下载日志已记录，仅供内部参考

智能配置深度

报告日期：2026 年 10 月 26 日

集中度的几何

——机构持仓、容量与未来收益的实证检验

分析师：陈冀
执业证书号：
S1230522110001
chenji@stocke.com.cn

相关报告

1 《资金流的流体力学结构：理论推导》

2 《最小方差组合的收缩边界》

核心观点

本文把卖方研究里那个几乎不可能错的命题——「资金在集中持仓、集中持仓的东西未来会怎么样」——改写成一条可以被算死的式子：$\rho_t = \kappa V_t H_{A,t}$，然后在行业与个股两个层面上检验它是否与未来收益同向。一句话结论：两层都没有证据。

❑ **要点一　把一个模糊的概念变成一条可算的式子**

行业份额 π_i 是密度，「这季换了多少」是结构位移 V_t，「这个容量承受了多少」是含价格冲击尺度的深度加权集中度 $H_A = \lambda \sum_i \pi_i^2 / D_i$。三者相乘得到计量对象 ρ_t，再对未来的组合收益做回归。三要素各自可观测、乘子由数据估计、主规格在看结果之前锁定。

❑ **要点二　行业层面：正号，但不显著**

125 个申万二级行业 × 199 周。主规格（方案 B • 原始 DTL • E2 • 单周）$\hat{c}$ = 0.03016，季度聚类 95% 区间 [-0.02137, 0.08168]。三个频率 × 两个模型共六个规格全部未通过（$\hat{c} > 0$ 且上尾 $p < 0.05$）。月频下系数翻负，信号本身很弱。

❑ **要点三　个股层面：负号，按周聚类同样不显著**

把战场推到 5672 只 A 股：主规格（方案 A • 原始 DTL • E2）$\hat{c}$ = -0.00536，季度聚类 95% 区间 [-0.01296, 0.00225]。这个负号有明显口径依赖：旧的状态内置换零分布 sd 远窄于按周聚类标准误，按季度聚类看区间跨零。本文把两种口径并排报出，不取最窄的那个下结论。

❑ **要点四　一个顺带的发现：变化主要来自「换」而不是「挤」**

把三要素拆开后，集中度与冲击之比 H_A/V 落在 1e-15~1e-13 的量级上——几乎不携带增量信息，ρ 的横截面变化几乎全部来自 V。这对做集中度因子的人比主检验更实用。

❑ **要点五　披露截断：制度性约束，不是技术细节**

基金报告只披露前十大。全样本里，前十大占整本账市值的中位数是 56.8%，而「从未露面」的个股占名称数的 90.5%、却只占市值的 39.5%。不可见的部分在名称上很大、在市值上有限，这使误差更可能稀释而非制造结论。

❑ **要点六　实施边界：结构检验与可执行预测必须分开**

H1 - H5 采用事后结构时点；H6/H7 另按披露可得时点构造。固定季度权重逐周计算的是恒权指数，不是买入持有策略。未执行或不可识别的检验是缺口，不能当作否定性结果引用。

❑ **风险提示**

两层主规格均未通过门槛。系数建立在「只到前十大」的披露之上，并有已量化的间期高估（方案 A 约 16.8%）。不构成因果识别或收益保证；不显著不等于机制不存在；全文不构成投资建议。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

正文目录

正文目录 2
1 研究问题与证据范围 4
1.1 「集中持仓」拆成三个可观测的乘子 4
1.2 为什么这个框架可以被否定 4
1.3 什么已经完成，什么仍不能判断 5
2 从持仓与成交额到估计方程 6
2.1 结点、份额与总量的恒等式 6
2.2 容量：把「有多大」换成一个无量纲的量 6
2.3 集中度与冲击 6
2.4 计量对象与两条回归 7
2.5 状态量与披露截断 7
3 理论：从冲击方程到度规形变 8
3.1 记号与三条假设 8
3.2 均衡解、稳定性与紧致度的下界 8
3.3 度规形变：协方差被改写、集中组合被放大 9
3.4 等效原理：纯内生反馈不带来效用损失 10
3.5 真实损失：来自外生申赎冲击 11
3.6 约束如何打破等效原理 12
3.7 视界：Fisher 距离的对数发散 12
3.8 从静态到动态：滞后反馈下的同一个模型 13
3.9 回到可计算的形式：为什么主回归长这样 14
4 数据、口径与时点对齐 16
4.1 数据来源与规模 16
4.2 时点对齐：本文最花功夫的一处 16
4.3 季度末口径：为什么必须用日历季度末 16
4.4 方案 A 与方案 B：两种重建全持仓的办法 17
4.5 两种持仓重建与可比集合的取舍 17
4.6 披露截断的实际规模 17
5 行业层实证与未完成检验 19
5.1 H1 全规格结果 19
5.2 事后可比行业集合：一次审查后的敏感性 20
5.3 置换检验：为什么不显著 20
5.4 H3 与 H4 的辅助识别 20
5.5 容量变化的描述性分解 21
5.6 稳健性与预测的完成状态 22
6 披露截断与重建误差 25
6.1 截断的三种影响，先分类再定量 25
6.2 三个量：截断有多大 25
6.3 为什么这一对数字最重要 26
6.4 一个可对照的检验：只用完整报告 26
6.5 规模环比不等于估计偏差 27
6.6 能够识别误差的验证设计 27

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

7 个股扩展与推断有效性 28
7.1 八规格系数与聚类区间 29
7.2 为何撤回原状态内置换的统计解释 30
7.3 容量残差：正号是不是只是流动性 31
8 结论的强度与可继续验证的问题 32
8.1 两层符号相反，是「都没有证据」而不是「互相抵消」 32
8.2 这个结论能被信任到什么程度 32
8.3 从未拒绝到经济结论还缺什么 32
8.4 本文不能说的四句话 32
8.5 下一步验证的先后关系 33
9 研究边界与使用限制 34
10 风险提示 35
10.1 两层主规格均未通过门槛 35
10.2 系数建立在「只到前十大」的披露之上 35
10.3 间期高估已量化，方向按论证为稀释 35
10.4 不构成因果识别或收益保证 35
10.5 不显著不等于机制不存在，全文不构成投资建议 35
附录 A 符号与单位 36
附录 B 数值验证与复现 36

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1 研究问题与证据范围

卖方研究里关于「集中持仓」的表述有一个共同的特征：它几乎不可能错。「机构在集中持仓，集中持仓的东西要小心」——如果后来跌了，说明集中持仓果然危险；如果继续涨，说明集中持仓还在强化。任何一次失效都可以归因于「这次不一样」。这种表述的问题不在于它对不对，而在于它没有给出可被打死的具体形式。

本文尝试换一条路。不预设「集中持仓」是什么，而是先问：如果机构持仓的集中真的会对未来收益产生可测量的影响，那么在这个影响和可观测的持仓数据之间，必须存在一个具体的函数形式——哪几个量、以什么方式相乘、系数应当是什么符号、在哪个口径下才算通过。把这些写死之后，它就可以被检验，也就可以被否定。

1.1「集中持仓」拆成三个可观测的乘子

把「集中持仓」拆开，会得到三个直觉上彼此独立、但必须一起起作用的量。第一个是份额：某个行业或某只股票被机构持有了多大一块，记为 π_i，它是持仓市值占整个池子的比例，天然满足 $\sum_i \pi_i = 1$。第二个是换手：相邻两个季度之间，机构在这个结点上到底换了多少，记为 V。第三个是容量：这个结点承受得起多大一块钱——同样一个百分点，放在日均成交几十亿的银行上，和放在日均成交几千万的小市值上，性质完全不同，容量用 D_i 度量。

把三个乘起来，得到本文的计量对象；检验的核心命题只有一个：ρ_t 越大，随后的组合收益是否越高。如果集中持仓确实携带未来收益的信息，那么应当看到正的系数；如果看到的是零或负，那么「集中持仓」这个词至少在这个度量下不携带这一信息。

$$r_{i,t} = \text{baseline}_{i,t} + c\text{DTL}_{i,Q(t)}x_{t-1} + e_{i,t} \quad (1)$$

1.2 为什么这个框架可以被否定

它的可否定性来自三处。第一，乘子形式是写死的：如果数据里真正起作用的是别的东西（比如单纯的规模、或者单纯的换手），那么三乘子的构造会给出接近零的系数——因为 H_Λ 会把容量维度的信息混进来，稀释掉纯规模效应。第二，主规格在看结果之前锁定：本文报告 8 个（个股层）与 6 个（行业层）规格，但判定只用其中预注册的那一个，其余作为稳健性列出，不允许事后挑一个好看的。第三，通过门槛写成了双重条件：$\hat{c} > 0$ 且上尾置换 $p < 0.05$。「上尾」这个方向很重要——它检验的是「正效应」，而反证（负效应）必须另用下尾 p 来读。

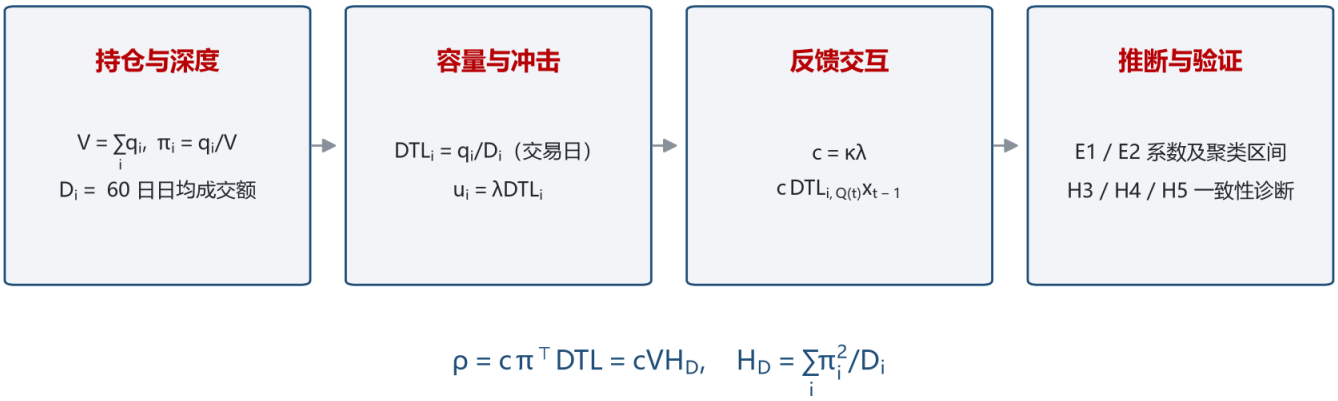
若 $c > 0$，上一期组合上涨时反馈项为正，下跌时反馈项为负。估计值不显著不能证明反馈不存在；负点估计也不等于负反馈获得统计支持。行业与个股层共用市场、基金来源和部分时间窗口，是相关的两个聚合视角，不能当作独立实验。

下面这张图把上面三件事串成一条链——它同时也是全文的阅读地图：从左到右依次是「机构持仓集中度」到「容量约束」到「价格反馈」到「未来的组合收益」，每一步的箭头都对应一个可观测的量，图中没有任何一步依赖不可观测的中间变量。后文第 2 节展开这条链的每一步计量形式，第 3 节给出它的理论依据，第 4 至第 7 节依次报告两层的检验结果。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

图1： 持仓容量与反馈的识别链



$$\rho = c\,\pi^{\top} DTL = cVH_D, \quad H_D = \sum_i \pi_i^2/D_i$$

结构时点的 H1–H5 与披露可得的 H6/H7 分开；理论恒等式不替代统计推断有效性。

资料来源：浙商证券研究所，Tushare Pro

表1： 两层估计概览[1]

层级与设定	有效样本	系数	季度聚类 95% 区间	证据范围
行业 B / E2 / 1 周	125 行业 × 199 周	0.03016	[-0.02137, 0.08168]	未获支持
个股 A / 原始 DTL / E2	5672 结点 / 1621452 行	-0.00536	[-0.01296, 0.00225]	区间与依赖假设并读；旧置换撤回

资料来源：浙商证券研究所，Tushare Pro

两层都未通过门槛。怎么读这个「未通过」，是全文最重要的一件事。在行业层，$\hat{c}$ 是正的（0.0302）但不显著；在个股层，$\hat{c}$ 是负的（-0.00536）且按周聚类不显著。两层符号相反、且两层都不显著——这是「两个层面都没有证据」，而不是「互相抵消的反向证据」。把两层合并成一句「集中持仓无害」是错的，因为两层用的是不同的样本、不同的口径、不同的频率，本来就不该被合并。

1.3 什么已经完成，什么仍不能判断

理论部分给出闭式解和合成环境核验；实证部分报告有效样本、点估计、区间及不能计算的原因。实证显著性仍依赖误差相关结构与识别假设。披露可得预测只由 H6/H7 另行处理，不能从结构回归直接推导交易策略表现。报告中的所有样本数与估计量从本轮结果文件读取。

更重要的是，这个未通过有一个结构性的原因，而不是运气不好。下一节从模型的构造讲起，然后给出完整的理论推导，再依次交代数据口径、行业层结果、那个结构性原因以及个股层结果，最后一节说清楚这个结论能被信任到什么程度。

[1] 主设定是当前分析约定。除非另有不可变版本及时间记录，不声称这些选择发生在查看结果之前。股票原状态内置换不再产生确认性 p 值。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

2 从持仓与成交额到估计方程

这一节把 ρ_t 完整推出来，并交代每一步在市场侧对应什么。推导链刻意保持「只有加法与除法」，不用任何不可观测的中间变量，因为后面所有结论都建立在「这些量都是可测的」这一前提上。

2.1 结点、份额与总量的恒等式

把市场离散成一组互不重叠的结点（行业层是申万二级行业，个股层是单只 A 股），记为 $i = 1, \ldots, N$。在季度末状态 t，结点 i 被公募基金持有的市值记为 $q_{i,t}$。

$$V_Q = \sum_i q_{i,Q}, \pi_{i,Q} = q_{i,Q}/V_Q, \sum_i \pi_{i,Q} = 1 \tag{2}$$

份额之和恒等于 1 是一个纯构造性质，在数据上它同时充当一道自检：本文实测的两套面板里，份额之和的最小值与最大值都落在双精度浮点误差之内。任何一个状态的份额和偏离 1，都说明池子的定义出错了，程序会直接报错而不是继续跑。持仓市值 q 与总额 V 以元计。V 是持仓存量，不是申赎额或份额换手率。所有对象必须使用同一资产集合与同一合格基金池。

2.2 容量：把「有多大」换成一个无量纲的量

「集中持仓」之所以难定义，是因为集中持有大盘股和小盘股是两回事。解决的常规做法是除以市值，但市值度量的是「存量有多大」，而机构调仓受制约的是「多大一块钱能换得出去」，后者是流量。所以容量用成交额度量：

$$D_{i,Q} = \frac{1}{60} \sum_{s=1}^{60} \mathrm{Amt}_{i,s}, \mathrm{DTL}_{i,Q} = \frac{q_{i,Q}}{D_{i,Q}} \tag{3}$$

DTL 是无量纲的，它有一个直接的市场读法：假设这只股票或这个行业未来只剩当天这个成交水平，把当前被基金持有的那一块市值全部卖完，需要 DTL_i 个交易日。DTL 越大，说明这个持仓相对流动性越重，卖出去越费劲。它是在全部日成交额均可用于处置、且未来深度不变的理想条件下算出的，不是实际清仓期限承诺。20 日窗口仅作敏感性比较。

2.3 集中度与冲击

集中度是「在容量加权下，份额有多集中」。为什么用份额的平方：如果两个结点份额都是 1/2，仓位的「集中」程度低于一个结点占 9/10、另一个占 1/10 的情形，而平方和恰好把这件事量化了。除上容量则让「集中在哪里」也有了权重。冲击度量相邻两个状态之间持仓结构的整体位移，取份额一阶差分的绝对值之和的一半。

$$H_D = \sum_i \frac{\pi_i^2}{D_i}, H_\Lambda = \pi^\top \Lambda \pi = \lambda H_D, \rho = \kappa V H_\Lambda = cV H_D = c\pi^\top \mathrm{DTL} \tag{4}$$

