

收益率时间间隔 对跟踪误差计算的影响

■黄瑞庆 林飞

跟踪误差是用来衡量指数基金相对目标指数偏离程度的计量指标。随着指数基金品种在我国证券市场的日渐增多,运作逐步走向规范,指数基金跟踪误差的计量方式成为指数化投资理论界及实务界越来越关心的问题。跟踪误差衡量公式的设计是指数化投资中必须解决的关键问题,它直接关系到跟踪组合的求解结果的优劣和指数化投资绩效评价的好坏。

一、跟踪误差的计量方法

指数化投资在国外已经有近30年的发展历史,跟踪误差指标的设计经历了由简单的收益率之差到相对复杂的标准差类指标的过程,跟踪误差本身的作用也由原先单纯揭示和描述偏离风险的功能慢慢向兼顾风险预示、风险控制的功能转变。近几年理论界的部分研究在跟踪偏离度指标的基础上又陆续提出了一些改进指标,如线性跟踪误差的表达式,试图用线性表达式来更精确的描述跟踪偏离风险给投资者带来的真实感受,但是由于采用的线性表达式在某些方面丧失了优良数学特性而未被广泛采用。

跟踪误差在我国目前最常用的计量方式,是用指数基金日超额收益时间序列的标准离差平方和加以计算的,也常被称为跟踪偏离度。我国现有华安180指数基金,融通深证100指数基金,公布跟踪误差使用的计算公式如式1所示:

$$S_e = \sqrt{\frac{\sum_{t=1}^T (e_t - \bar{e})^2}{T-1}} \quad (1)$$

其中 e_t 是超额收益率,即一定时间间隔的跟踪组合(跟踪指数的投资组合)收益率 R_{it} 和基准指数 R_{it} 收益率之差, $e_t = R_{it} - R_{it}$, T 为计算跟踪误差时选取的时间间隔的总个数,反映了采样的总时间跨度。

二、跟踪误差的年化处理

一般而言,对不同时间间隔收益率数据计算得出的跟踪误差是不能直接比较的,通常情况下,采用时间间隔越小的收益率时间序列数据,其最终计算出的跟踪误差更小,因为时间间隔越小的收

$\mu\Delta t$ 、标准差为 $\sigma\sqrt{\Delta t}$ 的正态分布,即认为股票在 Δt 时间间隔内的收益标准差同 $\sqrt{\Delta t}$ 是成正比的。该理论上进一步推广到跟踪误差的计算上,即得到式2中的年化调整方法。

