

投资策略报告 · v3 诚实修订版 · 2026 年 5 月

标普 500 的 衰退保险叠加策略

滚动前推样本外证据、稳健性，以及我们仍未知晓的部分

Vincent Wang / 王子秋

量化研究

研究区间 2000-01-01 → 2026-05-29 · 标的池 标普 500 (以 SPY 代理)

NBER 衰退区间（约占 8% 的交易日）

+12.15 pp/ 年

相对买入并持有

策略：- 16.16%/ 年

买入并持有：- 28.31%/ 年

非衰退区间（约占 92% 的交易日）

-0.57 pp/ 年

相对买入并持有（即保险的成本）

策略：+12.69%/ 年

买入并持有：+13.26%/ 年

这是一条保险型收益曲线，而非可持续的超额收益。夏普比率之差（0.586 对 0.43，点估计 +0.154）在 21 个样本外年份中，与零在统计上无法区分。

1. 问题与假设

为什么纯多头股票需要一层衰退叠加

2. 数据与关键取舍

58 → 14 → 3 个指标； 2009 年前的回补

3. 方法论

变换 → 意外值 → 复合 → 仓位权重

4. 事件研究证据

单次事件信号很弱；趋势才是超额收益所在

5. 样本内结果

五个候选；纯劳动力以微弱优势胜出

6. 滚动前推样本外

阈值层面的稳定性，并附置信区间

7. 诚实的稳健性

回补、固定阈值、归因、分解

8. 失效模式

2020 错失，2018/2022 属设计内

9. 局限与下一步

本研究并不主张的内容

三段行情贡献了长期回撤的绝大部分

标普 500 ， 2000–2026

2000 – 2002

互联网泡沫

–49%

2007 – 2009

全球金融危机

–57%

2020 年一季度

新冠冲击

–34%

在这几段行情中削减股票敞口——只要能避免大额的反复止损成本——便具有不对称的价值。

**能否用公开可得的周频
宏观数据，系统地识别出
应当削减股票敞口的时期？**

—— *同时在正常体制下保留大部分股票参与度* ——

H1 周频劳动力信号

初次申请失业金人数（初请）是频率最高的实时劳动力指标（周频，滞后 5 天）——高到足以据此行动。

H2 复合信号更稳定

由劳动力构成的复合信号（初请 + 非农 + 失业率）可能比任何单一劳动力指标都更稳定，在衰退体制下尤其如此。

H3 缓释尾部风险

将复合信号转化为仓位调节规则后，可在劳动力主导的衰退中显著降低回撤。

数据来源

可由 FRED API + yfinance 复现；以 SPY 复权收盘价作为 SPX 代理

数据类型	来源	覆盖范围
58 个宏观指标	FRED API	2000-01-01 → 2026-05-29
历史版本发布日期	FRED ALFRED	2009-05-28 起（2009 年前为估算）
标普 500 （SPY 复权收盘价）	yfinance	2000-01-04 起；已做股息调整
VIX 及波动率指数	FRED + CBOE CSV	2000 年起
国债收益率	FRED	2000 年起

策略以 SPY 总回报为基准；100%/125%/150% 仓位假定可通过保证金或期货实现（v3 暂未建立融资成本模型）。

指标筛选：58 → 14 → 3

从 FRED 全量指标到最终交易模型的漏斗

58

FRED 指标

初始指标库：通胀、劳动力、
经济活动、景气、住房、
货币、市场、贸易。



14

事件研究集

在单次事件分析（第 4 节）中，
发布日期数据可靠且对市场有
显著反应的指标。



3

最终劳动力复合

初请 + 非农 + 失业率。
劳动力聚焦源自样本内
五策略对比。

58 → 14 这一步偏探索性；14 → 3 则由样本内对比给出依据。下游的滚动前推只验证阈值的稳定性，并不验证上游的指标筛选——详见“局限”。

2009 年前初请发布日期的回补

一个可能起承重作用的假设

问题所在

FRED 中 ICSA 的真实历史版本发布日期只追溯到 2009-05-28。

ICSA 原始序列可上溯至 1967 年，但发布日期档案没有。

我们的方案

用确定性的 5 天滞后，回补 490 个 2009 年前的发布日期。

已验证：2009 年后 832 次发布中，818 次滞后恰为 5 天（98.3%）。

第 7 节给出评审所要求的敏感性检验。预告：在更保守的滞后假设下，2008 年的贡献不仅成立——而且更好。

流程概览

从原始指标到组合权重的四个步骤

1

变换

因指标而异
(同比 / 环比 / 原值)

2

意外值 z 分数

以滚动均值为基线
滚动标准化

3

复合

劳动力信号 z 分数的
等权求和

4

仓位

阶梯函数：
 $z > +d \rightarrow 0\% \dots z < a \cdot 2 \rightarrow 150\%$

滚动前推只包裹第 3、4 步：复合信号事前固定，阈值 d 与 a 则在每个样本外年份仅用此前数据重新校准。

步骤 1 • 因指标而异的变换

朴素的滚动均值意外值在单调趋势序列上会失效

指标类别	变换	理由
价格指数（CPI、PCE、PPI）	同比 %	市场定价的是变化率，而非水平
就业水平（非农、JOLTS）	环比差分	新闻报道的是环比变化
经济活动（工业产出、零售）	环比 %	增长率才是相关指标
比率（失业率、GDP 季环比）	原始水平	本身已是类变化形式
周频流量（初请）	原始水平	属流量变量，而非存量