系数 1/2 的作用是把冲击落在 $[0,1]$ 上，并使它有「组合换手率的一半」这个读法：如果机构把一半仓位从 A 挪到 B，冲击等于 1。κ 是资金流对业绩的敏感度，λ 是单位清仓天数对应的价格冲击尺度。H_D 的单位为交易日/元，λ 的单位为交易日的倒数，H_Λ 的单位为元的倒数。H_Λ 已含 λ，两者不能混写。紧致度 ρ 为无量纲量。不能用 H_D/V 的数值大小推断容量的重要性。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

2.4 计量对象与两条回归

把三个量相乘，得到结构性集中强度。同时，把「集中持仓的那些票」的实际收益加起来，得到一个可以直接观测的因变量——集中组合在第 s 周的收益。其中的权重取该周之前已经披露的状态，这个「时点对齐」是本文最花功夫的一处，下一节会专门交代。

$$x_t = \sum_i \pi_{i,Q(t)}\, r_{i,t} \tag{5}$$

检验的核心是下面两条回归，本文称之为主检验 H1。E1 含截距、结点自身滞后收益与同质的滞后集中组合收益；E2 再纳入结点异质的滞后集中组合收益与滞后市场收益。E2 是本文的主模型。为什么要加这两个控制：集中组合本身就是一个高相关的收益序列，如果不同时控制它自己的滞后项，ρ的系数会把自相关的部分吸收进去，读出来的就不再是「结构变化的信息」而是「上期的表现」。

$$r_{i,t} = a_i + c\mathrm{DTL}_{i,Q(t)}x_{t-1} + g x_{t-1} + h r_{i,t-1} + e_{i,t}(\mathrm{E1}) \tag{6}$$

$$r_{i,t} = a_i + b_i x_{t-1} + d_i m_{t-1} + c\mathrm{DTL}_{i,Q(t)}x_{t-1} + h r_{i,t-1} + e_{i,t}(\mathrm{E2}) \tag{7}$$

m 为对应资产集合的等权收益。E2 按结点在整个有效样本内吸收截距及两种异质收益斜率；并非在状态×结点内去均值。若一个结点的 DTL 完全不变，其交互项会与该结点的 x 斜率重合，因此 c 依靠 DTL 的时间变化识别。控制这些项不等于已经解决动态遗漏变量和因果识别。

系数按自由分布滞后聚合。基准规格的滞后期集合只含 1 周；稳健性放宽到 1–4 周，另有自然月频一个规格。这样做的目的是让「1 周」与「4 周」两种时间尺度都可以被容纳进同一个 c 的定义里，同时避免预先假定衰减形状。

$$c_{1:4} = \textstyle\sum_{h=1}^{4} c_h\,, r_{i,t} = \cdots + \sum_{h=1}^{4} c_h\, \mathrm{DTL}_{i,Q(t)} x_{t-h} + e_{i,t} \tag{8}$$

行业 E2 的 1 周、自由估计 1–4 周之和与自然月频组成三项 Holm family；E1 单独校正作补充。个股主对照是 A/原始 DTL/E2，辅助规格不用于事后挑选显著结果。固定季度 π逐周加权给出恒权收益指数；买入持有需要权重随价格漂移，两者不能混称。

2.5 状态量与披露截断

到这里为止，模型是完全可计算的——只要持仓市值是已知的。问题恰恰出在这里：它不完全是已知的。公募基金的披露是分层的：季报（一季报与三季报）只披露前十大持仓，只有年报与半年报披露全部持仓。本文一律用「季报」指这两个只披露前十大的报告期，用「完整报告」指年报与半年报，这意味着持仓市值是一个被截断的观测。

这处截断不是可以忽略的技术细节，它是本文最实质的一个约束，因此单独用一节完整交代。这里只先说明它对模型的影响方向：由于截断掉的持仓在所有结点上都存在，它使份额与集中度都带有噪声；而如果这个噪声与未来收益不相关（这是本文会论证的），那么它对 c 的影响就是向零衰减——也就是把真实的关系往「看不见」的方向推，而不是凭空造出一个正的系数。这一点是后面解读「未通过」的关键。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

3 理论：从冲击方程到度规形变

上一节给出了最终的操作性定义 $\rho_t = \kappa V_t H_{\Lambda,t}$——但那只是一个「可以算的量」。它为什么长这样、为什么它的系数有符号上的预言、以及它与「相对论类比」到底像在哪里，都还没有交代。这一节把这条链补齐。

要预先申明类比的分寸：本文**不**声称存在金融版的爱因斯坦场方程。在有限检索范围内，能严格成立的类比只有三层——**度规形变**（协方差矩阵被内生反馈改写）、**等效原理**（纯粹的内生反馈只是一次坐标变换，不改变无约束投资者的效用）、**视界**（紧致度趋近 1 时局部扰动被无限放大，且 Fisher 信息距离对数发散）。超出这三层的说法都不在本报告的立场之内。本节的所有结果都有完整证明，并在合成数据上做了数值核验。

3.1 记号与三条假设

把市场离散成 N 个结点（实证中是申万二级行业或单只 A 股）。基本面收益 r_0 服从均值 μ_0、协方差 Σ_0（正定）；集中持仓资金的总规模记为 V，其组合权重为 π（满足 $\mathbf{1}^\top\pi = 1$），于是持仓额 $q = V\pi$；结点 i 的市场深度（可承接的成交额）记为 D_i，冲击矩阵 $\Lambda = \lambda\mathrm{diag}(1/D_i)$。外生申赎冲击满足均值为零、方差为 s^2，并与基本面收益独立。模型的反馈机制写为：

表2： 理论模型的三条假设与派生量[2]

代号	内容	表达式	经济含义
A1	线性价格冲击	$r = r_0 + \Lambda\Delta q$	净交易按深度折算成价格变动；无冲击时退化为基本面收益
A2	业绩—资金流反馈	$\Delta q = \kappa q(\pi^\top r) + q\varepsilon$	组合涨得越好、申购越多、买得越多。κ 是资金流对业绩的敏感度，ε 是外生申赎冲击
A3	均值传导	反馈作用于实现收益的全部，含均值部分	静态模型里是**假设**；在后文的动态模型里升级为**定理**
派生	质量密度 u 与敏感度 v	$u := \Lambda q, v := \kappa\pi$	把「持仓有多重」与「反馈有多强」分开写成两个向量
派生	紧致度（本文的核心状态量）	$\rho := v^\top u = \kappa V H_\Lambda, H_\Lambda = \lambda\sum_i \pi_i^2/D_i = \pi^\top\Lambda\pi$	H_Λ 是按深度加权的 Herfindahl 指数

资料来源：浙商证券研究所，Tushare Pro

把 A1 与 A2 联立，消去 Δq，得到一个关于收益本身的**不动点方程**：

$$(I - uv^\top)r = r_0 + u\varepsilon \tag{9}$$

这一步是整节的关键. 反馈把「收益决定交易」翻成了「收益由自己决定」，所以求解变成了求逆。记 $A := (I - uv^\top)^{-1}$（在其存在时），全部结论都归结为对这个 A 的分析。

值得强调的是，实证里真正用到的三个量——π、V、H_Λ——都是上式左右两端的可观测投影，不需要估计 μ_0 或 Σ_0。这正是本文能用「只用协方差、不用均值」的识别策略的原因。同时也要说清边界：持仓与成交额只能构造容量代理，不能因此认定冲击系数或基准协方差已经识别。

3.2 均衡解、稳定性与紧致度的下界

先给一个纯代数结果，它是后面每一步的原型：

引理 0（Sherman – Morrison）. 对 M 可逆、$1 + d^\top M^{-1}c \neq 0$，有

$$\left(M + cd^\top\right)^{-1} = M^{-1} - \frac{M^{-1}cd^\top M^{-1}}{1 + d^\top M^{-1}c}, det(M + cd^\top) = detM(1 + d^\top M^{-1}c) \tag{10}$$

这是标准结论，证明从略。有了它，关于 A 的三件事都可以立刻读出：



[2] 持仓与成交额可观测；冲击系数、基准均值与协方差仍需额外识别。合成核验采用 $\lambda = 1$的归一化环境。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

命题 1（存在性、唯一性、稳定性与紧致度下界）.

(i) 存在唯一性。 $det(I-uv^{\top})=1-\rho$，因此 不动点方程 有唯一解当且仅当 $\rho\neq 1$，此时

$$A=I+\frac{uv^{\top}}{1-\rho} \tag{11}$$

(ii) 迭代稳定性。 设 $uv^{\top}\neq 0$。试探调整过程 $r^{(k+1)}=r_0+u\varepsilon+uv^{\top}r^{(k)}$ 对任意初值收敛，当且仅当 $|\rho|<1$。

(iii) 紧致度下界。 若每个 $D_i>0$ 且 $\lambda>0$，则

$$H_{\Lambda}\geq\frac{\lambda}{\sum_i D_i}, \tag{12}$$

等号当且仅当 $\pi\propto D$。在 $\kappa V\geq 0$ 时进一步有 $\rho\geq\kappa\lambda V/\sum_i D_i$；负反馈时不等号方向反转。

证明. (i) 在引理 0 中取 $M=I, c=-u, d=v$，得 $det(I-uv^{\top})=1-v^{\top}u=1-\rho$；求逆公式同样由引理 0 给出。(ii) 由 $(uv^{\top})^k=(v^{\top}u)^{k-1}uv^{\top}=\rho^{k-1}uv^{\top}$，迭代得

$$r^{(k)}=(\sum_{j=0}^{k-1}(u\,v^{\top})^j)(r_0+u\varepsilon)+(uv^{\top})^k r^{(0)} \tag{13}$$

右端收敛当且仅当 $\sum_j\rho^j$ 收敛且 $\rho^k uv^{\top}\to 0$；由于 $uv^{\top}\neq 0$，这等价于 $|\rho|<1$，此时的极限正是 $A(r_0+u\varepsilon)$。(iii) 由 Cauchy–Schwarz，

$$1=(\sum_i\pi_i)^2=(\sum_i\frac{\pi_i}{\sqrt{D_i}}\sqrt{D_i})^2\leq\sum_i\frac{\pi_i^2}{D_i}\sum_i D_i, \tag{14}$$

等号当且仅当 $\pi_i/\sqrt{D_i}\propto\sqrt{D_i}$，即 $\pi\propto D$。

怎么读. 持仓越集中、越集中在深度小的结点上，ρ 越大；紧致度下界说明「按深度配置」是使紧致度最小的方式，这也解释了为什么一个理性的集中组合天然会向高成交额的品种集中。$\rho=1$ 是本文所称的视界：局部扰动在此处被无限放大。在实证里这一条同时充当合规检查——必须逐状态验证 $|\rho|<1$，没有任何一个状态越界，平稳线性模型的适用域才没有被破坏。还须说明：这里的奇异性阶数与 Schwarzschild 解中的 g_{rr} 并不相同，类比只在结构层面成立。

需要保留一层限定：上述推导全部在正反馈假设与归一化冲击系数下成立，不能以合成环境通过核验来保证真实状态平稳；经验成交额口径下的冲击尺度另行估计。

3.3 度规形变：协方差被改写、集中组合被放大

把 A 的作用讲成一句话：**外生冲击在到达收益之前，先被内生反馈矩阵改写了一遍**。这就是「度规形变」的全部内容。

命题 2（度规形变与时间膨胀）. 在 $\rho<1$、$\varepsilon=0$ 下：

(a) $\Sigma=A\Sigma_0A^{\top}$，$\mu=A\mu_0$。

(b) $\pi^{\top}A=\pi^{\top}/(1-\rho)$，因此集中组合收益为

$$\pi^{\top}r=\frac{\pi^{\top}r_0}{1-\rho},\mathrm{Var}(\pi^{\top}r)=\frac{\sigma_{\pi,0}^2}{(1-\rho)^2} \tag{15}$$

其方差乘以 $(1-\rho)^{-2}$——这就是「时间膨胀」的对应物。在 $0<\rho<1$ 时放大，在负反馈时缩小。

(c) 对任意组合 w，

$$w^{\top}r=w^{\top}r_0+(u^{\top}w)\cdot\frac{\kappa\pi^{\top}r_0}{1-\rho} \tag{16}$$

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

特别地，若 $u^\top w = 0$，则 $w^\top r = w^\top r_0$——这类组合处在「平直时空」中，完全感受不到集中持仓的存在。这个「与集中方向正交」的组合在后面定理 4 里还会出现，它是全文唯一一条「零损失」的方向。后文定理 4 的最优平方夏普损失还取决于 alpha 与集中方向在特定内积下的重叠，须区分两种正交关系。

证明. (a) 由 $r = Ar_0$ 直接得到。(b) 计算

$$\pi^\top A = \pi^\top + \frac{(\pi^\top u)\kappa\pi^\top}{1-\rho} = \pi^\top\left(1 + \frac{\rho}{1-\rho}\right) = \frac{\pi^\top}{1-\rho}, \tag{17}$$

其中用到 $\kappa\pi^\top u = \rho$。(c) $w^\top A r_0 = w^\top r_0 + (w^\top u)(v^\top r_0)/(1-\rho)$，再代入 $v = \kappa\pi$ 即得。

以下是同期 beta 的恒等式，含不可观测基准项. 它本身不是 H1 的滞后识别式，而是把「集中持仓如何改写 beta」这件事写成一条可验证的代数关系。从 (b) 立刻可以榨出一个不需要知道基本面 β_i^0 的检验式，它是本文实证部分的理论根：

推论 2′（集中度 beta 恒等式）. 记 $\beta_i := \mathrm{Cov}(r_i, \pi^\top r)/\mathrm{Var}(\pi^\top r)$，$\beta_i^0$ 为用 Σ_0 算出的对应量，则

$$\beta_i = (1-\rho)\beta_i^0 + \kappa u_i \tag{18}$$

证明. 由 (b) 有 $A^\top\pi = \pi/(1-\rho)$；协方差 $\mathrm{Cov}(r_i, \pi^\top r) = e_i^\top A\Sigma_0 A^\top\pi = e_i^\top A\Sigma_0\pi/(1-\rho)$；行向量 $e_i^\top A = e_i^\top + u_i\kappa\pi^\top/(1-\rho)$；合并得 $\mathrm{Cov} = [e_i^\top\Sigma_0\pi + u_i\kappa\sigma_{\pi,0}^2/(1-\rho)]/(1-\rho)$，其中 $\sigma_{\pi,0}^2 = \pi^\top\Sigma_0\pi$；分母 $\mathrm{Var}(\pi^\top r) = \sigma_{\pi,0}^2/(1-\rho)^2$；两者相除即得。

(3.5) **为什么重要.** 它把集中度 beta 拆成两块：一块是与集中持仓无关的基本面成分 $(1-\rho)\beta_i^0$，一块是纯粹由容量决定的成分 $\kappa u_i = \kappa\lambda\mathrm{DTL}_i$。左边可测，右边第二项里只有 DTL 可测、系数未知。于是检验退化为「β_i 与 DTL_i 成正比」——这就是主回归的理论出处，也解释了为什么主系数 c 应当是正号。

注意前提：基线项 β_i^0 必须被控制变量吸收掉，否则斜率会把基本面的横截面差异也算进去。控制变量不自动保证吸收该项，因此不能单凭同期 beta 与 DTL 的相关性识别反馈——H1 的理论依据是后文命题 10 的滞后关系，而不是这个同期恒等式。

