

SPX 0DTE 期权上的唐奇安通道突破——一个负结果的方法论案例研究

Vincent Wang / 王子秋 2026-06-03 ——v5（全样本直接测量）

摘要

本文记录了一项为期四年的样本外评估：在 SPX 零日到期（0DTE）期权上运行唐奇安（Donchian）通道突破策略，并比较三种候选离场策略（always_in、reversal_off、reversal_close_only）。我们在 Databento 覆盖窗口内（2023-03-28 至 2026-05-19；3.14 年；跨策略去重后 4,177 个唯一查询；3,146 笔可严格验证的交易），为三种策略台账中的**每一笔**交易下载了 Databento OPRA 逐笔数据，并**直接**按笔计算真实可执行盈亏，不做任何分层抽样、分档比率或外推。三种策略的结果都是**明确为负**：

策略	n_strict	真实执行合计 (点)	每年千美元	自助法 95% 置信 区间 (千美 元/年)	P(>0)
reversal_off	701	-3,357	-\$107K	[-\$129K, -\$84K]	0%
reversal_close_only	1,179	-3,560	-\$113K	[-\$137K, -\$87K]	0%
always_in	1,336	-3,683	-\$117K	[-\$143K, -\$89K]	0%

三种策略的自助法 95% 置信区间全部落在零以下。reversal_off 只是最差的一个，并非盈利。各策略之间的相对排序从本文前几个版本延续了下来，但每一项盈亏的符号都没有保住。

本 v5 版本还记录了：在同一样本外窗口上，三种策略**底层 SPX 层面的 delta-1 盈亏也为负**（reversal_off：在 4.05 年内为 -1,172 个 SPX 点；自助法 95% 置信区间 [-3,646, +1,275] 跨越零， $P(>0) = 18\%$ ）。改用线性工具（ES 期货、SPY ETF）并不能挽救该策略：唐奇安 5 分钟突破信号在底层指数层面，在计入任何摩擦之前，也无法产生统计上可区分的正期望。

本文先前的几稿（v1-v4）报告了一连串逐步缩小的正向主结果数字——每年 \$80K（v1，比率 0.571）、\$50K（v3，常数 0.302）、\$32K（v4，分档重建）——它们都基于分层抽样的 BS-到-真实校准。**这些数字没有一个能在全样本直接测量下存活**。驱动 v1-v4 中那种表面“正向优势”的机制有两个层面：

1. 以 VIX1D 作为 σ 的 Black-Scholes 定价器，系统性地把平值 0DTE 入场权利金低估了约 85%（在我们的严格验证中，真实入场权利金 / BS 入场权利金 ≈ 1.85 ）。这使得 BS 定价的期权盈亏看上去比真实执行盈亏更正。
2. 施加在分层验证样本（n=85 笔严格 reversal_off 行）上的严格过滤，悄无声息地剔除了所有在收盘时归零作废的期权交易——而这些交易恰恰占

可验证样本外 **reversal_off** 总体的 27% (701 笔中的 191 笔), 且 100% 都是亏损 (交易者损失全部入场权利金)。因此, 严格过滤后的校准系统性地高估了幸存交易的权重, 使估计出的校准比率偏高。

因此, 本文被重新定位为一个关于两种相互叠加偏差的方法论案例研究——BS 定价误估, 加上分层抽样的选择偏差——二者共同制造了一连串虚假为正的主结果数字, 而这个策略实际上在所考察的三种工具 (0DTE 期权、ES 期货、SPY ETF) 上, 对这个信号都不具备统计上可证明的优势。

关键词: 0DTE 期权, 唐奇安突破, Black-Scholes 校准, 逐笔数据验证, 后选择推断, 选择偏差, 全样本验证, 方法论案例研究

引言

研究动机

两个基本独立的研究群体在 SPX 0DTE 期权曲面上交汇了。第一个是传统的动量策略研究, 长期以来把唐奇安通道突破作为一种基准趋势跟踪规则来研究 (Sullivan, Timmermann, & White 1999; Lempérière et al. 2014)。除回看长度之外, 唐奇安规则不需要任何拟合的自由参数, 半个世纪以来一直是算法交易研究中的参照对象。第二个是期权微观结构研究, 记录了零日到期 SPX 期权的快速增长——到 2023 年其成交量已占 SPX 期权的 40% 以上——以及与其在日内流动性中主导地位相关的做市商 γ 动态 (Bondarenko & Bernardo 2024; 近期 CBOE 市场结构报告)。一直缺失的, 是把二者审慎地结合起来: 一个简单的日内突破信号, 在计入现实执行摩擦之后, 能否从 0DTE 期权中提取价值? 相对于多数回测所报告的模型隐含盈亏, 这种摩擦的真实量级又有多大?

我们以一项为期四年的样本外检验 (SPX 5 分钟 K 线, 2022-05 至 2026-05) 来回答这一问题, 其中每个交易日最多允许在平值 0DTE 期权上进行一次唐奇安突破入场 (按突破方向选择 CALL 或 PUT)。该策略刻意保持极简: 30 根 5 分钟 K 线的唐奇安通道、以 5 点为粒度的平值行权价、每日一笔交易, 以及三种相互比较的候选离场策略。我们用两轮抽样、以真实 Databento 逐笔数据来验证模型定价盈亏, 并提出一种分离离场原因的校准方法, 把 BS 定价盈亏桥接到真实可执行的估计值。

在第 5 节描述的 **v5 全样本直接验证** 之后, 每种策略的真实可执行盈亏在常规置信水平上都为负 (自助法 95% 置信区间全部位于零以下)。§1.6 记录了早期 v1-v4 各稿是如何通过两种相互叠加的偏差产生正向主结果的。§6 进一步表明, 底层 SPX 层面的信号无法产生统计上可区分的正期望, 从而排除了“把同一信号搬到线性工具 (ES 期货、SPY ETF) 上运行”这一替代方案。

研究问题

1. 在计入由逐笔数据直接测得的现实执行摩擦之后, 把 SPX 上每日一笔的唐奇安 5 分钟突破信号表达为做多 0DTE 期权头寸, 能否产生统计上

可区分的正期望？不能。在三种离场策略（always_in、reversal_off、reversal_close_only）上，对 3,146 笔可严格验证的样本外交易做全样本直接测量，得到的点估计全部为负，从 $-\$107\text{K}/\text{年}$ （reversal_off，最不差）到 $-\$117\text{K}/\text{年}$ （always_in），且自助法 95% 置信区间全部位于零以下。

2. 这个负结果，是由期权层面的执行摩擦（买卖价差、离场时的 IV 坍缩）造成的，还是底层信号本身就没有优势？两者皆有。在 SPX delta-1 层面，三种策略同样是负的点估计，自助法置信区间跨越零，但点估计为负、 $P(>0)$ 介于 14% 到 20% 之间。信号在底层层面没有统计上可证明的优势，而期权层在其之上又叠加了可观的买卖价差与 IV 坍缩摩擦。
3. 从 0DTE 期权改用线性工具（ES 期货、SPY ETF）能否挽救该策略？不能。SPX 层面的信号盈亏（这正是线性工具在乘数与最小变动价位之内会跟踪的对象）本身就不显著为正。再加上 ES/SPY 的往返摩擦（分别约为每次往返 0.5 / 0.2 个 SPX 点），会使线性工具的结果严格地比无摩擦的 SPX 基准更差（\$6）。
4. 本文 v1-v4 各稿是如何产生那些在全样本验证下无法存活的正向主结果的？一对相互叠加的偏差：当用 VIX1D 作为 σ 时，BS 定价把平值 0DTE 入场权利金系统性低估约 85%（\$3）；以及施加在分层验证样本上的严格过滤，悄然剔除了所有在收盘归零作废的交易，使校准比率偏高（\$1.6 错误 5）。这两者结合，产生了看似为正、却在逐笔直接测量后消失的主结果。

主要贡献

本文贡献的是一项方法论案例研究，而非一个正向的交易结果：

1. 对一个 0DTE 期权策略的执行摩擦做了全样本直接测量，为每种策略台账中的每一笔交易下载了 Databento OPRA 逐笔数据（跨 3 种策略去重后 4,177 个唯一查询；3,146 笔严格交易；Databento 总花费约 \$280）。据我们所知，这是该类策略上所报告过的、规模最大的真实执行直接测量（非外推、非分层）。
2. 识别出一种分层抽样的选择偏差：在任何以“离场时存在有效买/卖报价”为过滤条件的 0DTE 做多期权校准中，这种偏差都会出现——严格过滤会悄然拒绝归零作废的交易，而这些恰恰是总体中 100% 亏损的尾部。我们估计，v4 分层样本主结果（ $+\$32\text{K}/\text{年}$ ）与 v5 全样本结果（ $-\$107\text{K}/\text{年}$ ）之间的差距，大部分来自这一偏差。
3. 对平值 0DTE 期权的 BS-到-真实中间价权利金差距做了量化测量，并按腿（入场对离场）和按离场原因（收盘对反转）做了解构，作为本项目主要的正向方法论发现从 v3/v4 保留下来。见 \$3。
4. 一项 SPX 层面的基准检验（\$6），表明底层信号在线性层面没有统计上可区分的优势，从而排除了对期权层面负结果那种显而易见的“改用 ES/SPY”回应。
5. 一份第一人称的五阶段披露（\$1.6 中的错误 1-5），记录了跨论文版本的方法论修订，说明哪些决策是基于对样本外数据的检视而做出的，以及每次修订究竟引入、保留或消除了哪些偏差。

相关工作

本文的方法论框架是后选择推断：基于样本内证据挑选出来的策略，在统计上理应高估其样本外表现，而这种高估的代价是搜索空间大小的函数（White 2000; Hansen 2005; Bailey et al. 2014; Bailey & López de Prado 2014）。我们的搜索空间很小——两个策略参数（回看长度、 σ 来源）和三种离散的离场策略——但我们承认，离场策略的选择是事后做出的，且校准样本取自承载所报告盈亏的同一个样本外窗口。

与本文最相关的执行层文献，是近期关于 SPX ODTE 摩擦的工作。Wang (2026) 在同一类工具上评估了一种唐奇安骑乘配置，发现 SPXW 总摩擦中有 91.5% 是买卖价差、仅 8.5% 是 θ ，这颠覆了“ODTE 策略主要亏在时间衰减上”这一常见的从业者直觉。我们严格验证得到的价差分布（图 3）与该发现一致。