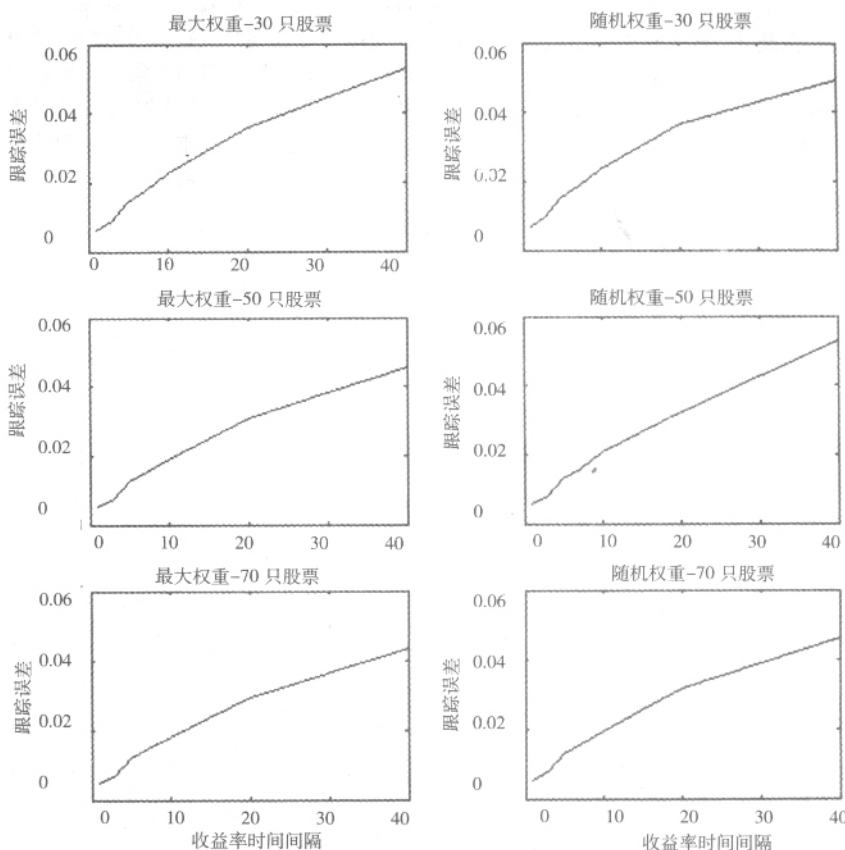


图1 根据不同时间间隔收益率数据计算的跟踪误差

益率,不论是指数的收益率或跟踪组合的收益率,其均值与方差都会越小,因而导致了跟踪误差也越小,为了让不同频率收益率数据计算的跟踪误差具有可比性,通常要对跟踪误差做年化处理,即不论其采用何种频率数据,将其统一换算成以年为时间间隔的跟踪误差指标,做年化处理后的跟踪误差公式,证券市场上股票价格行为模型通常用维纳过程(Winener processes)来表达,该理论认为股票 Δt 时间内的收益率 $\Delta S/S$ 服从均值

$$\bar{S}_e = \sqrt{\frac{\sum_{t=1}^T (e_t - \bar{e})^2}{T-1}} \cdot \sqrt{M} \quad (2)$$

其中 M 为一年时间跨度包含时间间隔的个数。

观察式2可以得知,年化处理的基本原理就是在式1的基础上乘以一个年化因子 \sqrt{M} ,使得不同时间间隔收益率数据计算的跟踪误差都具有可比性。按以上公式,计算同一跟踪组合相对于同一基准指数的跟踪误差,如果分别采用

