

“订单块”并无信息： 对一套聪明钱概念（SMC） 趋势延续与回踩策略的预注册证伪

管世宇 IRIS

EROS 量化实验室

内部审阅稿

摘 要

社交媒体流行的“聪明钱概念”（Smart Money Concepts, SMC）交易法主张：价格在大级别结构破坏（BOS）后回踩“订单块”区域时会反转并延续趋势，可据此挂限价单、设紧止损以博取高盈亏比。这套方法几乎从未接受过无未来函数、计真实成本的严格检验。本文将这套主观表述无歧义地机械化，在 19 个 OKX USDT 永续、2021 年 1 月至 2026 年 7 月（5.5 年、365 万根 15 分钟 K 线）上，以预注册门槛、分层随机对照与真实成本回测对其系统检验。核心发现分三层。（i）字面成立：价格回踩订单块后先走 +1 ATR 的概率为 55.5%（90% 置信区间排除 50%），“订单块容易反转”并非错觉。（ii）归因错误：该反转率完全由“等待的回撤深度”解释——立即入场恰为 50.3%，同尺寸随机区域反而更高（59.2%）——与订单块这一价位是否“有记忆”无关；相对随机匹配区域，订单块的增量效应量为 -0.134 ATR（区间排除 0）。（iii）量级不够：反转的毛期望仅为单次往返交易成本的 68%。48 格真实成本回测全数为负，样本外与多重检验折价（deflated Sharpe = -3.41 ）一致，全期净值 $1.0 \rightarrow 0.28$ 。三组补充实验表明：结论对 bootstrap 区块长度稳健（24/24 检查通过）、在第二交易所（Binance）冻结复制下同向成立、且在“ ± 1 次往返成本”的经济等价界内统计等效于零——即这是一次有统计功效支持的证伪，而非样本不足。研究过程中还发现并修正了一个会翻转结论的“成交 K 约定”错误（修正前订单块虚假夏普高达 3.41），本文将其作为技术策略回测的方法论案例一并报告。

关键词： 订单块，聪明钱概念（SMC），技术分析，事件研究，预注册，等价检验，交易成本，加密货币永续合约

1 引言

1.1 研究问题

“聪明钱概念”（Smart Money Concepts, SMC）是 YouTube、加密社群与零售交易教育中最流行的入场方法之一。其标准流程是：先在大级别（如 4 小时）确认趋势方向；等待中级别出现“结构破坏”（break of structure, BOS）并以一段“位移”（displacement）确认动能；再把位移起点前最后一根反向 K 线标记为“订单块”（order block），在该区域挂限价单等价格回踩，配合紧止

损博取高盈亏比。支持者的核心说法是：订单块是“聪明钱”建仓留下的足迹，价格回踩此处会因未成交的挂单而反转。

这类方法的传播极广，但几乎从未接受严格、无未来函数、计真实成本的实证检验。与既有技术分析文献不同，本文的对象不是抽象的移动平均或区间突破规则^[1,2]，而是零售社群中一套具体、可操作、被大量交易者实际采用的方法。本文问一个可证伪的问题：**在同等级别趋势状态、同一波动率、同一回撤深度下，价格回踩订单块之后的前向收益，是否显著优于回踩一个匹配的对照区域？**若答案为否，则“订单块有信息”这一主张即被证伪，无论其反转率多高。

1.2 文献锚点

技术分析的有效性争论由来已久。Brock 等发现移动平均与区间突破规则在道琼斯指数上有温和的超额收益^[1]；Lo 等用核回归为技术形态建立了统计推断框架^[2]。在加密市场，Urquhart 报告比特币收益存在显著自相关、市场弱式无效，为技术分析留出了空间^[3]。但这一文献传统始终伴随着数据挖掘的阴影：Sullivan 等用 White 的现实检验表明，Brock 等的规则经多重检验校正后不再显著^[4,5]；Harvey 与 Liu 系统论证了回测中的选择偏差与过拟合如何虚增夏普^[6,7]；Bailey 与 López de Prado 提出的 deflated Sharpe 比率可对多次试错造成的夏普虚增直接折价^[8]，本文在回测阶段直接使用。McLean 与 Pontiff 进一步表明，因子收益在发表后系统性衰减，相当部分是数据挖掘的产物^[9]。

方法上，本文借用两个工具把“零结果”升级为“有功效支持的证伪”。其一是**等价检验**(equivalence testing)：Lakens 主张用两个单侧检验 (TOST) 在预设的最小有价值效应 (smallest effect size of interest, SESOI) 界内直接检验“效应等价于零”，而非把“不显著”误读为“无效应”^[10,11]。本文把 SESOI 定为单次往返交易成本：小于成本的效应即使真实存在，也不足以覆盖一次进出的费用，经济上等同于无效应。其二是**平稳自助法** (stationary bootstrap) ^[12]，作为日历月区块自助法的非等长对照，用于检验结论对区块长度这一主观选择是否稳健。

与既有文献相比，本文的差异有三：对象是零售社群的具体交易方法而非抽象规则；采用**事件研究先于盈亏回测**的两阶段设计，先在效应量层面给出判定；并且把“证伪”本身当作目标——本文的全部设计（预注册、冻结门槛、分层随机对照、等价检验）都服务于一个问题：“如果这套方法是错的，我们能否有说服力地说明它错在哪里”。

1.3 贡献

本文的贡献有三。**第一，基础设施**：一套通过 31 项自动化测试（含截断历史重放逐位复现、故意泄漏的反向对照）的多周期结构引擎，把 SMC 语言转为可计算、无未来函数的研究对象，可复用于任何“多周期结构 + 回踩”类假设（第 2 节）。**第二，方法流程**：预注册 + 冻结门槛 + 分层随机对照 + 月区块自助法 + deflated Sharpe + 事件研究先于盈亏的完整证伪流程；本文在事件研究阶段 (P2) 即得到否定答案，盈亏回测 (P3) 仅是确认（第 3 节）。**第三，发现与案例**：对流行方法给出“字面成立 / 归因错误 / 量级不够”的三层判定；外加一个会翻转结论的“成交 K 约定”错误案例，对各类技术策略回测均有警示意义（第 4–6 节）。第 7 节讨论机制解释、方法论教训与局限。

2 数据与策略的机械化

2.1 数据

标的为 19 个 OKX USDT 永续核心币（全程存在、无幸存者筛选；晚上市的 10 个币单列子样本，不进主结论）。样本窗口 2021-01-29 至 2026-07-25，共 66 个日历月、约 365 万根 15 分钟 K 线。结构在三个周期上组织：4 小时为大级别（HTF，定趋势），1 小时为中级别（MTF，定 BOS、位移与订单块），15 分钟为小级别（LTF，作回踩检测与路径还原的基底）；1 小时与 4 小时均由 15 分钟聚合。BTC 另有 1 分钟数据，仅用于校准“同根 K 线内多空同时触及”的保守假设，不参与信号。资金费率以 Binance 永续资金费作 OKX 代理（两所费率高度相关），计入持仓成本。表 1 汇总数据与结构存量。

表 1: 数据与结构存量清单

标的	OKX USDT 永续，19 个核心币	
样本窗口	2021-01-29 – 2026-07-25（66 个日历月）	
基底 K 线	15 分钟，约 365 万根	
结构周期	HTF = 4h, MTF = 1h, LTF = 15m	
BOS 事件	41,620	
顺势合格订单块	9,460	
首次回踩成交（主分析样本）	6,662	
HTF 趋势明确占比	上升 31.1%，下降 30.0%（震荡 39% 不开仓）	
订单块数 / BOS 总数	1.000（每个 BOS 必能回溯到一个订单块）	

注：结构存量来自 P1 验收。“订单块数 / BOS 总数 = 1.000”意味着“存在订单块”是 BOS 的自动推论而非筛选条件（第 2.4 节）。数据来源见复现包中的 P1 结果。

2.2 SMC 语言的严格定义

SMC 的原始表述是口头的、含糊的；要检验它，先要把它无歧义地写下来。表 2 给出核心概念的可计算定义（完整定义与参数取值见附录 A）。要点有三。其一，一切结构点（摆动点、BOS）都带**确认延迟**：一个摆动点要等其右侧 L 根 K 线走完才“成立”，事件流按确认时刻而非 K 线标签排序处理。其二，订单块被操作性地定义为“位移腿起点前最后一根反向 K 线”，区域边界取该 K 线的最高价与最低价（影线口径，主）或开盘价与收盘价（实体口径，附录），宽度以 ATR 归一并设上限。其三，订单块有明确的生命周期：被首次触碰、超时、趋势翻转或被收盘击穿即告终结，第二、三次触碰单独记录，绝不与首次混入同一统计。