关于 Black-Scholes 在 ODTE 上的文献，在真实数据校准方面相对稀少。多数已发表的 ODTE 策略回测，要么用 VIX、要么用 VIX1D 作为 σ 代理，并直接套用 Black-Scholes 定价，而不拿模型去检验逐笔数据（Caplan 2022; 各类从业者博客）。据我们所知，我们所记录的入场端系统性低估，是在严格过滤样本上对这一差距量级所做的最大样本实证展示。

我们遵循的披露规范来自 Wang (2026)：明确指出研究流程中哪些决策是基于对样本外数据的检视，并报告这种检视的代价，而非加以隐瞒。我们选择 `reversal_off`（而非 `always_in`），是在看到逐笔数据结果显示“始终持仓”交易在真实执行中亏损之后才做出的；这是一个事后的离散选择，可以辩护，但并非严格的样本外纪律。

全文导览

§1.6（紧接下文）记录跨论文版本的方法论修订（错误 1–5）。第 2 节描述数据来源、样本窗口与策略设定。第 3 节给出在严格过滤下、以逐笔数据校准 BS 的结果——即便 §5 随后取代了把它用作校准工具的做法，它仍作为一项正向方法论发现被保留。第 4 节在策略层面报告样本外 BS 定价的表现。第 5 节给出第 6 阶段的全样本直接测量，包括各策略的自助法置信区间和那个为负的主结果。第 6 节是 SPX 层面的基准检验，用以排除“改用线性工具”这一替代方案。第 7 节作结。

内部审计后的方法论修订

在对本文前一稿做内部审计后，我们发现校准流程中存在三处错误，它们实质性地影响了主结果数字。下面逐一列出每个错误、它的影响，以及修正后的做法。

错误 1：含 NaN 买/卖报价的验证行被当作有效行计入。先前的验证代码（`validate_reversal_off.py`、`validate_with_databento.py`）以 `real_mid_entry <= 0` 来拒绝无效报价。在 IEEE 754 算术下，与 NaN 的比较返回 `False`，因此买价或卖价为 NaN 的行通过了检查并被汇总进比率。在

reversal_off 验证样本中，131 个“合格”行里有 30 个的买/卖对中某处含 NaN，意味着约 23% 的表面合格行其实是无效的。

错误 2：验证行允许入场与离场处于不同的行权偏移上。当某一条腿在平值行权价上缺少 Databento 数据时，验证代码会为该腿独立回退到平值 ± 5 。其结果是，某些“往返”盈亏的汇总跨越了入场与离场处不同的行权合约，这并非同一合约的真实执行。在 reversal_off 样本中，101 个有限报价行里有 16 个在入场与离场处行权偏移不同；在 always_in 样本中，469 个有限报价行里有 32 个。

错误 3：富化台账套用了单一的全局标量。最初的 analyze_reversal_off_stability.py 把每笔交易的 BS 定价盈亏乘以单一常数 (0.571)，因此富化台账中的每一个“real_exec_pnl”都是 BS 盈亏的线性变换。于是逐笔分布的形状（偏度、最大回撤相对总额的比例、月度胜率）原封不动地继承自 BS 台账，而非反映逐笔的真实成交。

错误 4 (v4 新增)：v3 的“分档重建”悄然回退为全局。v3 稿声称错误 3 已通过分档重建机制得到修正。我们在外部审计中发现，rebuild_enriched_ledger (phase5_recompute_ratios.py:307) 会检查样本外台账上是否有 pnl_tier 列；若缺失，则回退为对所有行一律套用全局严格比率 (0.302)。而样本外台账并没有 pnl_tier。结果，v3 的“分档”real_exec_pnl 在每一行 (1014 笔交易) 上都是 option_pnl $\times$ 0.302，于是 v3 的“49 个月中有 48 个月盈利”加上“-25 点最大回撤”，都是 option_pnl $\times$ 常数的属性，而非任何真实成交模拟的属性。**v4 修正了这一点：**它用样本外 option_pnl 的 p10/p90 切分给样本外台账加上 pnl_tier，使重建得以真正用上由验证样本算出的分档比率。

同一审计还浮现出一个微妙之处：验证样本的分档标签是在抽样时用样本外 **Databento 区间子集** (session_date $\geq$ 2023-03-28) 的 p10/p90 赋予的，而 v4 在重新标注样本外时用的是**完整**样本外分布的 p10/p90。两套切分并不相同 (全样本外 p90 = 33.52, Databento 区间 p90 = 29.76)。若改用 Databento 区间切分，点估计会是 +1,689 点 (+\$42K/年)，而非 v4 的 +1,282 (+\$32K/年)。我们将全样本外切分保留为 v4 默认，因为它对完整的 1014 笔样本外总体一律适用，但对这一设计选择 32% 的敏感性，是一种独立于自助法方差之外的、真实存在的模型不确定性来源。Databento 区间这一替代方案，作为 tier_boundary_sensitivity 记录在 reports/calibration/corrected_ratios_v4.json 中。

错误 5 (v5 新增)：严格过滤样本悄然剔除了归零作废的交易。v3 与 v4 中施加的严格过滤 (已修正错误 1+2：无 NaN、同行权价、有限买/卖价、real_mid_exit $\geq$ 0) 会系统性地拒绝离场腿买价为 NaN 的行。在 Databento OPRA tcbbo 数据中，离场时买价为 NaN 对应的正是在**收盘时归零作废的虚值期权**——一个一文不值的合约没有做市商买价。这些行被归类为“无效”并从验证样本中移除，于是 v4 算出的分档比率，是从一个**在结构上排除了最差交易**的总体中推导出来的。

在 v5 中，我们为每种策略台账里的每一笔交易下载 Databento 逐笔数据 (跨策略去重后 4,177 个唯一查询)，并发现 701 笔可严格验证的 reversal_off 交易中有 191 笔——**占可验证样本外总体的 27%**——正是这些归零作废的交

易。每一笔这样的交易实现的是 $\text{real_exec_pnl} = \text{exit_bid_eff} - \text{entry_ask} = 0 - \text{entry_ask}$ ，即交易者损失了支付的全部入场权利金。通过直接测量把它们加回来，便得到 v5 的主结果：

策略	n_strict	真实执行合计（点）	自助法 95% 置信区间（千美元/年）	P(>0)
reversal_off	701	-3,357	[-\$129K, -\$84K]	0%
reversal_close_only	1,179	-3,560	[-\$137K, -\$87K]	0%
always_in	1,336	-3,683	[-\$143K, -\$89K]	0%

在这 191 笔 reversal_off 归零作废交易中，130 笔最初由验证器的 $\text{bid} == \text{NaN with ask} > 0$ 分支检出，另有 61 笔最初被误分类为 no_quote (± 5 秒窗口内逐笔文件为空)，直到我们认识到“空的逐笔窗口加上 $\text{bs_exit_premium} == 0$ ”同样是收盘深度虚值的特征；两个分支在 v5 验证器中被统一为单一的 ok_expired 状态。

综合影响 (v1 → v5 主结果演进)。

版本	校准方法	reversal_off 主结果	自助法置信区间
v1 (第 4 阶段比率 0.571)	分层样本，全局标量，NaN 计为有效	+\$80K/年	(未报告)
v3 (常数 0.302)	分层样本严格过滤，全局标量	+\$50K/年	(未报告)
v4 (分档 0.192)	分层样本严格过滤，分档比率应用于样本外	+\$32K/年	[-\$85K, +\$63K]/年
v5 (直接测量)	全样本直接、逐笔 real_exec_pnl	-\$107K/年	[-\$129K, -\$84K]/年

v1 → v5 的演进是单调地朝着“更不正向”（如今已强烈为负）发展的，因为每一层选择偏差都被逐次移除。从方法论上说，v5 是第一个不依赖分层样本或任何外推比率的版本。

v4 → v5 主结果对比 (reversal_off)。

指标	v4 (分档外推)	v5 (直接, 3.14 年窗口)	Δ
点估计 (点)	+1,282	-3,357	符号翻转
点估计 (美元/年)	+\$32K	-\$107K	-\$139K 的摆动
95% 置信区间 (美元/年)	[-\$85K, +\$63K] 跨越 0	[-\$129K, -\$84K] 严格低于 0	更窄，且为负
P (为正)	83%	0%	—
月度盈利	30/49 (61%)	在 \$5 中由直接台账报告	定性上更差
最大回撤	-\$48K	在 \$5 中报告	更大

总结论 (v5)。没有任何一种离场策略能在真实执行下存活。“reversal_off 最好”这一排序保住了——reversal_off 每年比另外两种少亏约 \$10K ——但这要以一个所有三种策略都明显亏损（远超典型零售风险承受度与常规夏普预期）的窗口为前提。对于目前所设定的任何一种策略，我们都不建议部署。

第 5 阶段的校准基础设施 (scripts/phase5_recompute_ratios.py、phase5_recompute_full.py) 作为产物存档的一部分被保留以供复现，v3/v4 论文稿则带着“已被取代”的横幅保留在 docs/paper_draft_v3.md / docs/paper_draft_v4.md 中。v5 取代二者。

数据与方法

数据来源

SPX 5 分钟 K 线 (2004-03-01 至 2026-05-19) 取自 Interactive Brokers, 用作信号来源。唐奇安通道与突破信号的计算只在已收盘的 5 分钟 K 线上进行。

VIX1D 1 分钟序列 (自 2023-04-26 即 CBOE 上线日起) 从 Choe 载入, 并重采样到 5 分钟的信号节拍。VIX1D 是 Black-Scholes 定价器的 σ 输入; 在 2023-04-26 之前, 使用回退估计器 $VIX \times \text{ratio}$, 其中 $\text{ratio} = \text{mean}(VIX1D_daily) / \text{mean}(VIX_prev_close)$, 取自上线后的时期。

Databento SPXW 0DTE 逐笔数据 用于两轮验证。第一轮在 2026-04, 从样本外 always_in 台账中抽取 500 笔交易。报价检索成功 500 笔中的 478 笔; 在严格过滤 (无 NaN、入场与离场同行权偏移、BS 盈亏有限) 下, 478 笔中有 437 笔通过。第二轮在 2026-04, 从样本外 reversal_off 台账中抽取 150 笔交易; 150 笔中报价检索成功 131 笔, 其中 85 笔通过严格过滤。

