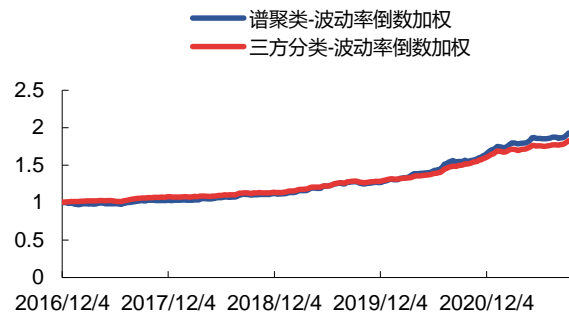
资料来源：华宝证券研究创新部

图 17：模型 $q = 4, p = 10$ 波动率倒数合成净值



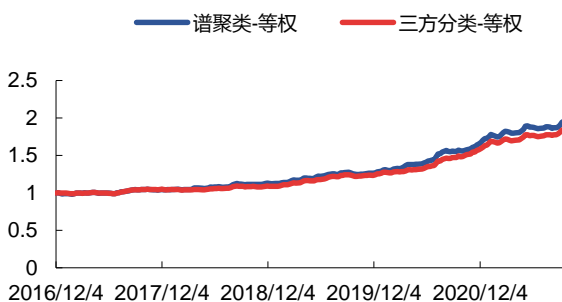
资料来源：华宝证券研究创新部

图 18：模型 $q = 4, p = 8$ 等权合成净值



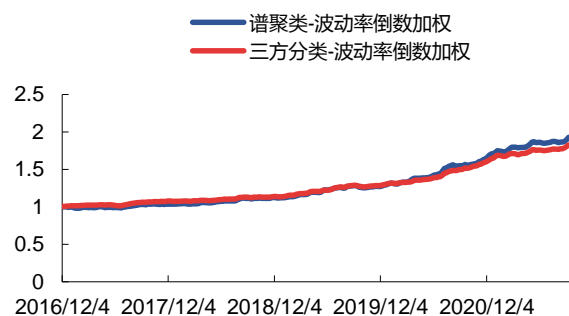
资料来源：华宝证券研究创新部

图 19：模型 $q = 4, p = 8$ 波动率倒数合成净值



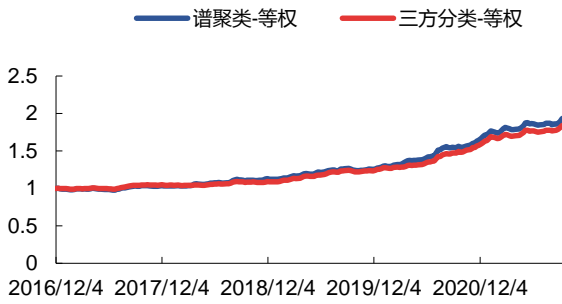
资料来源：华宝证券研究创新部

图 20：模型 $q = 3, p = 12$ 等权合成净值



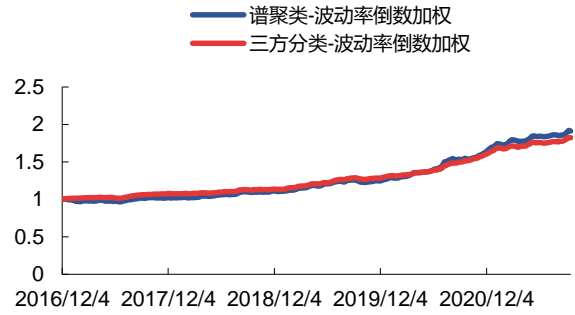
资料来源：华宝证券研究创新部

图 21：模型 $q = 3, p = 12$ 波动率倒数合成净值



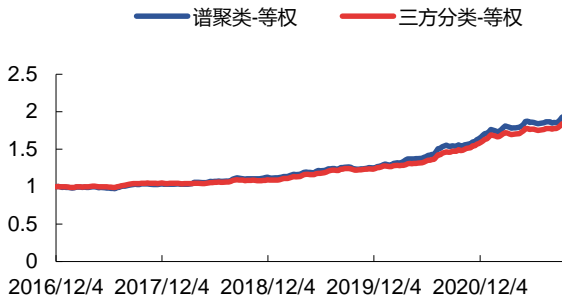
资料来源：华宝证券研究创新部

图 22：模型 $q = 3, p = 10$ 等权合成净值



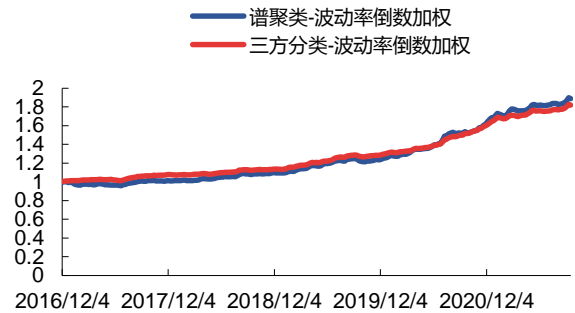
资料来源：华宝证券研究创新部

图 23：模型 $q = 3, p = 10$ 波动率倒数合成净值



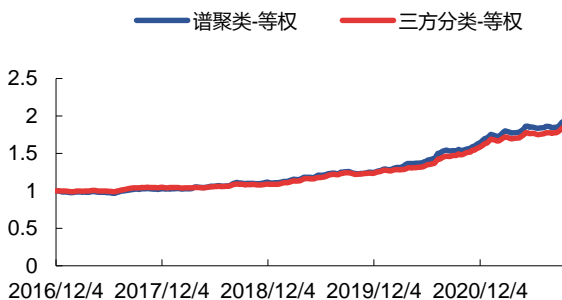
资料来源：华宝证券研究创新部

图 24：模型 $q = 3, p = 8$ 等权合成净值

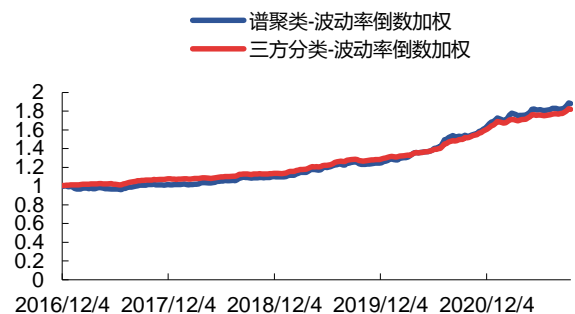


资料来源：华宝证券研究创新部

图 25：模型 $q = 3, p = 8$ 波动率倒数合成净值



资料来源：华宝证券研究创新部



资料来源：华宝证券研究创新部

4. 结论

本文从收益率相关系数的角度对 CTA 私募基金净值数据构建了谱聚类方法，通过截面虚拟变量回归验证了此方法得到的分类对收益率的解释力度强于市场上第三方提供的分类标准，