若不做这些变换，CPI 的单调趋势会产生一个永远为正、却毫无经济含义的“意外值”。

步骤 2 · 意外值 z 分数

对每条变换后的序列计算；窗口因频率而异

```
expected_t    = rolling_mean(transformed_{t-1 ... t-6})

surprise_t    = transformed_t - expected_t

surprise_z_t  = (surprise_t - μ_roll) / σ_roll
```

信号	预期窗口	标准化窗口	频率
初请 意外值 / 趋势	6 周 / 4 周对比前 4 周	52 周	周频
非农 环比差分	1 个月	36 个月	月频
失业率 水平 + Δ 萨姆	5 年滚动	5 年滚动	月频

步骤 3a • 初请复合信号

意外值 z + 趋势 z ，等权

```
surprise_z    = z((claims_t - rolling_mean(t-1..t-6)), 52w)
trend_z        = z((4w MA - prior 4w MA), 52w)
composite_JC   = 0.5 · surprise_z + 0.5 · trend_z
```

20 日前向收益（1,354 次事件，2000–2026）

大幅好于预期（初请下降）：+1.02% · 大幅差于预期（初请上升）：−0.05%

朴素 t 统计量 = +2.84 ($p=0.005$)。HAC / 分块自助法修正属下一步工作。

该信号在 2009–2026 子样本中单独看并不可见——回补是必需的。

步骤 3b • 纯劳动力复合信号

三个信号等权，且统一定号使 $z > 0 =$ 劳动力恶化

$$\text{composite_z_labor} = \frac{1}{3} \cdot \text{signal_JC} + \frac{1}{3} \cdot \text{signal_NFP} + \frac{1}{3} \cdot \text{signal_UNRATE}$$

初请

意外值 + 趋势复合信号（见上一页）。

非农

环比就业变化 z 分数的相反数——就业增长越低 $\Rightarrow z > 0$ 。

失业率

$\frac{1}{2}$ 水平 $z + \frac{1}{2}$ 相对过去 12 个月最低点变化的 z （受萨姆规则启发）。

步骤 4 • 仓位调节规则

带两个学习阈值（d, a）的符号化阶梯函数

复合 z 区间	仓位权重	体制标签
$z > 2 \cdot d$	0%（现金）	劳动力强烈恶化
$d < z \leq 2 \cdot d$	50%	风险升高
$a \leq z \leq d$	100%（中性）	正常体制
$2 \cdot a \leq z < a$	125%	劳动力改善
$z < 2 \cdot a$	150%	强劲扩张

滚动前推学习 (d, a)。在 21 个样本外年份中，它收敛到 $d \in \{0.5, 0.6\}$ 、 $a = -0.5$ （100% 的年份）。

第 7 节显示，固定阈值（ $d=0.6, a=-0.5$ ）与滚动前推结果一致——阈值优化几乎没有增加任何东西。

单次事件信号很弱

标普 500 对指标意外值的 1 日反应——前四名

指标	事件数	好于-差于预期价差（1 日）	相关系数
非农	301	-0.87%	-0.107
核心 CPI	289	+1.29%	+0.116
失业率	301	+0.05%	+0.046
初次申请失业金人数	1,354	-0.09%	-0.040

