

Investment Strategy Report · v3 honest revision · May 2026

Recession-Insurance Overlay for the S&P 500

Walk-forward OOS evidence, robustness, and what we still don't know

Vincent Wang

Quantitative Research

Research period 2000-01-01 → 2026-05-29 · Universe S&P 500 (SPY proxy)

The Honest Headline

What this strategy does, in one slide

NBER recession windows (~8% of days)

+12.15 pp/yr

vs Buy & Hold

Strategy: -16.16%/yr

Buy & Hold: -28.31%/yr

Non-recession windows (~92% of days)

-0.57 pp/yr

vs Buy & Hold (cost of insurance)

Strategy: +12.69%/yr

Buy & Hold: +13.26%/yr

This is an insurance payoff profile, not a persistent alpha. Sharpe difference (0.586 vs 0.43, point estimate +0.154) is not statistically distinguishable from zero in 21 OOS years.

Today's Agenda

9 sections · ~35 minutes

- 1. The problem & hypothesis** Why long-only equity needs a recession overlay
- 2. Data & critical choices** 58 → 14 → 3 indicators; the pre-2009 backfill
- 3. Methodology** Transform → surprise → composite → position weight
- 4. Event-study evidence** Per-event signals are weak; trend is the alpha
- 5. In-sample results** Five candidates; labor-only wins by a hair
- 6. Walk-forward OOS** Threshold-level stability, with confidence intervals
- 7. Honest robustness** Backfill, fixed-threshold, attribution, decomposition
- 8. Failure modes** 2020 missed, 2018/2022 by design
- 9. Limitations & next steps** What this study does NOT claim

Three Episodes Drove Most of the Long-Term Drawdown

S&P 500, 2000–2026

2000 – 2002

Dot-com

–49%

2007 – 2009

Global financial crisis

–57%

2020 Q1

COVID shock

–34%

Reducing equity exposure during these episodes — if it can be done without major whipsaw cost — has asymmetric value.

Can publicly available, weekly-frequency macro data systematically identify periods when equity exposure should be reduced?

— while preserving most equity participation in normal regimes —

Hypothesis

Three claims this study tests

H1 Weekly labor signal

Initial Jobless Claims is the highest-frequency real-time labor indicator (weekly, 5-day lag) — high enough to act on.

H2 Composite stability

A labor composite (JC + NFP + UNRATE) may be more stable than any single labor indicator, especially across recession regimes.

H3 Tail-risk mitigation

Translated into a position-size rule, the composite can meaningfully reduce drawdowns during labor-led recessions.

Data Sources

Reproducible from FRED API + yfinance; SPY adjusted close as SPX proxy

Data type	Source	Coverage
58 macro indicators	FRED API	2000-01-01 → 2026-05-29
Vintage release dates	FRED ALFRED	2009-05-28 onwards (pre-2009 estimated)
S&P 500 (SPY adjusted close)	yfinance	2000-01-04 onwards; dividend-adjusted
VIX, volatility indices	FRED + CBOE CSV	2000+
Treasury yields	FRED	2000+

Strategy is benchmarked against SPY total return; 100% / 125% / 150% positions assumed implementable on margin or via futures (no financing model in v3).

Indicator Selection: 58 → 14 → 3

A funnel from full FRED coverage to the final trading model

58

FRED indicators

Initial library: inflation, labor, activity, sentiment, housing, monetary, market, trade.



14

Event-study set

Indicators with reliable release-date data and meaningful market reactions on per-event analysis (Section 4).



3

Final labor composite

JC + NFP + UNRATE.
Labor focus emerged from in-sample 5-strategy comparison.

The 58 → 14 cut is exploratory; the 14 → 3 cut is informed by in-sample comparison. The downstream walk-forward validates threshold stability only, not the upstream selection — see Limitations.

Pre-2009 Jobless Claims Release-Date Backfill

A potentially load-bearing assumption

The problem

FRED's true vintage release dates for ICSA only begin 2009-05-28.

The raw ICSA series extends to 1967, but release-date archive does not.

Our solution

Backfill 490 pre-2009 release dates using the deterministic 5-day lag.

Validated: 818 of 832 post-2009 releases have lag = exactly 5 days (98.3%).

Section 7 reports the sensitivity test the reviewer asked for. Preview: the 2008 contribution survives — and improves — under more conservative lag assumptions.

Pipeline Overview

Four steps from raw indicator to portfolio weight

1

Transform

indicator-specific
(YoY / MoM / raw)

2

Surprise z-score

rolling-mean baseline
rolling normalization

3

Composite

equal-weighted sum of
labor signal z-scores

4

Position

step function:
 $z > +d \rightarrow 0\% \dots z < -a \rightarrow 150\%$

Walk-forward wraps Steps 3 and 4: the composite is fixed ex ante; thresholds d and a are recalibrated each OOS year on prior data only.

Step 1 · Indicator-Specific Transforms

A naive rolling-mean surprise fails on monotone-trending series

Indicator class	Transformation	Rationale
Price indices (CPI, PCE, PPI)	Year-over-year %	Market prices the rate, not the level
Employment levels (NFP, JOLTS)	Month-over-month diff	News reports MoM change
Activity (IndProd, Retail)	Month-over-month %	Growth rate is the relevant metric
Rates (Unemployment, GDP_QoQ)	Raw level	Already in change-like form
Weekly flow (Jobless Claims)	Raw level	Flow variable, not a stock

Without these transforms, CPI's monotone trend produces a perpetually positive "surprise" with no economic meaning.