表 2: SMC 核心概念的可计算定义（摘要）

概念	可计算定义
摆动点 (pivot)	左严格大于、右不小于的局部极值（等高取第一个），经 ATR 幅度过滤形成高低交替的 ZigZag；右侧 L 根 K 线收盘后才确认。
趋势状态 (HTF)	最近两高两低构成 HH-HL 为上升、LH-LL 为下降，其余为震荡；震荡不开新仓。
BOS (结构破坏)	收盘价突破“尚未被此前 BOS 消费”的最近已确认摆动极值。
位移 (displacement)	BOS 腿上净位移 $\geq d_{\min}$ ATR 且方向效率 $\geq \text{eff}_{\min}$ ；不满足则记为“无位移 BOS”（仍入表，用作对照）。
订单块 (order block)	位移腿起点前最后一根反向 K 线；区域为其高-低（影线口径）或开-收（实体口径），宽度 $\leq w_{\max}$ ATR。
生命周期	自 BOS 收盘激活；被首次触碰 / 超时 / 趋势翻转 / 收盘击穿即终结；首次触碰与第二、三次触碰分列。

注：完整定义、参数取值与理由见附录 A 与设计冻结文档。所有参数在跑数前冻结，取上游方案候选区间的中位而非端点。

2.3 因果性约束与 31 项测试

多周期策略最常见的未来函数，是在粗周期标签上直接做 asof 合并，让一根尚未走完的 4 小时 K 线提前可见。本文用**唯一一条因果性约束**根除它：任何周期、标签为 t 、长度为 D 的 K 线，其派生量（OHLC、ATR、摆动点、BOS 等）的可用时刻一律为 $\text{avail} = t + D$ ；在小级别时刻 u 做任何决策，只允许使用 $\text{avail} \leq u$ 的量。所有高周期特征表都带一列 `avail_at`，对齐到小级别网格时只做 `merge_asof(timestamp, avail_at)`，并断言合并后 $\text{avail_at} \leq \text{timestamp}$ 。

围绕这条约束，引擎通过了 31 项自动化测试，分四类（清单见附录 B）：结构自洽；因果性（含截断历史重放逐位复现、以及一个**故意引入泄漏的反向对照测试**——它保证因果性测试不是恒真的空测试）；成交与度量语义；对照组与主样本锚定于同一事件。这套测试是本研究可信度的基础。

2.4 一个结构性事实：“存在订单块”不构成筛选

P1 验收给出一个必须前置说明的事实：订单块数/BOS 总数 = 1.000。也就是说，**每一个 BOS 都必然能回溯到一根反向实体 K 线**——“找到订单块”这个听起来像在识别稀缺结构的动作，实际上是 BOS 的恒真推论，本身不构成任何筛选。真正起筛选作用的是位移质量与趋势对齐。这一点决定了过滤效应的拆解必须**分层**进行（第 4.2 节）：若不分层，趋势的贡献会被错误地归因于订单块。

2.5 一处会翻转结论的“成交 K 约定”错误

在报告任何结果之前，先记录一个过程中抓出、若不修正将**完全翻转结论**的实现错误，因为它对各类技术策略回测都有警示意义。首版的首次穿越解析**允许成交 K 线自身的最高价去触**

及目标价。在 15 分钟分辨率下这是不可判定的：多头成交发生在该 K 下探到区域近端的时刻，而同一根 K 的最高价完全可能出现在最低价之前，也就是出现在“成交”发生之前。

此处偏差只惠及订单块，因为订单块的止损半径 R 只有 0.91 ATR（回撤规则为 3.0 ATR），“ $1R$ 的距离”小到能在单根 15 分钟 K 线内走完。后果是：修正前 order_block|1R|h48 的成交 K 内触及目标占比高达 24.5%（回撤规则仅 0.15%），毛夏普 3.41、胜率 55.5%，是全表唯一表现突出的组合，与事件研究的否定结论直接冲突；修正后毛均值转负、夏普 -0.86，冲突消失。本文留档一条经验法则：凡是“只有最紧止损那一档特别好”的结果，先查成交 K 约定。

3 研究设计：先判定效应量，再评估盈亏

3.1 核心问题与派生问题

研究不预设订单块有预测能力，而是围绕核心问题 Q0（订单块是否有独立信息）组织一组派生问题：Q1 大级别结构过滤本身有无价值；Q2 中级别 BOS 在趋势过滤之上有无增量；Q3 位移强度与效率有无增量；Q4 订单块这一具体区域相对同宽度任意区域有无增量（即 Q0 的核心）；Q5 首次回踩是否优于第二、三次；Q6 等待小级别 CHOCH 确认相对直接挂单的净期望是升是降；Q7 结构退出是否优于固定 R；Q8 上述任何优势扣除成本后是否仍然存在。各派生问题按过滤层级逐级拆解，每增加一层都必须回答两个问题：这一层加进来之后，前向期望增加了多少，代价是损失多少样本。

3.2 对照组 A-E

证伪的关键是对照。五组对照全部与主样本锚定于同一事件，其中 D 组为分层随机伪区，是 Q0 的直接检验（表 3）。D 组按（币种、日历月、HTF 趋势状态、ATR 扩展窗分位桶）分层匹配，并构造与真实订单块同几何尺寸的伪区域，每个真实事件配 20 个随机抽样，给出对照分布而非单点。

表 3: 对照组 A-E 的定义与各自回答的问题

代号	构造	回答的问题
A	趋势确认后立即市价入场，不等回踩	等回踩到底提高期望，还是只提高了成交交易的表面盈亏比？
B	自近期高点回撤 {0.5, 1.0, 1.5} ATR 触发	订单块是否优于普通的波动率回撤入场？
C	回踩 EMA20 / EMA50	订单块是否优于最朴素的趋势回踩工具？
D	随机伪订单块：同分层内随机取 K 线，构造同 ATR 宽度、同等尺寸的伪区域，走完全相同的触碰与前向测量流程	区域本身有没有独立信息？（Q0 的直接检验）
E	位移起源区但不要求 BOS	拆开“位移”与“结构突破”各自的贡献

注：全部按（币种、日历月、HTF 趋势、ATR 扩展窗分位桶）分层匹配；分层用的 ATR 分位为扩展窗分位，不用全样本分位回填，否则对照组本身引入未来信息。主结论量 Δ_{simple} 取 A/B/C 七组对照中最强者（对订单块最不利的保守设定）。

3.3 预注册与门槛

事件研究 (P2) 的四道门槛在任何结果产生之前冻结: G2-0 功效 (样本量 ≥ 500); G2-1 信息性 (Δ_D 的 5% 分位 > 0); G2-2 优于简单规则 (Δ_{simple} 点估计 > 0 且 5% 分位 > -0.05); G2-3 可交易性 (订单块毛效应量 $\geq 3 \times$ 往返成本 = 0.42%)。四种结局 (从“值得继续”到“订单块被证伪”) 对应的事后动作均在事前规定。回测阶段 (P3) 的 48 格网格在 import 时由 `assert len == 48` 硬性固定, 选择顺序“design 选完才读 validate/test”由代码强制; 但 P3 的预注册 Markdown 是看到结果后补写的, 效力低于 P2, 已如实声明降级 (第 3.6 节)。

3.4 统计口径

全部不确定度用日历月区块自助法估计: 2000 次重抽、全币种联合抽整月、共享抽样矩阵使组间对比配对、随机种子 0xA57E 冻结; 报 90% 置信区间, 不报 p 值星号; 所有子样本同时报事件数。回测阶段用 deflated Sharpe 比率折价 48 格搜索^[8]。需要事先说明**本研究统计功效的边界**: 趋势策略收益由右尾主导且币间高度相关, 有效独立样本约等于独立行情段数 (量级 100–200)。因此本研究能够可靠回答的是“效应量大不大、方向稳不稳”, 而不是“夏普是 1.8 还是 2.1”。

3.5 为何事件研究先于盈亏回测

本文刻意把事件研究放在盈亏回测之前, 理由有三。(a) 事件研究用到**每一次触碰**而非仅成交的交易, 统计功效高得多; (b) 它把信号与执行解耦, 可以把效应归属到单一环节, 否则趋势、位移、订单块、退出四个未定量纠缠在一起无法归因; (c) 退出模型可从同一张路径表事后计算, 省掉一半回测。本项目的经历表明这个顺序至关重要: 若先建回测再找信号, 很可能止步于第 2.5 节那个 3.41 的虚假夏普。整体流程见图 1。