初请样本量最大，但单次事件的反应最小。它的价值在于跨多周的趋势提取——而非事件当日的波动。

初请信号真正所在之处——是趋势，而非单次读数

同一份数据，两种不同的解读方式

单次事件反应

1,354 次周频事件（2000–2026）

好于-差于预期价差，1 日 SPX：-0.09%

相关系数：-0.04

单次事件信号基本上是噪声。

多周趋势

以 z 分数表示的 4 周均线变化。

20 日 t 检验： $t = +2.84$ ， $p = 0.005$

大幅好于预期 20 日收益：+1.02%

大幅差于预期 20 日收益：-0.05%

注：该 t 检验未处理重叠收益——详见“局限”。

样本内策略对比（2000–2026）

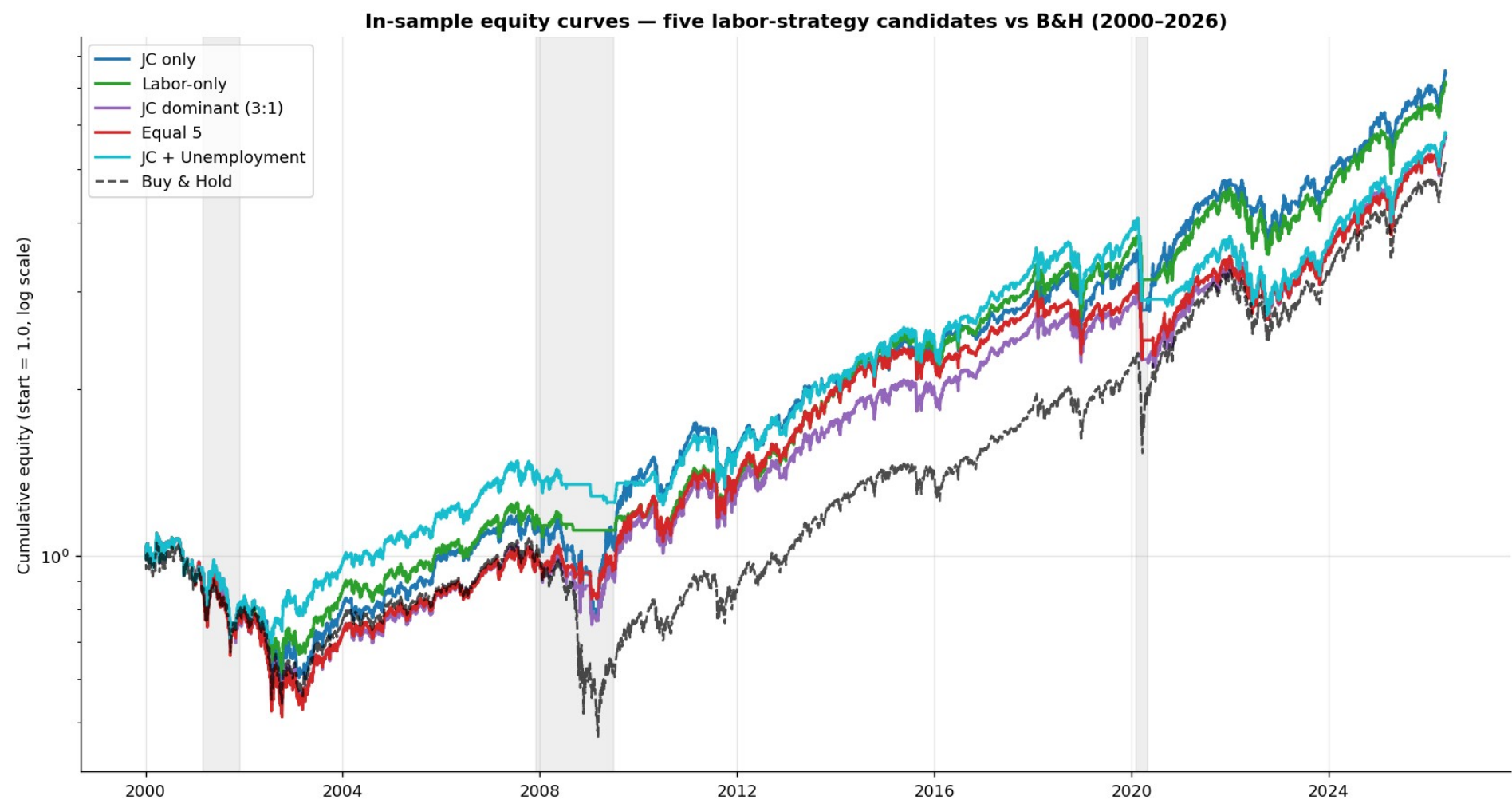
五个候选加买入并持有——仅为样本内拟合，非样本外

策略	CAGR	夏普	索提诺	最大回撤
仅初请	+7.92%	0.43	0.51	−47.5%
纯劳动力	+7.73%	0.47	0.55	−44.2%
初请主导（3:1）	+6.82%	0.39	0.47	−52.2%
五信号等权	+6.86%	0.39	0.48	−53.4%
初请 + 失业率	+6.67%	0.40	0.48	−43.1%
买入并持有	+6.37%	0.32	0.40	−56.5%

纯劳动力夏普 0.47 以微弱优势胜过单信号初请（0.43），也优于买入并持有（0.32）。加入非劳动力信号（五信号等权）反而更差。选定纯劳动力作为最终策略族，依据正是这一样本内对比。

样本内净值曲线——五个劳动力候选 vs 买入并持有

替换 v2 第 28 页——使用正确的五策略图



灰色竖带：NBER 衰退。净值取对数刻度，起点 = 1.0。

滚动前推方法论

扩张窗口，21 个样本外年份（2006–2026）

- 初始训练窗口：2000–2005（6 年）。
- 此后每一年，仅用此前全部数据重新校准 (d, a) ，再应用于当年。
- 网格搜索 20 种组合： $d \in \{0.4, 0.5, 0.6, 0.7, 0.8\}$ ， $a \in \{-0.4, -0.5, -0.6, -0.7\}$ 。
- 优化目标：索提诺比率。
- 滚动前推所验证的：阈值层面的稳定性。
- 滚动前推不验证的：指标筛选、变换选择、复合权重、仓位规则形态——这些都由研究者事前固定。