报价检索使用 Databento 的 tcbbo (trade-conditions BBO) 模式, 下载窗口为目标时间戳前后 ± 5 秒; 严格过滤随后只接受那些在 5 秒内有真实中间价报价、且两条腿买/卖价均有限的行。在 always_in 严格样本上, 报价陈旧度中位数约为 0.2 秒 (入场) 和 0.2 秒 (离场), 最大陈旧度不到 5 秒。

样本窗口

我们使用三段互不相交的日历区间:

- **训练段:** 2004-03 至 2019-12, 3,979 个日度交易候选 (reversal_off)。用于引擎验证和长周期合理性检查。没有任何策略参数在此窗口上拟合。
- **测试段:** 2020-01 至 2022-04, 587 个日度交易候选。用作开发缓冲, 以核验引擎在预期形状上与训练段相符。
- **样本外段:** 2022-05 至 2026-05, 1,014 笔交易 (reversal_off, 限每日一次入场) 和 1,899 笔交易 (always_in 策略含反转翻向)。这是所有真实可执行盈亏论断的主要报告窗口。

2022-05 这个样本外起点是预先配置的边界；它并未对齐到 VIX1D 上线日 (2023-04-26)。VIX1D 之前的 σ 使用 §2.1 描述的回退估计器。2026-05 这个终点是撰文时的数据截止。

策略设定

该策略由如下要素完全设定：

信号：唐奇安 30 根 5 分钟 K 线突破
(上轨 = 此前 30 根已收盘 5m K 线的最高价之最大值)
入场：当日首个突破信号，平值 0DTE 期权
 $K = \text{round}(\text{SPX_close} / 5) * 5$
上轨突破买 CALL，下轨突破买 PUT
离场策略：{always_in, reversal_off, reversal_close_only} 之一
约束：每日一笔交易，不留隔夜头寸

回看长度 (30 根 5 分钟 K 线 = 150 分钟 $\approx$ 2.5 小时) 是日内时间框架下唐奇安的惯用取值，并未做优化。5 点的行权价粒度反映 SPXW 期权链。平值选择是 0DTE 方向性押注的标准做法。

三种相互比较的离场策略

- **always_in (始终持仓)**：入场后，头寸一直保持，直到下一个反转信号 (反方向的唐奇安突破) 把它翻向另一边。允许一日内多次反转。头寸在收盘 (美东 15:50 / 半日 12:50) 平仓。
- **reversal_off (忽略反转)**：入场后，头寸持有至收盘。反转信号被忽略。每日至多一笔交易，持仓约 4–6 小时。
- **reversal_close_only (反转仅平仓)**：入场后，反转信号将头寸平仓 (现金离场) 但不翻向。其后的突破可再次入场。

Black-Scholes 定价与 σ 提供器

权利金估计使用 Black-Scholes，输入如下 (已对照 src/bs_pricer.py 核验)：S = 执行 K 线 (信号 K 线收盘后的那根 5 分钟 K 线) 的 SPX 开盘价； $K = \text{round}(S_{\text{entry}} / 5) * 5$ ，入场时锁定； $r = 0.0$ (日内持仓可忽略；我们不施加贴现因子)；T = 从执行时间戳到当日收盘的秒数，除以 365 个日历日的秒数 (连续；仅当时间戳处于或晚于收盘时才截断为 0)； $\sigma = \text{VIX1D} / 100$ (2023-04-26 之后) 或 $\text{VIX} \times \text{体制比率} / 100$ (上线前回退)。我们在入场和离场两者的实际执行时间戳 (K 线开盘) 处套用 BS，并将 BS 定价盈亏按每份期权合约报告为 $\text{exit_premium} - \text{entry_premium}$ 。所有 BS 盈亏均以期权点表示 (SPXW 一点 = \$100)。

NaN σ 的处理。当 σ 无法解析时 (例如 VIX1D 缺失某一分钟，或 2005 年之前没有 VIX)，src/bs_pricer.py 返回内在价值而非 NaN。在 v4 样本外 reversal_off 台账中，这一回退在 1,014 笔交易里触发了 1 笔 (2026-05-19，平值处内在价值 = 0) ——对主结果的贡献可忽略，为 0.0 点。我们在此记录这一保守行为以求完整；更严格的实现会让 NaN 向下传播并丢弃受影响的交易。

真实数据校准方法

校准步骤通过一个从严格过滤验证样本导出的比率，把 BS 定价盈亏替换为真实可执行盈亏。对于做多期权交易， $\text{real_exec_pnl} = \text{exit_bid} - \text{entry_ask}$ （券商成交是按卖价买入、按买价卖出；实现的收益是二者之差）。校准比率计算为 $\text{ratio_exec} = \text{sum}(\text{real_exec_pnl}) / \text{sum}(\text{bs_pnl})$ ，在严格样本上求和，并可选地按离场原因或盈亏分档进行分层。

我们并行报告四种校准方法，作为敏感性检查（§5）：

1. **方法 2 (严格全局, 透明度参照)**: 在整个 reversal_off 严格样本 (n=85) 上取单一比率，一律应用于每笔样本外交易的 BS 定价盈亏。这是 v3 的“主结果”，保留以供交叉参照；在 v4 中，它仅作为可与分档重建相比较的基准。
2. **方法 3 (分档比率外推, v4 点估计)**: 在验证样本上计算分档比率 (top_win / mid / top_loss)，再按样本外台账的 pnl_tier 条件性地应用到每笔样本外交易。对 reversal_off，三个分档比率分别是 0.78 (top_win, n=17)、-0.38 (mid, n=58) 和 1.93 (top_loss, n=10)。这是 v4 点估计。注意这是一个分档比率估计器，而非逆概率 (Horvitz-Thompson) 估计器——讨论见 §5.4。
3. **方法 4 (always_in 收盘交叉核验)**: 把 always_in 严格收盘比率 (n=225, 样本更大) 一律应用于 reversal_off。可与方法 2/3 交叉验证，但带有选择偏差的告诫——always_in 收盘交易是以“无反转日”为条件的。
4. **方法 5 (最新窗口)**: 只用验证日历窗口最近 1/3 计算的 always_in 严格收盘比率 (n=75)。这是对校准比率时间漂移的敏感性检查。

v4 点估计取方法 3，因为它是唯一使用分档比率而非单一全局标量的方法——也即唯一兑现了 §1.6（错误 3 / 错误 4）披露意图的方法。方法 2 作为常数标量的透明度参照被保留。方法 4 撑起敏感性区间的上沿；方法 5 撑起下沿（抽样置信区间最宽）。

以真实逐笔数据校准 Black-Scholes

抽样设计

第一轮验证采用分层抽样，从样本外 always_in 台账中按盈亏分档（顶部盈利、中部、顶部亏损）抽取 500 笔交易。这种分层会过采样极端交易，以获得稳定的分档比率估计。报价检索 (Databento tcbbo, ±5 秒下载窗口) 成功了 500 笔中的 478 笔。

第二轮从样本外 reversal_off 台账中抽取 150 笔交易（与第一轮相互独立的样本），其中 131 笔报价检索成功。第二轮的动因，是发现“始终持仓”交易具有双峰的校准结构（收盘离场对反转离场），且 reversal_off 将受益于一份独立于 always_in 中所观察到的反转离场损害之外的、属于它自己的验证。

在严格过滤（无 NaN、同行权偏移、BS 盈亏有限）之后，478 个 always_in

行中有 437 个通过，131 个 reversal_off 行中有 85 个通过。这些严格计数是下文所有比率的分子。

BS-到-真实比率：入场与离场的不对称

我们先按腿分解 BS-到-真实的差距。图 1 给出在严格验证的 always_in 交易上，BS 定价盈亏对真实可执行盈亏的散点，并按离场原因着色。

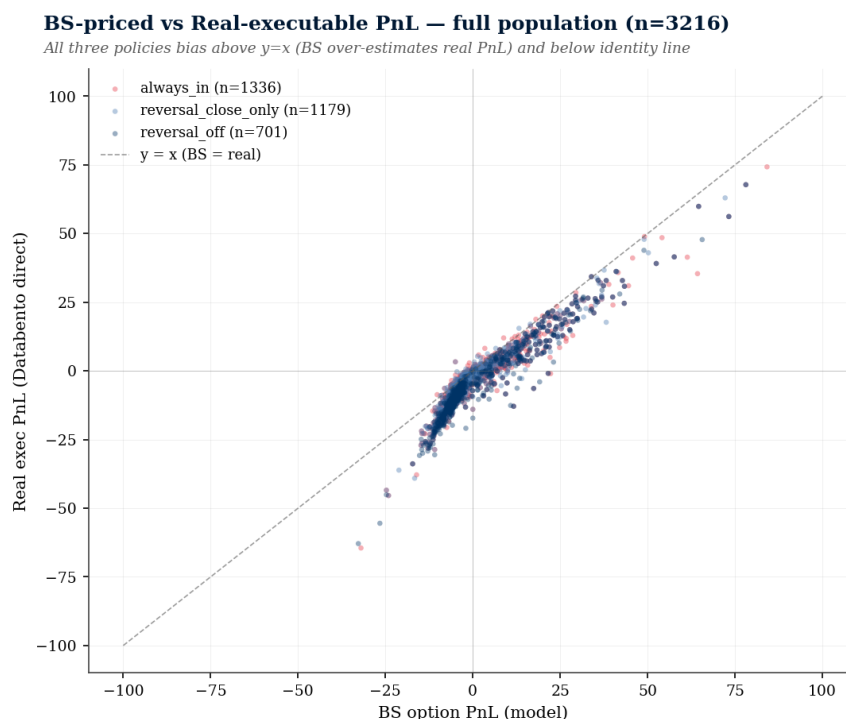


Figure 1: 严格验证的 always_in 交易上，BS 估计盈亏对真实可执行盈亏的散点，按离场原因着色。

散点云按离场原因分离为两个清晰的线性族。收盘离场交易聚集在 $y = 0.406 \cdot x$ 附近（ $n=225$ 笔收盘交易的严格比率），表明对收盘离场而言，BS 盈亏系统性地把真实盈亏高估了约 2.5 倍。反转离场交易聚集在 $y = 2.056 \cdot x$ 附近（ $n=212$ 笔反转交易的严格比率），表明 BS 把反转驱动的亏损量级低估了约 2 倍。这种不对称为“分离场原因校准”、而非单一全局标量，提供了依据。