Step 2 · Surprise z-Score

Applied to each transformed series; windows differ by frequency

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

```
surprise_t = transformed_t - expected_t
```

```
surprise_z_t = (surprise_t -  $\mu_{\text{roll}}$ ) /  $\sigma_{\text{roll}}$ 
```

Signal	Expectation window	Normalization window	Frequency
JC surprise / trend	6 weeks / 4w-vs-prior-4w	52 weeks	weekly
NFP MoM diff	1 month	36 months	monthly
UNRATE level + Δ Sahm	5y rolling	5y rolling	monthly

Step 3a · Jobless Claims Composite

Surprise z + trend z, equal-weighted

```
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-day forward return (1,354 events, 2000-2026)

BigBeat (claims fell): +1.02% · BigMiss (claims rose): -0.05%

Naive t-stat = +2.84 (p=0.005). HAC / block-bootstrap correction is next-step work.

Signal is not visible in the 2009-2026 sub-sample alone — the backfill is required.

Step 3b · Labor-Only Composite

Three signals, equal-weighted, signed so $z > 0$ = labor deterioration

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

JC

Surprise + trend composite (previous slide).

NFP

Negated z-score of MoM payroll change — lower job growth $\Rightarrow z > 0$.

UNRATE

$\frac{1}{2}$ level-z + $\frac{1}{2}$ z of change from trailing 12-month minimum (Sahm-rule inspired).

Step 4 · Position-Sizing Rule

Symbolic step function with two learned thresholds (d, a)

Composite z range	Position weight	Regime label
$z > 2 \cdot d$	0% (cash)	Strong labor deterioration
$d < z \leq 2 \cdot d$	50%	Elevated risk
$a \leq z \leq d$	100% (neutral)	Normal regime
$2 \cdot a \leq z < a$	125%	Labor improving
$z < 2 \cdot a$	150%	Strong expansion

Walk-forward learns (d, a). Across 21 OOS years it converges to $d \in \{0.5, 0.6\}$ and $a = -0.5$ (100% of years).

Section 7 shows fixed thresholds ($d=0.6$, $a=-0.5$) match the walk-forward result — the threshold optimization adds essentially nothing.