并且发现根据此方法合成的产品净值在收益表现上优于根据三方分类合成的产品净值，因此可以为 CTA 私募基金的划分及相应的分散化配置提供一定指导。由于净值数据不规则，虽然通过数据清洗得到了可比的收益率，但是不免破坏和损失了一些信息。当数据质量更高时，可以期待文中基于谱聚类的 CTA 私募基金分类方法会有更好的效果。

(感谢实习生张帅对本报告的贡献)

风险提示及免责声明

- ★ 华宝证券股份有限公司具有证券投资咨询业务资格。
- ★ 市场有风险，投资须谨慎。
- ★ 本报告所载的信息均来源于已公开信息，但本公司对这些信息的准确性及完整性不作任何保证。
- ★ 本报告所载的任何建议、意见及推测仅反映本公司于本报告发布当日的独立判断。本公司不保证本报告所载的信息于本报告发布后不会发生任何更新，也不保证本公司做出的任何建议、意见及推测不会发生变化。
- ★ 在任何情况下，本报告所载的信息或所做出的任何建议、意见及推测并不构成所述证券买卖的出价或询价，也不构成对所述金融产品、产品发行或管理人作出任何形式的保证。在任何情况下，本公司不就本报告中的任何内容对任何投资做出任何形式的承诺或担保。投资者应自行决策，自担投资风险。
- ★ 本公司秉承公平原则对待投资者，但不排除本报告被他人非法转载、不当宣传、片面解读的可能，请投资者审慎识别、谨防上当受骗。
- ★ 本报告版权归本公司所有。未经本公司事先书面授权，任何组织或个人不得对本报告进行任何形式的发布、转载、复制。如合法引用、刊发，须注明本公司出处，且不得对本报告进行有悖原意的删节和修改。
- ★ 本报告对基金产品的研究分析不应被视为对所述基金产品的评价结果，本报告对所述基金产品的客观数据展示不应被视为对其排名打分的依据。任何个人或机构不得将我方基金产品研究成果作为基金产品评价结果予以公开宣传或不当引用。

适当性申明

- ★ 根据证券投资者适当性管理有关法规，该研究报告仅适合专业机构投资者及与我司签订咨询服务协议的普通投资者，若您为非专业投资者及未与我司签订咨询服务协议的投资者，请勿阅读、转载本报告。