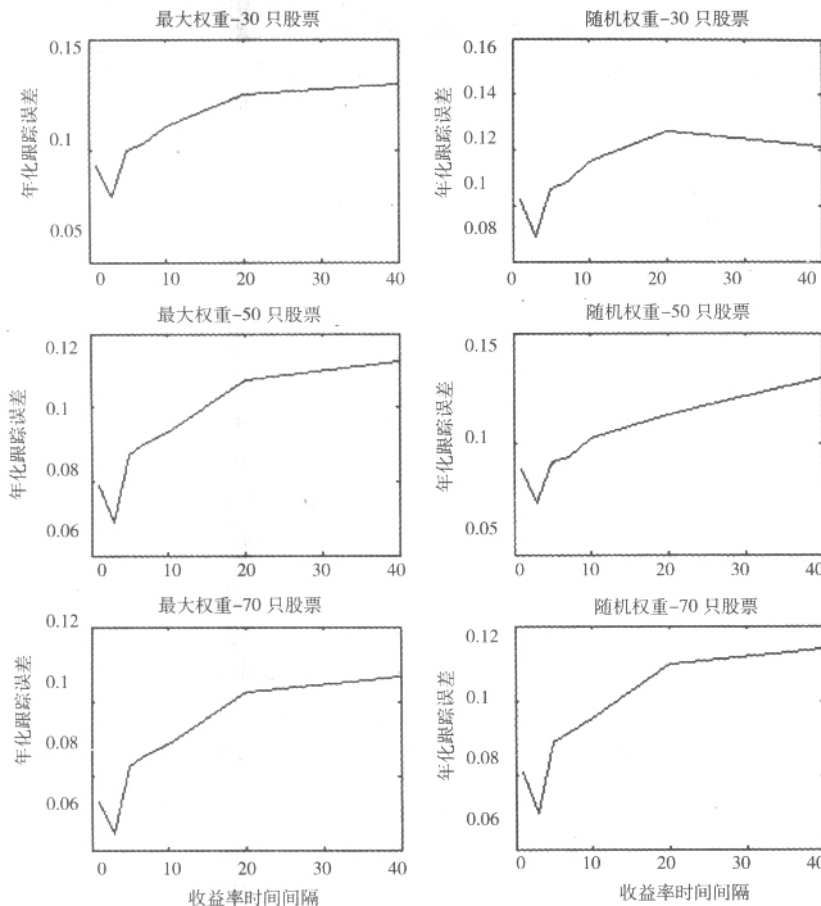


图2 年化后的跟踪误差

几种不同时间间隔的收益率数据,那么经过年化处理之后,其跟踪误差应该是大致相等的。

三、对不同时间间隔收益率数据计算出跟踪误差的实证分析

为了验证同一跟踪组合其跟踪误差和收益率数据时间间隔是否满足式2描述的关系,笔者用我国证券市场的数据做了相应的实证分析。数据选用了1997年前上市的100家公司(上海67只、深圳33只)从1997年1月1日到2002年1月1日的数据,按流通股本加权、连锁计算的通行指数编制方法,构造了“统一100”指数,并从100只股票中按最大权重样本抽样复制法和简单随机抽样复制法,分别抽取30只、50只、70只股票构造跟踪统一指数的跟踪组合,之后按式1和式2考察了该组合分别采用1天、3天、5天、7天、10天、20天、40天时间间隔收益率数据计算的跟踪误差。我们将最终结果绘制成图1和图2。

图1是未经过年化处理的跟踪误差和收益率数据时间间隔相关性的示意

图。观察图1不难发现,无论是用最大权重法还是随机抽样法复制指数,同一跟踪组合的跟踪误差随收益率数据的时间间隔扩大而不断扩大,跟踪误差在这里表现出来的规律和式2理论上应有的规律是一致的。

而图2是经过年化处理跟踪误差与收益率数据数据时间间隔相关性示意图。按公式2,同一跟踪组合,不论其采用何种时间间隔的收益率数据计算跟踪

误差,只要该跟踪误差经年化调整后,应该有大致相同的计算结果。但图2清晰的显示:式2并没有完全实现其“统一不同时间间隔收益率数据计算跟踪误差指标的对比口径”的目的,简而言之,式2对跟踪误差指标的年化调整,并没有完全调整到位。

四、收益率和跟踪误差时间序列自相关对跟踪误差计算的影响

笔者分析,年化调整效果不理想的主要原因是收益率和跟踪误差的时间序列都存在自相关。1988年Lo.A.W和A.C.MacKinlay发表论文“Stock Prices Do Not Follow Random Walks Evidence From a Simple Specification Test”,该文的一个重要观点是:对于一个随机游走时间序列 $x_t, x_t = x_{t-1} + \mu + \varepsilon_t$,且 ε_t 服从 $N(0, \sigma_0^2)$ 的正态分布,如果该时间序列存在着序列自相关,那么不同时间间隔数据计算出来的方差将不具有可比性,除非按如下公式进行调整:

$$\frac{\hat{\sigma}^2(q)}{\hat{\sigma}_a^2} \cong 1 + \frac{2(q-1)}{q} \hat{\rho}(1) + \frac{2(q-2)}{q} \hat{\rho}(2) + \dots + \frac{2}{q} \hat{\rho}(q-1) \quad (3)$$

其中 $\hat{\sigma}_a^2$ 和 $\hat{\sigma}^2(q)$ 分别为间隔为1期和间隔为q期的时间序列采样数据估计出的 σ_0^2 值, $\hat{\rho}(k)$ 为k阶自相关系数的估计值。 $\hat{\sigma}_a^2$ 和 $\hat{\sigma}^2(q)$ 的估计式分别为:

$$\hat{\sigma}_a^2 = \frac{1}{nq-1} \sum_{k=1}^{nq} (x_k - x_{k-1} - \hat{\mu})^2$$

$$\hat{\sigma}^2(q) = \frac{1}{q(nq-q+1)(1-\frac{q}{nq})} \sum_{k=q}^{nq} (x_k - x_{k-q} - q\hat{\mu})^2$$