Per-Event Signals Are Weak

1-day S&P 500 reaction to indicator surprises — top 4

Indicator	N events	Beat–Miss spread (1d)	Correlation
NFP	301	−0.87%	−0.107
Core CPI	289	+1.29%	+0.116
Unemployment	301	+0.05%	+0.046
Initial Jobless Claims	1,354	−0.09%	−0.040

Initial Jobless Claims has the largest sample but the smallest per-event reaction. Its value lies in trend extraction across multiple weeks — not in event-day moves.

Where the JC Signal Actually Lives — the Trend, not the Print

Same data, two different ways of reading it

Per-event reaction

1,354 weekly events (2000-2026)

Beat–Miss spread on 1-day SPX: -0.09%

Correlation: -0.04

Per-event signal is essentially noise.

Multi-week trend

4-week MA change as z-score.

20-day t-test: $t = +2.84$, $p = 0.005$

BigBeat 20d return: $+1.02\%$

BigMiss 20d return: -0.05%

Note: t-test ignores overlapping returns — see Limitations.

In-Sample Strategy Comparison (2000–2026)

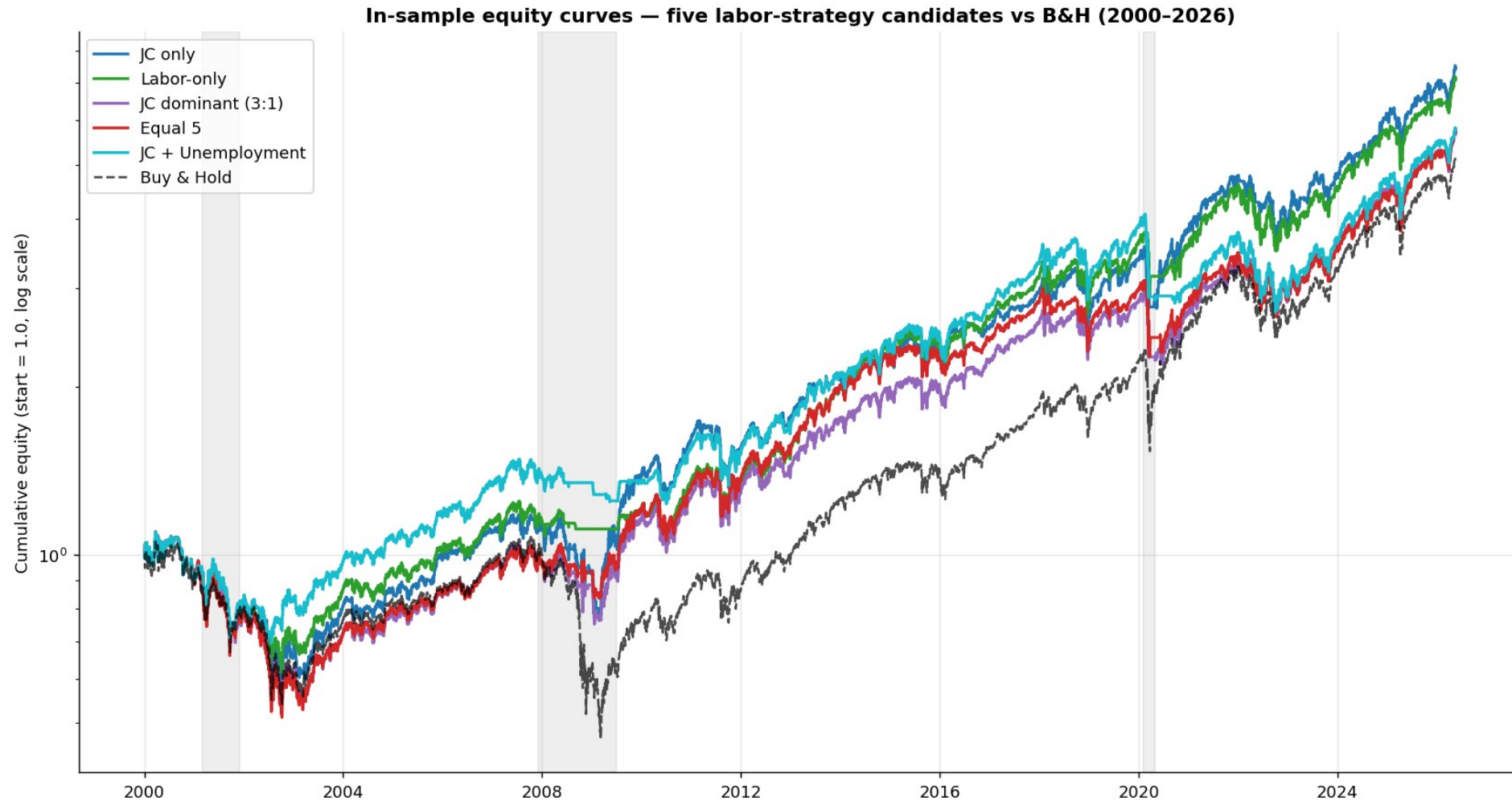
Five candidates plus Buy & Hold — in-sample fit only, not OOS

Strategy	CAGR	Sharpe	Sortino	Max DD
JC only	+7.92%	0.43	0.51	−47.5%
Labor-only	+7.73%	0.47	0.55	−44.2%
JC dominant (3:1)	+6.82%	0.39	0.47	−52.2%
Equal 5 (all signals)	+6.86%	0.39	0.48	−53.4%
JC + Unemployment	+6.67%	0.40	0.48	−43.1%
Buy & Hold	+6.37%	0.32	0.40	−56.5%

Labor-only Sharpe 0.47 narrowly beats single-signal JC (0.43) and improves on B&H (0.32). Adding non-labor signals (Equal 5) makes it worse. Selection of labor-only as the final family is informed by this in-sample comparison.

In-Sample Equity Curves — Five Labor Candidates vs B&H

Replaces v2 slide 28 — uses the correct 5-strategy figure



Grey vertical bands: NBER recessions. Log-scale equity, start = 1.0.

Walk-Forward Methodology

Expanding-window, 21 OOS years (2006–2026)

- Initial training window: 2000–2005 (6 years).