滚动前推样本外核心指标（2006–2026）

策略 vs 同期样本内拟合 vs 买入并持有

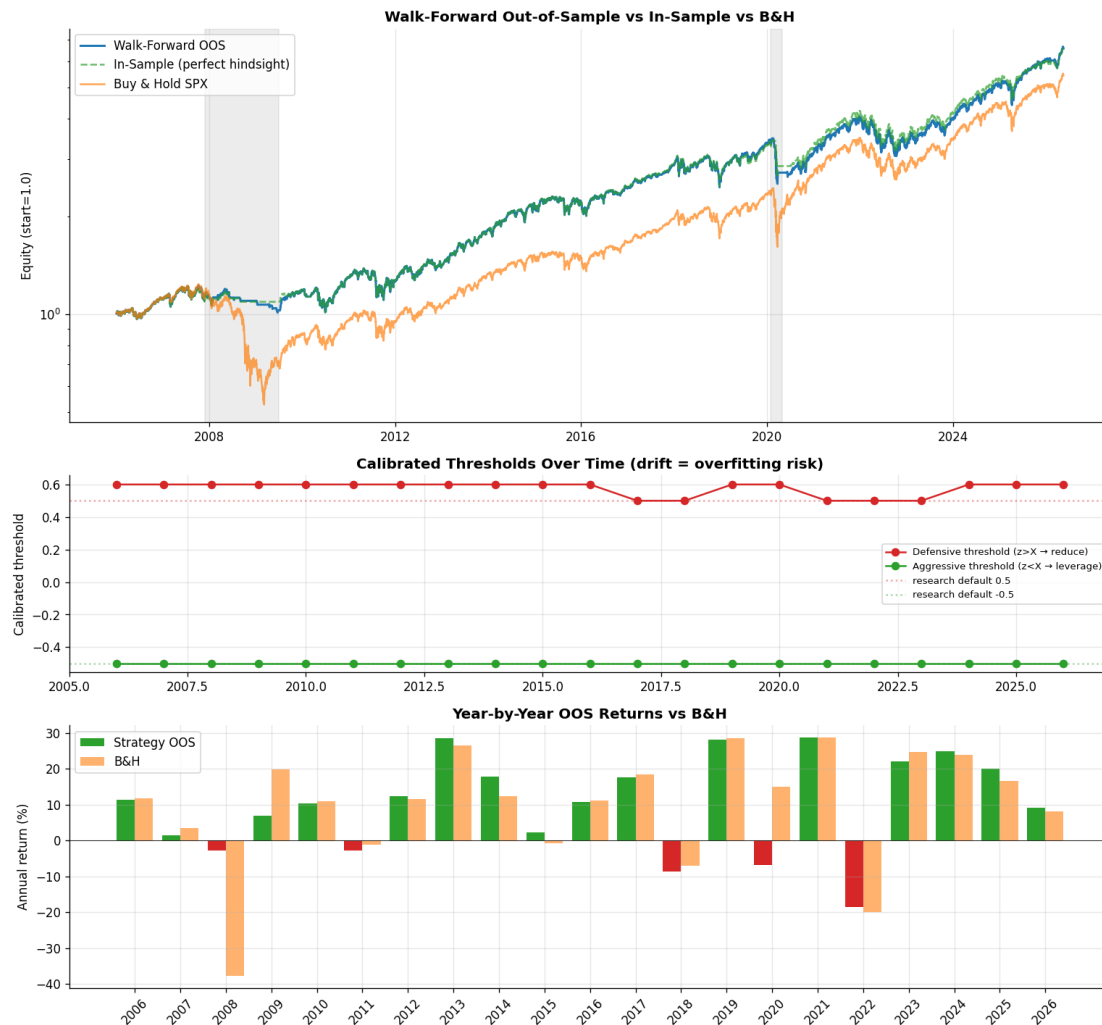
	CAGR	夏普	索提诺	最大回撤
滚动前推样本外	+9.75%	0.586	0.67	-27.6%
样本内拟合（样本外区间）	+9.72%	0.593	0.67	-24.4%
买入并持有（样本外区间）	+8.72%	0.432	0.52	-57.4%

夏普这个数字要谨慎看待

0.586 只是点估计。在 21 个样本外年份下，平稳分块自助法给出 95% 置信区间 = [0.232, 0.977]。策略减买入并持有的夏普之差为 +0.154，置信区间 = [-0.119, +0.481]。单侧 $p = 0.18$ 。我们无法在 5% 水平上拒绝“夏普等于买入并持有”。详见第 7 节。