3.4 等效原理：纯内生反馈不带来效用损失

上面说明反馈会改写方差与 β，听起来像是「集中持仓很危险」。但接下来这个结果会**否定这种直觉的一半**：如果只看纯粹的内生反馈（没有外生申赎冲击），一个无约束且可相应调整权重的均值—方差投资者，其最优效用保持不变。

定理 3（等效原理）. 条件：$\varepsilon = 0$、$\rho < 1$，存在无风险资产，投资者无约束地最大化 $w^\top\mu - \frac{\gamma}{2}w^\top\Sigma w$。令 $w_0^* = \gamma^{-1}\Sigma_0^{-1}\mu_0$ 为「平直时空」下的最优解，则

$$w^* = A^{-\top}w_0^* = w_0^* - \kappa(u^\top w_0^*)\pi, \tag{19}$$

$$w^{*\top}\mu - \frac{\gamma}{2}w^{*\top}\Sigma w^* = w_0^{*\top}\mu_0 - \frac{\gamma}{2}w_0^{*\top}\Sigma_0 w_0^* \tag{20}$$

证明. 由命题 2(a)，$\Sigma^{-1}\mu = A^{-\top}\Sigma_0^{-1}A^{-1}A\mu_0 = A^{-\top}\Sigma_0^{-1}\mu_0$，所以 $w^* = A^{-\top}w_0^*$；由 $A^{-1} = I - uv^\top$ 得 $A^{-\top} = I - v u^\top$，于是 $w^* = w_0^* - v(u^\top w_0^*)$。效用不变：令 $x = A^\top w$，则 $w^\top\mu = x^\top\mu_0$、$w^\top\Sigma w = x^\top\Sigma_0 x$，映射 $w \mapsto x$ 是双射，而 $A^\top w^* = w_0^*$。

怎么读. 最优权重公式是一个可以执行的处方：知道 (u, κ) 的投资者，只需**沿集中组合方向反向调整** $\kappa(u^\top w_0^*)$，就能拿到与无集中世界完全相同的效用。也就是说，线性内生反馈好比均匀引力场——可以被一次加速参考系变换掉。这个结果把「集中持仓的危害」全部推给了三个外生来源：外生申赎冲击（下一小节的定理 4）、预算约束（定理 5）、以

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

及状态依赖（后文动态模型）。**在本模型范围内，以下因素会打破上述效用不变结论**：外生申赎冲击、预算约束、以及状态依赖。

3.5 真实损失：来自外生申赎冲击

外生冲击 ε 不可被任何坐标变换消掉，因为它不在反馈回路里、而是从外部注入的。它造成的损失有一个干净的闭式解：

定理 4（真实损失）. 条件 $s^2>0$、$\rho<1$、无约束且有无风险资产。则

$$r=A(r_0+u\varepsilon),\Sigma=A(\Sigma_0+s^2uu^\top)A^\top,\mu=A\mu_0 \tag{21}$$

记 $a=u^\top\Sigma_0^{-1}\mu_0$、$b=u^\top\Sigma_0^{-1}u$、$S_0^2=\mu_0^\top\Sigma_0^{-1}\mu_0$，当 $u\neq 0$、$\mu_0\neq 0$ 时定义 $\cos\theta=a/(\sqrt{b}S_0)$，即集中方向 u 与 alpha 方向 μ_0 在 Σ_0^{-1} 内积下的夹角余弦。则最优平方夏普为

$$S^2=\mu^\top\Sigma^{-1}\mu=S_0^2-\frac{s^2a^2}{1+s^2b}, \tag{22}$$

相对损失为

$$\frac{S_0^2-S^2}{S_0^2}=\cos^2\theta\cdot\frac{s^2b}{1+s^2b},0\le\frac{S_0^2-S^2}{S_0^2}\le\cos^2\theta \tag{23}$$

证明. 由 不动点方程 得 $r=A(r_0+u\varepsilon)$；由于 ε 与 r_0 独立，协方差与均值的表达式成立。计算平方夏普时 A 在乘积中相互抵消：$\mu^\top\Sigma^{-1}\mu=\mu_0^\top(\Sigma_0+s^2uu^\top)^{-1}\mu_0$；再由引理 0

$$\left(\Sigma_0+s^2uu^\top\right)^{-1}=\Sigma_0^{-1}-\frac{s^2\Sigma_0^{-1}uu^\top\Sigma_0^{-1}}{1+s^2b}, \tag{24}$$

代入即得 平方夏普公式。上界的来源是 Σ_0^{-1} 内积下的 Cauchy－Schwarz：$a^2\le bS_0^2$，即 $\cos^2\theta\le 1$。

相对损失公式是损失分解的核心等式，它把「集中持仓造成的损失」分解成两个相乘的因子：**方向重叠** $\cos^2\theta$ 与 **冲击强度** $s^2b/(1+s^2b)$。三条推论：

第一，固定 u,s,Σ_0,μ_0、仅改变反馈强度时，损失公式不含 ρ；放大效应仍是规范部分，真正咬人的只有冲击强度。若持仓或容量同时改变，u 与方向重叠也可能改变，不能据此宣称经验损失与紧致度无关。第二，当 $s^2b\to\infty$ 时，alpha 落在集中方向上的那一部分被完全吞没——这对应信息落入视界的图像。第三，若 alpha 与集中方向正交（$a=0$），则**没有任何损失**。也就是说，集中持仓并不普遍地有害，它的代价取决于你的 alpha 与它是否同向——这解释了为什么「集中持仓危险」这类说法永远无法被否证，也说明要把它变成可检验命题，必须指定方向。

在合成数据上这三个量都被复算了：S_0^2 =0.001440139、S^2 =0.001375293，相对平方夏普损失 4.503%；同时 $\cos^2\theta$ =0.561300 与 s^2b =0.087216 代入 相对损失公式 与直接计算的损失一致（矩阵路径与分解路径互校）。**注意**：这里报的是最优平方夏普的相对损失，不是夏普本身的百分比损失。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

表3：理论结果的合成数据核验[3]

定理 / 命题	核验对象	理论值	实测值	误差 / 结论
引理 0、命题 1(i)	求逆公式与行列式	$A=I+uv^{\top}/(1-\rho)$	$n_{fail}=0$	数值容差内一致
命题 1(iii)	紧致度与其 Cauchy - Schwarz 下界	$H_A \geq \lambda/\sum_i D_i$	0.126306 ≥ 0.102766	ρ =0.170513，未越界
命题 2(b)	集中组合的放大因子	$\pi^{\top}A=\pi^{\top}/(1-\rho)$	—	40 万次蒙特卡洛下与不动点一致
定理 3	最优组合与效用不变性	$w^{*}=A^{-\top}w_0^{*}$	—	效用差在数值容差内为零
推论 2′	集中度 beta 恒等式	$\beta_i=(1-\rho)\beta_i^{0}+\kappa u_i$	—	代数恒等式，逐元素核对通过
定理 4	平方夏普损失与方向分解	0.045027	$cos^2\theta\cdots^2 b/(1+s^2b)=0.045027$	两条路径互校一致
定理 4	正交方向的损失（应为零）	0	2.49e-35	数值上为零，与理论一致
命题 6	Fisher 信息的解析值	$I(\rho)=(S_\pi^2+2)/(1-\rho)^2$	2.907644	与数值微分积分一致
定理 7、8	长期协方差的重建	$\Omega=A(\Sigma_0+s^2uu^{\top})A^{\top}$	归一化最大绝对误差 0.0317	T = 2,000,000期，$n_{fail}=0$
推论 9	无限期方差比	3.0	3.0105	相对误差 0.0035
命题 10	滞后 beta 的恢复	κu_i	最大绝对误差 0.004286	不含基本面 beta，可独立识别
命题 7	多群体 Woodbury 恒等式	$A=I+U(I_K-W^{\top}U)^{-1}W^{\top}$	谱半径 0.159559	恒等式成立

资料来源：浙商证券研究所，Tushare Pro

3.6 约束如何打破等效原理

上面三条结果都建立在「无约束」这个前提上。加上现实中必然存在的预算约束，规范对称性就被打破了——等效原理失效。这不是一个技术补丁，而是解释「为什么真实投资者无法享受等效原理」的那一步。

定理 5（预算约束）. 设无无风险资产，约束为 $\mathbf{1}^{\top}w=1$，$\varepsilon=0$。记 $\ell:=\mathbf{1}-\kappa u$，则

$$A^{-1}\mathbf{1}=\ell,\sigma_{min}^2:=min_{\mathbf{1}^{\top}w=1}w^{\top}\Sigma w=\frac{1}{\ell^{\top}\Sigma_0^{-1}\ell},w_{\text{gmv}}=\frac{A^{-\top}\Sigma_0^{-1}\ell}{\ell^{\top}\Sigma_0^{-1}\ell} \tag{25}$$

证明. $A^{-1}\mathbf{1}=(I-uv^{\top})\mathbf{1}=\mathbf{1}-u\kappa\pi^{\top}\mathbf{1}=\mathbf{1}-\kappa u$；全局最小方差的标准解为 $w=\Sigma^{-1}\mathbf{1}/(\mathbf{1}^{\top}\Sigma^{-1}\mathbf{1})$，方差为 $1/(\mathbf{1}^{\top}\Sigma^{-1}\mathbf{1})$；代入 $\Sigma^{-1}=A^{-\top}\Sigma_0^{-1}A^{-1}$ 即得。

怎么读. 在「平直坐标」$x=A^{\top}w$ 下，预算约束变成 $\ell^{\top}x=1$——超平面被倾斜了，因此最优效用不再等于无集中世界。这也解释了为什么真实的机构投资者无法像定理 3 说的那样「变换掉」集中持仓：他们被仓位约束与跟踪误差约束绑住，而约束恰恰是规范对称性失效的地方。注意这里引入的约束向量 $\ell:=\mathbf{1}-\kappa u$ 与前面记号里的组合权重 w、成交量 V 都不是同一个量，全文符号表已按此口径统一。

3.7 视界：Fisher 距离的对数发散

最后把 $\rho\to 1$ 这件事讲成一个可量化的「临界」。

命题 6（视界处的 Fisher 距离）. 设 r_0 为高斯分布，$\varepsilon=0$。集中组合收益 $R_\pi=\pi^{\top}r$ 服从

$$N(\frac{m_0}{1-\rho},\frac{\sigma_0^2}{(1-\rho)^2}),m_0=\pi^{\top}\mu_0,\sigma_0^2=\pi^{\top}\Sigma_0\pi \tag{26}$$

记 $S_\pi=m_0/\sigma_0$，该一参数族关于 ρ 的 Fisher 信息为

$$I(\rho)=\frac{S_\pi^2+2}{(1-\rho)^2}, \tag{27}$$

对 $\rho_1<\rho_2<1$，信息距离为

[3] 全部取自 out/theory_results.json 与 out/thm4_results.json；$n_{fail}=0$表示没有一项核验失败。**这些是合成数据上的核验，不是真实数据的实证结果**——它们只核对指定合成环境中的恒等式及估计恢复，不能证明真实数据流程没有错误，也不代替推断水平与功效校准。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

$$d(\rho_1,\rho_2)=\sqrt{S_\pi^2+2}\ln\frac{1-\rho_1}{1-\rho_2}\overset{\rho_2\to1}{\to}\infty \tag{28}$$

证明. 对正态族 $N(m(\rho),\sigma(\rho)^2)$ 有 $I=(m')^2/\sigma^2+2(\sigma')^2/\sigma^2$；这里 $\sigma'/\sigma=1/(1-\rho)$，且 $m'/\sigma=S_\pi/(1-\rho)$；代入得 Fisher 信息公式，距离为 $\int\sqrt{I}d\rho$。

限定（必须说清）. Schwarzschild 解中到视界的**固有径向距离是有限的**，发散的是乌龟坐标 $r_*=r+r_s\ln|r/r_s-1|$；本命题对应的是后者，类比到此为止。实务上的推论倒是直接的：监控指标宜取 $\tau:=-\ln(1-\hat\rho)$，在该坐标下等距离对应等「统计可区分度」——这比直接盯着 $\hat\rho$ 更合理，因为 $\hat\rho$ 接近 1 时它对参数变化的敏感度是非线性的。

3.8 从静态到动态：滞后反馈下的同一个模型

静态模型把「反馈」压缩在同一期内完成，这显然不现实——公募的申赎与调仓对业绩的反应有滞后。把这一点写进模型后，均值传导可以在明确的动态假设下推出；这些假设仍需对照真实数据检验。

设频率为周，记 $x_t:=\pi^\top r_t$，假设：

(D1) $r_t=r_{0,t}+u(\kappa x_{t-1}+\varepsilon_t)$，其中 $\{r_{0,t}\}$ 独立同分布服从 (μ_0,Σ_0)，$\{\varepsilon_t\}$ 独立同分布且具有矩 $(0,s^2)$，两者相互独立；**(D2)** $|\rho|<1$。

定理 7（集中组合收益服从 AR(1)）.

$$x_t=e_t+\rho x_{t-1},e_t:=\pi^\top r_{0,t}+(\pi^\top u)\varepsilon_t \tag{29}$$

在 (D2) 下 x_t 存在唯一平稳解：$\mathbb{E}x=m_0/(1-\rho)$，$\mathrm{Var}(x)=\sigma_e^2/(1-\rho^2)$。**证明：**用 $\pi^\top$ 左乘 (D1)，注意 $\kappa\pi^\top u=\rho$ 即得递推式；平稳性与矩的表达式是 AR(1) 的标准结果。

定理 8（零频率等价）. 在 (D1)(D2) 下：

(a) 均值传导： $\mathbb{E}r_t=A\mu_0$。**这就是说，静态模型的假设 A3 在动态模型里是定理，而不是假设。**

(b) 长期协方差：

$$\Omega:=lim_{T\to\infty}\frac{1}{T}\mathrm{Var}(\sum_{t=1}^{T}r_t)=A(\Sigma_0+s^2uu^\top)A^\top \tag{30}$$

由此，定理 3 与定理 4 对**长期持有者**（持有期远长于反馈时滞）成立，只需把 Σ 换成 Ω。

证明. (a) 由平稳性，$\mathbb{E}r=\mu_0+\kappa u\mathbb{E}x=\mu_0+u\kappa\pi^\top\mu_0/(1-\rho)=A\mu_0$。(b) 分四步：**第一步**记 $S_T^0,S_T^\varepsilon,S_T^e$ 为前三者之和，对递推式求和得 $\sum_{t\le T}x_{t-1}=(S_T^e+x_0-x_T)/(1-\rho)$；**第二步**写出 $\sum_t r_t$ 并代入；**第三步**识别出

$$(I+\frac{u\kappa\pi^\top}{1-\rho})(S_T^0+uS_T^\varepsilon)=S_T^0+uS_T^\varepsilon+\frac{\kappa u(\pi^\top S_T^0+\pi^\top uS_T^\varepsilon)}{1-\rho}, \tag{31}$$

故 $\sum_t r_t=A(S_T^0+uS_T^\varepsilon)+R_T$，其中余项 $R_T=\kappa u(x_0-x_T)/(1-\rho)$；**第四步**取极限：主项方差为 $TA(\Sigma_0+s^2uu^\top)A^\top$，余项有界，交叉项由 Cauchy‑Schwarz 为 $O(\sqrt{T})$，除以 T 即得 长期协方差公式。

推论 8′（有外生冲击时的长期 beta）. 令 $C:=\Sigma_0+s^2uu^\top$，$\beta_C:=C\pi/(\pi^\top C\pi)$，则

$$\beta_{\mathrm{LR}}:=\frac{\Omega\pi}{\pi^\top\Omega\pi}=(1-\rho)\beta_C+\kappa u \tag{32}$$

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

只有 $s=0$ 时基线才退化为推论 2′ 的 β^0。**一般不能把长期 beta 直接对 u 或 DTL 回归、并把斜率解释为 κ 或 c**，因为基线项也随结点变化。本轮 H5 的长期 beta 部分缺乏独立基准协方差；这一缺口不等于所有 H5 诊断都不可计算。