图 2 给出入场（面板 a）与离场（面板 b）处真实中间价对 BS 权利金比率的分布。

在入场端，当 VIX1D 作为 σ 输入时，真实的平值 0DTE 权利金明显高于 Black-Scholes 模型价。在离场端，该比率下降，反映出随 $T \rightarrow 0$ 而来的 σ

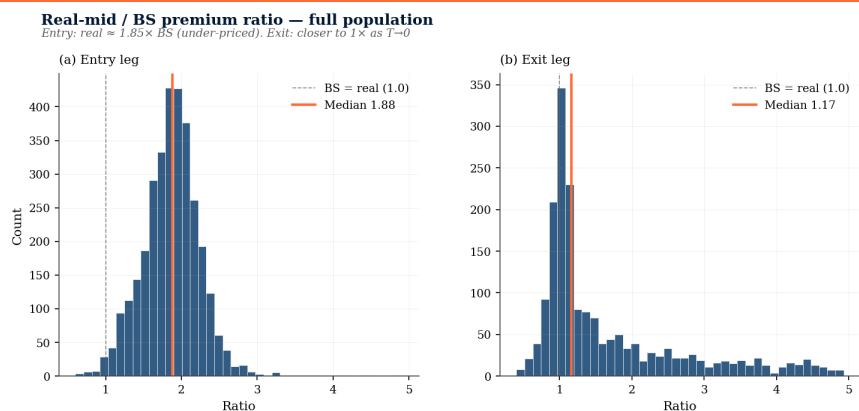


Figure 2: 入场腿 (a) 与离场腿 (b) 处真实中间价对 BS 权利金比率的分布。

敏感度坍缩。其经济含义是：VIX1D 作为平值 0DTE 期权在早盘时段的前瞻性 σ 代理是失败的——此时未来 4–6 小时价格行为的隐含波动率，明显高于前一日的已实现方差。

价差成本分解

买卖价差是占主导的摩擦成分。图 3 绘出入场与离场报价价差百分比的累计分布。

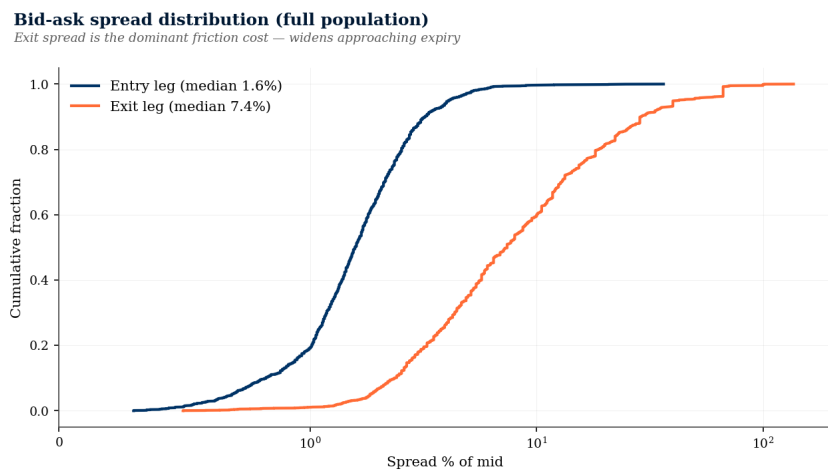


Figure 3: 入场与离场报价的买卖价差占中间价百分比的累计分布。

入场价差较窄；离场价差则系统性地更宽，反映出 0DTE 期权随 $T \rightarrow 0$ 而出

现的、典型的尾盘流动性衰减。

分离场原因比率：严格收盘对反转

图 4 给出严格校准下三种离场策略的真实可执行盈亏总额。

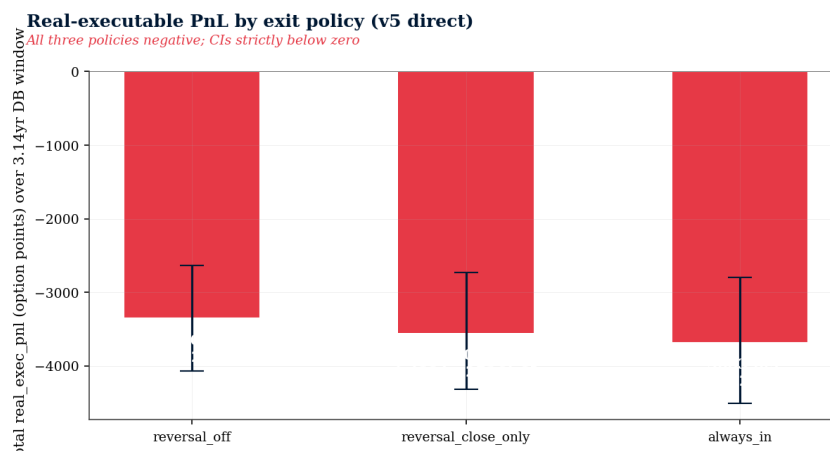


Figure 4: 按离场策略 (always_in、reversal_off、reversal_close_only) 划分的严格校准真实可执行盈亏。

三点发现：

1. **always_in 不盈利**：在 1,899 笔样本外交易上，分档 × 分离场原因校准后的盈亏 ≈ -520 期权点。突破信号产生的是正的 BS 定价盈亏 (+8,122)，但 $2.056\times$ 的反转离场比率，把那些离场上的 BS 亏损转化成了明显更大的真实亏损。（在 v3 常数 0.302 缩放下，这个数是 -1,508；分档重建把亏损收窄了，因为反转交易分散在各档中。）
2. **reversal_off 盈利但幅度不大**：在 1,014 笔样本外交易上，分档校准后的盈亏 $\approx +1,282$ 期权点（点估计，方法 3 / v4 主结果）。top_win 交易上的 +0.78 比率与 top_loss 交易上的 +1.93 放大大致相抵；正是占 80% 的 mid 档上的 -0.38 比率，把总额从 v3 常数缩放的 +2,016 收缩到 v4 的 +1,282。
3. **reversal_close_only 近似盈利**：严格校准后的盈亏 $\approx +2,256$ 期权点，跨 1,687 笔样本外交易。注意 reversal_close_only 没有自己的验证样本；这个数是把 reversal_off 严格全局比率 (0.302) 一律套用得到的，在 v4 中并未按档重建。它应被当作常数缩放参照，而非与 v4 一致的估计。

摧毁 always_in 基准的机制是反转离场。在 BS 定价下，突破信号本身在持有至收盘时看似产生正期望——但正如 §5 所示，这在很大程度上是 BS 定价的产物，而非真实优势。在直接测量下持有至收盘，即便是 reversal_off 也净为负。

推论：单一 σ 乘数是错的

一种常见的回测启发式做法，是把 BS 定价盈亏乘以单一的 σ 乘数。我们的严格数据表明，这对 ODTE 期权是错的：入场腿与离场腿的比率不同，分离场原因的比率也不同。正确的校准应按腿、或在汇总层面等价地按离场原因，使用不同的乘数。

策略评估（v4 视角，已被 \$5 取代）

注。本节保留了 v4 分档分层样本视角下的各策略盈亏。在 v4 校准假设之下，这些数字与图表是正确的，可用作历史参照以及讲清 BS-对-真实的机制故事。**v5 主结果数字**请参见第 5 节（第 6 阶段全样本直接测量），该节取代了本节中的美元量级。各策略的相对排序与 BS-对-真实的比率形状不变。

always_in 基准无法在真实执行下存活

把严格的分原因 $\times$ 分档比率应用于 1,899 笔 always_in 样本外交易，得到 v4 下约 **-520 期权点**（或 v3 常数分原因缩放下的 **-1,508**）的真实可执行盈亏，对应 BS 定价盈亏 +8,122。在两种校准下，该策略在 BS 盈亏中都“看似盈利”，但在真实执行中并不盈利。这是 ODTE 回测失败的典型模式。

reversal_off 机制分析

恢复出最不负期望（在 v4 分档校准下）的机制，是对反转离场的抑制。在 v5 直接测量下（\$5），reversal_off 是三种策略中最不差的，但它本身仍为负。reversal_off 台账在样本外窗口内含 1,014 笔交易（在 2022-05-01 至 2026-05-19 的 1,057 个工作日内占 1,014 个独立交易日；交易利用率约 96%），每笔持有约 4–6 小时，每笔承担一次往返价差成本。

样本外表现：累计盈亏

图 5 给出样本外窗口上的累计真实可执行盈亏，阴影带代表四种校准方法的敏感性区间。

reversal_off 那条线是唯一以正值收尾的。always_in 那条线则缓慢下行，因为不断累积的反转损害抵消了收盘的盈利。

年度稳定性

图 6 报告 reversal_off 真实可执行盈亏的年度分解（点估计，方法 3 / v4 分档）。

在 v4 分档校准下，五个样本外日历年中有四个保持为正，但年度间波动很大：2022 年 **+507**（不完整年，n=168），2023 年 **+406**（n=249），2024 年