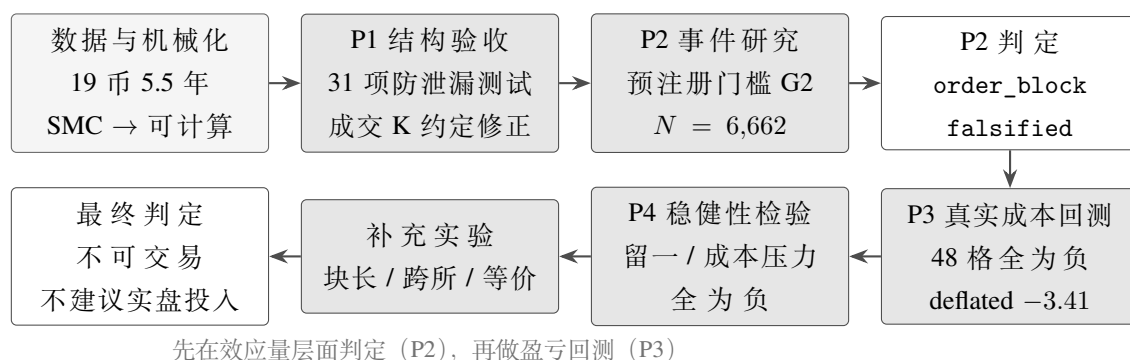


图 1: 研究流程。事件研究 (P2) 在效应量层面先行给出判定, 盈亏回测 (P3) 仅作确认; 每一环节均有对应的事前约束: 预注册门槛、冻结网格、分层随机对照与多重检验折价。

3.6 多重检验与诚实披露

研究纪律以代码而非自律强制, 逐行登记在试验台账: P2 事件研究的 `n_configs` 因看过影线与实体两个区域口径而如实记为 2; P3 网格记为 48; 两个事后 (POST-HOC) 分析 (实体口

径、CHOCCH 确认) 明确标注、不与预注册结果混列。P3 预注册为事后补写、效力降级, 已声明。这些披露与第 7 节的局限一并构成本文证据定级的框架。

4 事件研究结果 (P2)

4.1 主判定

事件研究的检验统计量是**成交后 12 小时的顺势收益**, 以成交时的 1 小时 ATR 归一。订单块相对随机匹配区域的增量效应量为 $\Delta_D = -0.1343 \text{ ATR}$ (90% CI $[-0.2079, -0.0542]$, $N = 6,662$); 相对同一 BOS 时刻最强简单对照的增量 $\Delta_{\text{simple}} = -0.0186$ (CI $[-0.0941, +0.0240]$ 跨零)。订单块自身毛效应量 $\mu_{\text{OB}} = 0.0199 \text{ ATR} = 0.0246\%$, 仅为单次往返成本 ($14\text{bp} \approx 0.113 \text{ ATR}$) 的 18%。四道预注册门槛过一道 (表 4), 判定为 `order_block_falsified`。G2-0 通过这一点很关键: **这是一次有统计功效支持的否认, 而不是“样本不够”**。

表 4: P2 门槛判定表

门槛	预注册判据	实测	结论
G2-0 功效	$N \geq 500$	$N = 6,662$	通过
G2-1 信息性	5% 分位 (Δ_D) > 0	$\Delta_D = -0.1343$, CI 不过 $[-0.2079, -0.0542]$	不过
G2-2 优于简单规则	点估计 > 0 且 5% 分位 > -0.05	$\Delta_{\text{simple}} = -0.0186$, CI 不过 $[-0.0941, +0.0240]$	不过
G2-3 可交易性	$\mu_{\text{OB}} \geq 3 \times \text{往返成本} = 0.42\%$	$\mu_{\text{OB}} = 0.0246\%$	不过

注: 最强简单对照为 B 组 1.5 ATR 回撤 ($\mu = +0.0385$)。成本基准: taker 进 5bp + taker 出 5bp + 每边 2bp 滑点 = 往返 14bp, 以 ATR/price 中位数 0.01238 折算为 0.113 ATR。稳健性指标: 缩尾均值 -0.0086 、中位数 -0.0799 、按币中位数 -0.0187 、正币数 9/19。

解读结果前需要先说明两点。其一, 订单块的点估计 $+0.0199$ 看似为正, 但 CI 宽到跨零、缩尾均值与中位数均为负、19 个币只有 9 个为正——**连“订单块本身有正期望”这一最弱主张都没有得到支持**。其二, Δ_D 有识别性缺陷: D 组未匹配“距最近摆动极值的距离”, 订单块位于位移腿起点 (即价格已经走过一段之后回头寻找的位置), 随机伪区则可能落在趋势中段, 故 Δ_D 混入了“位移腿起点不如趋势中段”这一位置效应。因此本文以 Δ_{simple} 为主结论量——A/B/C 组锚定在同一个 BOS 时刻, 是严格的事件内对比, 唯一差异是“等哪个价位”; Δ_D 降级为参照基准。 Δ_D 仍显著为负这一事实意味着: 同尺寸随机区域的 12 小时顺势收益并不比订单块差, 且 19/19 个币都为正——后者的正均值主要来自 2021–2026 加密市场自身的漂移, 而非任何入场方法的作用。

4.2 逐级过滤检验: 效应来自哪一层

把 SMC 的三层过滤逐级叠加 (每层是上一层的子集), 各层“通过后立即入场”的 12 小时顺势收益见表 5 与图 2。六层里没有任何一层的置信区间排除零: 点估计从 $+0.003$ 升至 $+0.028$

再回落到 +0.006，起伏全部落在噪声量级。把这组逐级结果当作“位移幅度有用”或“订单块有害”的证据，都是过度解读。

表 5: 逐级过滤检验 L1-L6: 每层过滤通过后立即入场的 12 小时顺势收益

层级	N	μ (ATR)	90% CI	相对上一层增量
L1 HTF 趋势	557,831	0.0029	[-0.0414, 0.0465]	—
L2 + BOS	12,235	0.0141	[-0.0639, 0.0920]	+0.0112
L3 + 位移幅度	10,732	0.0278	[-0.0554, 0.1133]	+0.0137
L4 + 位移效率	10,731	0.0274	[-0.0558, 0.1131]	-0.0004
L5 + 订单块存在	9,458	0.0058	[-0.0842, 0.0973]	-0.0216
L6 + 等到首次回踩	6,662	0.0199	[-0.0761, 0.1102]	+0.0141

注：L1-L5 为该层过滤通过即在 BOS 收盘立即入场，L6 为再多等一次回踩成交。
 两点附带发现：`min_efficiency` 过滤器只淘汰 1 个样本 (10,732 $\rightarrow$ 10,731)，是一个失效参数；L4 $\rightarrow$ L5 的样本缩减来自区域宽度等质量门槛，而非订单块概念本身（“存在订单块”恒成立，见第 2.4 节）。

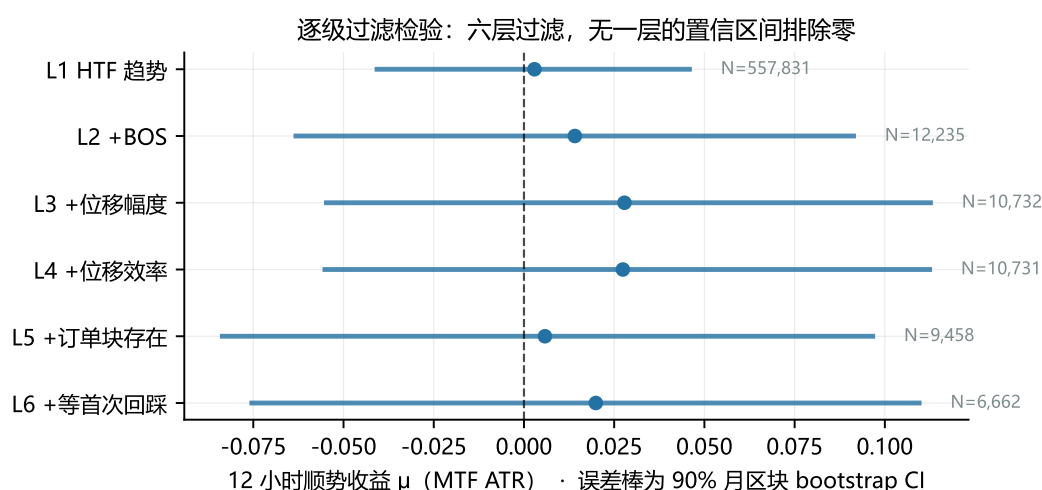


图 2: 逐级过滤检验 L1-L6 的点估计与 90% 月区块自助法置信区间。六层过滤无一层的区间排除零，即“顺大级别趋势 + 等 BOS + 等位移 + 找订单块 + 等回踩”这条链上，没有任何一环的效应量与零可区分。