- Each subsequent year, recalibrate (d, a) on all prior data only; apply to current year.
- Grid search 20 combinations: $d \in \{0.4, 0.5, 0.6, 0.7, 0.8\}$, $a \in \{-0.4, -0.5, -0.6, -0.7\}$.
- Optimization target: Sortino ratio.
- What WF validates: threshold-level stability.
- What WF does NOT validate: indicator selection, transform choice, composite weights, position-rule shape — these are fixed ex ante by the researcher.

Walk-Forward OOS Headline Metrics (2006–2026)

Strategy vs in-sample fit (same period) vs Buy & Hold

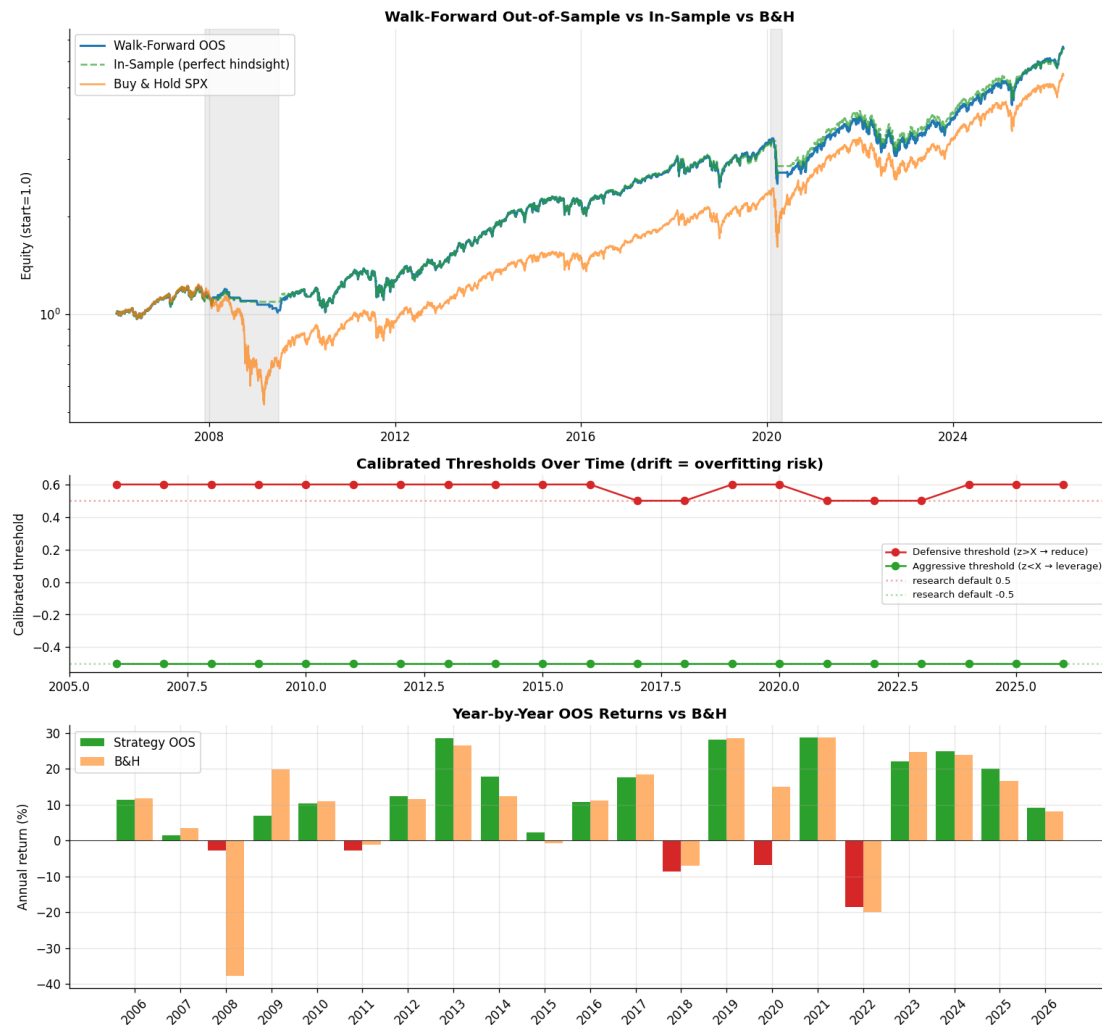
	CAGR	Sharpe	Sortino	Max DD
Walk-forward OOS	+9.75%	0.586	0.67	−27.6%
In-sample fit (OOS pd)	+9.72%	0.593	0.67	−24.4%
Buy & Hold (OOS pd)	+8.72%	0.432	0.52	−57.4%

Read the Sharpe number with caution

0.586 is a point estimate. With 21 OOS years, stationary block-bootstrap gives 95% CI = [0.232, 0.977]. The strategy minus B&H Sharpe difference is +0.154, CI = [−0.119, +0.481]. One-sided $p = 0.18$. We cannot reject "Sharpe equals B&H" at the 5% level. See Section 7.

Walk-Forward OOS Equity Curve

Strategy (OOS) vs in-sample fit (same period) vs Buy & Hold



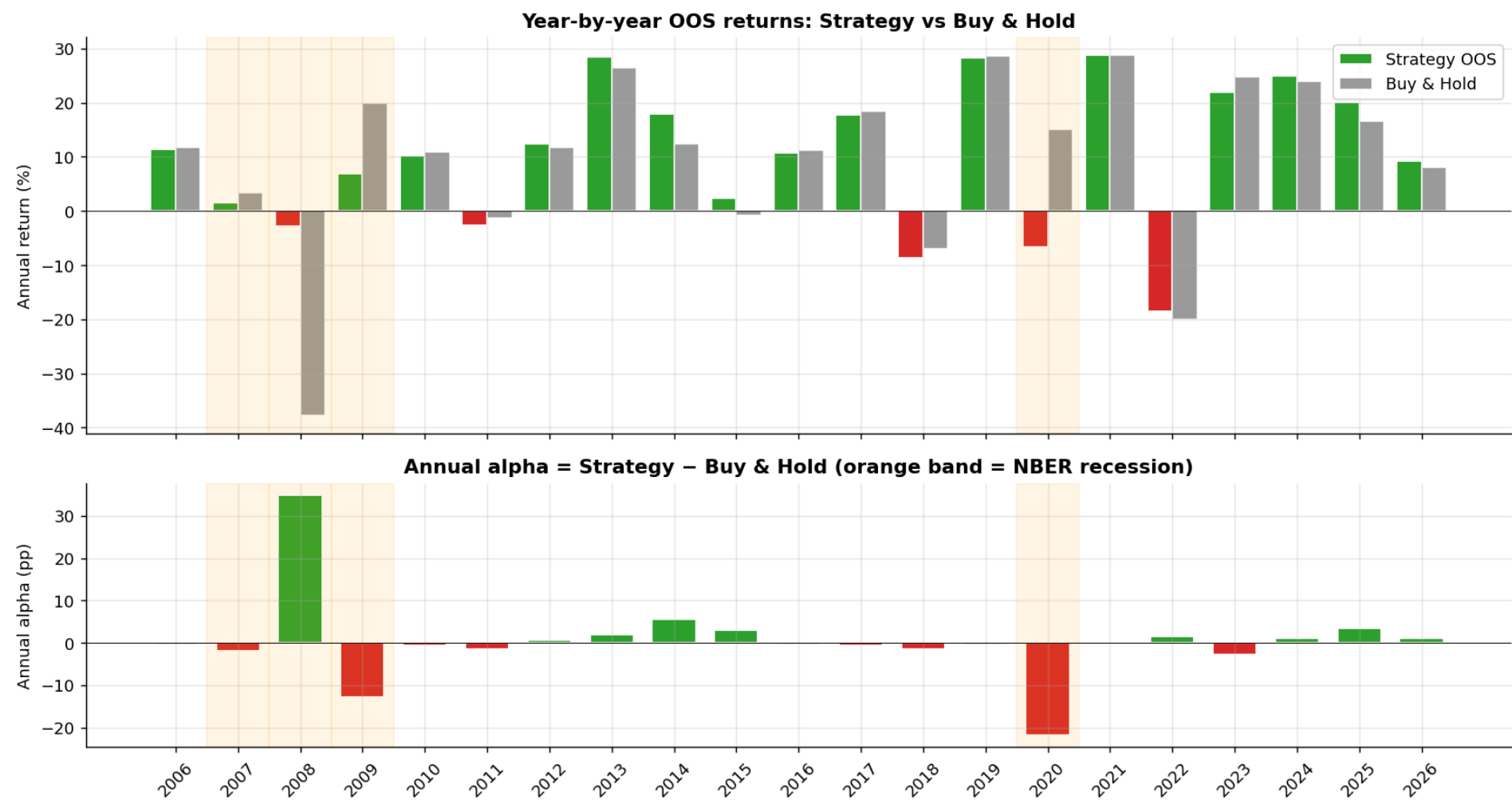
WF equity, calibration drift, year-by-year (auto-generated)

Two observations

- OOS line tracks in-sample line closely → threshold-level stability
- Both diverge from B&H during 2008-09; the 2009 recovery gives back ≈13 pp
- Sharpe-decay from in-sample 0.593 → OOS 0.586 is small (0.007)

Year-by-Year OOS Returns and Alpha

Where the cumulative result actually comes from



Orange bands: NBER recessions. 2026 is partial (through May).

Robustness 1 · Pre-2009 Backfill Sensitivity

How load-bearing is the 5-day lag assumption?

Scenario	2008 strat	2008 B&H	OOS Sharpe	Max DD
Base (5d lag)	−2.84%	−37.74%	0.586	−27.6%
Lag +1d (6d)	+5.24%	−37.74%	0.627	−24.4%
Lag +2d (7d)	+5.95%	−37.74%	0.618	−24.4%