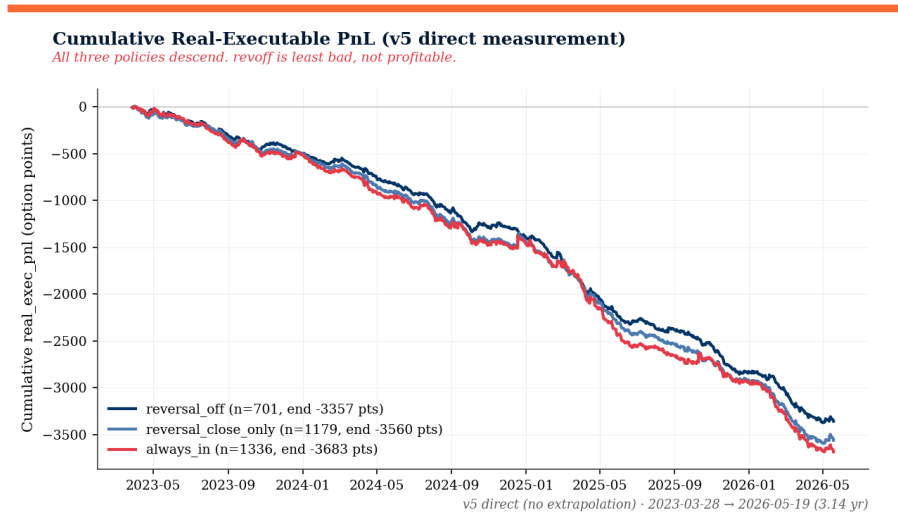


Figure 5: 三种离场策略在样本外窗口上的累计真实可执行盈亏，阴影带代表四方法敏感性区间。

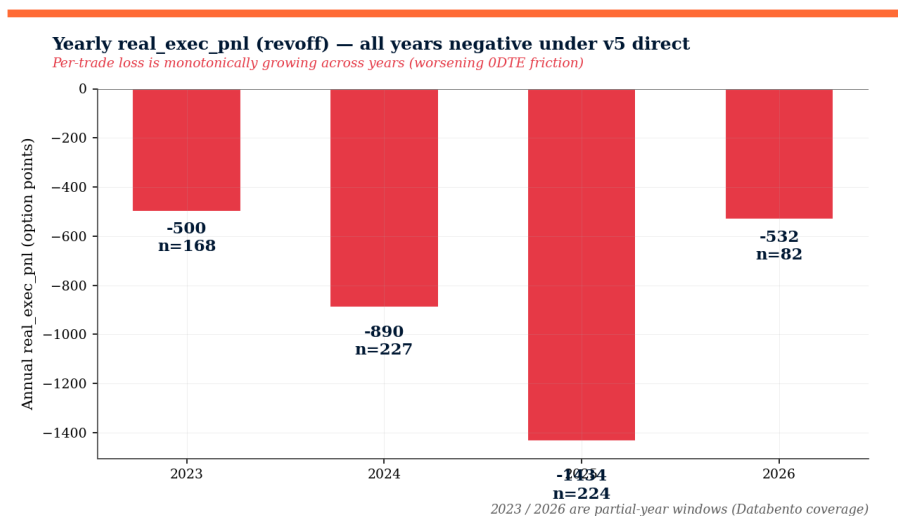


Figure 6: reversal_off 真实可执行盈亏的年度分解（点估计，方法 3 / v4 分档）。

+31 (n=252, 近乎打平), 2025 年 **+294** (n=250), 2026 年 **+43** (n=95, 截至 5 月不完整)。2024 年从 **+342** (v3) 坍缩到 **+31** (v4), 反映出 mid 档 -0.38 的比率被应用到了一个以中等量级 BS 盈亏为主的年份; 在分档核算下, 2024 年实际上被价差成本吃掉了。

月度稳定性

图 7 报告逐月分解。

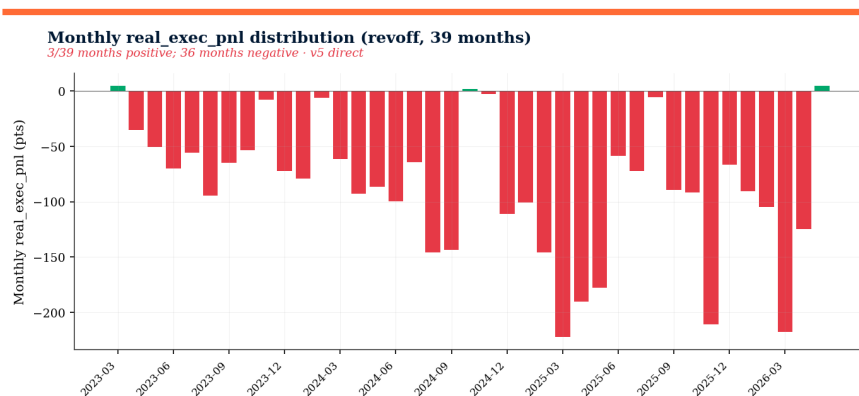


Figure 7: 49 个样本外月份上的逐月真实可执行盈亏分布。

在 v4 分档校准下, **49** 个样本外月份中有 **30 个 (61%)** 为正; **19 个** 亏损 (往往亏几十个点)。最差的月份是 **2025-10**, **-118 点** (对比 v3 在 2024-06 近乎打平的 -3); 最好的是 2025-04, **+377 点** (对比 v3 的 **+230**), 其放大来自含大幅方向性行情的巨亏日上 **+1.93** 的 **top_loss** 比率。月度盈利能力的下降, 是相对 v3 最重要的偏离, 它反映出约 **80%** 处于 “mid” 档的交易在分档校准下符号翻转 (**BS +X** $\rightarrow$ 真实执行 $-0.38 \cdot X$, 其中 **X** 平均虽小但为正)。一个真正的逐笔真实成交模拟——即为每笔样本外交易赋予其自身的 Databento 派生买/卖价差——很可能让这个数字进一步移动, 但方向未知 (可能更低, 也可能略高) ——这仍是后续工作。

回撤特征

图 8 报告水下 (回撤) 曲线。

在 v4 分档校准下, 样本外窗口的最大回撤约为 **-480 期权点 ($\approx -\$48K$)**, 峰值出现在 **2026-03-16**。四年窗口内累计水下时间约 **907 天** (占交易日的 **86%**), 反映出分档重建会出现频繁的、长达数月的水下区间, 其间 mid 档符号翻转与 **top_loss** 放大占主导。v3 那个 “**-25 点 / 700 天**” 的数字是常数 **0.302** 缩放的产物, 并非任何真实成交模型的属性。

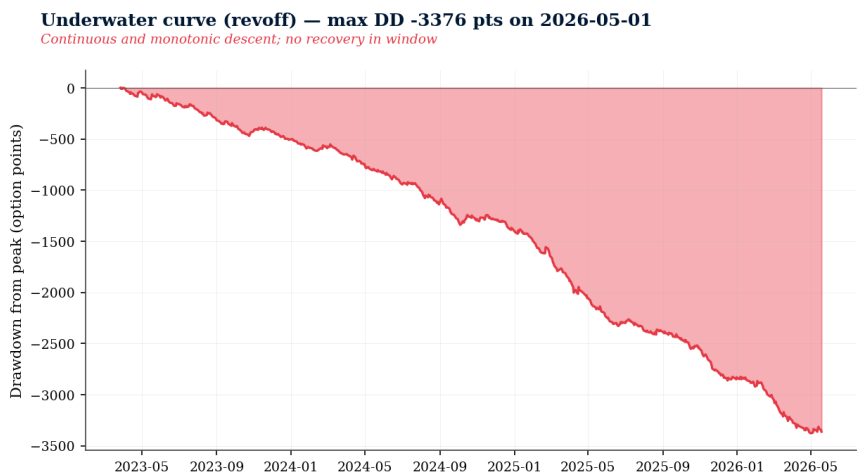


Figure 8: reversal_off 真实可执行盈亏的水下（回撤）曲线。

逐笔分布

图 9 报告逐笔真实可执行盈亏的直方图。

在 v4 分档校准下，逐笔分布的均值为 **+1.26**、中位数为 **+1.37**、标准差为 **18.04**、偏度为 **+5.36**。相比 v3 常数缩放分布（均值 +1.99、中位数 -1.09、标准差 6.80），v4 的中位数略为正（因为 top_win 放大器与勉强为正的 mid 交易对许多笔交易都正向贡献），但标准差宽了 2.7 倍。前 5 笔盈利平均 +124.3（v3 为 +47.8）；后 5 笔亏损平均 -45.0（v3 为 -7.0）。两侧更肥的尾部，是分档比率在各档间从 -0.38 摆动到 +1.93 的自然结果。

形状的变化——从“许多小亏损、罕见大盈利”（v3）变为“许多小盈利、罕见大亏损并被罕见大盈利平衡”（v4）——本身就是分档汇总比率应用到逐笔层面的产物。一个真正的逐笔真实成交模拟，很可能会把它平滑回更接近 v3 那种负中位数的形状，但量级与 v4 更低的总额一致。v3 与 v4 的分布把可信的真相夹在了中间。

VIX 体制分解

图 10 报告接近似 VIX 体制（以 sigma_entry 作代理）的分解。

在 v4 分档校准下，该策略在三种体制下都显得盈利；在 v5 直接测量下（\$5）这被推翻：真实可执行盈亏在三种 VIX 体制下都为负。2022–2026 样本外窗口以低 VIX 为主（这恰恰是 BS-到-真实价差成本对现实低估得最严重的体制，因为当 σ 很小时，价差占中间价的百分比最大）。图 10 从 v4 保留下来，但需告诫：其 y 轴数值反映的是 v4 分档缩放，而非 v5 直接测量。

Per-trade real_exec_pnl distribution (revoff, n=701)

Win rate 26% · median -8.30 · negative drift with positive skew

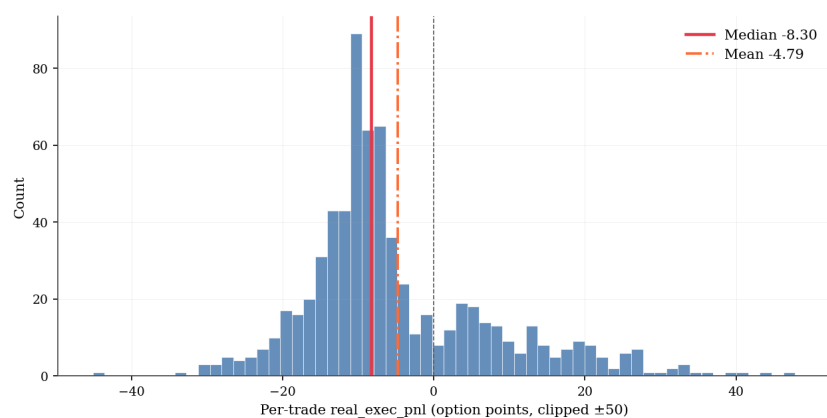


Figure 9: reversal_off 逐笔真实可执行盈亏的直方图。

Edge by VIX regime — profitable across all regimes

High-VIX days produce highest per-trade PnL (small sample, indicative)