4.3 反转频率检验：字面主张对不对

前述分析测的都是**收益**，而 SMC 的字面主张更窄——价格到了订单块更容易反转——这是一个关于**频率**的主张，完全可能成立而策略仍不赚钱，值得单独测。统计量为：从成交价出发沿 15 分钟路径前行，记录 +1 ATR 与 -1 ATR 谁先到。选择该统计量的理由有四：无信息基准恰好是 50/50，可独立解读；统计量对称，不依赖止损定义； ± 1 ATR 远在区域之外，并非“区域是否有效”的另一种问法；ATR 单位与尺度无关，八种规则可放在同一坐标轴上比较。结果见表 6

与图 3。

表 6: 反转频率与成本对照表（按反转率降序）

入场规则	<i>N</i>	反转率	90% CI	完全穿透	毛期望	净期望
D 随机同分层伪区	118,649	59.24%	[58.55, 59.93]	27.96%	0.1375	+0.0251
C EMA50	7,682	56.71%	[55.38, 58.07]	—	0.0936	-0.0188
订单块	6,666	55.54%	[54.11, 56.97]	30.30%	0.0761	-0.0364
B 1.5 ATR 回撤	9,061	53.25%	[52.05, 54.47]	—	0.0530	-0.0595
C EMA20	8,812	52.61%	[51.60, 53.65]	—	0.0321	-0.0803
B 1.0 ATR 回撤	9,299	50.75%	[49.55, 51.88]	—	0.0110	-0.1015
A 立即入场	9,460	50.28%	[49.09, 51.50]	—	0.0036	-0.1088
B 0.5 ATR 回撤	9,442	49.99%	[48.84, 51.15]	—	-0.0015	-0.1139

注：反转率为“已判定”样本（剔除同 K 双穿与超时未到）中先到 +1 ATR 的比例；毛/净期望为 ± 1 ATR 对称括号的全样本口径（赢 +1、输 -1、同 K 记亏、超时记 0），成本为单次往返 14bp = 0.1124 ATR（未含资金费）。订单块与随机伪区之反转率差为 -3.69 个百分点，90% CI [-5.08, -2.35]（配对抽样）。八种规则中唯一净为正的是不可交易的随机对照。

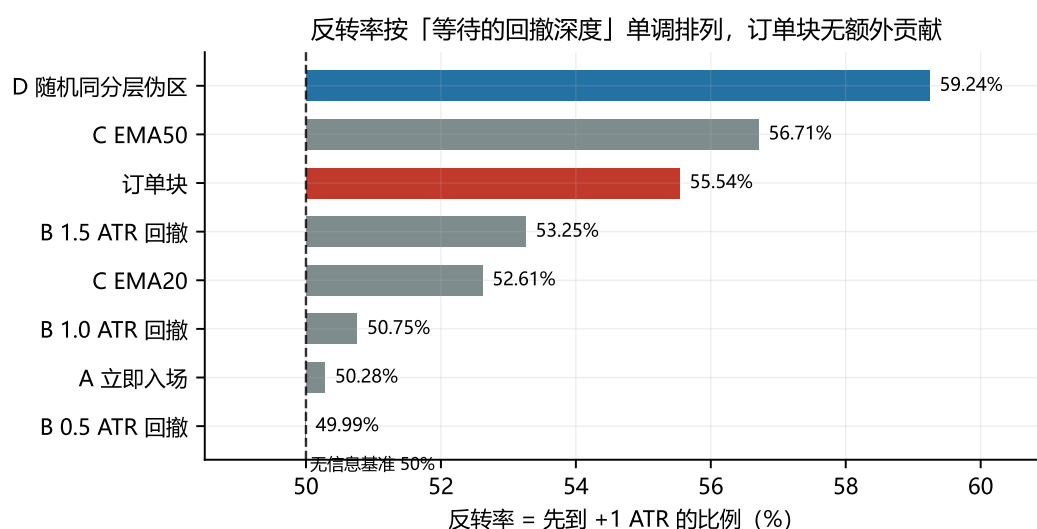


图 3: 反转率按“等待的回撤深度”单调排列。立即入场恰为 50.28%（测量方法无偏的自检）；订单块的反转率恰好处于其回撤深度所预测的水平，无额外贡献；同尺寸随机伪区反而反转得更频繁。虚线为无信息基准 50%。

以下四点必须放在一起看。第一，订单块的反转率确实显著高于 50%（55.54%，CI 排除 50）——这是全项目唯一一个字面支持该方法的结果。第二，但这不是订单块本身的贡献：整张表按“等待的回撤深度”单调排列，立即入场恰好落在 50.28%（这同时是测量方法无偏的自检），订单块的反转率恰好处于其回撤深度所预测的水平。机制在于测量原点——等得越深，+1 ATR 距离价格刚刚走过的位置越近，越容易到达。第三，同尺寸随机区域反转得更频繁（59.24%，差 -3.69 个百分点，CI 排除 0）；且订单块被价格完全穿透的比例（30.30%）甚至略高于随机区域

(27.96%)，即订单块并不比随机选取的同尺寸区域更“坚固”。**第四，也是最关键的一点，是幅度**：反转的毛期望仅 0.0761 ATR，而单次进出的成本为 0.1124 ATR。

4.4 反转为何仍不赚钱：成本表

表 6 的最后一列是全项目最容易理解的一个数：订单块的反转是真的，毛期望 +0.0761 ATR，但单次进出的成本为 0.1124 ATR——**真实存在的优势只有成本的 68%**。八种规则里唯一净为正的是随机对照 (+0.0251 ATR)，而它不可交易；可交易的七种全为负。这与第 4.1 节的 G2-3（毛效应量 = 往返成本的 18%）是同一件事的两种算法，差别只在出场模型，结论同向。

4.5 机会成本与收益集中度

等待回踩是有代价的。9,460 个顺势合格区域里 **29.5% 在终止前从未被回踩**；这批被放弃的行情并非没有启动——12 小时内 87.4% 已顺势走了 ≥ 1 ATR，中位数 2.92 ATR（24 小时为 93.7%、4.05 ATR）。这解释了为什么 L2（BOS 后立即入场）与 L6（等回踩）在均值上打平：**等回踩换来的更优入场价，恰好被漏掉的那批最流畅的趋势所抵消**——在竞争性市场中，更优的价格与更低的成交概率本质上是同一笔交换。

收益集中度是全篇最关键的一个数：全样本 $\mu = +0.0199$ ，剔除最好的 1%（67 笔）后 $\mu = -0.1176$ ，**这 67 笔贡献了总和的 685%**。即主口径那个“看起来是正的”均值完全由 1% 的极端样本支撑，其余 6,595 笔合计显著为负。任何声称该方法有效的说法，都必须先解释为什么可以期望那 67 笔重复出现。此外，即便是最终盈利的那一半样本，也有一半曾先逆行超过 0.75 ATR——“精准入场”的说法在数据上不成立。

4.6 两条方向有利但仍不足的线索

有两处结果的方向对该方法有利，必须如实报告，并说明为何仍然不够。**其一，实体区域口径**：用开-收（实体）替代高-低（影线）定义订单块，每个指标都更优—— μ 从 0.0199 升至 0.0474， Δ_{simple} 从 -0.019 翻正为 +0.009（实体订单块均值首次略超最强简单对照）， Δ_D 仍为 -0.103（CI 排除 0）。但它仍不过任何门槛（毛效应量为往返成本的 42%），且这是同一批数据上的第二个口径（台账如实记 `n_configs=2`）——两次尝试下看到四分之一标准差的漂移，完全在预期之内。**其二，等待小级别 CHOCH 确认**：成交率降至 83.2%，入场价高 0.43 ATR，换来 12 小时毛均值 $0.0181 \rightarrow 0.0462$ ATR（毛/成本 $0.18 \rightarrow 0.42$ ），单笔机会的净亏损从 -0.092 收窄到 -0.054 ATR。方向对，但改进量的 CI 跨零 ($+0.028$, $\text{CI} [-0.042, +0.093]$)，且这是事后分析（台账标注 POST-HOC）。两条线索的毛/成本都停留在 0.42，距打平仍差一倍多，且改进的模式同为“少交易、在行情启动后才进场”——与第 5 节网格最优配置位于参数边界是同一种模式。

5 真实成本回测结果 (P3-P4)