Lag +5d (10d)	+4.51%	−37.74%	0.620	−24.4%
Lag −2d (3d)	−3.64%	−37.74%	0.525	−27.6%
No backfill	−1.26%	−37.74%	0.544	−30.3%

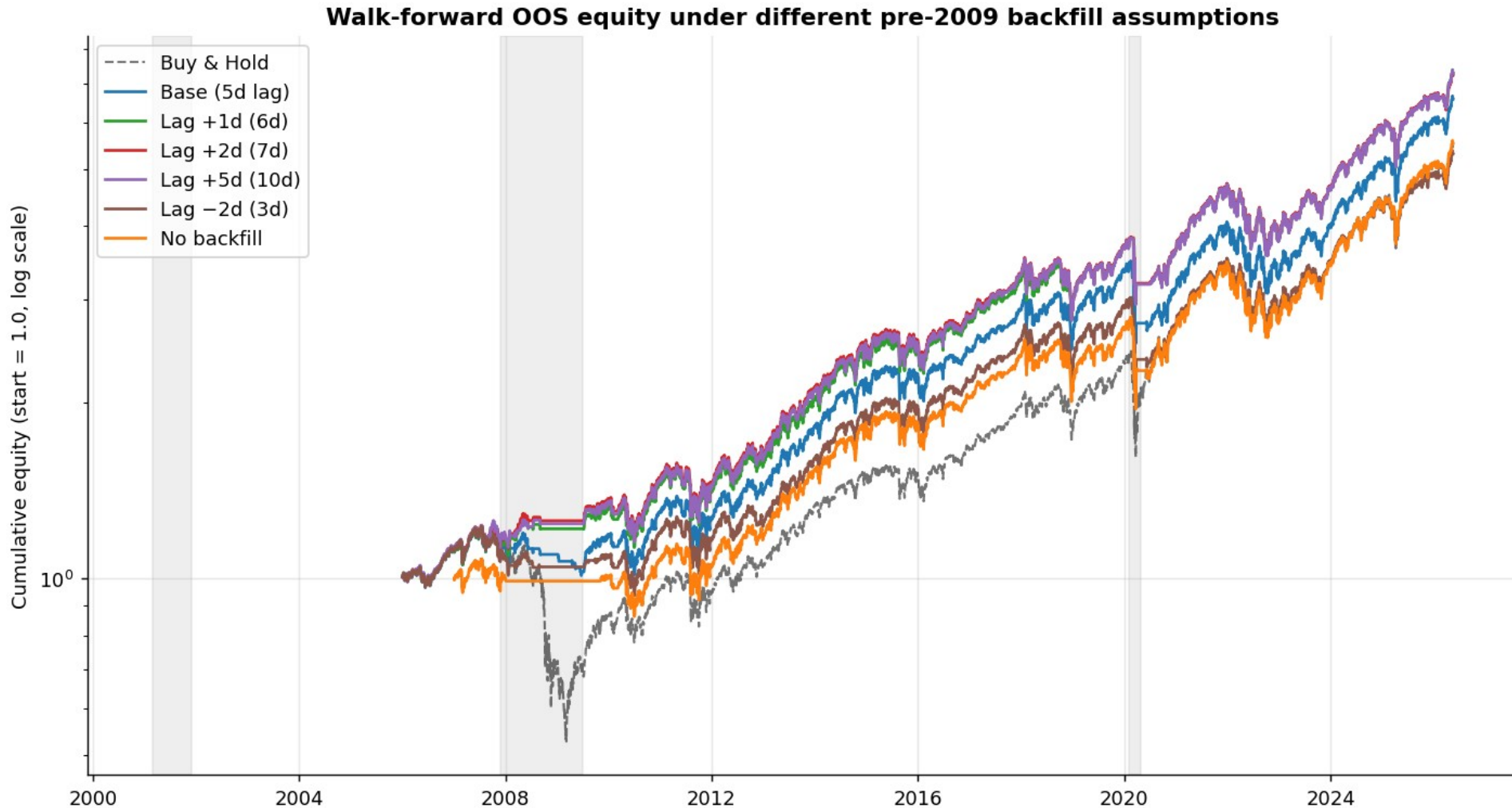
What we found

- 5-day lag is actually one of the more conservative scenarios
- Longer lag (6–10d) improves 2008 outcome and full-period Sharpe
- Even with no backfill at all, OOS Sharpe stays 0.544

Conclusion: the pre-2009 backfill assumption is robust. The headline 0.586 is not the product of an artifact.

Robustness 1 · Backfill Sensitivity — Equity Curves

Six WF runs overlaid; all rise above Buy & Hold by 2010+



Grey bands: NBER recessions. Log-scale equity, start = 1.0.

Robustness 2 · Sharpe Confidence Interval

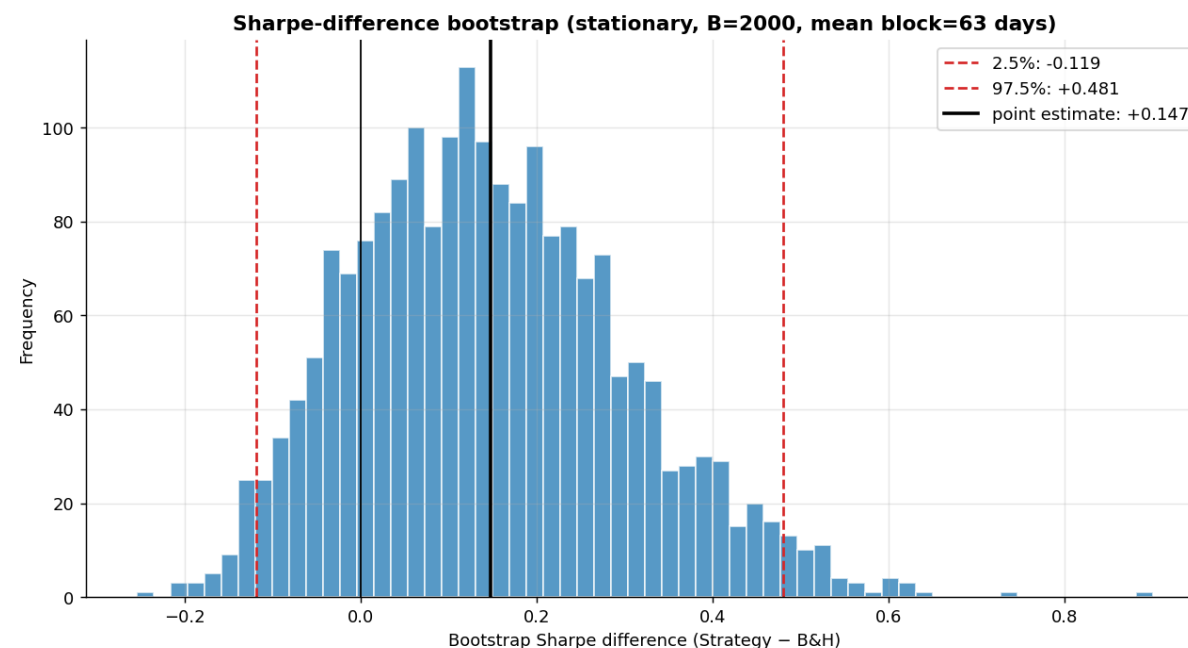
Stationary block bootstrap, $B = 2,000$, mean block 63 trading days

Metric	Point	95% low	95% high
Strategy OOS Sharpe	0.586	0.232	0.977
Buy & Hold OOS Sharpe	0.432	0.042	0.865
Sharpe diff (Strat – B&H)	+0.154	−0.119	+0.481

Sharpe difference test

One-sided $p = 0.18$

Cannot reject H_0 : strategy and B&H have equal Sharpe at conventional 5% level.



*Bootstrap distribution of Sharpe difference.
Black = point estimate; red dashes = 95% CI bounds.*

Robustness 3 · Fixed-Threshold vs Walk-Forward

Does threshold optimization add value, or is it theatre?

Variant	CAGR	Sharpe	Sortino	Max DD
Walk-forward (learned)	+9.75%	0.586	0.67	−27.6%
Fixed $d=+0.5$ $a=-0.5$	+9.72%	0.593	0.67	−24.4%
Fixed $d=+0.6$ $a=-0.5$	+9.93%	0.596	0.69	−27.6%
Fixed $d=+0.7$ $a=-0.5$	+10.08%	0.591	0.68	−27.6%
Fixed $d=+0.6$ $a=-0.4$	+9.67%	0.568	0.66	−30.8%
Fixed $d=+0.6$ $a=-0.6$	+9.56%	0.587	0.67	−27.0%

Fixed thresholds ($d=0.6$, $a=-0.5$) match or beat the walk-forward result. Interpretation: (a) strategy is insensitive to threshold choice within reasonable range — no overfitting; (b) walk-forward step contributes nothing over fixed thresholds.

