

标准研究

2024 年第 5 期（总第 37 期）

证标委秘书处

2024 年 8 月 8 日

证券公司指标命名标准化管理策略

【摘要】证券公司在指标应用实践中会遇到业务术语使用不统一、关联关系不清晰、同名不同义和同义不同名等问题。指标命名标准化管理为指标命名制定了统一的业务术语、维度项和维度顺序，可以形成风格一致、术语规范、结构清晰和相互可比的指标集。同时，指标命名标准化管理也有局限性，指标名称仅能近似表达指标业务含义，需要结合指标需求扎口管理机制和一致性检查机制，通过事前约束和事后检查进一步规避、排查和修正指标问题。

关键词：指标命名 标准化 维度 相似度

一、证券公司指标管理痛点

（一）业务术语使用不统一

业务术语的用法不统一，当使用业务术语时有时会使用全称，有时会使用简称，甚至简称的简写方式也不统一。业务术语在指标名称中的位置不统一，对于相同业务含义的指标，即便指标名称中所使用的业务术语完全一致，当业务术语的相互位置变动时也很容易被错认为不同的指标。

（二）指标关联关系不清晰

指标之间的关联关系有“衍生关系”和“组合关系”两种，“衍生关系”指的是下游指标只有一个上游指标，下游指标由上游指标限制一个或多个统计维度生成；“组合关系”指的是下游指标有多个上游指标，下游指标由上游指标运算生成。当指标命名较为随意时，从指标名称中不能直观体现指标之间的关联关系，经过一段时间的人事更迭后，指标关联关系大概率仅能从文档或代码中获取，不利于知识传递和指标使用。

（三）指标名称和统计口径不一一对应

当指标名称命名不规范时，业务交叉单位很容易分别提出统计口径一致但指标名称不一致的指标需求。在指标开发缺乏统一管理的情况下，上述指标需求很可能会交由不同的技术团队实现，从而无法及时发现指标“同义不同名”问题。同样的，由于命名粗放导致维度信息体现不完整，不同业务

单位提交的指标开发需求中会出现指标名称一致但统计口径不一致的情况，若无法及时发现并修正，上线后会出现“同名不同义”问题。

指标名称是指标应用的抓手，为一定程度上解决上述问题，可以考虑从指标命名标准化入手，形成风格一致、术语规范、结构清晰的指标命名方法。

二、指标命名管理原则

（一）术语规范

指标名称中要使用规范的业务术语，需要注意的是，术语规范重在内部统一，包括中英文名称、业务含义达成一致。在制定业务术语规范的过程中以内部业务使用情况为主，以外部行业或教材通用术语为辅，尽量保障业务应用连贯性，降低改造成本。

（二）信息充分

指标名称中要体现主要的统计主体信息、统计维度信息和计算方式信息等。主体信息、维度信息和计算方式信息要与统计口径相适配，根据统计口径的变动进行增改。不同的指标通过主体信息、维度信息和计算方式信息进行区分，统计口径调整后要同步调整指标名称。

（三）结构固定

指标名称中的统计主体信息、统计维度信息和计算方式信息的相互顺序要固化，避免由于指标名称中信息要素顺序

不一致导致的“同义不同名”问题。证券公司可以基于现有指标清单，归纳本公司信息要素类型，固化信息要素相互顺序，规范指标命名方式。

指标命名标准化管理的核心思路是“将指数难度问题变为多项式难度问题”。可以将所有的指标名称看成一个集合，指标集合的元素数量是指标各维度可用词汇的叉乘，数量级是指数，因此想要通过制定“指标标准”穷尽指标名称是不现实的。通过归纳信息要素类型、制定各类信息要素中的术语规范，可以将管理难度从指数量级降低到多项式量级，极大降低管理成本。

三、指标命名管理方法

（一）指标命名信息维度归纳

参考 SDOM（证券期货行业数据）模型数据项分类和指标应用实际情况，可以将证券公司指标的信息维度归类为主体维度、财会维度、业务维度、品种维度、账户维度、渠道维度、事件维度、计算方法维度、统计频率维度、计量单位维度和统计范围维度。

（二）指标命名信息维度示例

主体维度即指标的计量对象。证券公司典型的主体维度有客户主体、员工主体、交易对手主体和信息系统主体等。需要注意的是为了与其他信息维度保持独立性和区分度，主体维度主要包括自然人、机构、地区和信息系统等“非业务”

对象。

财会维度即指标中涉及的往来款项和资产负债数据。证券公司典型的财会维度有资产、负债、净资产、应付应收款、佣金收入、利息收入、营业收入、毛利润、净利润、手续费、税费、市值、面值和成本等。