5.1 48 格全负与多重检验折价

P2 否定的是“订单块这一特定价位”，并未否定“顺势回撤入场”这个更一般的做法。因此 P3 去掉订单块层、保留“HTF 趋势 + BOS + 回撤入场”框架，用真实成本 (14bp 往返 + 3bp/日资金费)、真实并发约束与样本外切分，把 48 格配置完整回测。结果 **一致性为负**：48 个配置里 0 个在 design 期有正的 `mean_R`，0 个有正的夏普。按 design 期夏普选中的配置为 `atr_pullback_1|time|h96`，其 design 期夏普 -0.056 ，**deflated Sharpe 为 -3.41** ——在一个 48 格网格上搜索，即使每一格都毫无优势，仅凭挑选最优者也平均能获得 $+3.35$ 的虚增夏普，而实际挑出的最好一格是 -0.056 ，比“无优势”的零假设还差得多（表 7）。

表 7: P3 design 期 (2021–2023) 48 格前 8 名 (按夏普排序) 与多重检验折价

配置	笔数	mean_R	胜率	夏普	最大回撤	总 R
atr_pullback_1 time h96	3,373	-0.0111	38.7%	-0.056	-67.3%	-37.6
atr_pullback_0.5 time h96	3,428	-0.0153	38.8%	-0.114	-67.1%	-52.5
atr_pullback_0.5 time h48	4,112	-0.0209	42.5%	-0.282	-67.5%	-86.0
atr_pullback_1 time h48	4,062	-0.0348	40.5%	-0.443	-73.2%	-141.3
atr_pullback_1 time h24	4,529	-0.0363	41.7%	-0.534	-74.3%	-164.6
atr_pullback_0.5 time h24	4,611	-0.0358	42.6%	-0.965	-65.0%	-164.9
order_block time h96	3,327	-0.1308	19.5%	-1.075	-95.6%	-435.1
atr_pullback_0.5 3R h48	4,140	-0.0428	42.5%	-1.168	-71.9%	-177.3

注：48 格中 `mean_R` > 0 与夏普 > 0 的配置均为 0。deflated Sharpe = -3.41 (48 格横截面夏普离散度推出的 $\mathbb{E}[\max \text{夏普} | \text{无优势}] = +3.35$)。网格最优配置位于参数边界 (最长 horizon、无目标价、纯时间出场)，呈现“少交易就少亏”的模式。订单块是最差的入场规则 (夏普 -1.07 、回撤 -95.6%)；其紧止损使成本占掉 0.16 个 R，而回撤规则仅占 0.05–0.07。

5.2 尺度无关的横向对比：结构筛选的价值

各入场规则的止损半径不同 (订单块 0.91 ATR vs 回撤规则约 3 ATR)，`mean_R` 不可跨规则比较。把毛收益、成本、净收益全部换算成尺度无关的 ATR、出场统一为 `time|h96`、成本统一 $14\text{bp} + 3\text{bp/日}$ ，得到全项目最清晰的一张表 (表 8 与图 4)：**成本是一条水平线 (约 0.146 ATR)**，九种规则没有一种的毛优势达到这条线；最接近打平的是随机对照，订单块排倒数第二。

表 8: 九种入场规则的尺度无关横向对比（同一出场、同一成本，单位 ATR）

入场规则	笔数	毛收益	成本	净收益	毛/成本
D 随机同分层伪区	118,593	0.1507	0.1519	-0.0012	0.99
C EMA50	7,679	0.0714	0.1458	-0.0744	0.49
B 1.0 ATR 回撤	9,292	0.0543	0.1467	-0.0925	0.37
B 1.5 ATR 回撤	9,054	0.0519	0.1467	-0.0948	0.35
A 立即入场	9,453	0.0378	0.1459	-0.1081	0.26
B 0.5 ATR 回撤	9,435	0.0363	0.1461	-0.1098	0.25
E 位移无 BOS	51,986	0.0291	0.1453	-0.1162	0.20
订单块	6,660	0.0282	0.1466	-0.1184	0.19
C EMA20	8,805	-0.0077	0.1428	-0.1505	-0.05

注：逐笔换算为 ATR，不施加并发约束（随机对照有 20 倍过采样，比的是逐笔期望）。随机对照不是可交易策略（未匹配“距摆动极值的距离”，优势里混着趋势中段的位置优势与市场漂移），它的作用只是标定“结构性筛选相对随机入场有多少增量”的参照基准——答案是负值。

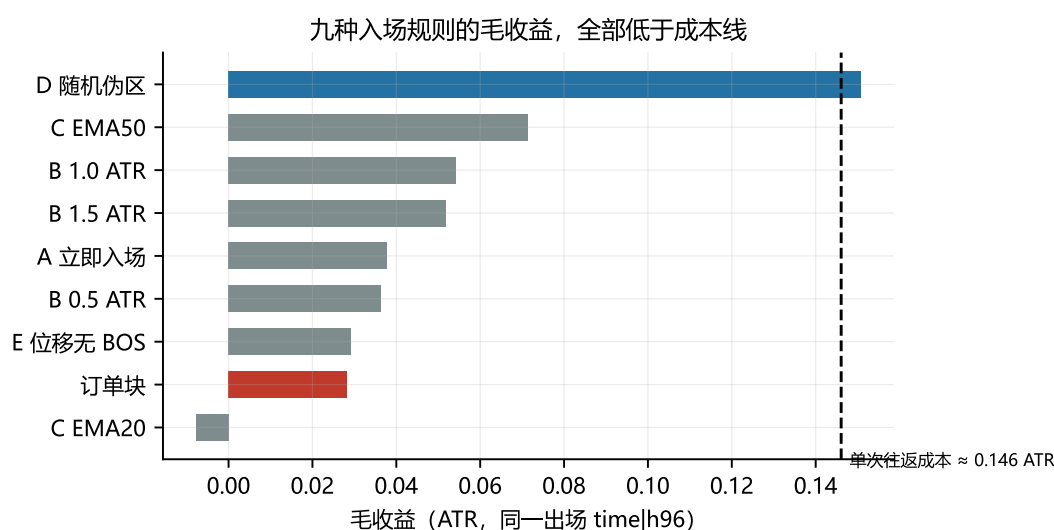


图 4: 九种入场规则的毛收益（ATR，同一出场 time|h96）对照成本水平线。没有一种规则的毛优势达到成本；最接近打平的是不可交易的随机对照，订单块排倒数第二。

5.3 样本外与成本压力

选中配置在样本外继续为负：validate 期（2024–2025） $\text{mean_R} = -0.063$ 、bootstrap $p(\text{mean_R} \leq 0) = 0.943$ （门槛要求 < 0.10 ）；test 期（2026）同样为负。但该框架并非完全没有信息：零成本对照下，顺势回撤入场的毛期望为 $+0.026 \text{ R/笔}$ （毛夏普 $+0.39$ ）。问题在于打平的条件（表 9）：要约 7bp 往返成本且资金费为零才刚好打平。

表 9: 成本压力测试：选中配置 atr_pullback_1|time|h96 在不同成本假设下

往返成本 (bp)	资金费 (bp/日)	mean_R	夏普	总 R
7	0	-0.0006	+0.063	-3.9
7	3	-0.0047	+0.013	-29.0
7	6	-0.0087	-0.038	-54.2
14 (基准)	3	-0.0310	-0.315	-193.3
25	3	-0.0724	-0.828	-451.4
40	3	-0.1289	-1.525	-803.4

注：零成本对照的毛 mean_R = +0.0257、毛夏普 +0.390。打平需约 7bp 往返且零资金费：7bp 往返意味着双边 maker 成交且零滑点，但该配置为时间出场（24 小时后无条件平仓），maker 无法保证成交；资金费为零在永续上不可能（持仓 24 小时跨 3 个资金费结算点）。故在永续合约上，“有信息但收益被成本吞噬”这一出路并不成立。

这个区分决定后续动作：若“毫无信息”，应放弃该方向；若“有信息但收益不足以覆盖成本”，应寻找更便宜的执行方式。而 7bp 往返 + 零资金费在永续合约上不可达，因此后一条路同样走不通。六道 G3 门槛 G3-0 通过，G3-1 (deflated Sharpe -3.41)、G3-2 (validate $p = 0.943$)、G3-3 (25bp 下 mean_R -0.072)、G3-4 (剔除 top 1% 后 -0.123) 全不过，G3-5 形式上满足但无信息量（邻域 9 格同为负）。