Robustness 4 · The 2008 Dependency

Drop 2008 → strategy underperforms Buy & Hold

Scenario	Strategy CAGR	B&H CAGR	Alpha	Strat Sharpe
Full OOS	+9.75%	+8.72%	+1.03pp	0.586
Drop 2008	+10.44%	+11.91%	−1.47pp	0.616
Drop 2008-2009	+10.64%	+11.49%	−0.85pp	0.614
Drop 2020	+10.68%	+8.40%	+2.28pp	0.680
Drop 2008 and 2020	+11.47%	+11.74%	−0.27pp	0.717

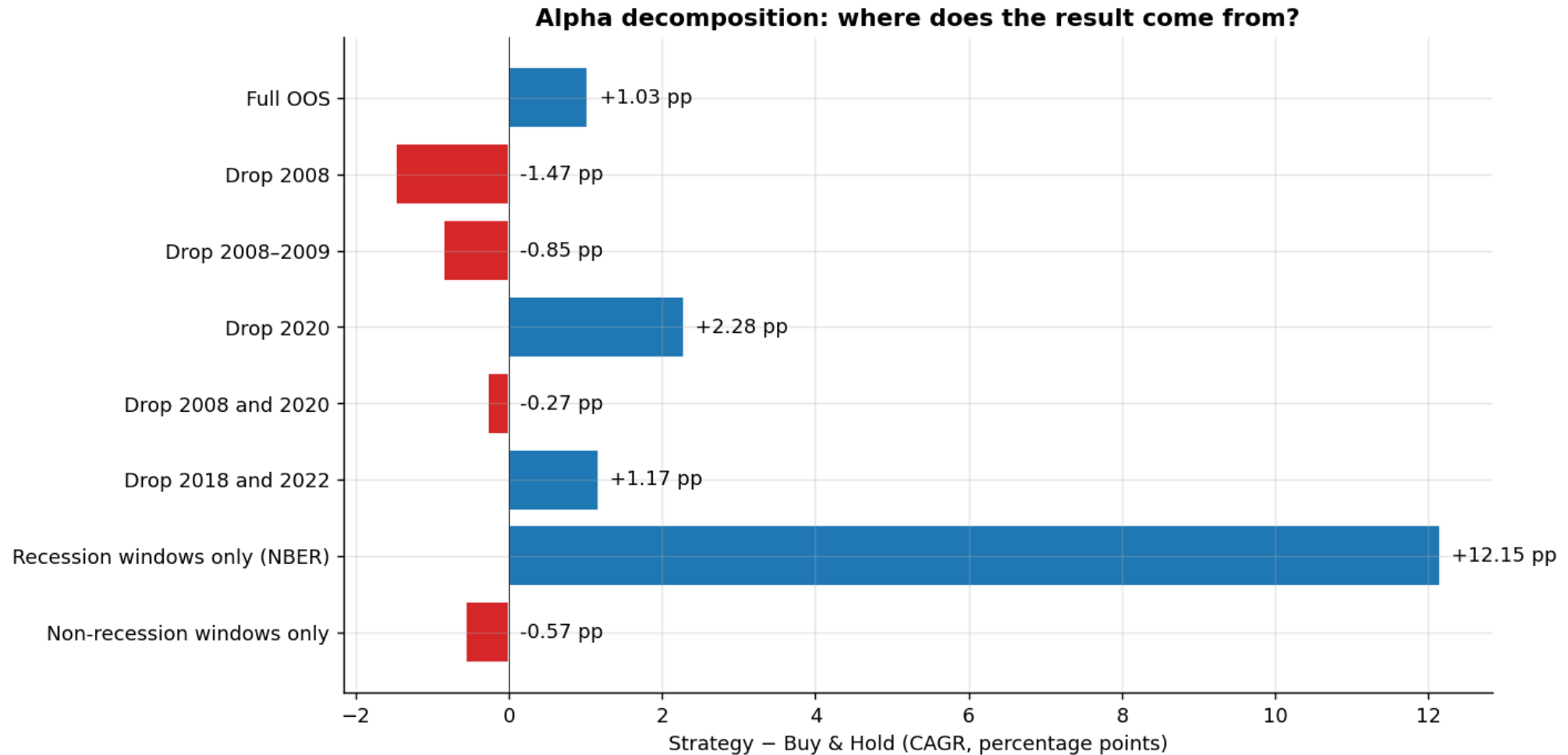
Without 2008, the overlay underperforms Buy & Hold by 1.5 pp/yr.

Without 2020, the overlay outperforms by 2.3 pp/yr.

Net 1.03 pp/yr alpha = 2008 (+35 pp) − 2009 (−13 pp) − 2020 (−22 pp) + small.

Robustness 5 · Insurance Decomposition

NBER-recession days vs non-recession days



Recession-only and non-recession-only rows are the key insurance evidence.

2020 — A Real Failure, Not Just Late

Annual alpha = -21.82 pp (strategy -6.7% vs B&H $+15.1\%$)

What happened

Labor signal triggered defensive position during March 2020 (correctly). But by the time the position re-risked, the Fed-supported rebound had run. The strategy captured the COVID drawdown but missed most of the recovery.

What this means

Slow-rebound regimes are a documented weakness — drop 2020 and the alpha rises to $+2.3$ pp/yr. The COVID episode is the single largest detractor in the OOS sample.

Earlier framing of "1-week lag" understated this; the annual attribution makes the magnitude visible.

Failure-Mode Taxonomy

What this strategy can and cannot do, by regime

Regime	Strategy behavior	Sample episode
Labor-led recession	Reduces exposure ahead, captures alpha	2008 (+35 pp)
Sudden exogenous shock	Reacts to labor lagged → misses rebound	2020 (−22 pp)
Inflation / Fed valuation drawdown	Stays invested (labor still healthy)	2022 (−1.4 pp)
Late-cycle correction	Stays invested (labor still healthy)	2018 (−1.6 pp)
Normal expansion	100% or 125% participation	2010s, 2024–25