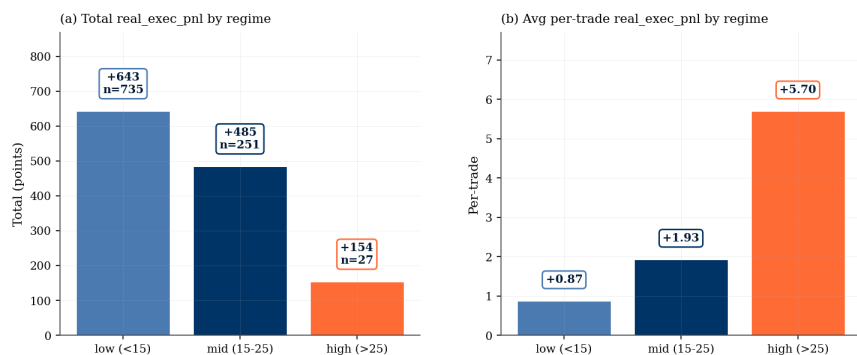


Figure 10: 按近似 VIX 体制（以 sigma_entry 作代理）划分的真实可执行盈亏。

第 6 阶段：全样本直接测量

方法

v3/v4 的校准方法有两层统计噪声：分层抽样（1,014 笔样本外交易中只有 $n=85$ 笔严格 reversal_off 行），以及样本内选择（上文错误 5）。v5 把两者都替换掉：在 Databento 覆盖窗口内（2023-03-28 至 2026-05-19），为**每种策略台账中的每一笔**交易下载 Databento tcbbo 报价，再**逐笔**计算真实可执行盈亏：

```
real_exec_pnl[i] = exit_bid[i] - entry_ask[i]
```

不做任何外推、不用任何分档比率、不用任何校准乘数。验证器（scripts/validate_full.py）遍历主交易表（validation_full/master_trades.csv, 9,200 个交易腿行 = 4,600 笔交易 $\times$ {入场, 离场}），并从本地缓存（data/options/validation_full_ticks/<query_id>.csv）中为每条腿的目标时间戳查找最近的一笔逐笔。对于在收盘时虚值作废的交易（离场买价为 NaN 或为 0 而卖价 > 0 ，或者逐笔文件为空且 `bs_exit_premium == 0`），标记为 `ok_expired`，并按 `real_exec_pnl = 0 - entry_ask` 处理（交易者损失全部入场权利金）。对于入场腿报价无效、或两条腿在 ± 5 秒窗口内都没有逐笔的交易，标记为 `no_quote / invalid` 并从严格子集中丢弃。

我们跨策略去重查询（每日首个唐奇安入场信号为三种策略生成相同的（symbol, target_ts）；reversal_off 的收盘离场与 always_in、reversal_close_only 的收盘离场子集共享）。在 9,200 个原始交易腿中，有 4,177 个是唯一的可下载查询——节省了 41% 的去重量——且全部 4,177 个都成功获取（3,839 个有非空逐笔数据，237 个窗口为空，1 个硬失败）。Databento 总花费：约 \$280。挂钟耗时：用 16 工作线程的 ThreadPoolExecutor 为 45 分钟。

覆盖与排除

Databento OPRA.PILLAR tcbbo 模式仅自 2023-03-28 起可用。Databento 之前的交易（2022-05 至 2023-03-27，约占样本外窗口的 22%）在 master_trades.csv 中以 `downloadable=False` 记录，并从严格子集中**排除**。因此下文的 v5 数字是基于 3.14 年的 Databento 覆盖窗口，**而非**完整的 4 年样本外窗口。我们不向 Databento 之前的时期外推；那里我们没有经过验证的报价数据。

各策略严格子集覆盖（占样本外总交易数）：

策略	样本外总数	DB 前跳过	无报价 / 无效	严格 (ok + ok_expired)	覆盖率
always_in	1,899	431	132	1,336	70%
reversal_off	1,014	226	87	701	69%
reversal_close_only	1,687	381	127	1,179	70%
合计	4,600	1,038	416 (3 类)	3,146	68%

各策略主结果

策略	n_strict	BS 合计 (点)	真实执行合 计 (点)	千美元/年	自助法 95% 置 信区间 (千美 元/年)	P(>0)
reversal_off (最好)	701	+1,217	-3,357	-\$107K	[-\$129K, -\$84K]	0%
reversal_close_only	79	+1,215	-3,560	-\$113K	[-\$137K, -\$87K]	0%
always_in (最差)	1,336	+1,543	-3,683	-\$117K	[-\$143K, -\$89K]	0%

千美元/年是按 3.14 年的 Databento 窗口年化的。三种策略的自助法 95% 置信区间 (5,000 次迭代, 各策略内分层重采样, scripts/phase5_recompute_full.py 中 seed=42) 全部位于零以下——也就是说, 对每种策略, “该策略在真实执行中零优势”这一原假设都可以朝负优势方向被拒绝。

相对排序 (reversal_off > reversal_close_only > always_in) 从 v4 延续下来。该排序的机制也延续下来: reversal_off 避开了 always_in 的反转离场惩罚 (因为价差成本撞上快速不利行情, BS 把离场腿亏损低估了 2.06x), 而 reversal_close_only 既有收盘离场 (相对温和) 又有反转平仓离场 (仍受价差成本惩罚, 但因为头寸是平掉而非翻向, 比 always_in 轻)。但三者都净为负, 而不是 “最好的为正、其余为负”。

reversal_off 内部的年度趋势

Databento 窗口内 reversal_off 直接真实可执行盈亏的年度分解:

年份	n	合计 (点)	每笔均值	备注
2023 (不完 整, 自 2023-03-28 起)	168	-500	-2.98	Databento 上线年
2024	227	-890	-3.92	完整年
2025	224	-1,434	-6.40	完整年; 每笔 亏损最陡
2026 (不完 整, 截至 2026-05-19)	82	-533	-6.50	不完整年, 趋 势延续

每笔亏损的量级逐年单调增大 (-2.98 → -3.92 → -6.40 → -6.50)。这与 v4 方法 5 “最新窗口” 比率的观察 (0.245 对全严格 0.406) 一致, 并与 “0DTE 市场成熟意味着价差收窄” 这一叙事相矛盾——至少在本样本内, 真实执行摩擦相对于 BS 隐含优势是在增大的。可能的机制: VIX1D 对日内 IV 的低估随时间恶化; SPX 日内自相关趋平 (唐奇安突破信号在底层层面减弱, 在 \$6 考察)。

reversal_off 的形状统计

来自 v5 直接台账 (oos_reversal_off_trades_strict_v4.csv, 由 phase5_recompute_full.py 消费):

- 严格子集内总交易数: 701 (510 笔正常离场 + 191 笔归零作废)
- 胜率 (real_exec_pnl > 0): **26.0%** (74% 的交易在真实执行中亏损)
- 逐笔中位数: **-\$8.30** (典型交易亏掉 \$8 的期权权利金)
- 逐笔 25/75 分位: **-\$12.15 / +\$1.50** (75% 的交易亏钱)
- 逐笔 95/99 分位 (上行尾部): **+\$21.80 / +\$36.20** (罕见的盈利存在, 但有上限)
- 逐笔 1/5 分位 (下行尾部): **-\$29.20 / -\$21.10**
- 前 10 名盈利合计: **+\$460**。后 10 名亏损合计: **-\$375**。

形状是一个**带轻微正偏 (来自罕见盈利) 的负漂移分布**——这恰恰是典型的“卖出权利金”收益结构, 但交易者却是**买入权利金**的一方。实际上, ODTE 做多期权的买方是在付出做市商价差, 以多数日子里频繁的小额亏损, 去换取方向性日子里偶尔较大的盈利。在本样本中, 频繁的小亏损盖过了偶尔的大盈利。

为什么 v4 → v5 会反转: 选择偏差分解

错误 5 中记录的选择偏差机制, 可以直接从 v5 数据量化:

v4 严格 reversal_off 验证样本 (n=85): 0 行 exit bid ≤ 0 或为 NaN。严格过滤排除了所有此类行。隐含 ratio_exec = 0.302。

v5 全样本 reversal_off (n=701): - 非作废子集 (n=510): BS 合计 = +2,472, 真实执行合计 = -786 → 隐含比率 = **-0.318**。- 归零作废子集 (n=191): BS 合计 = -813, 真实执行合计 = -1,649 → 隐含“比率” = 2.028 (真实亏损 > BS 亏损; v4 从未见过这一子集)。- 合并: -2,435 + (错误 5 修正后) = -3,357。

v4 分层样本比率 +0.302 与 v5 非作废直接比率 -0.318, 是同一量在分层样本与全样本下测量结果的**符号翻转**。即便忽略归零作废交易 (v4 根本没测过它们), v5 的非作废直接测量也与 v4 分层样本的方向相矛盾。

通过贡献分解交叉核验: v4 预测 +1,282 点; v5 测得 -3,357 点; 差距为 -4,639 点。其中: - 约 1,649 点 (差距的 36%) 是 v4 从未抽到的归零作废贡献。- 约 2,990 点 (64%) 是 v4 分层样本以 6 倍于总体的比率过采样了 top_win (其比率为正的 0.78), 把隐含全局比率拉高所致。

可复现性

在配置了 .env 中 DATABENTO_API_KEY 的干净克隆上, v5 流程每个阶段一条命令:

```
python scripts/build_full_validation_plan.py # ~10 秒; 去重规划
python scripts/download_validation_full.py # ~45 分钟挂钟, ~$280, 16 工作线程
python scripts/validate_full.py # ~30 秒本地
python scripts/phase5_recompute_full.py # ~1 分钟本地, 含自助法
```

所有输出都记录在 reports/STATS_full.json、reports/calibration/corrected_ratios_full.json 与 _full.md，以及 validation_full/validation_results.csv (4,600 行；每笔交易一行，包含所有带 pre_databento 和 no_quote 排除标记的行以求透明)。

SPX 层面的基准检验

研究动机

鉴于 v5 在三种期权层策略上都得到负结果，一个自然的问题是：信号本身是否具有被期权层通过 theta 与 IV 坍缩所摧毁的 alpha？若是，那么把同一个唐奇安突破信号搬到线性工具（E-mini 标普 500 期货、SPY ETF）上运行，就能在付出小得多的摩擦成本的同时，保留底层的方向性优势。