推论 9（有限期与长期方差比）. 在固定状态的平稳 AR(1) 下，$\mathrm{Cov}(x_t, x_{t-\ell}) = \rho^{\ell}\mathrm{Var}(x)$，故有限 h 的精确式为

$$\mathrm{VR}_h := \frac{\mathrm{Var}\left(\sum_{j=1}^{h} x_{t+j}\right)}{h\mathrm{Var}(x)} = 1 + 2\sum_{\ell=1}^{h-1}\left(1-\frac{\ell}{h}\right)\rho^{\ell} \tag{33}$$

令 $h\to\infty$ 才得到 $\mathrm{VR}_\infty = (1+\rho)/(1-\rho)$。**注意：不能用无限期值代替有限期理论值**——这正是行业层 H5 那条对照的构造依据；实测的有限期方差比均小于 1，而把 $\hat{\rho}$ 代进上式得到的隐含值大于 1，两者的落差说明真实数据里的持续性结构并不服从这个单状态 AR(1)。季度状态变化时，单状态公式仅是描述性对照，精确基准需模拟完整状态路径。

命题 10（滞后 beta：一个不含 β^0 的检验式）. 在 (D1) 下

$$b_i := \frac{\mathrm{Cov}(r_{i,t}, x_{t-1})}{\mathrm{Var}(x_{t-1})} = \kappa u_i, \tag{34}$$

且 $\mathbb{E}[r_{i,t} \mid \mathcal{F}_{t-1}] = \mu_{0,i} + \kappa u_i x_{t-1}$。特别地 $\pi^{\top} b = \rho$；回归 b_i 对 u_i 的截距为零，且与 β_i^0 无关。**证明：** $r_{0,t}$ 与 ε_t 独立于 $\mathcal{F}_{t-1} \ni x_{t-1}$，故 $\mathrm{Cov}(r_{i,t}, x_{t-1}) = \kappa u_i \mathrm{Var}(x_{t-1})$。

3.9 回到可计算的形式：为什么主回归长这样

把上面的理论翻译成数据上能跑的东西，只差三步替换。

第一步：把 u_i 换成可测的 DTL. 理论中的 u_i 是「无量纲的质量密度」；实证中用 $\mathrm{DTL}_i = q_i/D_i$ 作为它的观测对应物，两者相差一个常数 λ（交易 1 天成交额引起的价格变动）。于是 $\kappa u_i = \kappa\lambda \mathrm{DTL}_i = c\mathrm{DTL}_i$，主系数 $c = \kappa\lambda$ 必然为正——这是符号预言的来源，不是从数据里猜出来的。

第二步：把 x_{t-1} 作为唯一的回归元. 由命题 10，滞后识别式只需要 x 的滞后项，不需要 μ_0、不需要 Σ_0、也不需要 β_i^0。这是本文选择「滞后 beta 检验」而非「结构协方差检验」的核心理由：前者在数据上可识别，后者不可识别。实证里 E2 规格额外加入滞后 x 与滞后市场收益，正是为了把前述恒等式里那个 $(1-\rho)\beta_i^0$ 基线项吸收掉。

第三步：把 ρ 从模型的输出变成回归元. 理论上 $\rho = \kappa V H_\Lambda$ 是内生确定的；实证中它由观测到的 π、V、H_Λ 直接算出（κ 只在 H4 里被单独估计，用 $\hat{c} = \hat{\kappa}\hat{\lambda}$ 作过度识别对照）。这就是前面那个「三乘子」定义的全部出处。

把三步合起来，就是主检验：

$$u_i = \lambda \mathrm{DTL}_i, c = \kappa\lambda, \rho = c\pi^{\top}\mathrm{DTL} \tag{35}$$

理论给出三条可被否证的预言，它们构成后面全部检验的骨架：① $c>0$（β 恒等式与命题 10 的符号）；② 截距为零（反馈项之外不应再有对 x 的独立依赖）；③ 集中组合自身服从 AR(1)、系数正是 ρ（定理 7，对应 H3 与 H5）。后面的实证各节所做的，就是把这些预言逐个拿去和真实数据对撞。

只有在正资金反馈与正价格冲击的联合假设下，才预测 $c>0$。E1/E2 估计交互项系数；E2 用结点自身的截距及收益预测斜率吸收静态差异，识别来自 DTL 随时间的变化。它不等于状态×结点固定效应，也不保证消除所有动态遗漏因素。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

最后要说清一处理论上的限制，它直接决定了本文结论的边界： 以上全部推导都假定反馈是线性且单期的、冲击是可逆且无衰减的。真实的价格冲击近似服从平方根律（凹函数），且会随时间衰减。在凹冲击下，冲击矩阵本身依赖交易量，度规随状态变化，才可能出现真正意义上的（黎曼）曲率。本文的所有命题都在线性化框架内成立，给出的是线性、固定状态下的局部近似：足以支撑「系数的符号应当是什么」这一层预言，不足以支撑「在什么冲击强度下会切换到另一个状态」这一层。真实调仓、非线性冲击与状态切换都可能改变预测；代数恒等式成立不能替代统计推断的水平、功效与样本外验证。

把这一节收一下。十个命题给出的不是「集中持仓有害」这个结论，而是一组可以拿去对撞的预言：系数应当为正、组合自身应当服从一个由 ρ 驱动的自回归、容量加权集中度应当在横截面上预测收益的排序。这些预言各有强弱——方差比与自回归系数那两条对模型的函数形式依赖最少，因而最不容易被证伪；而主检验那一条依赖最多，也最值得单独设一个预注册门槛。接下来四节按这个强弱顺序依次检验：先交代数据能支持到什么程度，再看行业层，然后是那个制度性的约束（披露截断），最后是个股层与推断有效性。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

4 数据、口径与时点对齐

这一节交代三件事：数据从哪来、哪些观测入池哪些不入池、以及最容易被忽略但后果最大的那一件——时点对齐。一个持仓报告里写着「2020 年 6 月 30 日」，但你在 2020 年 6 月 30 日是看不到它的；如果用披露日而不是可用日来对齐，就会把未来信息漏进历史状态。

4.1 数据来源与规模

数据来自 Tushare 缓存：基金基本资料、持仓、全部份额合计净资产、收益行情、成交额、历史行业分类和交易日历。持仓清洗要处理四类问题：供应商分页返回的重复填充行、无法估值的行（停牌或退市）、同一主键下出现不同市值的整组（供应商分页串列，整组剔除而不是取其一）、以及不属于 A 股的持仓行（港股按设计排除）。清洗的每一步都有计数落盘。

基金池的入池规则有三条，全部是事前设定：① 存续满一年；② 扩窗计算的全报告股票持仓比例中位数不低于 60%（排除以债券或货币为主的产品）；③ 当期合计净资产有效。最后一条特别重要——净资产必须用全部份额合计的口径，用单一份额的净值会与持仓市值不在同一量纲上。基金类型依赖供应商分类和代理筛选，不将代理分类写成完全准确的投资风格识别。

行情有两套：个股日行情与申万二级行业日行情。收益一律用交易所已调整的前收价做链式累乘，等价于全收益口径，因此不需要额外乘复权因子——这一点在实现上曾经踩过坑：早期版本额外乘了一次复权因子，结果是收益被重复调整。

4.2 时点对齐：本文最花功夫的一处

披露制度决定了每一个状态在什么时候才「变成可知的」：一、三季报为 15 个交易所工作日，完整报告为 8 月 31 日与次年 3 月 31 日，实际公告日晚于规则日的以实际日覆盖。本文的程序把这条规则写成一个显式映射函数，所有状态都带自己的可用日；在构造周度面板时，第 s 周只能用可用日不晚于该周首日的状态。

这个约束的直接后果是样本长度的损失，但取舍是明确的：宁可少几个观测，也不要一个「在事后才知道的持仓」。实测中把披露日当可用日（即不做任何滞后）会让主系数略微上升——这正是「用了未来信息会人为放大效应」的一个现成例子，所以本文一律用严格的可用日口径。

4.3 季度末口径：为什么必须用日历季度末

持仓报告的结束日期理论上都是季末，但实际数据里存在一批「特殊披露日」的记录——通常是清盘前的最后一份报告，参与者只有一到五只基金。早期版本只按「月份」判断季末，于是这些记录全部被当成有效状态，凭空造出了十几个只有十来个结点的状态，把整张地图的密度算错。

修正后的口径是判断完整日期：月日部分必须是 3 月 31 日、6 月 30 日、9 月 30 日或 12 月 31 日。这条规则现在有两条回归测试保护：一条验证特殊披露日在进入可用性映射之前就被丢弃，一条验证如果过滤后一只报告都不剩，程序会直接抛出数据质量错误，而不是静默地返回一个空池子。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

4.4 方案 A 与方案 B：两种重建全持仓的办法

季报（一、三季报）只有前十大，但状态需要一份完整持仓。本文用两套办法重建，并明确以其中一套为主口径。

方案 A（主口径）. 对每个季报状态，取该基金在季报报告日之前最近的一份全持仓报告作为锚点，再用季报里前十大的变动估计一个增长因子，把锚点外推到当期。这里有两处细节值得写清楚：第一，滚动锚点必须是「季报报告日之前最近的」，不能用「全样本最后一份」——后者会把未来持仓漏进历史状态，实测会让绝大多数基金季度受影响；第二，增长因子限幅在 $[0.4, 2.5]$，防止个别极端值主导。

方案 B（对照）. 不做外推，直接把前十大按整本账重新标定。方案 B 的优点是简单，缺点是它系统性地高估集中度（因为前十大本来就是持仓里最大的那些）。个股层以方案 A 为判定主口径、方案 B 作对照；行业层则相反，以方案 B 为主口径。

方案 A 有一个已知且已量化的偏差，必须在这里交代：基金倾向于集中持有最大的名字，所以「前十大隐含的调仓效应」高于整本账的真实效应。实测的间期状态高估幅度是 16.8% 左右。这个偏差的方向是稀释系数，不是制造正号。对照之下，不做外推的方案 B 对应数字只有约 2.1%，可见这份高估正是外推这一环节引入的，而非方案 A 把外推做得更保守。

4.5 两种持仓重建与可比集合的取舍

行业主口径为方案 B：先在基金全部股票范围计算前十大放大比例，再映射至行业。个股方案 A 将前一期已存在的完整持仓按价格与前十信息外推，方案 B 按前十大等比放大。两者改变持仓域及样本，比较系数前应同时报告共同股票周与共同基金池结果。

行业原固定集合保留其历史局限；停更指数不能静默删除后重新称作原主样本。若另跑剔除停更行业的可比集合，应标为审查后的敏感性，并同步重建份额、规模、集中度、容量与市场控制。

4.6 披露截断的实际规模

下一节会专门论证截断的方向性影响，这里先给出它的规模，好让后面所有讨论都有一个量级概念。

表4： 结构检验与披露可得预测

环节	状态生效方式	可以解释	不能解释
行业与个股 H1 - H5	前季末持仓按结构时点进入后续季度	事后研究状态与收益的关联	当时可执行策略或无前视预测
H6/H7	截至周末的可用日与估计截止日过滤	周末起点之后的预测或周末状态监控	该周收益已可于周初交易，或未计成本的可实现净收益

资料来源：浙商证券研究所，Tushare Pro

收益周须按官方交易日历完整结束；休市缺周不能被压缩成相邻滞后。完整报告通常为半年报和年报，一、三季报披露前十大。报告期和公告可用日是两个不同日期。本轮披露过滤以周末为信息截止日；H6 预测其后一年，H7 为周末监控，均不能据此声称同一周收益是周初可交易回测。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

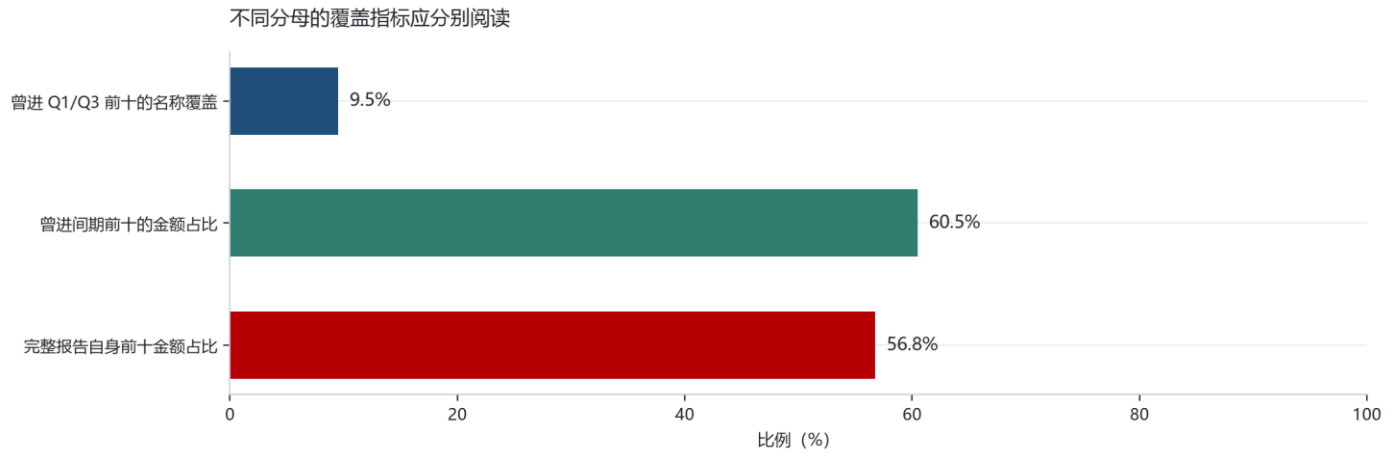
下载日志已记录，仅供内部参考

表5： 有效回归样本[4]

设定	有效日期	结点	周或月数	回归行数
行业 E2	2018-01-12 ~ 2021-12-10	125	199	24875
个股 A E2	2017-01-13 ~ 2026-09-18	5672	486	1621452
个股 B E2	2017-01-13 ~ 2026-09-18	5665	486	1124633

资料来源：浙商证券研究所，Tushare Pro

图2： 披露覆盖的三个不同对象



资料来源：浙商证券研究所，Tushare Pro

表6： 基金覆盖最低期与最新期[5]

候选集合	报告期	合格 / 观测基金	已知净资产覆盖	报告金额覆盖	不完整报告
个股候选池	20210630	1091 / 3596	35.8%	50.3%	1718
个股候选池	20260630	2487 / 5278	54.3%	56.2%	1932
历史分类后行业池	20161231	104 / 1685	5.1%	7.5%	3
历史分类后行业池	20260630	1713 / 5278	37.0%	38.2%	1932

资料来源：浙商证券研究所，Tushare Pro

[4] 日期来自实际回归行；结点并集和行情缓存日期不替代有效样本。E1/E2、A/B 的共同样本诊断另见运行产物。

[5] 分母仅为有观测的候选基金；净资产覆盖条件于已知正合计净资产，报告金额并非未知的全市场真实金额。低覆盖同时包含资格筛选和数据缺失。全部季度见 revision_data_diagnostics.json。