滚动前推样本外净值曲线

策略（样本外） vs 同期样本内拟合 vs 买入并持有



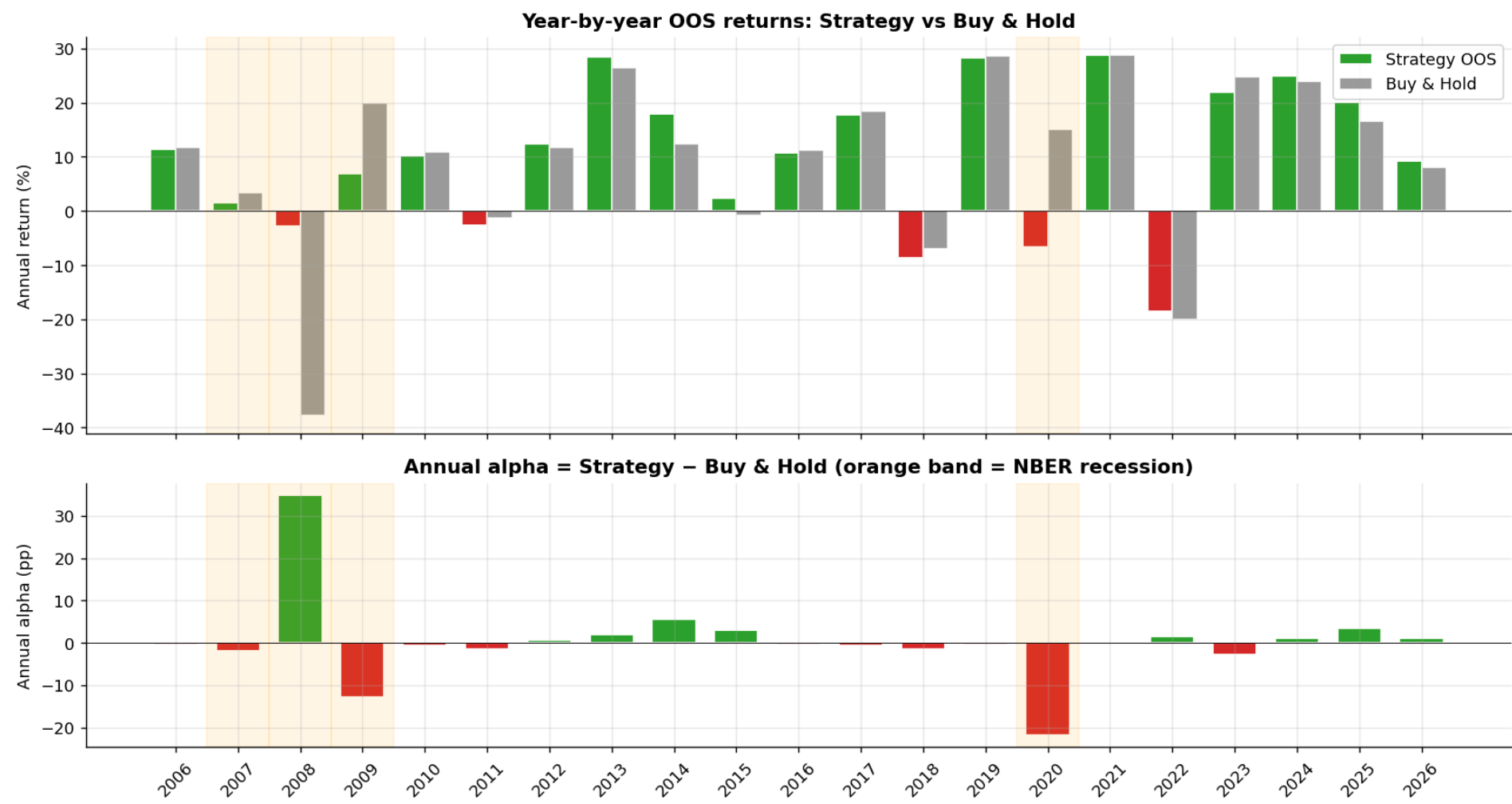
滚动前推净值、校准漂移、逐年表现（自动生成）

两点观察

- 样本外曲线与样本内曲线高度贴合 → 阈值层面稳定
- 两者均在 2008-09 期间偏离买入并持有；2009 年的反弹回吐约 13 pp
- 夏普从样本内 0.593 → 样本外 0.586 的衰减很小（0.007）

逐年样本外收益与超额收益

累计结果究竟从何而来



橙色竖带：NBER 衰退。2026 年为部分年份（截至 5 月）。

稳健性 1 • 2009 年前回补敏感性

这个 5 天滞后假设到底有多承重？

情景	2008 策略	2008 买持	样本外夏普	最大回撤
基准（5 天滞后）	-2.84%	-37.74%	0.586	-27.6%
滞后 +1 天（6 天）	+5.24%	-37.74%	0.627	-24.4%
滞后 +2 天（7 天）	+5.95%	-37.74%	0.618	-24.4%
滞后 +5 天（10 天）	+4.51%	-37.74%	0.620	-24.4%
滞后 - 2 天（3 天）	-3.64%	-37.74%	0.525	-27.6%
不做回补	-1.26%	-37.74%	0.544	-30.3%