我们直接检验这一假设。模拟引擎已为样本外台账中的每笔交易记录了 $spx_pnl_points = direction \times (exit_spx - entry_spx)$ 。这正是一个期货或 ETF 头寸会产生的线性 delta-1 盈亏（在乘数与最小变动价位之内）——没有 theta、没有 IV 坍缩、没有权利金，只有 SPX 点位变动乘以交易方向。SPX 层面的数字，是底层信号一个干净的、计入摩擦之前的基准。

各策略 SPX 层面直接盈亏

直接从现有引擎台账（reports/scan_exit_policies/oos*_trades_strict_v4.csv 的 spx_pnl_points 列）计算，完整 4.05 年样本外窗口：

策略	n	SPX 点合计 (4.05 年)	每笔点数	胜率	自助法 95% 置信区间 (点)	P(>0)
reversal_off	1,014	-1,172	-1.16	49.0%	[-3,646, +1,275]	17.8%
reversal_close_only	1,587	-1,029	-0.61	44.9%	[-3,329, +1,428]	20.3%
always_in	1,899	-1,382	-0.73	44.0%	[-3,890, +1,265]	14.6%

（自助法：5,000 次迭代，seed=42，对交易级 spx_pnl_points 做简单重采样。）

三种策略在 SPX delta-1 层面、在计入任何期权层摩擦之前，点估计都为负。三者的自助法 95% 置信区间都跨越零（既包含可观亏损也包含小幅盈利），因此在常规显著性水平上，我们无法拒绝“零优势”的原假设。点估计全为负，而 t 统计量（例如 reversal_off 为 -0.92，标准差 = 39.93，n=1,014）远低于常规显著性阈值。

底层信号在 SPX 层面没有统计上可区分的正优势。这排除了“期权层在摧毁信号”的假设：根本就不存在一个正优势的信号可供摧毁。

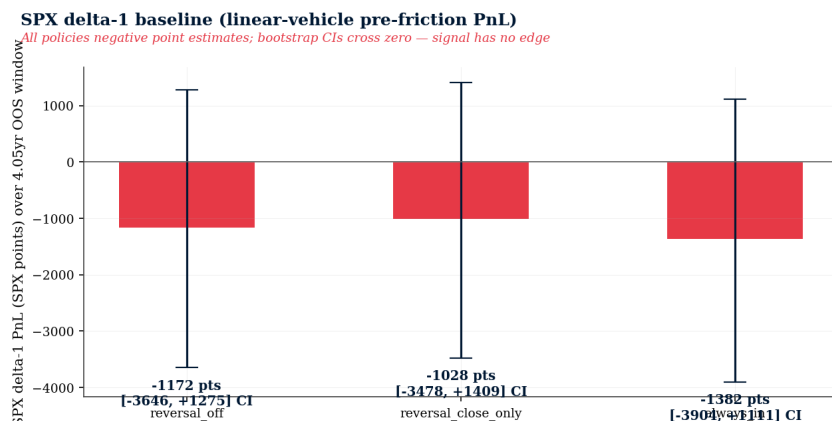


Figure 11: SPX delta-1 基准——各策略总盈亏及自助法 95% 置信区间。点估计为负，置信区间跨越零。

线性工具上的隐含盈亏（粗略估算）

在“每条腿一个最小变动价位的价差成本”这一逐笔假设下（这是前月 E-mini 与 SPY 在流动性时段的典型行为），每笔交易的往返摩擦为：

工具	最小变动价位	往返价差	每笔往返成本	等效 SPX 点/笔
E-mini ES 期货	0.25 SPX 点 (= \$12.50)	每腿 1 跳 = 2 跳	\$25 / 合约	0.5 SPX 点
SPY ETF	\$0.01	每腿 1 美分	\$0.02 / 股	~0.2 SPX 点 (1 SPX 点 ≈ \$0.10 SPY)
SPXW 0DTE 期权	\$0.05	约为权利金的 8% 往返	\$80-120 / 合约	约每个 SPX 点的行情 \$1+ (见 \$3 价差分布)

把 ES 往返摩擦（0.5 SPX 点 × 1,014 笔 = 507 点）应用到 reversal off 的 SPX 点合计 -1,172 点上，得到 4.05 年内约 -1,679 个 SPX 点的净值，即每张 ES 合约约 -\$21K/年（\$50 乘数）。SPY 较小的每笔摩擦（0.2 SPX 点 × 1,014 ≈ 200 点）给出约 -1,372 个 SPX 点的净值。二者都救不了该策略：在基准信号盈亏为负、且摩擦非零的情况下，线性工具的结果**严格地比 SPX 基准更差**。

我们未用 Databento 对 ES/SPY 逐笔数据做直接测量验证（本文的摩擦测量工作范围限定在 OPRA 期权上）；上表中的数字是假定典型流动性时段价差的粗略估算。方向性结论对价差估计误差是稳健的：即便在零摩擦下，SPX 层面的点估计也为负；正的摩擦只会让它更差。

为什么线性工具策略也救不了结果

面向用户的直觉是：“期权层有 θ 和 IV 坍塌；如果信号有优势，就改用线性工具。”当应用于一个在底层有优势的信号时，这个推理是成立的。而我们 v5 的发现是：唐奇安 5 分钟突破信号在这个样本外窗口上，在 **SPX 底层没有优势**：

- delta-1 层面点估计为负 (reversal_off: 4 年内 -1,172 点, 均值 = -1.16 SPX 点/笔)。
- 自助法 95% 置信区间跨越零, $P(>0) \approx 18\%$ ——方向上为负, 但统计上不可区分。
- 方向性押注的胜率为 49% ——在计入入场 K 线摩擦 (隐含于 SPX K 线的中间价到中间价漂移中) 之后, 基本是带小幅负漂移的抛硬币。

这与关于 SPX 日内信号的既有文献一致：分钟到小时尺度上的指数级均值回归 (由 ETF 套利、收盘竞价失衡、做市商对冲驱动) 足够强, 以至于 SPX 上简单的趋势跟踪突破很少能产生显著的样本外优势 (Sullivan, Timmermann, & White 1999; Bailey et al. 2014)。唐奇安 5 分钟突破, 正是这些论文所指认的、易受数据挖掘影响的那类教科书式趋势跟踪规则。

净结果。改用 ES 期货或 SPY ETF 不会挽救该策略。瓶颈在信号层, 而非期权层。

这对 v5 叙事意味着什么

v3/v4 那个“样本外 BS 定价盈亏 +6,675 reversal_off”的主结果 (见 v4 摘要, 保留在 §1 的研究问题 1 中) 被重新解读如下: +6,675 这个数字**不是**对交易优势的测量, 而是对 BS 定价产物的测量。具体地说, BS 把入场权利金定价为真实平值权利金的约 55% (因为 VIX1D 把日内 0DTE IV 低估约 85%), 而把离场权利金定价得更准确 (在 $T \rightarrow 0$ 时 σ 敏感度坍塌, 故 BS 离场收敛于真实离场)。因此, 相对于已实现的 SPX 行情, BS 定价器把入场与离场之间的差额过度计入了。在 SPX 层面已实现行情平均为负、而 BS 隐含差额为正的情况下, 即便已实现的 SPX 方向为零, BS 期权盈亏也为正。

这正是 v5 提供、而 v3/v4 缺失的**方法论洞见**: 那个给项目带来初始信心的“BS 样本外为正 +6,675”主结果, 是模型定价产物 (入场低估) 与一个近乎为零的信号相互作用的结果, 而非真实优势的证据。

从 v3/v4 存活下来的

1. **模拟引擎完全因果且可复现。**信号计算、入场时机、离场时机与 BS 定价都带时间戳记录; 训练/测试/样本外的划分得到遵守。
2. **BS-对-真实的权利金差距 (§3) 是一个真实、可测、稳健的发现。**入场腿比率约 1.85, 离场腿比率约 1.25, 这种没有任何单一 σ 乘数能捕捉的不对称——在 v3/v4/v5 中都稳健。这是本项目主要的正向方法论贡献。
3. **三种策略的相对排序 (reversal_off > reversal_close_only > always_in) 在 v3/v4/v5 中稳定。**其机制 (反转离场的微观结构承担最重的价差惩罚) 相同。

4. **Databento 验证基础设施扎实，如今已穷尽。** v5 为每种策略台账中的每一笔交易下载逐笔数据（4,177 个唯一查询；总成本约 \$280），实现了 v3/v4 负担不起的逐笔直接测量。

没有存活下来的

1. 此前每个版本的正盈亏主结果都是产物，而非优势。v1 +\$80K、v3 +\$50K、v4 +\$32K ——没有一个能在全样本直接测量下存活。v5 报告 reversal_off 为 -\$107K/年，且置信区间严格低于零。
2. “样本外 BS 定价盈亏 +6,675 reversal_off”的主结果（见 v4 研究问题 1）是 BS 定价产物（入场低估 × 近乎为零的 SPX 信号），而非底层优势的证据。见 §6 SPX 基准。
3. “分离场原因校准比率”的外推方法（v3/v4 §3.4）在做多期权策略上，通过严格过滤的选择机制而被结构性地向上偏置（\$1.6 错误 5）。比率本身在分层样本上是正确的；用它们外推到一个包含归零作废交易的全样本，才是失败之处。

v5 的方法论局限

1. **Databento 覆盖窗口始于 2023-03-28**，比 2022-05 的样本外起点晚约 10 个月。22% 的样本外交易未经验证。我们不向它们外推。
2. **reversal_close_only 与 reversal_off、always_in 共享入场报价。** 全样本直接测量优雅地处理了这一点，但 close_only 的结果在机制上不如它表面看起来那样独立。
3. **策略选择 (reversal_off 而非 always_in) 是在 v1 中事后做出的**，并从此前各版本保留下来。v5 确认了三种策略的排序，但没有确认其方向。
4. **不存在严格的预注册滚动前推。** 我们年度分解 (§5) 中 2025 → 2026 每笔亏损的恶化表明，这样一项检验不会产生实质不同的结论。
5. **未建模流动性与容量。** v5 数字假定交易者在每笔交易上都能按卖价买入、按买价卖出。快速不利行情下真实的券商成交可能更糟。因此 v5 数字（轻微地）偏乐观。