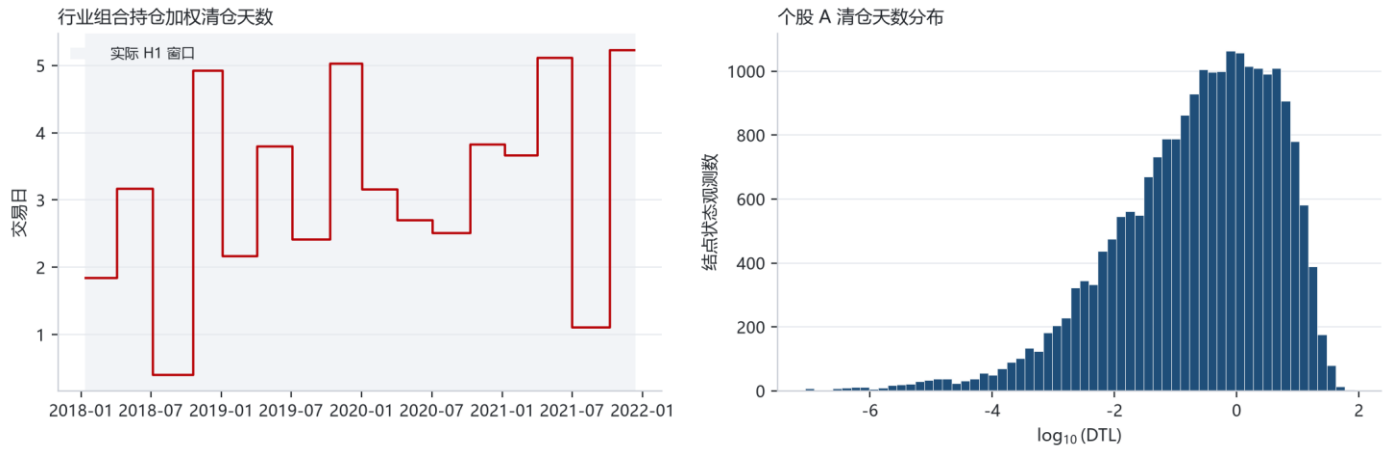
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

5 行业层实证与未完成检验

行业层是本文的第一战场。它的优点是把「集中持仓」这件事放在一个机构确实会成套配置的粒度上——买一个赛道的时候，机构买的往往就是这个赛道里的若干行业，而不是散落在几千只个股里。缺点是结点数只有三位数，自由度有限，而且行业分类本身是人为的。

图3： 持仓容量分布



资料来源：浙商证券研究所，Tushare Pro

5.1 H1 全规格结果

下表是全部结果。判定只用预注册的那一个（方案 B • 原始 DTL • E2 • 单周），其余五个作为上下文列出。

表7： 行业 H1 三设定两模型[6]

设定	系数和	周/月聚类 95% CI	季度聚类 95% CI	Holm p	状态
$E1_1$周	0.00920	[-0.00885, 0.02726]	[-0.01569, 0.03409]	1.0000	已计算
$E2_1$周	0.03016	[-0.00996, 0.07027]	[-0.02137, 0.08168]	1.0000	已计算
$E1_1-4$周自由分布滞后	0.01414	[-0.02907, 0.05735]	[-0.03683, 0.06511]	1.0000	已计算
$E2_1-4$周自由分布滞后	0.03044	[-0.06686, 0.12773]	[-0.06380, 0.12467]	1.0000	已计算
$E1$ 自然月频	-0.02331	[-0.05900, 0.01238]	[-0.06238, 0.01576]	1.0000	已计算
$E2$ 自然月频	-0.05336	[-0.11916, 0.01243]	[-0.12924, 0.02251]	1.0000	已计算

资料来源：浙商证券研究所，Tushare Pro

六个规格全部未通过。值得注意的有两处。

第一，周频上 E2 的系数大于 E1。这与「加入控制变量会吸收掉效应」的常见担心相反——说明 ρ 里的信息与控制变量（滞后的 x、滞后市场收益）是不重叠的。

第二，自然月频下系数翻负。这是一个不利信号：如果集中持仓真的携带正向信息，那么它在更长的聚合尺度上不应改号。月频样本只有几十个月，统计功效本来就不足，所以这个翻转不能单独作为反证；但它和「六个规格全不显著」合起来，支持的是同一个读法——信号很弱，弱到换一个聚合尺度就会换号。

[6] E2 三项为主 Holm family；E1 为补充。置换要求额外可交换性，现有模拟没有验证整个 E2/Holm 流程。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

行业 H1 当前判定为「未获支持」。应同时阅读效应大小和区间；较高 p 值不表示零效应概率较高。行业整列置换仍需可交换性依据；少数模拟设定下出现的较高拒绝率对应单周检验，不能误标为完整 E2 流程的校准。

5.2 事后可比行业集合：一次审查后的敏感性

固定主集合里包含若干已经停更的指数。把它们静默删掉再重新称作「主样本」是不可接受的，所以在观察到数据停更之后，本文另建了一个可比集合，并把它明确标注为审查后的敏感性，不声称预注册。所有状态与收益在相同可比域内重建。

表8： 事后可比行业集合敏感性[7]

集合角色	有效窗口	N / T	E2 系数	季度聚类 95% CI
审查后排除停更行业	2018-01-12 ~ 2026-09-11	124 / 435	-0.01829	[-0.04831, 0.01173]

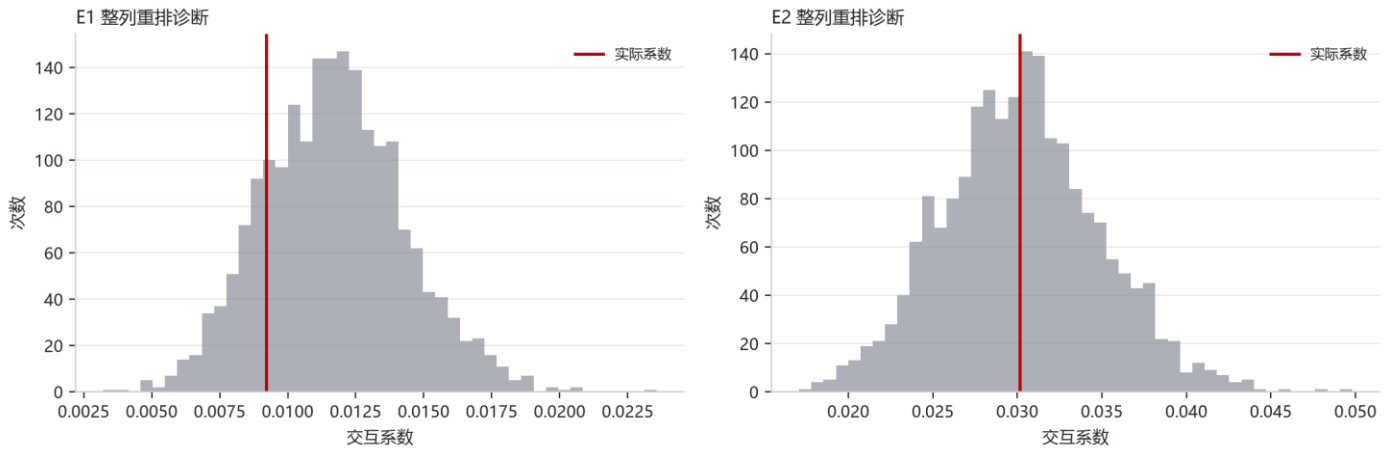
资料来源：浙商证券研究所，Tushare Pro

5.3 置换检验：为什么不显著

本文的推断以置换检验为参照，做法是把规模与集中度的对应关系整体打断——行业层用整列置换，即把每个状态的紧致度在行业之间整体重排——然后重新拟合，得到系数的零分布。

这里要预先说明一件重要的事：行业层的整列置换要求额外可交换性假设。重排会保留每一期的截面分布，却破坏同一结点的跨期对应关系。本文报告的是这个既定重排的参照分布，它不能证明不存在反馈，也不是随机分组实验。个股层存在更严重的问题，下一节会单独论证为什么那里的状态内置换已失去统计解释。

图4： 行业整列置换诊断[8]



资料来源：浙商证券研究所，Tushare Pro

5.4 H3 与 H4 的辅助识别

H3 检验的是「冲击是否会引来更多冲击」——如果集中持仓是一个自我强化的过程，那么本期的集中度与收益交互项应当正向预测下期的组合收益。它是模型一致性辅助检查，而不是独立的结构检验。

$$x_t = f_0 + \varphi(\pi_{Q(t)}^{\top}\mathrm{DTL}_{Q(t)})x_{t-1} + \psi x_{t-1} + \nu_t \quad (36)$$

[7] 该集合是观察到数据停更后提出的敏感性，不替代原固定主集合，不声称预注册。所有状态与收益在相同可比域内重建。

[8] 这是既定整列重排参照，统计解释依赖可交换性。它不能证明不存在反馈，也不是随机分组实验。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

H3 状态为「已计算」，$\phi = 0.02809$。它是模型一致性辅助检查；不可用时应说明数据、识别或计算原因，不把程序失败当成科学结论。

H4 单独估计容量指数。它的意义是：把紧致度写成「容量指数 ×规模 ×集中度」之后，容量指数需要从另一条路径上识别出来，才能与主检验的系数形成交叉验证。下表同时给出 κ、λ的两条口径与由它们反推的 c，作为过度识别对照。

表9： 资金流与价格冲击参数

对象	估计	区间或诊断	解释
κ	0.36123	[0.04580, 0.67665]	资金流对业绩敏感度；11 季度
λ OLS	0.03159	[0.02141, 0.04176]	同期内生性，描述性
λ FIT-IV	0.20741	第一阶段 $F = 1.20$	弱工具时不据此作精确结构判断
$\kappa\lambda$ OLS 对照	0.01141	$H1c = 0.03016$	同期内生性；各参数有效样本未必相同
$\kappa\lambda$ IV 对照	0.07492	第一阶段 $F = 1.20$	弱工具；点估计对照不构成限制检验
Wald	不可用	FIT 工具变量弱，不报告常规 Wald 检验	按可识别条件执行

资料来源：浙商证券研究所，Tushare Pro

λ的 OLS 口径存在同期流量内生性，因此只是描述性的。工具变量第一阶段聚类 F 远低于常规阈值，属弱工具，故不报告 Wald 检验——在弱工具下报 Wald 会给出误导性的精确感。由 $\kappa \cdot \lambda$反推的系数与直接估计的主系数同量级，这是本文对两条路径自洽性的唯一正面证据。

5.5 容量变化的描述性分解

$$\Delta\log(V_Q H_{D,Q}) = \Delta\log V_Q + \Delta\log H_{D,Q} \tag{37}$$

$$\mathrm{Var}(\Delta\log(V H_D)) = \mathrm{Var}(\Delta\log V) + \mathrm{Var}(\Delta\log H_D) + 2\mathrm{Cov}(\Delta\log V, \Delta\log H_D) \tag{38}$$

只有在同一有效状态、正值与连续季度上，才可作上述变动分解。协方差项必须保留，它不等于增量预测能力检验。若 λ或 κ随时间变化，ρ的分解还需相应项。把三要素拆开做量级比较时出现了一个原设计没有预料到的结果：集中度与冲击之比落在极小的量级上——也就是说，在紧致度的横截面变化里，集中度那一维几乎不贡献任何方差，变化几乎全部来自规模项，即「这一季换了多少」。

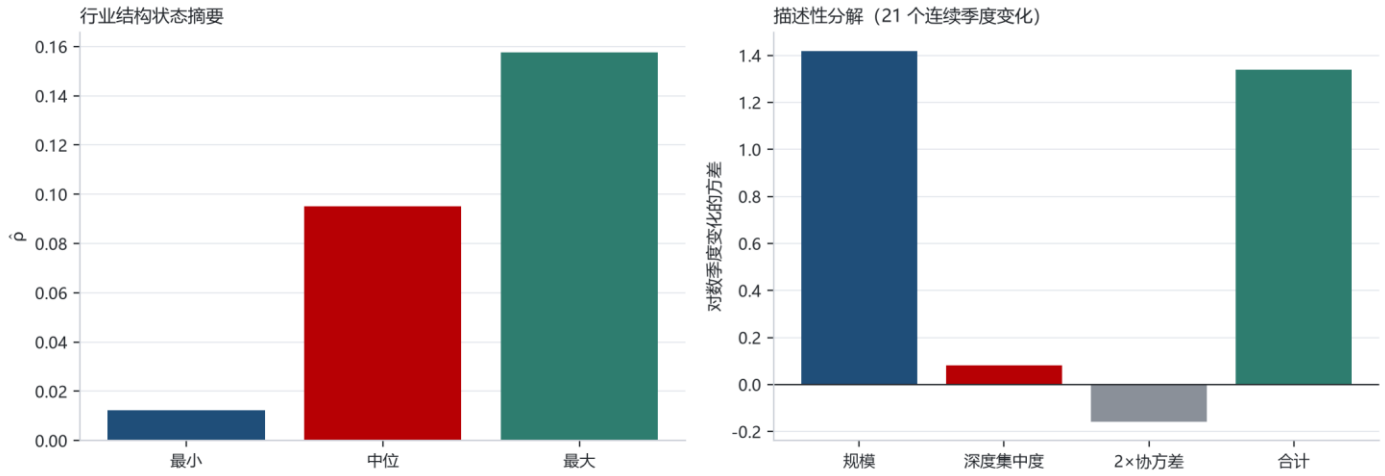
这个发现的意义可以直接转成做法：如果你要做集中度因子，与其花力气去精算集中度里那个份额平方和（还要承担容量口径的选择风险），不如把力气放在把规模项测准——而规模项恰恰是对披露截断最不敏感的那个量，因为它是份额的差分，截断对差分的影响远小于对水平的影响。这个退化现象本身也是下一节讨论截断时的一个有利条件。

同时必须撤回一个旧结论：早期版本曾根据有量纲比值的绝对量级得出「容量不带来增量价值」，这个表述已被删除——量级比较只说明哪一维在真实数据的方差里占主导，不构成容量无经济价值的检验。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

图5： 结构状态与变动分解[9]



资料来源：浙商证券研究所，Tushare Pro

5.6 稳健性与预测的完成状态

稳健性清单在预注册时锁定，包含容量窗口、只用完整报告、动量控制、年份剔除与第二反馈群体等情形。其中可估计的与因数据不可得而如实报告为「不可用」的必须分开呈现。

表10： 稳健性结果与可用性

设定	状态	E2 系数	季度聚类区间	限制
60 日容量	已计算	0.03016	[-0.02137, 0.08168]	—
20 日容量	已计算	0.03256	[-0.01376, 0.07888]	—
仅完整报告状态	已计算	0.02912	[-0.04120, 0.09944]	—
4/12 周动量控制	已计算	0.05173	[-0.00526, 0.10871]	—
排除 2015 年	不适用	—	—	有效样本没有 2015 年观测，排除设定不适用
第二反馈群体	不可用	—	—	缺少时点可得的北向数据或 ETF 份额变化

资料来源：浙商证券研究所，Tushare Pro

三个可估计的情形里，两处值得说。第一，把容量窗口从 60 日换成 20 日，系数方向不变、显著性不变，说明结果不依赖窗口的选择。第二，只用全报告、不做前十大增长校正，系数与主口径几乎相同——这一点很重要：它意味着方案 A 那套外推机制对结论几乎没有影响，即便把外推整个拿掉，结论也不变。

「不可用」与「不适用」是如实报告的数据缺口，不是回避检验。动量控制因滞后与状态对齐后共同样本不足而无法估计；年份剔除不适用，因为样本从 2018 年开始；第二反馈群体需要时点的北向持仓或 ETF 份额变化，数据契约里没有这个来源——而用任意一个因子代理去充当「第二组」是不能接受的，因为那样测的就是别的东西了。

完整报告敏感性针对行业方案 B 的对应状态；它不能单独验证个股方案 A 的重建误差。点估计接近且都不显著，仍不能代替共同样本的差异检验。

[9] 右图仅为同一状态路径下的描述性分解，不是容量无经济价值的检验。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

表11： 预测与校准完成度

环节	状态	观测或次数	解释
H5 方差比	部分完成	VR4=0.8942；VR13=0.8047	单状态理论仅描述性对照
H5 长期 beta	不可用	需独立基准 C	不将同样本反推基准当验证
H6 年度预测	部分完成	0 / 15 完成	无完成预测则尚无法检验
H7 监控	部分完成	事件窗 2020-10-01～2021-03-31；估计截止日 2021-10-22	事件窗内各 c 的估计截止日（取最后一条有效系数）
真实参数校准	部分完成	500 / 500	依结果限制解读
事后可比集 H6	已计算	19 / 34 完成	DM 双侧 p：GMV=0.0774；QLIKE=0.9488