式3表明:如果时间序列 x_t 不存在序列相关,即相关系数都为0,间隔为1期和间隔为q期的时间序列采样数据估

表1 超额收益的自相关系数的计算结果

时间间隔	30 跟踪组合跟踪误差			50 跟踪组合跟踪误差		70 跟踪组合跟踪误差	
		自相关系数	显著水平	自相关系数	显著水平	自相关系数	显著水平
1	P ₁	0.07	0.015	0.086	0.003	0.094	0.001
	P ₂	0.024	0.036	0.025	0.008	0.039	0.002
3	P ₁	0.168	0	0.163	0	-0.529	0
	P ₂	0.059	0	0.041	0	0.153	0
5	P ₁	0.16	0.01	0.153	0.014	0.137	0.029
	P ₂	0.11	0.008	0.078	0.023	0.094	0.029
7	P ₁	0.171	0.022	0.155	0.036	0.143	0.054
	P ₂	0.165	0.006	0.154	0.013	0.159	0.016
10	P ₁	0.274	0.002	0.282	0.002	0.279	0.002
	P ₂	0.081	0.007	0.09	0.005	0.107	0.004
20	P ₁	0.152	0.229	0.143	0.256	0.153	0.224
	P ₂	-0.234	0.083	-0.214	0.121	-0.227	0.092
40	P ₁	-0.194	0.265	-0.164	0.345	-0.167	0.336
	P ₂	-0.081	0.481	-0.021	0.635	-0.069	0.58

计出的单位方差 σ_a^2 应该是相等的。如果时间序列 x_t 存在序列相关,那么 $\hat{\sigma}_a^2$ 和 $\hat{\sigma}^2(q)$ 的比值仅和1至 $q-1$ 阶的自相关系数有关。该式已经明确的告诉了我们时间序列自相关对不同时间间隔数据估计出单位方差相互比较的影响及其影响的形式,为我们对比不同频率收益率数据估算出来的单位跟踪误差给出了有益的提示。

五、跟踪组合、指数收益率时间序列自相关实证分析

为了研究跟踪组合、指数收益率及超额收益率时间序列的自相关,我们同样选用了上面研究中使用的数据,分别计算了1天、3天、5天、7天、10天、20天、40天频率下,指数收益率时间序列、包含30、50、70个样本股的最大权重跟踪组合的超额收益率时间序列数据的第一阶和第二阶自相关系数及其显著性水平。计算结果整理如表1所示。

观察表1不同频率下不同样本股个数跟踪组合超额收益的自相关系数的计算结果,可以发现1到10天间隔的超额收益数据绝大部分都表现出了非常显著的一二阶正自相关系数,特别是间隔为3天和间隔为10天的数据,均表现出较大的正自相关系数,且其正自相关系数的显著性指标均不超过0.007,其中3天间隔的超额收益数据其显著性指标甚至小于0.001。

按式3给出的结论,如果单位间隔数据存在一定程度的正相关,那么时间间隔为 q 单位的数据计算出来的时间序列单位方差 $\hat{\sigma}^2(q)$ 和单位间隔数据直接计算出来的方差 $\hat{\sigma}_a^2$ 的比值会大于1,进一步可以得到的更具有实用价值的推论是:对同一时间序列而言,在某一个时间间隔上的数据的正自相关系数越大,显著性水平越高,那么由该时间间隔数据自身计算出来的单位时间序列单位方差将会越小;反之,某一时间间隔上的数据有绝对值越大且越显著的负自相关系数,那么由该时间间隔数据计算出来的单位方差会越大。因此,表1中关于超额收益时间序列自相关的计算结果进一步解释了为什么不同频率跟踪误差数据经过调整仍然无法和20天40天数据达成一致的原因:正是由于1天、3天、5天、7天、10天时间间隔上的数据存在一个显著的正相关,所以使得该数据尽管经过式2的频率调整,其跟踪误差计算结果通常还会比其它频率(20天、40天)上的小。重新回顾图2,可以发现3天和10天的数据其调整后的跟踪误差均比较小,特别是3天的数据在图形上形成了一个明显的凹点,而这和3天和10天跟踪误差数据自相关系数特别大与特别显著的实证分析结论是完全吻合的。