业务维度即指标中涉及的业务条线。证券公司典型的业务维度有经纪业务、信用业务、自营业务、财富管理业务、产品代销业务、投资银行业务、研究业务、托管外包业务和投资顾问业务等。需要注意的是，为了保障区分度和独立性，这里的业务条线多指前台业务，并不包括风险管理、运营管理、绩效管理和投融资审批管理等涉及多业务交叉的中后台管理业务。

品种维度即指标中涉及的金融工具。证券公司典型的品种维度有股票、债券、基金、衍生品、资管计划、利率和汇率等。

账户维度即指标中涉及账户类型。证券公司典型的账户维度有资金账户、证券账户、银行账户、养老金账户、基金TA账户和衍生品合约账户等。

渠道维度即指标中涉及的交易或营销渠道。证券公司典型的渠道维度有网上、柜台、一级市场、二级市场、境内市场和境外市场等。

事件维度即指标中涉及事情和事项。证券公司典型的事

件维度有交易事件、融资事件、违约事件、结算事件、营销事件、签约事件、风险事件、操作事件和发售事件等。

计算方法维度即指标中涉及的统计方法。证券公司典型的计算方法维度有平均、加总、计数、折算、占比、同比、环比、净值、排序和极值等。

计算频率维度即指标中涉及的统计频率。证券公司典型的计算频率维度有实时、每交易日、每自然日、每周、每月、半年、每年和近 12 个月等。

计量单位维度即指标中涉及的标量。证券公司典型的计量单位维度有人民币元、人民币万元、人民币亿元、港元元、港元万元、美元元、美元万元、基点、个数、笔数和百分数等。

统计范围维度即指标中涉及的取数边界。证券公司典型的统计范围有公司级、部门级、分公司级、业务团队级和营业部级。

在完成指标名称信息维度归纳后，需要固化各信息维度的相互顺序，各信息维度之间可以添加分隔符，指标中不涉及的信息维度可以为空。

四、指标命名信息维度分类分级

在指标命名管理实践中，除固化指标信息维度外还可以对各信息维度中的维度项进行分类分级，一定程度上提升指标命名的精准度和灵活度，尽量保障指标名称对指标业务含

义的准确表达，减少由于“粒度”导致的信息遗失。

信息维度的分类分级就是对维度项进行“聚类”，每个“聚类”是一个树形结构，树形结构中的下级节点是上级节点的子类，每个下级节点只对应一个上级节点。以品种维度为例，一个典型的三级聚类树形结构见表 1。

表 1 三级聚类树形结构

股票	主板	
	科创板	
	创业板	
	新三板	基础层
		创新层
基金	公募基金	ETF
		LOF
		REITS
	私募基金	私募债权基金
		私募股权基金
债券	国债	
	企业债	中小企业私募债
		城投债
		产业债
	政策性金融债	
衍生品	公司债	
	期货	
	互换	收益互换
		CDS
	期权	
	远期	

从表 1 中可以看到，维度项层级越高，“粒度”越细，“信息量”越多，在指标命名过程中可以根据指标的业务口径选用恰当维度项。需要注意的是，理论上由“聚类”形成的树形结构应达到“相互独立，完全穷尽”的状态，但在实践中这样的“理想状态”不利于指标命名管理。

由于每类维度可能会存在多个合理的分类视角，如基金至少有三种分类视角，公募私募、开放式封闭式、股基债基混合基，而且不同分类视角下的子类可能会有交叉，如ETF既是公募基金又是开放式基金。除此之外，指标的统计范围很可能是“跨类型”的，如统计股票和基金的交易量或统计股票和债券的交易量。为了应对这些应用场景，维度项的分类分级应当允许一定程度的重叠和交叉并且融合不同的分类视角。

在规范维度信息分类分级的同时也要对指标需求端进行约束，避免由于指标需求过度发散导致维度项数量的过度膨胀以及连带的分类分级管理成本的急剧增大，最终达到指标应用灵活性和指标命名维度项管理成本的协调发展。

五、指标命名维度项编码

为便于使用，除了规范指标命名外还应该规范指标编码，指标编码应可以体现指标的维度结构和分类分级，为量化评估各指标的“相似度”做铺垫。还是以品种维度为例，一个典型的编码结构见表2。

表2 典型的编码结构

股票			4010000
	主板		4010100
	科创板		4010200
	创业板		4010300
	新三板		4010400
		基础层	4010401
		创新层	4010402
基金			4020000

	公募基金		4020100
		ETF	4020101
		LOF	4020102
		REITS	4020103
	私募基金		4020200
		私募债权基金	4020201
		私募股权基金	4020202
债券			4030000
	国债		4030100
	企业债		4030200
		中小企业私募债	4030201
		城投债	4030202
		产业债	4030203
	政策性金融债		4030300
	公司债		4030400
衍生品			4040000
	期货		4040100
	互换		4040200
		收益互换	4040201
		CDS	4040202
	期权		4040300
	远期		4040400