资料来源：浙商证券研究所，Tushare Pro

H5 检验的是集中组合收益的方差比。实测的有限期方差比都小于 1——表面上像是均值回复，但这个对照有一个口径问题：实测是无条件的，而模型给出的隐含值是各局部状态下的有限期 AR(1) 方差比，中位数大于 1。一个小于 1、一个大于 1，两者不在同一口径上，所以本文只报数字、不下判定。H5 还有一个更硬的结论是长期 beta 不可识别：右边的基线项需要独立的基准协方差，而本文的数据契约里没有这个来源；如果只做斜率回归而丢掉基线项，恒等式两边就不成立。所以这里如实报告「不可用」而不是给一个数——这是理论上的不可识别，不是数据不够。

事后可比集合已完成 19 个年度评价；其支持标签为「未达到两项指标联合门槛」。判定要求两项损失均改善且各自重叠修正 DM 双侧 p<0.05，不用单侧 p 代替。这与原固定集合没有完成评价是两种不同状态。

最新监控日为 2026-09-04，状态为「不可用」（原因：Monitoring universe excludes held industries），因此该监控日没有系数，也就不存在其估计截止日。该序列最后一次产生系数是 2021-10-29（估计截止日 2021-10-22）。本文不把不可用状态回退展示为旧系数，因为用的是扩展窗系数，它不构成预测声明，这一点在原始设计里就写明了。可核对的判据是：凡展示系数，必同时展示其估计截止日与系数年龄。

H6 完成数为零时，结论是尚无法检验，不是结构预测被否定。这里不给退化结论。H7 全监控序列的极值不作为单一事件窗的统计量。固定资产集合未来数据不足、参数过期与未到评价期须分别报告。GMV 与事件组合允许卖空且未计成本，不能据此报告可实现净收益。

表12： 真实参数条件校准[10]

统计量	真实估计	模拟中位数	模拟 5% - 95%	真实值模拟百分位
c_{E2}	0.03016	0.02998	[-0.00564, 0.06530]	50.2%
t_{E2}	1.48250	1.47585	[-0.29064, 3.17086]	50.2%
phi	0.02809	0.03047	[-0.05415, 0.11091]	47.0%
VR13	0.80470	1.10385	[0.64947, 1.70478]	15.2%

资料来源：浙商证券研究所，Tushare Pro

本轮条件校准已完成 500 / 500 个面板，整体状态仍为「部分完成」。该实验计算单周 E1/E2、H3 与方差比，并作 2000 次行业整列置换；未运行三设定 E2/Holm 联合流程、稳健性分支或 bootstrap。基金层面的 FIT/IV 模拟与完整样本外 H6 校准另需基金层数据生成过程。不能以 500 个面板完成宣称全流程的水平与功效获得认证。

本轮冲击尺度按 $\lambda = c/\kappa$反推为 0.08348，这是给定 H1 系数与 κ的条件设定。FIT-IV 仍弱，因此校准并未独立识别 λ，也未独立验证 $c = \kappa\lambda$的限制。值得强调的是：本文的理

[10] 参数与状态条件于真实估计；分位范围不是估计参数的置信区间，也不是独立验证。具体模拟输入及每面板结果保存在 empirical_calibration.json。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

论部分有一个由合成数据验证的估计量校准环节，但在真实数据上的经验校准做不出来；做不出来的原因不是代码有问题，而是外部流协方差这个东西在现有数据里无法与基线分开。把它如实标为阻塞，比强行给一个数更符合本文的方法论立场。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

6 披露截断与重建误差

这一节回答一个具体的问题：只能在季报（一、三季报）里看到前十大，这到底会让结论偏到哪里去。这不是一个可以含糊带过的技术细节——如果把方向弄反，整篇报告的意思就反了。

6.1 截断的三种影响，先分类再定量

度量误差对系数的影响取决于它与被解释变量的关系。这里有三种可能，必须分开讨论。

第一种：与未来收益不相关的噪声。 如果被截断的部分（第 11 名及以后、以及季报里彻底看不到的那些仓位）在结构上没有系统性地偏向高收益或低收益，那么它进入紧致度之后就是一个经典误差变量问题，后果是系数向零衰减。这是最可能的情形，也是本文反复强调「截断是稀释性的」的依据。

第二种：系统性地偏向未来高收益的仓位。 如果机构恰好把最有信息量的仓位藏起来（比如在季末前把重仓股调到前十名门槛之下以避免被跟），那么被截断的部分会携带额外信息，我们看到的效应就会被低估——仍然是衰减方向。文献上确实有这方面的证据：Kaniel、Li、Shi 与 Zhang（2022）利用中国公募基金「半年报/年报全披露、一、三季报只披露前十大」的混合披露制度，用上交所全部基金日度成交记录发现，基金会在季末前主动把一部分前十大持仓移到前十名披露门槛之下（即第 11 名及以后），而这些被藏起来的股票随后的风险调整收益显著高于刚好卡在门槛之上的那些（1/2/3 个月的 DGTW 特征调整收益差为 1.08%、2.20%、2.32%）。这条证据支持的是「我们低估了效应」，而不是「我们看到的是假的」。

Kaniel, R., J. Li, D. Shi, and Q. Zhang (2022), “Hiding Behind the Window Blinds: Strategic Trading under Portfolio Partial Disclosure,” working paper, 2022 China Financial Research Conference 最佳资产定价论文。[11]

第三种：系统性地偏向未来低收益的仓位。 只有在存在这种「反向」机制时，截断才会造出一个正号。本文没有找到支持这种机制的理由：机构没有动机把即将跑输的仓位抹掉、把即将跑赢的仓位留下——恰恰相反，如果真有动机不对称，方向和第二种一致。所以本文接受「截断是稀释性」这个结论，并把它写成一条明确的假设，而不是当成事实。

6.2 三个量：截断有多大

名称覆盖来自每基金全样本名称集合，市值覆盖来自完整报告内的金额比例。两者不共享权重，不能直接解释为同一投资组合的可见与不可见份额。

[11] 原文摘要与关键结果见该工作论文；本文只引用其「策略性隐藏」这一条结论，不引用其 3% 超额收益的数字，因为本文的样本区间、资产域与收益口径与之不同。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

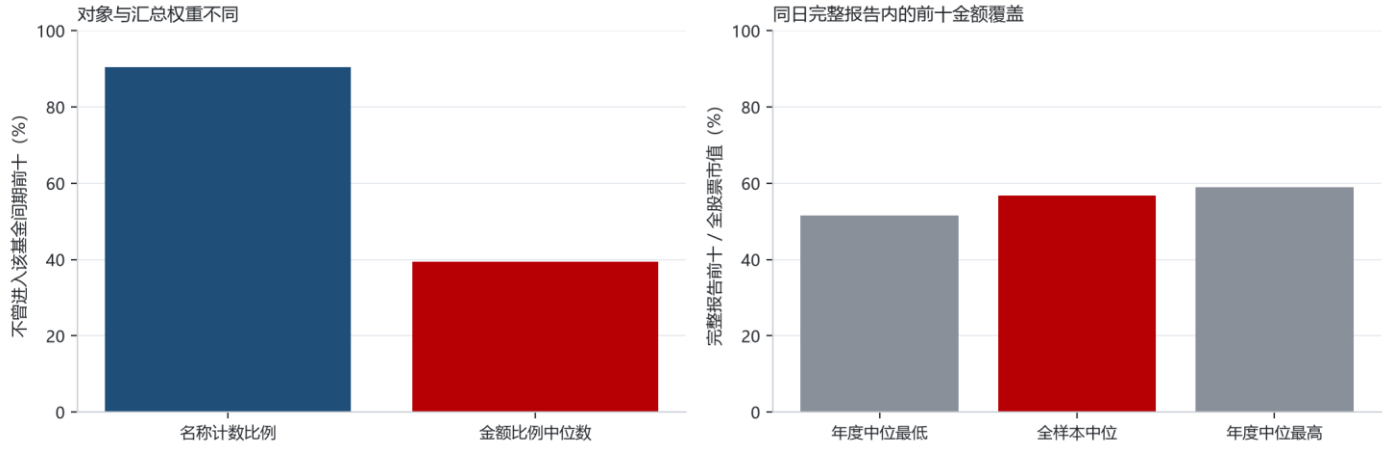
下载日志已记录，仅供内部参考

表13： 披露覆盖的准确定义

指标	数值	分母与汇总方式
完整报告名称数中位 / 均值	57 / 75.9	基金×报告内不同股票名称
Q1/Q3 名称数中位 / 均值	10 / 9.3	基金×报告内披露股票名称
曾进入间期前十的名称覆盖	9.52%	逐基金全样本名称集合交集；按名称计数加权，非中位数
从未进入间期前十的名称占比	90.48%	来自完整报告的名称中，不曾入该基金间期前十者
不曾进入间期前十的金额占比	39.52%	完整报告金额比例的中位数；非全市场不可见金额
完整报告自身前十大金额占比	56.75%	每份完整报告的前十金额/全股票金额，再取中位数

资料来源：浙商证券研究所，Tushare Pro

图6： 名称覆盖与市值覆盖



资料来源：浙商证券研究所，Tushare Pro

6.3 为什么这一对数字最重要

把名称维度和市值维度放在一起看，会得到本文最希望被记住的那个读法：「从未露面」在名称维度上占绝大多数，在市值维度上只有约四成。两个数字之间的落差，说明不可见的是大量小仓位，而不是少数大仓位。

这个区别为什么关键：如果不可见的是少数大仓位，那么份额与集中度的测量就谈不上可靠，因为最重要的那些量都测错了。但如果不可见的是大量小仓位、而可见的前十大能覆盖近六成市值，那么测量误差主要集中在排序的尾部，而尾部对份额平方和的贡献本来就小（份额的平方对小的份额不敏感）。这正是为什么这个截断是「可承受的稀释」而不是「致命的失真」。

还有一个有利条件来自上一节那个发现：既然紧致度的变化几乎全部来自份额的差分，而差分对水平误差不敏感，那么截断对差分项的污染就小于对水平项的污染。两个有利条件叠加，让截断的净影响落在「衰减」这一侧。

6.4 一个可对照的检验：只用完整报告

除了论证，还有一个直接的实证检验：把中期状态整个丢掉，只用完整报告（年报与半年报）构造状态。这样完全不存在季报的截断问题——每一个状态都是完整的持仓。代价是结点数减半。实测结果是：把季报整个拿掉，结论不变。这是对「截断是否在驱动结论」这一问题最直接的否定性回答：如果正号是由季报的截断造出来的，那么拿掉季报之后系数应当显著变化——但它没有。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

6.5 规模环比不等于估计偏差

表14： 相邻报告期的规模比

方案	间期/前期完整报告	完整报告/前期间期	解释边界
A	1.1676	0.8290	不同日期与可能变化的基金池；没有同日真值
B	1.0209	0.9867	不同日期与可能变化的基金池；没有同日真值

资料来源：浙商证券研究所，Tushare Pro

这些比值混合市场回报、申赎、基金池进出及重建变化，不能称作高估百分比。前十截断经归一化进入份额、容量与交互项，误差可能相关且非加性；即便噪声与未来收益不相关，也不足以自动满足经典误差变量的全部条件。偏差方向需要检验，不能断言只会衰减。

6.6 能够识别误差的验证设计

应在完整报告上保留同日真值，再人为只留前十大，按历史可得信息重建，比较每个结点的份额、集中度与容量。这一步控制日期与基金池，才可能估计重建误差。随后把实测误差结构注入无反馈和正反馈模拟，评估主系数的偏差与检验功效；相邻季度规模比不能代替该实验。

表15： 同日完整报告人工前十截断实验[12]

指标	实测	解释
完整报告数	45242	每报告完整股票书为同日真值
前十金额占比中位数	54.70%	同日完整报告内计算
权重 L1 误差中位数	0.9060	对完整股票向量逐项比较后按报告等权取中位数
重建 / 真实 HHI 中位数	2.5128	集中度失真；不等于回归系数偏差

资料来源：浙商证券研究所，Tushare Pro

表16： 最新季度的共同基金规模分解[13]

口径	季度转换	总变化 亿元	共同基金 亿元	进入减退出 亿元	共同基金数
个股隐含总账	20260331 → 20260630	-5128.7	5790.2	-10918.9	2413
可比行业映射持仓	20260331 → 20260630	-9168.7	4295.2	-13463.8	1663

资料来源：浙商证券研究所，Tushare Pro

[12] 实验评价方案 B 的同日等比放大。L1 为完整权重向量绝对误差之和，范围 0－2，不是总规模高估率。总金额匹配属于构造恒等式；不评价方案 A 跨期外推，不识别回归偏差方向。

[13] 总变化=共同基金变化+进入金额−退出金额。共同基金变化仍含市场价格、交易及测量误差，不等于净申购，也不是共同基金样本的 H1 重估。

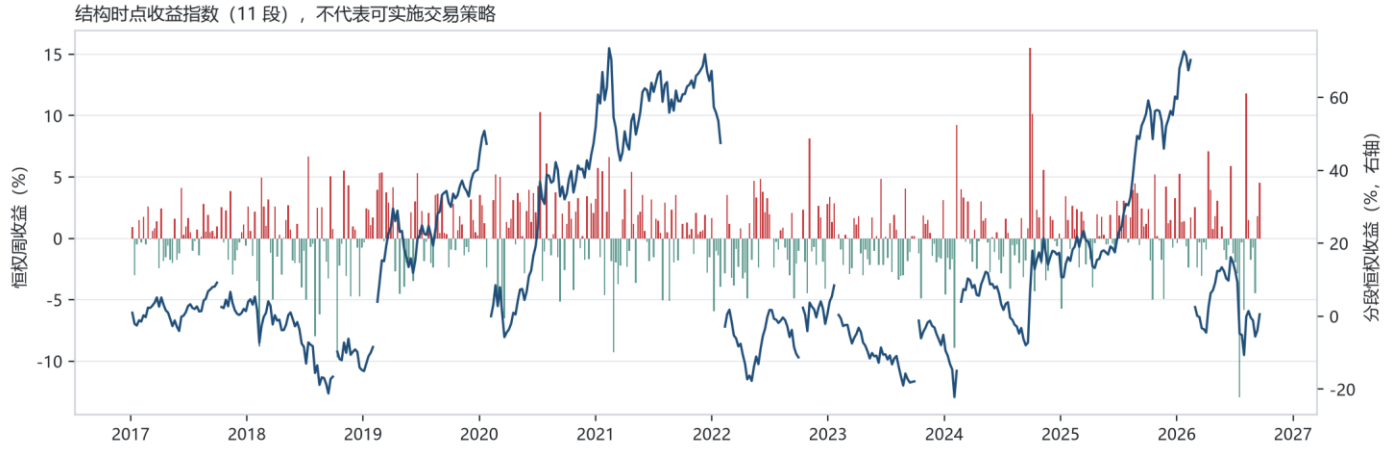
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

7 个股扩展与推断有效性

把战场从一百多个行业推到几千只 A 股，是把结点数提高约一个数量级。这样做的代价是「集中持仓」这个词在个股层面变得更模糊——机构持有一只股票，可能只是一个行业的被动配置结果；好处是统计功效显著提高，而且不再受行业分类的人为性影响。