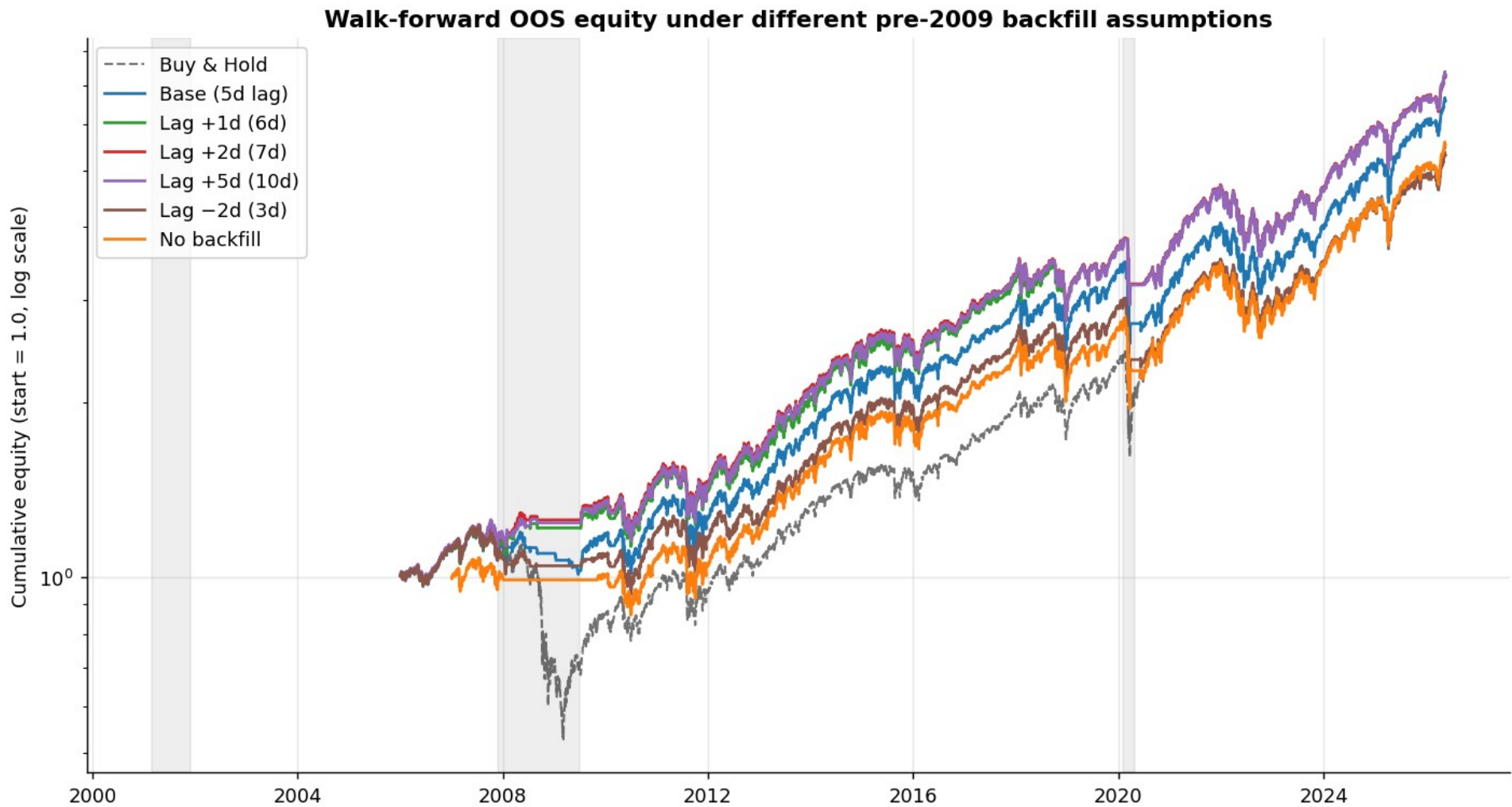
我们的发现

- 5 天滞后其实属于较保守的情景之一
- 更长滞后（6-10 天）反而改善 2008 结果与全期夏普
- 即便完全不回补，样本外夏普仍维持在 0.544

结论：2009 年前的回补假设是稳健的。核心结论 0.586 并非某种数据假象的产物。

稳健性 1 • 回补敏感性——净值曲线

六次滚动前推叠加；到 2010 年后全部升至买入并持有之上



灰色竖带：NBER 衰退。净值取对数刻度，起点 = 1.0。

稳健性 2 • 夏普置信区间

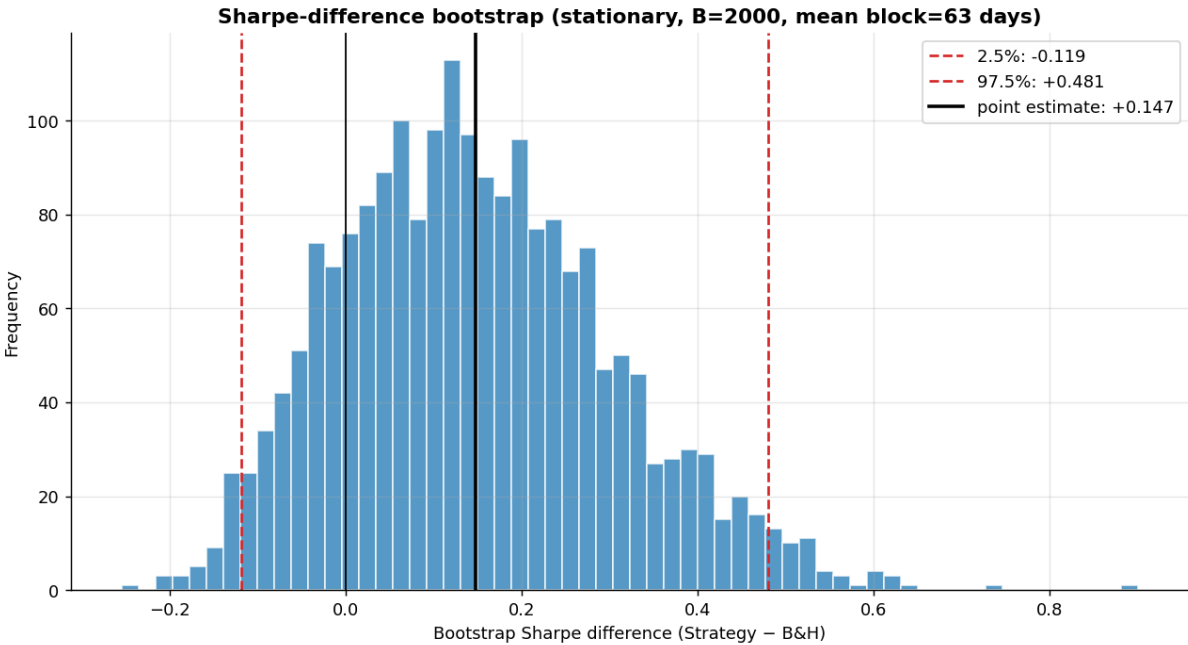
平稳分块自助法，B = 2,000，平均分块 63 个交易日

指标	点估计	95% 下限	95% 上限
策略样本外夏普	0.586	0.232	0.977
买入并持有样本外夏普	0.432	0.042	0.865
夏普之差（策略 - 买持）	+0.154	-0.119	+0.481

夏普之差检验

单侧 $p = 0.18$

无法在常规 5% 水平上拒绝 H_0 ：策略与买入并持有的夏普相等。



夏普之差的自助法分布。

黑线 = 点估计；红色虚线 = 95% 置信区间边界。

稳健性 3 • 固定阈值 vs 滚动前推

阈值优化究竟带来了价值，还是只是做做样子？

变体	CAGR	夏普	索提诺	最大回撤
滚动前推（学习得到）	+9.75%	0.586	0.67	-27.6%
固定 $d=+0.5$ $a=-0.5$	+9.72%	0.593	0.67	-24.4%
固定 $d=+0.6$ $a=-0.5$	+9.93%	0.596	0.69	-27.6%
固定 $d=+0.7$ $a=-0.5$	+10.08%	0.591	0.68	-27.6%
固定 $d=+0.6$ $a=-0.4$	+9.67%	0.568	0.66	-30.8%
固定 $d=+0.6$ $a=-0.6$	+9.56%	0.587	0.67	-27.0%