表 2 中采用了“1+2N”编码策略，其中编码首位（表 2 中编码首位是 4）指代品种维度，2 指代各层级中维度项“位数”（表 2 中各层级可包括 100 个维度项），N 指代“层级”（表 2 中 N 等于 3）。通过控制层级参数和位数参数可以匹配各信息维度的分类分级情况和各类中可容纳的维度项数量。

六、指标“相似度”计算

指标的分类分级信息定性体现了各维度项的“相似度”情况，同时指标编码保留了指标分类分级信息，可以考虑通过指标编码定量计算指标间“相似度”。以品种维度为例，指标单个信息维度的“相似度”可以参考图 1。

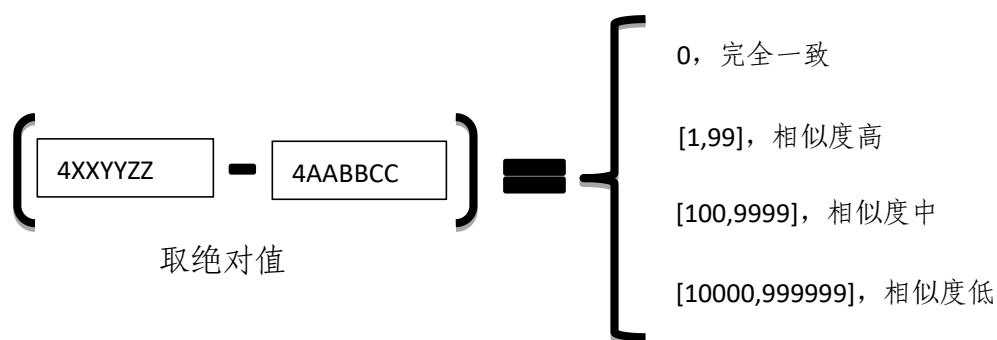


图 1

当根据图 1 中的相似程度赋予不同的系数时就可以得到“相似度”的量化数值，指标间“相似度”值等于各信息维度“相似度”值的加总，这里需要注意的是需要为指标间不一致的信息维度赋予设定的“相似度”值。计算公式如下：

$$\text{相似度值} = \sum_{i=1}^N D_i$$

其中，N 表示信息维度数， D_i 表示各维度的“相似度”值。

求得指标间“相似度”值后，可以很方便地对各指标的相似程度进行排序，当查询某指标时推荐其他“相似度”较高的指标，或在评估指标开发需求时协助判断是否有存量指标具有相似的统计口径。

七、指标命名管理有效性

（一）规范的指标名称

指标命名管理为指标名称制定了统一的“语素”和“语序”规范，理论上基于相同的指标定义会生成一致的指标名称，基本规避了业务术语使用不一致的问题。

（二）清晰的指标结构

指标命名管理为指标名称制定了清晰的指标结构，“衍生关系”体现为下游指标比上游指标在名称中增加了限制的维度信息，“组合关系”体现为下游指标计算方法等维度的业务解释包括了上游指标，从指标名称中即可判断指标之间的关联关系。

（三）一一对应的指标定义和名称

指标命名管理要求指标名称由指标定义提取而来，理论上不同的指标定义应提取出不同的维度项，从而生成不同的指标名称。

八、指标命名管理局限性

不能规避指标重复开发。证券公司中的典型指标开发流程是由需求单位向建设单位提出指标开发需求，不同的需求单位可能对应建设单位中不同的实施团队，这种情况下，存在业务交叉的需求单位很可能会关注同样的指标。当同样的指标开发需求被递交到不同的实施团队时，很可能会导致指标重复开发。

不能避免指标“同名不同义”。指标名称是指标业务定

义的“近似值”，加上不同人员对指标业务定义信息量的提取及对指标维度项的理解存在差异，这种情况下，当业务定义高度相似的指标经由不同人员处理时，很可能被赋予相同的指标名称。由于同样的原因，不同人员解读业务含义相同的指标时也可能被赋予指标不同的指标名称，同样不可避免“同义不同名”问题。

指标命名标准化管理是为指标名称的生成提供统一的方法，促使不同业务场景的指标形成风格一致、术语规范、结构清晰、相互可比的指标集。如果要解决“重复开发”“同名不同义”和“同义不同名”等指标管理痛点，需要配套建立指标需求扎口管理机制和一致性检查机制。指标需求扎口管理机制要做到所有待开发的指标和已开发指标的唯一性管理，指标开发仅在特定数据层开展且名称仅出现一次，其余数据层仅能引用已开发的指标，解决“重复开发”和“同名不同义”痛点。一致性检查机制要做到定期检查是否存在指标值一致但指标名不一致的指标，解决“同义不同名”痛点。

（兴业证券股份有限公司金融科技部总经理蒋剑飞、

大数据中心副总经理陈锋、项目经理邓颖锋供稿）