六、关于不同采样时间间隔跟踪误

差数据对比的总结

调整不同时间间隔收益率数据计算的跟踪误差,使其最终具有可比性是指数字化投资领域近年来比较关注的一个问题。总结以上的研究过程可以发现,与普通的随机过程不同时间间隔数据方差的调整方法不同,跟踪误差数据的年化调整较为复杂,绝不能用简单的乘以时间跨度开平方做为其最终调整因子,实证分析结果表明该方法效果并不理想,无法达到统一对比口径的目的。由于组合超额收益时间序列数据通常都存在着不同程度的自相关,在调整的时候需要计算不同时间间隔超额收益数据的自相关系数,然后根据3式进行调整。本文的最终结论和我们以往对跟踪误差年化调整的认识具有很大的不同:跟踪误差的年化调整公式并不是一成不变的,它将随着所要调整的数据的时间间隔不同、跟踪目标指数不同而不同。换句话说,如果没有把时间序列自相关的因素考虑清楚,那么对比不同时间间隔数据计算出的跟踪误差指标通常是没有意义的,跟踪误差数据年化处理后得出的指标也是不客观的,不能被直接作为评价跟踪组合相对目标指数年度偏离风险的衡量指标。

(作者单位/厦门大学计划统计系,
融通基金管理有限公司研究部)
(责任编辑/李友平)

(上接11页)拟费用不同。

由于纯粹按理论要求对污染物治理进行“定价”不具有可操作性,按实际发生的污染物治理价格测算的“绿色GDP”所反映的经济发展的负效应与事实情况相悖。如何切实合理地处理环境折耗的计量依据,并标准化,使具体的“绿色GDP”核算工作可操作,核算的结果具有政策指导意义是急需解决的问题。本文认为,维护成本法的环境折耗计量标准的核定可参照台湾的做法,以目前“最佳可行技术”下的污染防治单位成本作为环境折耗计量标准。

(四) 关于环境保护费用的处理

在联合国SEEA中,为政策分析的需要,环境保护费用被单独列出来,但并不影响GDP总量、“绿色GDP”或EDP的数值。台湾省、“湖南省统计局课题”的处理与联合国SEEA要求一致。而“中国

环境经济综合核算体系课题”将固体废物治理费、空气污染治理费、水污染治理费作为环境折耗;“北京市课题”将政府的环境保护与治理支出作为环境折耗。认为在最终消费范围,政府(以及家庭)的环境保护费用具有“社会成本”的含义,它是经济活动对环境影响的结果,因此,应从GDP中扣除。本文认为,政府环境保护费用、家庭的环境保护费用作为社会成本的确应该“内部化”,成为企业的成本,应从GDP中扣除。但应比较政府环境保护费用和企业交付给政府的环境治理费,如果前者大于后者,说明“外部成本”大于“内部成本”,两者之差应从GDP中扣除。具体处理上,可进一步研究。

(五)“绿色GDP”总量指标的公布应附说明

即使是基于SEEA核算模式测算

的“绿色GDP”,亦将由于核算范围不同,计量方法与计量依据不同,使核算的“绿色GDP”总量指标值大小差异极大,使核算结果不具有可比性,导致人们难以用“绿色GDP”总量指标评价经济的可持续发展状况,或难以衡量国家和区域的环境与经济的实质进步。因此,在给出测算的“绿色GDP”总量指标时,固定的展示原始资料和详细地描述核算范围、使用的计量方法、计量标准既是必要的,也是重要的,以便使用者能评价结果的误差。而且,随着最初数据的提供,才能给使用者一个推理的自由,如果仅给出最终的结果将令使用者迷惑。因此,展示“绿色GDP”总量指标时,应附有关说明。

(作者单位/浙江工业大学)
(责任编辑/亦民)