5.4 稳健性检验与净值曲线

稳健性检验全部一致为负：留一年 6 次全负；留一币 19 次全负（剔除 XRP 后结果反而最差，为 -0.053，说明 XRP 是少数对结果有正贡献的币种；“单一币种支撑全部盈利”的过拟合症状确实存在，但即便有它支撑，总体仍为负）；剔除最好 1% 交易后 mean_R 从 -0.031 恶化到 -0.123（亏损扩大 3-4 倍）。全期净值从 1.0 跌到 **0.282**，最大回撤 -82.6%，CAGR -20.7%（图 5）。全期 profit factor 0.936，胜率 38.8%，盈亏比 1.478，中位数交易 median_R = -0.310——盈利时确实赚得多，但中位数交易明确亏损。

6 补充实验

主研究 (P0-P5) 完成后，本文补做三组实验，正面回应三个遗留质疑：区块长度是主观选择、结论可能依赖单一交易所、零结果可能是低功效所致。三者均先写预注册再跑数，与主研究同一纪律；产物与台账逐行登记（附录 C）。

6.1 Bootstrap 区块长度敏感性

主口径的自助法区块为日历月，是主观选择。若结论只在某一块长下成立，证伪不牢。用六种块长（半月、1 月基准、2 月、3 月、6 月，外加平稳自助法）对同一批冻结数据重抽，重



图 5: 选中配置 atr_pullback_1|time|h96 的全期净值曲线，从 1.000 跌到 0.282（5.5 年亏掉约 72% 本金），最大回撤 -82.6%。净值曲线的绝对形状依赖被冻结的 max_concurrent 与 risk_per_trade，但判定用的 mean_R、胜率、盈亏比不依赖它们。

新估计各量的置信区间。四道预注册检查在六种块长下 24/24 全部通过 (Δ_D 的 CI 始终排除 0、 Δ_{simple} 的 CI 始终跨 0、反转率差 CI 始终排除 0、validate 期 $p(\text{mean_R} \leq 0)$ 始终 > 0.10)；即便在分辨率最差的 6 月档（仅 12 块）， Δ_D 的 CI 上界仍为 -0.037，仍排除 0。独立重写的重抽样器在 1 月档逐位复现主口径三个核心区间，两份实现互为验证（图 6）。

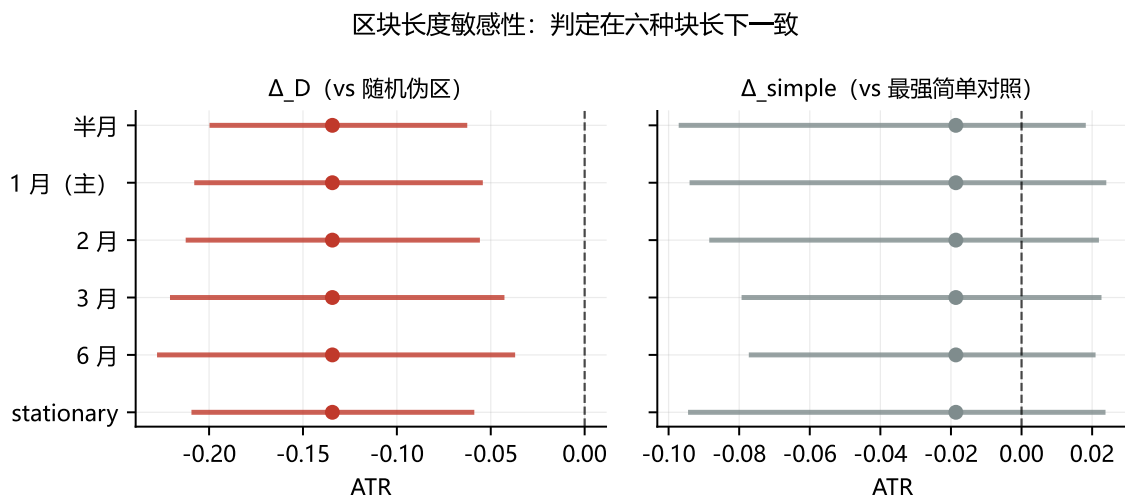


图 6: 区块长度敏感性。 Δ_D 与 Δ_{simple} 的 90% 置信区间在六种块长下判定一致： Δ_D 始终排除 0（红）， Δ_{simple} 始终跨 0（灰）。证伪结论不是月区块自助法的产物。

一个不完全符合预期的发现值得报告：随块长增加， Δ_D 的区间宽度单调增大（0.137 $\rightarrow$ 0.191，+39%），但 μ_{OB} 的宽度反而略降（0.186 $\rightarrow$ 0.163）。这说明序列相关存在于配对差值而非水平——跨币种与跨时间的共同成分体现在差值上，而不在水平上。它不改变任何判定，但事后印证了主研究把 Δ_{simple} （配对差值）而非 μ_{OB} （水平）定为主结论量是合理的选择。

6.2 第二交易所 (Binance) 冻结复制

单一交易所是主研究自列的缺陷之一。加密同一币种跨所走势高度同源，故做**冻结复制**：同一份配置、同一代码、同一种子、同一套门槛，唯一变量是价格数据源 (OKX 换 Binance)。BTC 从 Binance Vision 补下载 (2021-01 至 2026-06 共 192,672 根 15 分钟 K 线，零缺口零重复)；缺 BCH/ETC/TRX 三币 (预注册已声明，不替换)，共 16 币，取两所共有区间 2021-01-29 至 2026-07-08。不重跑 48 格网格 (那会引入 48 次额外的多重检验，却不回答新问题)，只复核完整 P2、反转检验与单个冻结 P3 配置。结果见表 10 与图 7。

表 10: Binance 冻结复制与 OKX 主口径对照

量	OKX (主, 19 币)	Binance (复制, 16 币)
首次回踩成交 N	6,662	5,583
μ_{OB} (ATR)	+0.0199 [-0.0761, +0.1102]	+0.0182 [-0.0841, +0.1187]
Δ_D (ATR)	-0.1343 [-0.2079, -0.0542]	-0.1640 [-0.2465, -0.0727]
Δ_{simple} (ATR)	-0.0186 [-0.0941, +0.0240]	-0.0477 [-0.1207, -0.0093]
反转率	55.54% [54.11, 56.97]	55.86% [54.32, 57.43]
反转率差 (OB - 随机)	-3.69pp [-5.08, -2.35]	-3.88pp [-5.42, -2.39]
完全穿透 (OB / 随机)	30.30% / 27.96%	29.33% / 27.09%
反转毛优势 / 成本	68%	76%
冻结 P3 配置全期 mean_R	-0.0310	-0.0196
判定词	order_block_falsified	order_block_falsified

注：Binance 在每个指标上都比 OKX 更不利，但两所 Δ_{simple} 的点估计相互落在对方的置信区间内，是相容的估计。 Δ_{simple} 的 CI 在 Binance 上排除 0 (方向更不利)，在 OKX 上跨 0；主结论以更保守的 OKX 跨零版本为准，Binance 版本作“复制不仅同向、且方向更不利”的补充证据。一处符号翻转如实披露：冻结 P3 配置在 Binance 的 design 期 mean_R 为 +0.0165，但不显著 ($p = 0.390$, CI 跨零)，validate / test / 全期全为负。

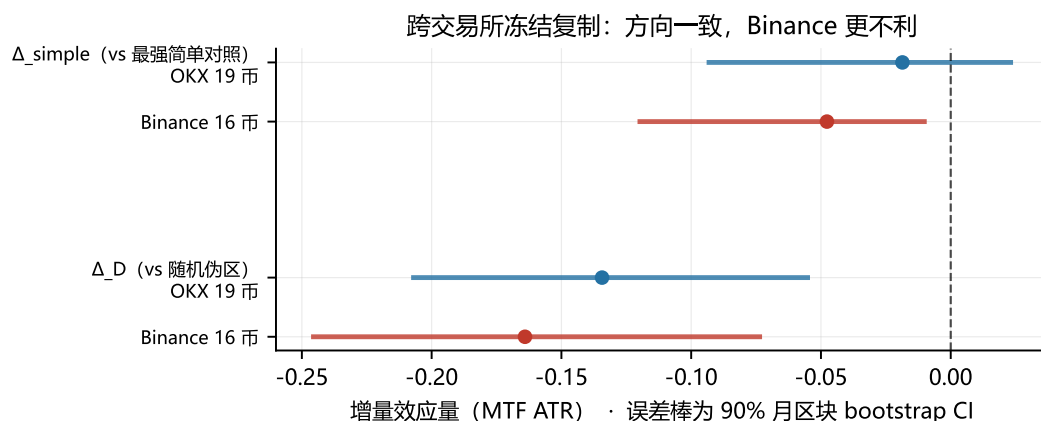


图 7: 跨交易所冻结复制。 Δ_D 与 Δ_{simple} 在两所方向一致，且 Binance 均更负。证伪不是 OKX 微观结构的产物。

预注册判定词在 Binance 上仍为 order_block_falsified，四道门槛逐项相同。反转检验

复现得尤为清晰：八个规则的**单调排序逐位相同**，立即入场的 50% 自检独立复现 (50.83%)，随机伪区仍反转更频繁 (差 -3.88pp , CI 排除 0)，反转毛优势仍仅为成本的 76%。唯一与 OKX 方向不同的数字——design 期 mean_R 翻正为 $+0.0165$ ——不显著 ($p = 0.390$)：一个贴近零的量在复制中变号，恰恰说明它贴近零，而非说明它有效。**本实验的定位是排除“交易所微观结构依赖”这一类失败模式，而非把证据量翻倍**：同一币种在两所由同一批套利者做市，且主研究的资金费本就用 Binance 数据代理，新增信息全部来自价格腿。