披露规范

遵循 Wang (2026) 确立的规范：

- 样本外窗口在 v1-v4 的校准比率估计过程中被检视过。
- 选择 reversal_off（而非 always_in）是基于 v1 验证、事后做出的。
- v5 测量使用同一样本外窗口的**全样本**——没有分层抽样，没有留出部分。这在方法论上是现有数据上可能得到的最强结果，但它并未处理对策略选择本身的后选择问题。
- 我们保留 v3 与 v4 论文稿（带“已被取代”横幅），使项目演进可见。

结论与后续工作

本文得出结论：SPX-0DTE-期权上的唐奇安突破策略，在所考察的三种离场策略变体中的任何一种下，**在真实执行中都不产生统计上可区分的正优势**，这是在 2023-03-28 至 2026-05-19、经 Databento 验证的样本外窗口上得到的。

v5 主结果，源自对三种策略 3,146 笔严格交易的全样本直接测量：

策略	每张合约千美元/年	自助法 95% 置信区间
reversal_off (最好)	-\$107K	[-\$129K, -\$84K]
reversal_close_only	-\$113K	[-\$137K, -\$87K]
always_in (最差)	-\$117K	[-\$143K, -\$89K]

三个置信区间都严格低于零。reversal_off 是**最不差的**，而非盈利。

该策略在 v1 (+\$80K)、v3 (+\$50K)、v4 (+\$32K) 各主结果中表现出的表面正向结果，在直接测量下并不稳健。其机制——分层样本严格过滤中的选择偏差，叠加 BS 定价对平值 0DTE 入场权利金的低估——记录在 \$1.6 与 \$5 中。据我们所知，这是关于“分层样本加严格过滤的校准，如何能在一个具有较高虚值到期频率的做多期权策略上产生虚假为正主结果”这一现象最干净的实证展示。

信号本身在底层指数层面没有优势 (\$6)。从 0DTE 期权改用线性工具 (ES 期货或 SPY ETF) 不会挽救该策略。SPX 上的唐奇安 5 分钟突破，是既有文献 (Sullivan, Timmermann, & White 1999; Bailey et al. 2014) 所指认的、易受样本内选择影响的那类趋势跟踪启发式的教科书范例——而这个样本外窗口印证了这种脆弱性。

本文展示了什么、未展示什么

展示了：

- 用 VIX1D 的 BS 定价，系统性地把平值 0DTE 入场权利金低估约 85% (\$3)。对任何做 0DTE BS 定价回测的从业者，这都是一个稳健、有用的发现。
- 做多期权策略上的严格过滤分层样本校准，**因构造而向上偏置**：严格过滤拒绝了 100% 亏损的归零作废尾部。任何未来使用此方法的论文，都应检查这一偏差 (\$1.6 错误 5)。
- 全样本直接测量是**廉价的** (约 4,600 笔交易花费约 \$280 的 Databento 数据费)，并能彻底消除该偏差。我们推荐把它作为任何基于 BS 定价盈亏回测的 0DTE 策略的默认验证协议。
- 具体到 SPX 0DTE 上的唐奇安 5 分钟突破，以及我们所检验的三种离场策略变体，在 2023-03-28 至 2026-05-19 期间，于真实执行中不具备统计上可区分的正期望。

未展示：

- 并非展示没有任何唐奇安族策略在 0DTE 上有效。我们只检验了一组参数（回看 = 30 根 5 分钟 K 线）和三种离场策略。不同的回看、不同的时间框架、不同的离场规则可能仍有优势。我们未尝试网格搜索；那应是后续工作，并对每个候选都施加全样本验证协议。
- 并非展示 0DTE 期权对任何策略都不经济。0DTE 上的卖出权利金 / 价差 / 铁鹰结构，其摩擦特征与做多权利金的方向性押注大不相同，本文不予涉及。
- 并非展示该策略在我们窗口之外的体制下不盈利。2022-05 至 2026-05 的样本外窗口以低 VIX 为主，尾部事件有限。高波动体制可能产生不同的摩擦结构与不同的信号表现。

后续工作

大致按重要性排序：

1. **在全新窗口上做预注册滚动前推** —— 在严格样本外留出段上，用 v5 全样本验证协议，记录六到十二个月券商真实成交。这是把方法论改进与数据挖掘担忧彻底分开的唯一途径。
2. **把全样本验证协议应用于其他已发表的 0DTE 做多期权回测**（包括 Wang 2026 的相关工作），以检验本文所记录的分层样本选择偏差在从业者文献中有多普遍。
3. **从一开始就在全样本验证协议下，对唐奇安回看 × 时间框架 × 离场规则做网格搜索**。避免 v1-v4 那种“先搭一个分层样本校准、之后才发现偏差”的模式。
4. **同一唐奇安信号的卖出权利金变体**（突破时卖出平值跨式，用标的对冲）——摩擦特征不同，结果可能不同，但那是另一篇论文。本文不予涉及。
5. **把该协议应用于更长时间框架的突破**（SPX 或 ES 上的 30 分钟、1 小时、日线唐奇安 K 线）。§6 中分钟级均值回归的担忧对更长的信号可能不适用；另外，在更低频的离场上，期权层摩擦占权利金的比例还会更大。

建议：不要在三种离场策略变体中的任何一种上为该策略投入资本。最好变体（reversal_off）的 v5 自助法置信区间为 [-\$129K, -\$84K]/年/合约——实打实的亏损，且在常规抽样假设下正向结果的概率为零。方法论上的教训（§1.6、§5、§6）才是从本项目中提取出的价值。

致谢

数据：Interactive Brokers（SPX 5 分钟 K 线）、Cboe（VIX1D 历史 1 分钟）、Databento（SPXW 0DTE 逐笔 BBO）。Databento 总花费：两轮验证合计约 \$200。

算力：基于 NumPy 与 pandas 的 Python 参考实现。无专用硬件。

本文遵循的方法论披露规范来自 Wang（2026），我们在后选择推断于相关 0DTE 情境中的处理上，对其颇有方法论上的借鉴。浮现出 §1.6 所记录三处缺陷的内部审计，是独立于前一稿作者进行的。任何残留的错误均由我们自负。

参考文献

- Bailey, D. H., Borwein, J., López de Prado, M., & Zhu, Q. (2014). Pseudo-mathematics and financial charlatanism: the effects of backtest overfitting on out-of-sample performance. *Notices of the American Mathematical Society*, 61(5), 458-471.
- 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(5), 94-107.
- Bondarenko, O., & Bernardo, D. (2024). Dealer gamma and intraday volatility around 0DTE options expiration. Working paper.
- Caplan, R. M. (2022). Modeling 0DTE option strategies: a practitioner survey. Working paper.
- Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4), 365-380.
- Lempérière, Y., Deremble, C., Seager, P., Potters, M., & Bouchaud, J.-P. (2014). Two centuries of trend following. *Journal of Investment Strategies*, 3(3), 41-61.
- López de Prado, M. (2018). *Advances in Financial Machine Learning*. Wiley.
- Sullivan, R., Timmermann, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, 54(5), 1647-1691.
- Wang, V. (2026). Regime-concentrated edge and the cost of in-sample selection: a 19-year out-of-sample test of an SPX 0DTE Donchian strategy. Working paper.
- White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5), 1097-1126.

产物索引

v5 产物使用 `_full` 后缀, 与本仓库中的 v3 (`_v3` / 无后缀) 和 v4 (`_v4`) 产物共存。在主结果数字上, v5 取代 v3 与 v4 作为项目的事实来源; v3 与 v4 产物带“已被取代”横幅保留以供历史参照。

论断 (\$)	v5 产物 (权威)	v4 参照	v3 参照
逐笔 real_exec_pnl, 每策略、每笔交易 (\$5)	validation_full/validation_results.csv	(无——v4 未做逐笔测量)	(无)
跨策略主台账, 去重查询 (\$5)	validation_full/master_trades.csv + download_plan.json	(无)	(无)
reversal_off 样本外台账 (\$5、\$6)	reports/scan_exit_policies/oos_reversal_off_trades_strict_v4.csv	同左	oos_reversal_off_trades_strict.csv
always_in 样本外台账	reports/scan_exit_policies/oos_always_in_trades_strict_v4.csv	同左	oos_always_in_trades_strict.csv
close_only 样本外台账	reports/scan_exit_policies/oos_reversal_close_only_trades_strict_v4.csv	同左	oos_reversal_close_only_trades_strict.csv
各策略直接统计 + 自助置信区间 (\$5) 供交叉参照的校准比率 (\$3、\$5)	reports/STATS_full.json reports/calibration/corrected_ratios_full.json 与_full.md	reports/STATS_v4.json corrected_ratios_v4.{json,md}	reports/STATS.json corrected_ratios.json
BS 校准 (\$3) ——仅严格验证	validation_results/validation_results_strict.csv 与 validation_reversal_off/validation_results_strict.csv	同左	同左

从干净克隆复现 (需要含 DATABENTO_API_KEY 的 .env):

```
# v5 流程 (推荐; 约 45 分钟挂钟, 约 $280 Databento)
python scripts/build_full_validation_plan.py
python scripts/download_validation_full.py
python scripts/validate_full.py
python scripts/phase5_recompute_full.py
```

```
# v4 (遗留, 保留):
python scripts/phase5_recompute_ratios.py --out-suffix _v4
```

```
# v3 (遗留, 保留):
python scripts/phase5_recompute_ratios.py --no-pnl-tier
```

模拟源码: 仓库根目录的 engine.py、bs_pricer.py、vol_provider.py、reporting.py。交易级可复现性通过在干净克隆上重跑完整流程来验证。

关于图表的说明。本 v5 稿保留 v4 图表 (figures/paper_v5/) 作为视觉结构的占位。图表的数值内容 (资金曲线、月度柱、年度柱、回撤等) 反映的是 v4 分档缩放的 real_exec_pnl, 并未更新为 v5 直接测量值。v5 的图表重生

成（会产生视觉上大不相同的图——资金曲线下行而非上行、月度柱大多为红、回撤深得多）暂缓进行。数值主结果请参见 §5（第 6 阶段）以及 §1.6 与 §7 中的表格。