The strategy targets labor-led recession risk specifically — not valuation compression or velocity shocks.

What This Strategy Is

The honest positioning

- Recession-insurance overlay — pays a small persistent cost in normal regimes for a large payoff in labor-led recessions
- Interpretable — every position change has an explicit labor-market explanation
- Cheap to monitor — three FRED series, weekly cadence
- Threshold-robust — fixed ($d=0.6$, $a=-0.5$) works as well as walk-forward learning
- Backfill-robust — sensitivity test confirms the 2008 result is not an artifact of the 5-day lag

What This Strategy Is NOT

Claims this study does NOT make

- Statistically distinguishable from Buy & Hold on Sharpe (one-sided $p = 0.18$; 21 OOS years insufficient)
- An all-weather alpha source — drop 2008 and the strategy underperforms B&H
- Robust to velocity shocks — 2020 alpha was -22 pp
- Free of upstream selection — only thresholds are walk-forward; indicator and composite choices are fixed ex ante
- Net of implementation cost — no transaction cost, cash yield, or financing-cost model yet

Comparison vs Alternative Approaches

Reference ranges, not apples-to-apples

Approach	Typical Sharpe	This strategy	Comments
Buy-and-hold S&P 500	~0.43	—	Benchmark (this study)
60/40 stock-bond	~0.45-0.55	—	Asset-class diversification
Trend-following CTA	~0.40-0.60	—	Regime-dependent
Labor-only overlay	—	0.586	Macro-event-driven (this study)
Vol-targeting	~0.55-0.65	—	Reacts to realized vol

Ranges shown are loose practitioner references, not controlled benchmarks. Honest comparison would build all five strategies on the same data, costs, and period — that exercise is in Next Steps.

Limitations

Seven caveats the audience should hold

Few recession episodes	21 OOS years contain 2 NBER recessions + 1 pandemic. Result is governed by 3 events.
2008 concentration	Drop 2008 → alpha goes negative (−1.5 pp/yr). The result is recession-event-dependent.