6.3 SESOI、经济等价检验与功效曲线

“零结果”最易受的质疑是“会不会只是功效不足”。本实验把“不显著”升级为“统计等效于零”，并定量区分“是没测出来，还是真没有”。SESOI 定为单次往返成本 0.113 ATR (由成本定义，非由数据定义)：小于成本的效应即使真实存在，也不足以覆盖一次进出的费用。等价检验 (TOST) 与功效曲线结果见表 11 与图 8。

表 11: 经济等价检验与功效/最小可检测效应 (MDE)

被检量	点估计	90% CI	TOST (± 0.113)	SESOI 处功效	MDE (80%)
Δ_{simple} (主结论量)	-0.0186	$[-0.0941, +0.0240]$	等价 ($p = 0.021$)	0.939	0.087
μ_{OB} (绝对水平)	+0.0199	$[-0.0761, +0.1102]$	等价 ($p = 0.046$)	0.654	0.139
Δ_D	-0.1343	$[-0.2079, -0.0542]$	不等价	0.757	0.120

注：SESOI = 单次往返成本 = 0.113 ATR。TOST 为 bootstrap 版两个单侧检验。在更严的 ± 0.05 ATR 界下， Δ_{simple} 与 μ_{OB} 均不等价 ($p = 0.318/0.299$)，如实报告。非优越性检验：bootstrap 分布 95% 分位 = 0.1102 < SESOI，在 5% 水平上拒绝“订单块赚得回一次往返成本” ($p = 0.046$, 系勉强拒绝)。功效用中心化 bootstrap 分布做非参数估计，正态近似作交叉校验。

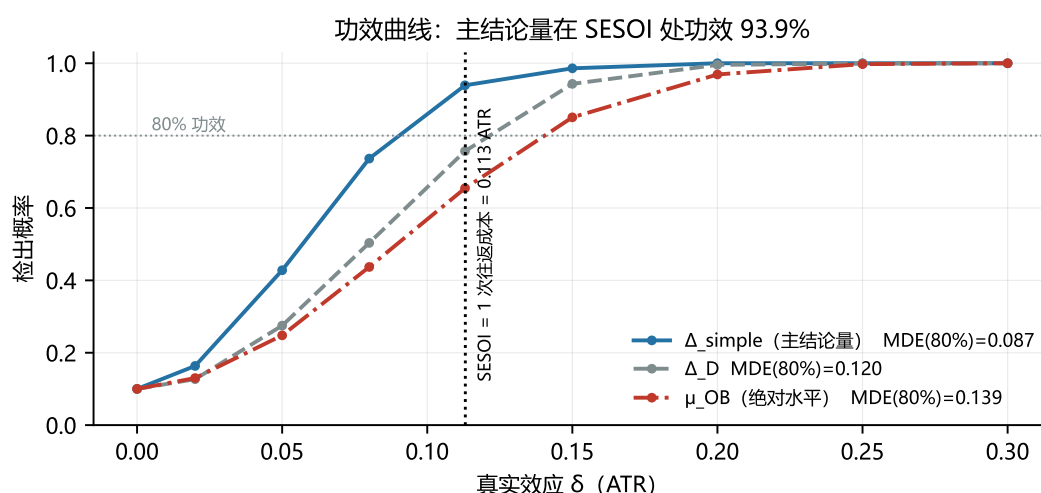


图 8: 功效曲线与 SESOI 竖线。主结论量 Δ_{simple} 在 SESOI 处功效 93.9% (MDE 0.087 < SESOI); 绝对水平 μ_{OB} 功效仅 65.4% (MDE 0.139 > SESOI)。对主结论量而言，“零结果 = 低功效”这一反驳被定量排除。

三道预注册检查过两道、明确不过一道，全部如实报告。**S3-1 通过**： Δ_{simple} 与 μ_{OB} 在 ± 1

次往返成本界内统计等效于零——措辞可从“无法与零区分”升级为“等价于零”。**S3-3 通过**：在 5% 水平上拒绝“订单块赚得回一次往返成本”（95% 分位 $0.1102 < 0.1131$, $p = 0.046$ ；但这是勉强拒绝，差 0.003 ATR，正文不写成“远低于”）。**S3-2 不过**：MDE 要求两个量都低于 SESOI, Δ_{simple} 过了 (0.087, SESOI 处功效 93.9%)， μ_{OB} 没过 (0.139, 功效仅 65.4%)。原因是结构性的： μ_{OB} 是单组水平量、承受全部市场共同波动， Δ_{simple} 是同一批 BOS 事件内的配对差值、共同成分被差分消除（标准误小 37%）——与第 6.1 节“差值序列相关强、水平序列相关弱”是同一件事的两面。这一条**限制了证伪的表述范围**：能说“相对最强简单对照没有超出成本量级的增量，且可拒绝它赚回一次往返成本”，**不能说**“已证明订单块的绝对收益小于成本”。主结论建立在前者之上，这也是主研究事前把 Δ_{simple} 而非 μ_{OB} 定为主结论量的又一条理由。

7 讨论

7.1 机制：为什么反转“真”却不赚钱

三层判定背后是一个统一的机制。反转率之所以“真”，是**测量原点**的几何后果：等得越深，+1 ATR 距离价格刚刚走过的位置越近，越容易到达——与订单块这一价位是否“有记忆”无关。起作用的是**等待**，不是**信息**。而等待之所以不赚钱，是因为在竞争性市场中，更优的价格与更低的成交概率本质上是同一笔交换（第 4.5 节）：29.5% 的区域从不被回踩，其中 87% 的行情已顺势走远，等回踩的机会成本恰好抵消其入场价优势。最后，即便把这个真实的、由深度驱动的反转优势完整兑现，其幅度也只有单次往返成本的 68%。“有反转”与“能赚钱”之间，差着一个成本量级。

7.2 方法论教训

三个教训对各类技术策略回测都有意义。其一，**成交 K 约定**（第 2.5 节）：凡是“只有最紧止损那一档特别好”的结果，先查成交 K 约定——本项目若不查这一步，会发表一个虚假夏普 3.41 的相反结论。其二，**对照组的识别性**： Δ_D 因未匹配“距摆动极值的距离”而混入位置效应，故必须以锚定同一事件的 Δ_{simple} 为主结论量；随机对照永远只是参照基准、不是策略。其三，**事后变量不可用作过滤器**：回踩深度表（“被完全穿越 → 深亏”）在成交后测量，“被完全穿越”几乎等价于“止损被触发”，在交易时点上不可知。

7.3 对 SMC 方法的三层判定

对 SMC 方法本身，分三层表述比一句“是骗局”更准确。**字面成立**：价格回到订单块之后确实更容易反转（反转率 55.5%，CI 排除 50），这不是错觉，而是可测量的真实现象。**归因错误**：这个反转率由“等待的回撤深度”解释，而非由“订单块这个价位有记忆”解释——立即入场 50.3%、随机同尺寸区域 59.2%，都更能说明问题。**量级不够**：反转的毛期望只有单次往返成本的 68%，即便归因正确也无法覆盖成本。“精准入场”亦被否定：即便最终盈利的样本，也有一半曾先逆行超过 0.75 ATR。