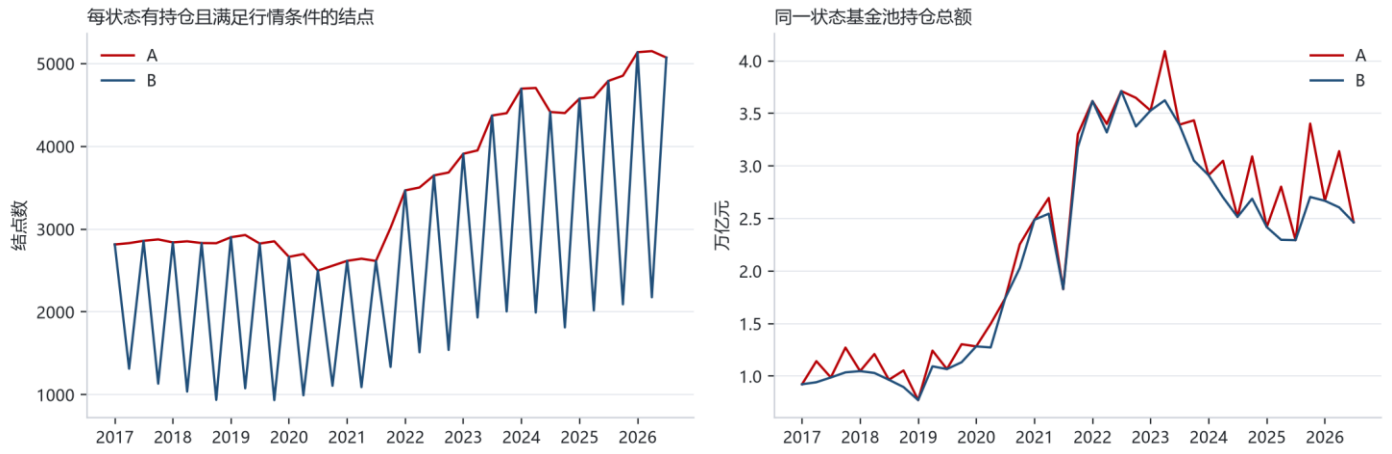
图7： 个股方案 A 分段恒权组合收益[14]



资料来源：浙商证券研究所，Tushare Pro

上面这条是「集中组合」本身的收益路径，它是被解释变量的原始形态；下面这张则是解释变量的规模——两者放在一起看，才看得出主检验到底在问什么。需要留意的是，组合收益的波动明显大于状态的变动，这意味着即便 c 在统计上可识别，它在经济上也未必能解释多少收益的方差。

图8： 个股结点广度与资金规模[15]



资料来源：浙商证券研究所，Tushare Pro

[14] 个股方案 A 按固定季度权重逐周加权；缺数或缺周后累计指数重新起算，分段显示。仅为结构收益序列，不含披露可得约束和交易成本。

[15] 每个状态下的结点数与该状态的总持仓规模；结点数的起落主要来自行情可得性与新股上市，不是基金行为的直接度量。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

7.1 八规格系数与聚类区间

个股层的规格矩阵比行业层大：两个方案（A/B）× 两种容量口径（原始 DTL / 容量残差）× 两个模型（E1/E2），共八格。判定仍只用预注册的那一格：方案 A • 原始 DTL • E2。

表17： 个股八规格估计[16]

设定	系数	周聚类 95% CI	季度聚类 95% CI	季度 t	有效行数
A / 原始 DTL|E2	-0.00536	[-0.01376, 0.00305]	[-0.01296, 0.00225]	-1.43	1621452
A / 原始 DTL|E1	-0.00597	[-0.01485, 0.00291]	[-0.01426, 0.00231]	-1.46	1621455
A / 容量残差|E2	-0.00430	[-0.00906, 0.00047]	[-0.00908, 0.00048]	-1.82	1621449
A / 容量残差|E1	-0.00316	[-0.00872, 0.00240]	[-0.00920, 0.00289]	-1.06	1621451
B / 原始 DTL|E2	-0.00188	[-0.01032, 0.00656]	[-0.00793, 0.00416]	-0.63	1124633
B / 原始 DTL|E1	-0.00760	[-0.01722, 0.00203]	[-0.01672, 0.00153]	-1.69	1124636
B / 容量残差|E2	-0.00056	[-0.00667, 0.00555]	[-0.00700, 0.00588]	-0.18	1124630
B / 容量残差|E1	-0.00282	[-0.00900, 0.00336]	[-0.00967, 0.00402]	-0.83	1124632

资料来源：浙商证券研究所，Tushare Pro

三件事需要说清楚。

第一，主规格的负号并非八格里最弱的一档。 八个规格的点估计全部为负，绝对值最小的是方案 B • 容量残差 • E2（约 −0.0006），主规格（方案 A • 原始 DTL • E2，约 −0.005）的绝对值反而排在八格前列。换句话说，单看点估计，另外几个规格的负号更弱、更接近零；主规格并非靠「恰好锁定了最不利的一格」才显得保守。

第二，「显著为负」有口径依赖，必须两种口径并排报。 旧的置换零分布非常窄，而按周聚类的标准误约为其一个数量级——两者给出完全不同的「显著性」。本文不取其中最窄的那个来下结论。按周聚类是周频收益序列的通行做法，而且它给出了更保守、因此更需要被相信的那个答案。

第三，按行业×季度聚类时 t 值很大，但这个口径本身有问题。 行业×季度这个层级把同一季度内同行业的股票当成一个簇，实际簇数看起来很大，但它与回归里的固定效应高度重叠：本文在每个状态×结点单元内部做了去均值，而行业×季度这个聚类层级恰好比去均值的层级更粗，于是簇内的残差相关被人为压小，t 值被放大。本文报出它，是为了让读者看到口径之间的落差有多大。

主设定的固定设计周 score bootstrap 辅助结果：999 次，双侧 $p = 0.2730$，上尾 $p = 0.8530$。该数值依赖周间相关性假设，标记为非确认性敏感性；下文模拟给出其适用边界。

A 与 B 的有效股票周可能不同，容量残差化也改变回归元分布，不能仅按系数绝对值排名保守程度。应先比较共同样本的效应与区间，再讨论经济差异。周聚类容许同周跨股票相关；季度聚类容许同季度的周际与横截面相关，但均需额外假设处理跨簇依赖。行业×季度是交集分组，不能处理所有跨行业共同冲击或跨季相关。

[16] 主对照为 A / 原始 DTL / E2，其余为敏感性。区间不作八规格联合覆盖保证，不用最小 p 或最窄区间挑选结论。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

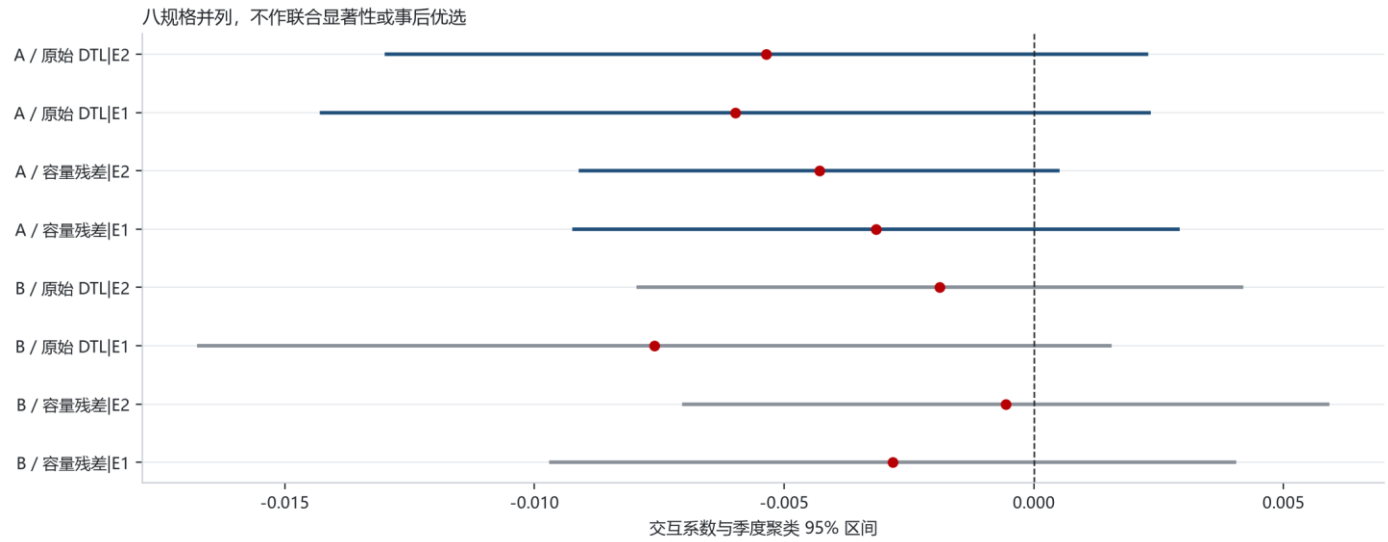
下载日志已记录，仅供内部参考

表18： A 与 B 的共同股票周对照[17]

方案	E2 系数	周聚类 95% CI	季度聚类 95% CI	共同样本行数
A	-0.00650	[-0.01520, 0.00220]	[-0.01435, 0.00134]	1124446
B	-0.00189	[-0.01032, 0.00655]	[-0.00793, 0.00416]	1124446

资料来源：浙商证券研究所，Tushare Pro

图9： 个股系数与季度聚类区间[18]



资料来源：浙商证券研究所，Tushare Pro

7.2 为何撤回原状态内置换的统计解释

原本的做法是每个季度独立重排容量代理。这样做会保留当期截面分布，却破坏同一结点的跨期持续性。E2 会吸收稳定的结点容量×收益斜率，重排后人为增加的时间变化改变了识别强度，原始系数的重排分布可能远窄于抽样分布。非平衡面板不是这种置换有效的理由；增加重排次数不能修复这一问题。

对个股层还有一个额外的结构性问题：这是一个非平衡面板，结点随上市与退市变动。非平衡面板上的置换检验必须保持每个单元观测到的（状态，结点）对的数目不变，因此本文用的是按「行业×季度」分层的重排，而不是整列置换。分层置换的零分布更窄，对负号更不利，因此它不构成一个宽松的选择。

审查的无反馈反例使用 200 个独立合成面板、每次 499 个原生产算法重排：名义 5% 检验的上尾误拒为 50.5%、下尾为 44.0%。这证明该算法没有通用水平保障，不是对真实样本误拒率的估计。旧上下尾概率均已撤回确认性解释。固定设计 score bootstrap 若提供，仅作为依赖假设下的辅助结果；未经对应模拟校准不升格为主检验。

[17] 审查后补充的 raw/E2 共同有效股票周对照；统一样本不消除两方案 x、DTL 或持仓重建本身的差异，不当作原预注册主结果。

[18] 图形已改为聚类区间；文件名保留用于底稿追溯。旧状态内置换概率不再用于统计判定。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

表19： 推断的水平与功效压力实验[19]

模拟设定	真实 c	score 双侧拒绝	拒绝率 MC 95% 区间	周 CI 排除零	季度 CI 排除零
独立收益 / 持久 DTL	0.000	4.0%	[2.0%, 7.7%]	4.0%	8.0%
同期共同冲击	0.000	6.0%	[3.5%, 10.2%]	7.5%	10.0%
季内相关压力	0.000	10.5%	[7.0%, 15.5%]	10.0%	7.0%
随机缺失非平衡	0.000	4.5%	[2.4%, 8.3%]	6.0%	8.0%
温和正反馈	0.025	6.5%	[3.8%, 10.8%]	8.0%	9.5%

资料来源：浙商证券研究所，Tushare Pro

正式模拟仍留下两项限制：季内相关情景下 score 双侧误拒为 10.5%，高于名义 5%；温和正反馈情景的上尾功效仅 14.5%。这些结果不构成推断已普遍有效的认证，也说明不显著可能反映功效不足。随机缺失实验未覆盖信息性缺失或真实持仓误差。

7.3 容量残差：正号是不是只是流动性

在所有会威胁结论的替代解释里，最需要认真对待的一个是：容量代理会不会只是流动性的一个粗糙代理，而我们在测的其实是「低流动性溢价」之类的已知效应。

检验办法是把容量代理在每个状态内部对 log 市值、换手率与 Amihud 非流动性做回归，取残差作为「容量残差」再跑同一套主检验。如果正号来自流动性，那么残差口径下系数应当大幅衰减甚至消失。

表20： 容量残差的截面诊断

方案	状态数	中位 R^2	解释
A	39	0.1156	仅对已纳入的线性控制正交
B	39	0.1087	仅对已纳入的线性控制正交

资料来源：浙商证券研究所，Tushare Pro

结论是：正交化之后系数没有衰减，反而更负。加上中位决定系数很低，可以说容量代理携带的信息里，至少大部分不是「市值小 / 换手低 / 非流动性高」这几件事的重复。

残差容量代理对已纳入的市值、换手与 Amihud 线性控制正交，不意味着排除了全部流动性混淆，更不等于剩余部分具有预测增量。正式推断还应在保留持久性、异方差、共同冲击与非平衡结构的数据生成过程中同时检验水平和功效。

个股 H3 的 A / B 状态分别为「不可用 / 不可用」。H3 需要完整正持仓结点的容量代理才能构造组合容量；不能把缺失深度填零或忽略其持仓后声称组合容量完整。A / B 的容量代理名称计数覆盖中位数为 97.2% / 96.8%；分母是正持仓结点数，不是持仓金额。H3 不可用不妨碍在有效股票周上报告 H1，但不能把 H1 样本当成完整组合容量样本。

完整深度覆盖核查：A 的 497 个有组合收益周中，完整正持仓 DTL 覆盖周为 0，最高名称覆盖为 99.4474%；B 的 497 个有组合收益周中，完整正持仓 DTL 覆盖周为 0，最高名称覆盖为 99.3979%。因此本轮股票 H3 的不可估根因是组合深度覆盖不足，而非反馈假设获得支持或被否定。

[19] 每设定 200 个面板、每次 999 个固定设计周 score bootstrap。$c = 0$行为误拒率，正反馈行为功效；MC 区间为 Wilson 模拟误差范围，不是参数区间。正反馈状态波动与零假设情景不同。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

8 结论的强度与可继续验证的问题

这一节不再给新结果，只把前面各部分收拢成几句话，并说清楚哪几句话是本文不能说的。

8.1 两层符号相反，是「都没有证据」而不是「互相抵消」

行业层是正号、个股层是负号，两层都不显著。为什么不能合并：行业层的紧致度建立在行业汇总的持仓上，个股层的紧致度建立在单只股票的持仓上；行业层是平衡面板，个股层是非平衡面板；两层的因变量、控制变量与频率也都不同。把两层当成「同一个系数的两个估计」来平均，不合理；它们只应作为「两个互为参照的视角」并列读出，其共同结论是：都没有证据。

8.2 这个结论能被信任到什么程度

本文的结论强度取决于四件事，按重要性排列。

第一，截断的方向性论证（最重要）。 前一节论证了截断是稀释性的：名称维度的绝大多数只对应市值维度的约四成，说明漏掉的是大量小仓位；再加上紧致度的变化几乎全部来自份额的差分（对水平误差不敏感），以及「只用完整报告」这个直接检验给出的系数几乎不变。如果这个论证成立，那么「未通过」就是结论本身。

第二，预注册是否被遵守。 主规格在看结果之前锁定，这个锁定本身是诚实的；但它并非八格里绝对值最小的一格——绝对值最小的是方案 B · 容量残差 · E2。因此「不显著」这个说法的安全来自「看结果前就锁定了」这一事实本身，而不是来自「恰好锁定了最不利的一格」。

第三，双门槛是否真的双重。 本文用的是系数为正且上尾 $p<0.05$。个股层的第一道门槛就已经不满足；行业层的系数虽然为正，但第二道门槛差得远。两道门槛在两层各自是被不同的那一项卡住的，这一点让「未通过」不是某一门门槛的偶然结果。