固定阈值（ $d=0.6, a=-0.5$ ）与滚动前推结果持平甚至更优。解读：（ a ）在合理范围内，策略对阈值选择不敏感——不存在过拟合；（ b ）相对固定阈值，滚动前推这一步没有贡献任何额外价值。

稳健性 4 • 对 2008 的依赖

剔除 2008 → 策略跑输买入并持有

情景	策略 CAGR	买持 CAGR	超额收益	策略夏普
完整样本外	+9.75%	+8.72%	+1.03pp	0.586
剔除 2008	+10.44%	+11.91%	-1.47pp	0.616
剔除 2008-2009	+10.64%	+11.49%	-0.85pp	0.614
剔除 2020	+10.68%	+8.40%	+2.28pp	0.680
剔除 2008 与 2020	+11.47%	+11.74%	-0.27pp	0.717

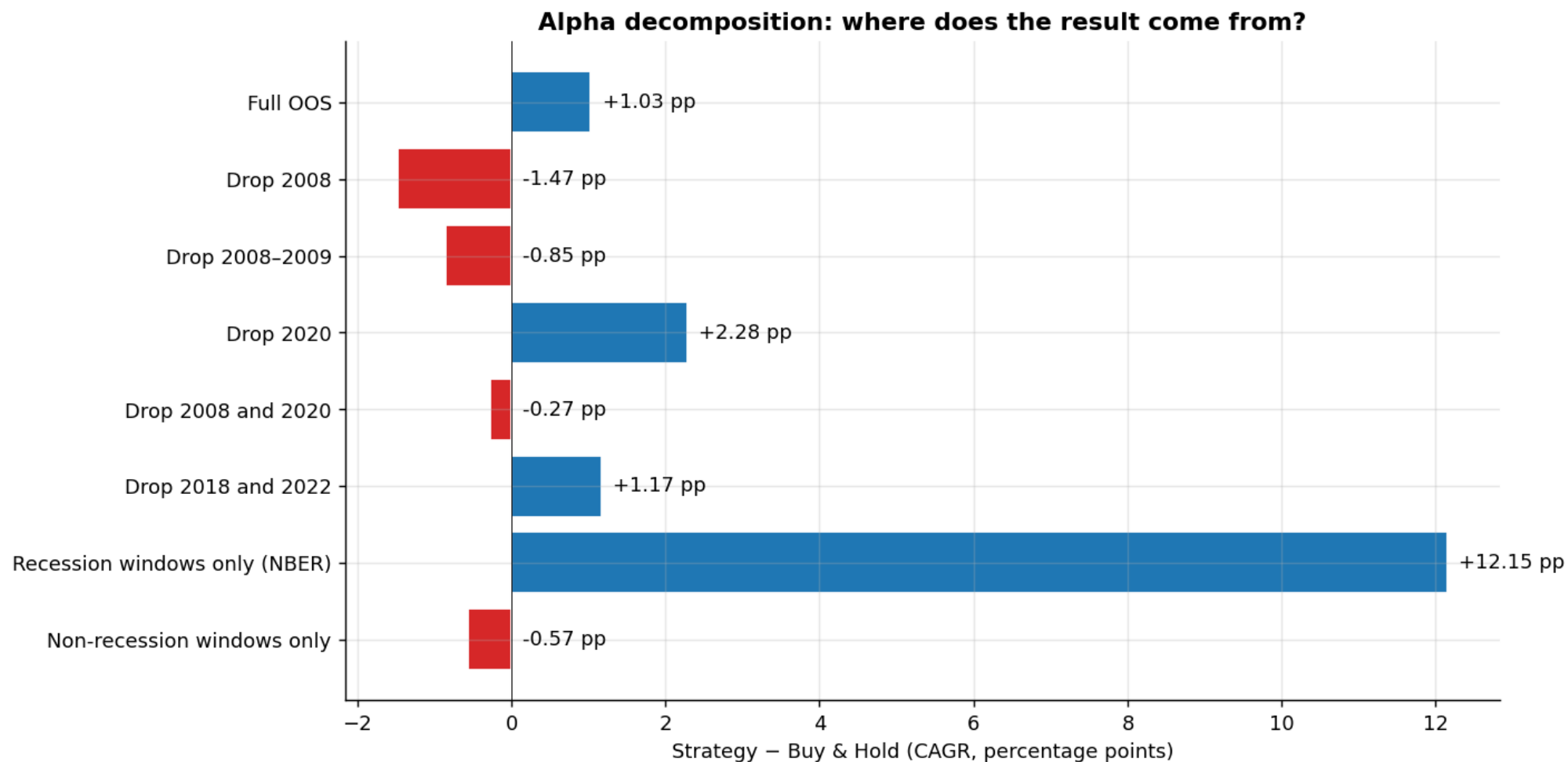
若没有 2008，这层叠加策略每年跑输买入并持有 1.5 pp。

若没有 2020，这层叠加策略每年跑赢 2.3 pp。

净超额收益 1.03 pp/ 年 = 2008（+35 pp）- 2009（- 13 pp）- 2020（- 22 pp）+ 少量。

稳健性 5 • 保险分解

NBER 衰退日 vs 非衰退日



仅衰退期与仅非衰退期这两行，是保险价值的关键证据。

2020 年——一次真实的失败，而不只是“晚了”