7.4 局限

本文的局限按重要性如实列出。其一，主证据来自单一交易所 OKX——第 6.2 节的 Binance 冻结复制排除了“微观结构依赖”这一类失败模式，但两所价格高度同源，复制是“排除一类失败”而非“独立证据”。其二，未做容量与冲击成本建模；14bp 含每边 2bp 滑点，在 19 个大市值永续上小仓位大致合理但未验证。其三，`max_concurrent=6` 与 `risk_per_trade=0.005` 是人为设定而非数据标定的（被冻结以避免用杠杆制造夏普），净值曲线的绝对形状依赖它们，但判定用的 `mean_R`、胜率、盈亏比不依赖。其四，P3 预注册为事后补写，效力低于 P2；实体口径是同一批数据的第二个口径。其五，可能的“区域质量”人工判断（如某种流动性缺口联合条件）未实现——若真实 SMC 交易者用了某个本文未实现的关键条件、恰好把 6,662 笔筛到几百笔高质量样本，结论可能被推翻；但本文口径可复算（配置见附录 D），这种质疑可以通过具体检验来验证，而不必停留在争论层面。

8 结论

订单块方法的核心主张**字面成立**——价格回踩后确实更易反转（55.5%），但**归因错误**——这是回撤深度而非价位记忆的作用，且**量级不够**——反转优势仅为单次往返成本的 68%。预注册事件研究（ $\Delta_D = -0.134 \text{ ATR}$ ，CI 排除 0）、48 格真实成本回测（全负，deflated Sharpe -3.41 ）、第二交易所冻结复制（同向且更不利）、经济等价检验（在 ± 1 次往返成本界内等效于零）一致判定：这套聪明钱概念的趋势延续与回踩策略**不具备可交易性，不建议投入实盘资金**。对方法更有普遍意义的教训是：若先建回测再找信号、且不查成交 K 约定，本文将得出一个虚假夏普 3.41 的相反结论并就此止步。

A 参数取值表

P2 只跑一个固定配置（事件研究不做配置选择，证据标准落在效应量上），参数取上游方案候选区间的中位而非端点，全部在跑数前冻结（表 12）。

表 12: 参数取值表（冻结值与理由）

参数	值	理由
L_{htf} / L_{mtf}	5 / 5	4h 约 280 摆动/年、1h 约 1100 摆动/年，样本量充足
k_{pivot}	0.5 ATR	候选区间下端；过高会把结构点滤到不足
β (BOS 缓冲)	0	纯收盘突破，假设最少
d_{min} / eff_{min}	1.0 ATR / 0.5	候选区间下端，不做苛求
M (向前搜索)	5	区间上端，减少无反向 K 线的流失
W (位移腿上限)	3	同上游方案
w_{max}	1.5 ATR	区间中位
E (过期)	40 根 MTF	区间中位 (约 1.7 天 @1h)
止损缓冲	0.25 ATR	区间中位
区域版本	wick (主) / body (附录)	主口径用影线，实体作稳健性附录

注：P3 参数预算上限为 48 个配置，网格在跑数前写入 P3 预注册并由 `assert` 固定。

B 防未来函数测试清单

引擎通过 31 项自动化测试，归纳为七类（表 13）。其中 T1（截断历史重放）是状态机防泄漏最有效的检验手段，T4 的故意泄漏反向对照保证因果性测试不是恒真的空测试。

表 13: 防未来函数测试清单（七类，共 31 项测试函数）

编号	测试	说明
T1	截断历史重放	对递增的 k 只喂 <code>data[:k]</code> ，断言所有 <code>avail_at ≤ t_k</code> 的事件行逐字段不变（容差 10^{-9} ）。
T2	摆动点确认延迟	合成数据：摆动点在右侧 L 根收盘前不得出现在任何输出。
T3	多周期对齐	合成数据：修改尚未收盘的 4h K 线后半段，断言此前所有 LTF 决策不变。
T4	故意泄漏反向对照	对“区域激活”布尔序列做未来扰动检验，保证因果性测试非恒真。
T5	表级不变量	<code>activated_at > ob_bar_ts</code> 等恒等式。
T6	无全样本统计量	断言所有分位/标准化用扩展或滚动窗口。
T7	成交顺序保守性	同根 K 线内入场价与止损价同触时取最不利顺序；BTC 用 1m 数据校准该假设的代价。

C 补充实验完整表格

表 14 给出区块长度敏感性的完整区间（对应第 6.1 节）。三个补充实验的完整产物、预注册与结果文档见复现包。

表 14: 区块长度敏感性: 六种块长下的 90% 置信区间

方案	块数	μ_{OB}	Δ_D	Δ_{simple}	p_{val}^a
半月	132	$[-0.0705, +0.1126]$	$[-0.1998, -0.0625]$	$[-0.0971, +0.0182]$	0.921
1 月 (主)	66	$[-0.0761, +0.1102]$	$[-0.2079, -0.0542]$	$[-0.0941, +0.0240]$	0.938
2 月	34	$[-0.0620, +0.1024]$	$[-0.2125, -0.0558]$	$[-0.0885, +0.0219]$	0.956
3 月	23	$[-0.0669, +0.1023]$	$[-0.2208, -0.0427]$	$[-0.0793, +0.0227]$	0.947
6 月	12	$[-0.0595, +0.1039]$	$[-0.2277, -0.0370]$	$[-0.0773, +0.0210]$	0.966
平稳 (均长 1 月)	66	$[-0.0739, +0.1118]$	$[-0.2094, -0.0588]$	$[-0.0945, +0.0238]$	0.943

注: 点估计在所有方案下逐位不变 (本实验只更换不确定度估计器, 不触碰数据)。半月档对反转率差标 n/a (明细表不带时间戳, 无法派生半月块, 预注册规定标 n/a 而非静默回退)。1 月档由独立重写的重抽样器逐位复现主口径区间。

^a $p_{val} = p(\text{mean_R} \leq 0)$, validate 期 (2024–2025)。

D 产物清单与复现指引

复现包 (随文在线附件) 包含: 设计冻结文档、P2/P3 两份预注册、P1–P3 结果文档、三份补充实验预注册与结果文档、多周期结构引擎源码 (`src/nemesis/artemis/`)、31 项测试、补充实验脚本、以及逐行登记的试验台账。主口径产物在 `outputs/artemis/`, 实体口径附录产物在 `outputs/artemis_body/`, Binance 复制产物在 `outputs/artemis_binance/`, 补充实验统计量在 `outputs/artemis/paper/`; 冻结配置为 `configs/artemis.canonical.yaml`。本文全部数据图由 `make_figs.py` 生成。台账关键行: `artemis_p2_event_study` (`n_configs=2`, 含实体口径)、`artemis_p3_grid` (`n_configs=48`)、`artemis_paper_s1_blocklen` / `_s2_binance_repl` / `_s3_sesoi_power` (各 `n_configs=1`, 冻结复制或纯后处理, 非新搜索)。

E 实验时间线

阶段	内容
P0	设计冻结: 把 SMC 方法机械化, 确立因果性约束与对照组设计。
P1	结构验收: 31 项测试通过; 发现并修正成交 K 约定错误; 确立 “订单块存在恒成立”。
P2	事件研究 (真事前预注册): 门槛判定 <code>order_block_falsified</code> 。
P3	真实成本回测: 48 格全负, deflated Sharpe -3.41 (预注册事后补写, 已声明)。
P4	稳健性检验: 留一年 / 留一币 / 集中度 / 成本压力, 全部一致为负。
P5	最终报告: 三层判定与 “不建议实盘投入” 的建议。
S1–S3	补充实验 (各自先写预注册): 块长敏感性、Binance 冻结复制、等价检验与功效。

参考文献

- [1] Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance* 47, 1731–1764.
- [2] Lo, A.W., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance* 55, 1705–1765.
- [3] Urquhart, A., 2016. The inefficiency of Bitcoin. *Economics Letters* 148, 80–82.
- [4] Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance* 54, 1647–1691.
- [5] White, H., 2000. A reality check for data snooping. *Econometrica* 68, 1097–1126.
- [6] Harvey, C.R., Liu, Y., 2015. Backtesting. *Journal of Portfolio Management* 42, 13–28.
- [7] Harvey, C.R., Liu, Y., Zhu, H., 2016. ...and the cross-section of expected returns. *Review of Financial Studies* 29, 5–68.
- [8] Bailey, D.H., López de Prado, M., 2014. The deflated Sharpe ratio: correcting for selection bias, backtest overfitting and non-normality. *Journal of Portfolio Management* 40, 94–107.
- [9] McLean, R.D., Pontiff, J., 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71, 5–32.
- [10] Lakens, D., 2017. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science* 8, 355–362.
- [11] Lakens, D., Scheel, A.M., Isager, P.M., 2018. Equivalence testing for psychological research: a tutorial. *Advances in Methods and Practices in Psychological Science* 1, 259–269.
- [12] Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–1313.
- [13] Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance* 25, 383–417.
- [14] Barber, B.M., Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* 55, 773–806.