Statistical power	Sharpe difference vs B&H not significant (95% CI includes 0; one-sided $p = 0.18$).
Threshold-only walk-forward	Indicator selection, transforms, composite weights are fixed ex ante — not nested OOS.
Vintage data only partial	ALFRED vintage release dates used; vintage values (especially NFP, UNRATE) not.
No cost / financing / cash model	Frictionless backtest. 25-50 bps cost + 0% cash + leverage financing not modeled.
U.S.-only	Generalization to ex-US markets needs weekly labor data; mostly unavailable.

Conclusion — Five Narrower Claims

What the current evidence supports

- The labor-only overlay improves OOS risk-adjusted performance over 2006–2026 — point estimate Sharpe 0.586 vs 0.43.
- Most value comes from labor-led recessionary drawdown mitigation, especially 2008 (+35 pp alpha).
- Threshold calibration is stable: fixed ($d=0.6$, $a=-0.5$) matches walk-forward output.
- Pre-2009 backfill assumption is empirically robust under ± 5 -day perturbations.
- Sharpe improvement is not statistically significant on 21 years (one-sided $p = 0.18$).

Next Steps — Ranked by Decision Value

Before any capital is deployed

- Vintage value reconstruction for NFP / UNRATE / ICSA using ALFRED
get_series_all_releases — removes the final remaining look-ahead risk
- Nested walk-forward — let the outer loop also choose the strategy family from the 5 candidates
- Full implementation cost model — transaction cost, cash T-bill yield, SOFR + 50bps margin financing
- Test Core CPI and yield curve as conditional filters (not additive components) for 2022-type regimes
- Pre-registered paper-trading window (minimum 6 months) before any live capital

Questions & Discussion

Vincent Wang · Investment Strategy Report (v3, honest revision) · May 2026

Reproducibility: FRED API key + ~10 minutes of compute · Analysis code in `/analysis/`