年度超额收益 = -21.82 pp (策略 - 6.7% vs 买持 +15.1%)

发生了什么

劳动力信号在 2020 年 3 月触发了防御仓位（这是对的）。但等到仓位重新加回风险时，美联储托底的反弹已经走完。策略躲过了新冠回撤，却错失了大部分反弹。

这意味着什么

快速反弹体制是一个有据可查的弱点——剔除 2020 后，超额收益升至 +2.3 pp/ 年。新冠这一段是样本外区间内最大的单一拖累项。

早先用“滞后 1 周”来描述低估了它的影响；逐年归因让这一量级清晰可见。

失效模式分类

按体制划分，这套策略能做什么、不能做什么

体制	策略表现	样例时段
劳动力主导的衰退	提前降低敞口，获取超额收益	2008 （ +35 pp ）
突发外生冲击	对劳动力的反应滞后 → 错失反弹	2020 （ - 22 pp ）
通胀 / 美联储估值回撤	保持持仓（劳动力仍健康）	2022 （ - 1.4 pp ）
周期尾部回调	保持持仓（劳动力仍健康）	2018 （ - 1.6 pp ）
正常扩张	100% 或 125% 参与	2010 年代、2024-25

这套策略专门针对的是劳动力主导的衰退风险——而非估值压缩或速率冲击。

- 衰退保险叠加策略——在正常体制下付出一笔小而持续的成本，换取劳动力主导衰退中的大额收益
- 可解释——每一次仓位变动都有明确的劳动力市场解释
- 监控成本低——三条 FRED 序列，周频更新
- 对阈值稳健——固定 ($d=0.6$, $a=-0.5$) 与滚动前推学习效果相当
- 对回补稳健——敏感性检验确认 2008 结果并非 5 天滞后造成的假象

这套策略不是什么

本研究并不主张的内容

- 在夏普上与买入并持有有统计学上的可区分性（单侧 $p = 0.18$ ；21 个样本外年份不足）
- 全天候的超额收益来源——剔除 2008 后，策略跑输买入并持有
- 对速率冲击稳健——2020 年超额收益为 - 22 pp
- 不受上游选择影响——只有阈值是滚动前推；指标与复合的选择均为事前固定
- 已扣除实施成本——尚未建立交易成本、现金收益或融资成本模型

与其他方法的对比

参考区间，并非严格可比

方法	典型夏普	本策略	说明
买入并持有标普 500	~0.43	—	基准（本研究）
60/40 股债	~0.45–0.55	—	资产类别分散
趋势跟踪 CTA	~0.40–0.60	—	依赖体制
纯劳动力叠加	—	0.586	宏观事件驱动（本研究）
波动率目标	~0.55–0.65	—	对已实现波动率作出反应

所示区间为宽泛的业界经验参考，并非受控基准。诚实的对比应在相同的数据、成本与区间上构建全部五种策略——该工作列在“下一步”。

衰退事件太少

21 个样本外年份只含 2 次 NBER 衰退 + 1 次疫情。结果由 3 个事件主导。

过度集中于 2008

剔除 2008 → 超额收益转负 (-1.5 pp/ 年)。结果依赖于衰退事件。

统计功效不足

夏普之差相对买入并持有不显著 (95% 置信区间含 0 ; 单侧 $p = 0.18$) 。

仅阈值做滚动前推

指标筛选、变换、复合权重均为事前固定——并未纳入嵌套样本外。

历史版本数据仅部分使用

使用了 ALFRED 历史版本发布日期；历史版本数值（尤其非农、失业率）未用。

无成本 / 融资 / 现金模型

无摩擦回测。25-50 个基点成本 + 0% 现金收益 + 杠杆融资均未建模。

仅限美国

推广到非美市场需要周频劳动力数据；大多无法获得。

结论——五项更克制的主张

当前证据所能支撑的内容

- 纯劳动力叠加在 2006–2026 期间改善了样本外风险调整后表现——夏普点估计 0.586 vs 0.43。
- 价值主要来自缓释劳动力主导的衰退性回撤，尤以 2008 为最（+35 pp 超额收益）。
- 阈值校准稳定：固定 ($d=0.6$, $a=-0.5$) 与滚动前推输出一致。
- 2009 年前的回补假设在 ± 5 天扰动下经验上稳健。
- 夏普的改善在 21 年样本上不具统计显著性（单侧 $p = 0.18$ ）。

下一步——按决策价值排序

在投入任何资金之前

- 用 ALFRED 的 `get_series_all_releases` 重建非农 / 失业率 / ICSA 的历史版本数值——消除最后残余的未来函数风险
- 嵌套滚动前推——让外层循环也从 5 个候选中挑选策略族
- 完整的实施成本模型——交易成本、现金 T-bill 收益、SOFR + 50 个基点保证金融资
- 将核心 CPI 与收益率曲线作为条件过滤器（而非加性成分）来检验，以应对 2022 型体制
- 预先登记的纸面交易窗口（至少 6 个月），方可投入实盘资金

提问与讨论

Vincent Wang · 投资策略报告 (v3 诚实修订版) · 2026 年 5 月

可复现性: FRED API 密钥 + 约 10 分钟算力 · 分析代码见 </analysis/>