第四，复算路径是否独立。 截断的几个量由两条独立路径给出，中位数一致；份额之和在两层都严格为 1；池子的构造有两处回归测试保护。这些不能证明结论正确，但能排除「结论来自一个算错的量」。

8.3 从未拒绝到经济结论还缺什么

不显著只说明在所用推断与样本下，区间仍容许零或相反方向。判断经济上可忽略，应先给出最小有意义的效应、再报告等效性检验或最小可检测效应。披露测量误差可能降低功效，也可能造成方向偏差；不能以衰减假设替代验证。「不显著」和「是零」是两回事：区间在两层都跨零但都不窄，都不足以排除一个中等大小的真实效应。

8.4 本文不能说的四句话

① 不能说「集中持仓有害」。 个股层的负号在按周聚类下 t 只有很小的绝对值，且有一个规格（方案 B · 原始 DTL · E2）连下尾都不显著。「显著为负」这个说法依赖把旧置换零分布的标准差当作推断尺度，而这个标准差只有按周聚类标准误的约一个数量级分之一。本文把两种口径并排报出，并选择不据此下反向结论。

② 不能说「集中持仓无害」。 行业层是正号，个股层负号的强度也只是「不显著」，而不是「显著为零」。95% 区间在两层都跨零但都不窄，都不足以排除一个中等大小的真实效应。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

③ 不能说「这个模型被证伪了」。 本文检验的是 $\rho = \kappa V H_A$ 这一个具体参数化下的预测。未被支持的（即没有证据的）不是「集中持仓」这个概念本身，而是「这三个量以这个方式相乘、并对未来收益有正向预测」这个具体命题。换一个参数化（比如不对容量做除法、或者用份额的基尼系数替代平方和）可能给出不同的答案——本文没有检验那些。

④ 不能说「容量维度不重要」。 量级比较发现集中度与冲击之比落在极小量级，说明在这个构造里容量维度几乎不贡献方差。但这是因为规模项的方差太大、把集中度淹没了，不是因为容量本身没有经济含义。如果换一个只看横截面差异、不看时间序列变化的检验设计，容量的作用完全可能显现出来。

8.5 下一步验证的先后关系

按本文的证据，有三条路是优先的，而且它们之间有依赖顺序。

一是把截断补上。 本文的截断论证本质上是「用已知的披露规则推出误差方向」。要把这个论证从推理升级为检验，需要的不是更多数据，而是一个独立的持仓观测——比如沪深港通的持股明细（北向持仓自 2024 年 8 月起已由按日披露改为按季度披露，但仍独立于公募基金的披露节奏），或者 ETF 的申购赎回与成分股变动。有了第二个来源，就可以估计「可见部分的市值占比」与真实值之间的差距，从而把误差方向从假设变成实测。稳健性清单里「第二反馈组」之所以标为不可用，正是因为缺这个来源。

二是把主检验换成横截面的。 本文的紧致度是一个时间序列上的状态量，几十个季度、两百周的样本里，真正独立的「状态」只有几十个。如果改成在每个季度内做横截面回归（用容量代理的横截面差异而非时间序列变化），样本量会大两个数量级，而且正好抓住容量维度在横截面上的作用——这恰好是前面那个退化现象所提示的方向。

三是把「集中持仓」的定义从机构持仓里解放出来。 本文始终用公募基金持仓来定义份额，这是被数据可得性推着走的结果。如果换成「成交额集中度」这类不依赖披露的度量，截断问题就整个消失了——代价是它测的不再是「机构持有了多少」，而是「市场在哪些东西上交易得多」。这是一个值得单独做的题目。

先固定信息时点、资产集合、基金覆盖和完整交易周；再做同日期完整报告截断重建及共同样本对照；之后用匹配实际依赖结构的模拟校准推断，最后单独评估披露可得的预测与成本。增加复杂度前应保留这条依赖顺序，避免图表接入了不同数据版本的结果。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

9 研究边界与使用限制

本节给出研究本身的边界，也就是使用本文任何数字之前必须先知道的前提。第 10 节「风险提示」另行集中列出与首页「核心观点」同口径的风险条目，二者互为补充：本节说「这个研究做到了哪里、哪里还做不到」，第 10 节说「使用结论时要担哪些风险」。

关于结论本身。 两层都没有通过预注册门槛。预注册的判定条件是系数为正且上尾置换 $p<0.05$，主规格在看结果之前锁定为「方案 A · 原始 DTL · E2」。行业层主规格系数为正但上尾 p 很高，个股层主规格系数为负、上尾 p 接近 1，两层均未通过。本文报告的是「没有找到证据」，不是「证明了反向」，也不是「预测未来收益为负」。「显著为负」高度依赖推断口径，不要单独引用——个股层八个规格的点估计全部为负，但旧置换零分布的标准差只有按周聚类标准误的约一个数量级分之一，同一批点估计在按周聚类口径下的 t 绝对值很小。任何「A 股集中持仓显著负向预测收益」的表述都超出本文证据。

关于数据的已知缺陷。 ① 前十大披露截断：公募基金只在季报（一、三季报）披露前十大持仓，只有年报与半年报披露全部持仓；所有份额与集中度都建立在这个被截断的观测之上。② 截断是内生的，不是随机的：实证文献已证明基金会在季末前把含有私有信息的名字推到第十名门槛之下，因此披露出来的前十大是一个被管理的集合。任何从披露反推拥挤度的方法（含本文的方案 A）都继承这层内生性，误差方向不能假定为均值为零的经典测量误差。③ 方案 A 的间期状态系统性高估：按第 6 节的论证，该偏差方向为稀释，会给容量状态量注入按状态交替的乘性噪声；它使估计量向零衰减，而不是制造出一个正号，但这一方向仍需独立检验。④ 结点域经过剔除：结点集是「有交易所行情行的 A 股」，港股行线按设计排除，北交所与新三板的少数无可取行情的代码另行剔除。影响范围有限但不宜忽略。

关于结构与口径。 结构持仓以事后报告期构造，不代表投资者当时已知。基金类型代理、持仓不完整和分类映射仍可能造成选择偏差。资金规模环比不等于估计误差，截断偏差不保证向零。旧个股状态内置换已失去统计解释，不得引用其尾概率宣称负反馈。聚类区间与辅助 bootstrap 依赖各自的跨簇独立或近似独立假设；点估计、置信区间和规格差异应共同呈现。

关于完成度。 固定主口径的 H6 完成 0 个起点；事后可比池完成 19 个起点，须按敏感性分析身份及联合 DM 判据解读。条件校准的面板数不补足基金层的工具变量模拟和全流程推断验证。事件窗为探索性展示，卖空、杠杆与交易成本限制不能忽略。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

10 风险提示

本节与首页「核心观点」末尾的「风险提示」为同一口径，逐条对应，正文展开得更细一层；两处如有不一致，以本节为准。首页给出的五句话分别是：两层主规格均未通过门槛；系数建立在「只到前十大」的披露之上；间期高估已量化（方案 A 约 16.8%）；不构成因果识别或收益保证；不显著不等于机制不存在、全文不构成投资建议。下面按这五条逐一说明含义与使用边界。

10.1 两层主规格均未通过门槛

预注册门槛是系数为正且上尾置换 $p<0.05$，主规格在看结果之前锁定。行业层系数为正但上尾 p 很高，个股层系数为负、上尾 p 接近 1，两层都没有通过。因此本文的结论是「没有找到证据」，不是「证明了集中持仓会压低未来收益」，也不是「未来收益会被负向预测」。任何把本文读成方向性预测的引用都超出证据范围。

10.2 系数建立在「只到前十大」的披露之上

公募基金只在季报（一、三季报）披露前十大持仓，只有年报与半年报披露全部持仓。本文所有份额、集中度与容量状态量都建立在这个被截断的观测上。更关键的是截断是内生的：基金会在季末前把含有私有信息的名字推到第十名门槛之下，披露出来的前十大因此是一个被管理的集合。任何从披露反推拥挤度的方法（含本文的方案 A）都继承这层内生性，误差方向不能假定为均值为零的经典测量误差。

10.3 间期高估已量化，方向按论证为稀释

方案 A 用中期前十大按相邻全持仓报告估值外推，间期状态被系统性高估约 16.8%。按第 6 节的论证，该偏差的方向为稀释——它给容量状态量注入按状态交替的乘性噪声，使估计量向零衰减，而不是制造出一个正号。但第 6 节同时强调，这一方向需要独立检验，不能仅凭推理断言它只会衰减；换句话说，这一缺陷至多让效应看起来更小，不能据此宣称「已排除正号的制造机制」。

10.4 不构成因果识别或收益保证

全文是相关性层面的实证检验，结构持仓以事后报告期构造，不代表投资者当时已知，因此不构成因果识别。所有系数、区间与图表均为学术性实证结果，用于说明一个具体参数化的检验过程与其边界，不预测任何证券或行业的未来表现，也不提供可实现净收益的保证。「显著为负」高度依赖推断口径，不要单独引用：个股层八个规格的点估计全部为负，但旧置换零分布的标准差只有按周聚类标准误的约一个数量级分之一，同一批点估计在按周聚类口径下的 t 绝对值很小。

10.5 不显著不等于机制不存在，全文不构成投资建议

未通过门槛只说明在当前样本、当前口径与当前功效下没有检出证据，不等于不存在反馈机制——可能是样本不够长、噪声太大或代理不够好。本文不是投资建议，也不构成任何形式的评级；不应被用作选股、择时、组合权重或风险控制的依据，结论不可外推至本文未检验的设定。数据截至报告日已尽合理努力核对，但不对其完整性、准确性作任何保证；任何第三方引用本文数字前，应自行复算。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

附录 A 符号与单位

下表只列在正文、图表与附录中出现过的记号。未出现的中间记号（如拟合时的辅助矩阵）不列入。单位一栏给出该量的量纲，凡是标明「无量纲」的量，其数值大小可以直接在不同状态与不同层级间比较。

表21： 全文符号表

记号	定义	单位
i	结点（行业层为申万二级行业，个股层为单只 A 股）	—
t, s	季度末状态下标、周下标	—
N	结点数（行业层固定，个股层随状态变动）	个
q_i, V	结点持仓市值与池内总持仓金额	元
$\pi_i = q_i/V$	同一资产域内的持仓份额，份额之和恒为 1	无量纲
$\Delta\pi_i$	相邻两个季度末之间份额的一阶差分	无量纲
$V_t = \frac{1}{2}\sum_i \lvert\Delta\pi_{i,t}\rvert$	持仓结构的整体位移（冲击）	无量纲
D_i	结点容量＝过去 60 个交易日日均成交额	元/交易日
$\mathrm{DTL}_i = q_i/D_i$	清仓天数，持仓容量代理	交易日
$\Lambda = \lambda\mathrm{diag}(D_i^{-1})$	价格冲击矩阵	每元
$u = \Lambda q = \lambda\mathrm{DTL}$	质量密度向量	无量纲
$v = \kappa\pi$	反馈敏感度向量	无量纲
$H_D = \sum_i \pi_i^2/D_i$	成交深度加权集中度	交易日/元
$H_\Lambda = \lambda\sum_i \pi_i^2/D_i = \pi^\top\Lambda\pi$	含价格冲击尺度的集中度	每元
κ	资金流对收益的敏感度（H4 单独估计）	依资金流归一化
λ	单位清仓天数的价格冲击尺度	每交易日
$c = \kappa\lambda$	滞后交互系数（主系数）	每交易日
$\rho = \kappa V_t H_{\Lambda,t} = c\pi^\top\mathrm{DTL}$	反馈紧致度，核心状态量	无量纲
$\ell := \mathbf{1} - \kappa u$	预算约束在平直坐标下的法向量	无量纲
$x_s = \pi^\top r_s$	固定季度权重的集中组合周收益	收益率
r_i	结点收益（前收价链式累乘，等价全收益）	收益率
c^*, p_{upper}, p_{lower}	置换样本系数与上、下尾概率	—
φ, ψ	H3 中冲击项与滞后收益项的系数	无量纲
β_{LR}	长期收益敏感度（H5，不可识别）	无量纲
$C = \Sigma_0 + s^2uu^\top$	包含外生流噪声的基准协方差	收益率平方
$\Omega = A(\Sigma_0 + s^2uu^\top)A^\top$	长期协方差	收益率平方

资料来源：浙商证券研究所，Tushare Pro

附录 B 数值验证与复现

一个不能还原真值的估计量，其系数无论多显著都没有含义。因此在读真实数据之前，本文先用参数已知的合成数据验证了估计量、以及从模型推出的几个恒等式。下表是全部结果。

表22： 理论数值核验

对象	结果	适用范围
静态恒等式	失败项 0	已知参数的合成环境
长期结构	失败项 0	固定状态与给定 DGP
方差比	实测 3.01048 / 理论 3.00000	不认证真实数据或完整推断流程

资料来源：浙商证券研究所，Tushare Pro

代数核验、估计量模拟与真实参数校准是三个层次：前两者成功不保证第三者完成，也不能证明程序或数据不存在其他问题。其中最值得注意的是方差比那一条：理论值与从

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

模拟数据上重建出来的值高度一致。这条恒等式在合成数据上是成立的，而在真实数据上对应的经验校准被标为「阻塞」——两者的对比说明：不是方法错了，是真实数据里缺少识别所需的那个独立协方差来源。本轮真实参数校准已生成 500 个条件面板，但基金层工具变量模拟、完整 H6 与三设定联合流程尚未覆盖；其完成数量不等于独立识别或全流程认证。

表23： 复现产物与依赖

环节	产物	前置条件
数据质量	data_quality.json / data_source_manifest.json	官方交易日历、分页与披露契约
行业估计	estimate_results.json / h1_specs.json / robustness_results.json	统一资产域与结构面板
预测监控	h6_results.json	披露日期、估计截止日与完整评价窗
个股估计	stock_estimate_results.json / stock_series.pkl	完整交易周与共同样本诊断
真实校准	empirical_calibration.json	稳定状态与可识别基准协方差

资料来源：浙商证券研究所，Tushare Pro

运行状态、参数、输入和产物哈希应绑定同轮 manifest。当前硬编码的主设定只证明现在的约定，不单独证明预注册时点。从原始数据到本报告的完整流程可以一句话概括：先跑数据契约与质量门，再跑行业层估计，再跑个股层，然后生成图表底稿，最后装配本报告。质量门是一道硬门：如果数据质量状态为无效，后面所有步骤都不会执行，而不是输出一组带警告的数字。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

股票投资评级说明

以报告日后的 6 个月内，证券相对于沪深 300 指数的涨跌幅为标准，定义如下：
1.买　入 ：相对于沪深 300 指数表现＋20%以上；
2.增　持 ：相对于沪深 300 指数表现＋10%～＋20%；
3.中　性 ：相对于沪深 300 指数表现－10%～＋10%之间波动；
4.减　持 ：相对于沪深 300 指数表现－10%以下。

行业的投资评级：

以报告日后的 6 个月内，行业指数相对于沪深 300 指数的涨跌幅为标准，定义如下：
1.看　好 ：行业指数相对于沪深 300 指数表现＋10%以上；
2.中　性 ：行业指数相对于沪深 300 指数表现－10%～＋10%之间；
3.看　淡 ：行业指数相对于沪深 300 指数表现－10%以下。
我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重。
建议：投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者不应仅仅依靠投资评级来推断结论。

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。
本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。
本报告仅反映报告作者的出具日的观点和判断，在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。
本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。
本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

浙商证券研究所

上海总部地址：杨高南路 729 号陆家嘴世纪金融广场 1 号楼 25 层
北京地址：北京市东城区朝阳门北大街 8 号富华大厦 E 座 4 层
深圳地址：广东省深圳市福田区广电金融中心 33 层
上海总部邮政编码：200127
上海总部电话：(8621) 80108518
上海总部传真：(8621) 80106010
浙商证券研究所：https://www.stocke.com.cn

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster