Stock–Bond Diversification in an Inflation-Driven Rates Regime

Regime-Aware Allocation and Treasury Curve Loss Attribution, 2007–2026

Vatsal Maniar

Research Memo June 2026

Abstract

For most of the four decades to 2020, one assumption did the heavy lifting in balanced portfolios: when equities fell, Treasuries would rise and cushion the loss. 2022 broke it. Stocks and nominal Treasuries fell together, and the 60/40 portfolio had one of its worst years on record. The episode is better read as a regime than an accident, which raises two questions: when does the Treasury hedge stop working, and why does it fail when it does. For the first question, a multi-asset allocation backtest over 2007–2026 runs a rule-based, regime-aware allocation against Equal Weight, global minimum variance, maximum-Sharpe, and SPY and 60/40 benchmarks on thirteen liquid ETFs. VIX, one-month-lagged CPI, and the rolling SPY–TLT correlation enter only as stress signals, never as holdings. The results do not show that any one rule dominates. The regime-aware rule is transparent and conditional, but it does not beat 60/40 on risk-adjusted return in this sample; what it does is make hedge behavior explicit, and in 2022 adding duration deepened the drawdown while cash and real assets defended. For the second question, a Treasury curve risk engine reprices synthetic Treasury books under inflation-regime curve shocks and decomposes the loss into duration, convexity, curve-shape, real-rate, and breakeven channels. Replaying the measured 2022 curve move through a long-duration hedge book produces a loss of roughly 19%, attributed almost entirely to real-rate repricing rather than inflation breakevens, and rule-based hedge overlays recover under a third of it. None of this implies the regime is forecastable. The contribution is measurement and attribution, not prediction.

Keywords: stock–bond correlation; regime-aware allocation; inflation; Treasury curve risk; duration and convexity; real yields; P&L attribution; 60/40.

1 Introduction and motivation

The 60/40 portfolio is, underneath, a bet on a correlation. Sixty percent equities and forty percent intermediate Treasuries behaves like a single diversified portfolio only as long as the two legs do not fall together, and for most of the period from the late 1990s through 2020 they did not. Treasuries rallied through the dot-com unwind, through 2008, and through the COVID crash. That negative stock–bond correlation is the entire reason a balanced portfolio could carry less risk than the weighted sum of its parts.

2022 is the year the arrangement stopped holding. As the Federal Reserve lifted the funds rate from near zero to above four percent, both legs sold off at once. The S&P 500 fell roughly 18%, long Treasuries fell considerably more,[1] and a standard 60/40 had one of its worst calendar years on record. The lesson is not that bonds are a poor asset. It is that the specific property investors pay bonds to provide, a hedge against equity drawdowns, is conditional on the macro regime, and the inflation regime is precisely the state in which it is weakest.

The mechanism is not subtle. When a growth shock drives markets, as in 2008 and 2020, the Fed cuts, bonds rally, and duration hedges equities. When inflation and the policy response to it drive markets, as in 2022, the same rising discount rate hits stocks and bonds together and pushes them down in step. The stock–bond correlation is not a constant of nature; it tracks whether the dominant shock is to growth or to inflation (AQR, 2023). Treat it as fixed and you will misjudge your hedge in exactly the regime where the misjudgment is most expensive.

As of mid-June 2026 this is a live question again rather than a historical one; the figures in this paragraph carry that as-of date. Headline CPI re-accelerated to about 4.2% year over year in May 2026, its highest reading since early 2023, driven mainly by an energy spike tied to a Middle East supply shock, while core inflation was milder, near 2.9%.[2] The FOMC held the funds rate at 3.50–3.75% at its June 17, 2026 meeting, and its statement attributed elevated inflation in part to energy supply shocks and "the conflict in the Middle East" while restating a commitment to price stability.[3] The ten-year Treasury yield sat near 4.45%.[4] Beneath the level, the ten-year term premium has, by the New York Fed's ACM estimate, turned positive for the first time since 2023,[5] and large fiscal deficits run through most 2026 fixed-income outlooks as a reason the long end stays under pressure (J.P. Morgan Asset Management, 2026). The precise term-premium figure is model-dependent and the exact figure should be read with caution, but the direction is corroborated across sources.

The correlation itself has come down from its 2022–2024 peak. By one estimate, State Street's, the rolling 12-month stock–bond correlation fell back toward roughly 0.16 in late 2025, and the 60/40 produced strong returns over 2023–2025.[6] That figure is one data-provider's estimate and, like the

[1] Calendar-2022 total returns computed from the study's daily ETF return data (Yahoo Finance adjusted closes): SPY about −18%, the long-Treasury ETF TLT about −31%, and intermediate Treasuries (IEF) about −15%.

[2] U.S. Bureau of Labor Statistics, *Consumer Price Index — May 2026*, `https://www.bls.gov/news.release/cpi.htm`.

[3] Federal Open Market Committee statement, June 17, 2026, `https://www.federalreserve.gov/newsevents/pressreleases/monetary20260617a.htm`.

[4] FRED series DGS10, 10-Year Treasury Constant Maturity, `https://fred.stlouisfed.org/series/DGS10`.

[5] Federal Reserve Bank of New York, *Treasury Term Premia* (ACM model; Adrian, Crump, and Moench, 2013), `https://www.newyorkfed.org/research/data_indicators/term-premia-tabs`; the level is model-dependent.

[6] State Street Global Advisors, "Mind on the Market" (October 2025), `https://www.ssga.com/us/en/institutional/insights/mind-on-the-market-03-october-2025`; a single-provider, secondary estimate of the rolling 12-month correlation.

term-premium reading above, should be read cautiously rather than as a settled number. It points to easing, not structural reversion. The 2026 re-acceleration is the reminder: as long as inflation can surprise upward, the diversification a balanced portfolio is counting on is not guaranteed to be present in the drawdown when it is most needed.

That gap is what the analysis below addresses. An earlier version of this work proposed a "Dynamic Correlation-Adaptive Hedge" and asserted that it would reduce drawdowns. The current version builds and tests it against three benchmark rules, and the honest result is narrower and more useful than the proposal: no single rule dominates, but a portfolio's regime behavior can be made explicit, the point at which the bond hedge fails can be measured, and the reason it fails can be decomposed at the level of the Treasury curve.

2 Research questions

The original draft asked one broad question: can an adaptive hedge respond in real time to correlation reversals and inflation without relying on forecasts? The revision keeps that spirit but splits it into questions that data can actually answer.

First, is stock–bond diversification conditional on regime, and can a transparent, rule-based allocation that reads inflation, correlation, and volatility signals behave differently across regimes without forecasting them? This is the *when*: when does the hedge weaken.

Second, when the bond hedge does fail, does a regime-aware rule actually help relative to simple alternatives, Equal Weight, global minimum variance, maximum Sharpe, SPY, and 60/40, once realistic constraints, transaction costs, drawdown paths, and tail loss are all accounted for? This is the test of whether the rule earns its added complexity.

Third, why does the bond hedge fail in an inflation regime, mechanically? If you take the Treasury leg on its own and reprice it under the actual 2022 curve move, how much of the loss is plain duration, how much is the curve reshaping, and how much is real yields repricing rather than inflation expectations? This is the *why*.

The first two questions are the allocation study; the third is the rates-attribution extension. None of them is a forecasting question. The aim is not to predict the next inflation shock, but to measure and attribute hedge behavior once a regime is already underway.

3 Data and methodology

The allocation study uses daily adjusted-close ETF data from 2007 through June 2026, the common window over which all thirteen funds and the signal series exist. Because the estimation lookbacks consume the first window of that history, the backtest is evaluated from 2008 onward: references to 2007–2026 denote the data span, while the 2008–2026 figures in the results tables denote the evaluation window. Thirteen liquid ETFs stand in as transparent public-market proxies for the major asset classes: equities (SPY, QQQ, IWM, EFA, EEM), high-yield credit and listed real estate (HYG, VNQ), Treasuries across the curve plus a T-bill cash proxy (TLT, IEF, SHV), and inflation and real-asset hedges (TIP, GLD, DBC). ETF proxies are imperfect, which the limitations revisit, but they are liquid, daily, and reproducible, and that matters more for a study about process than about specific holdings.

Three signals feed the regime engine, and none of them is ever held. The rolling 90-day SPY–TLT return correlation measures whether the stock–bond hedge is currently working. CPI year over year measures the inflation regime. The VIX level measures volatility stress. Two corrections to the original design matter here. First, VIX is a signal only. Spot VIX is not directly investable, and volatility-linked products behave very differently from the index, so they are out of scope; the strategy never allocates to VIX. The earlier draft listed VIX as an allocation target, which was a category error. Second, CPI is monthly and released with a lag, so the backtest uses CPI year over year lagged one month. The strategy never trades on an inflation print before its public release.

Returns are compared against two benchmarks, 100% SPY and a 60/40 SPY/IEF portfolio, and across four allocation rules: Equal Weight, global minimum variance (GMV), maximum Sharpe, and the regime-aware rule. Equal Weight is the naive baseline, GMV and maximum Sharpe are the standard optimizers, and the regime-aware rule is the object of interest. All four run on identical data, the same monthly rebalance, the same constraints, and the same cost model, so differences in behavior trace to differences in inputs rather than to differences in plumbing.

Inputs (means, covariances, signals) are estimated on rolling six-month, one-year, and three-year lookbacks, reported primarily at one year. Portfolios rebalance monthly at the month-end trading day, are long-only, and are capped at 35% per asset. Each unit of turnover pays five basis points, measured against drifted pre-trade weights so that even Equal Weight pays realistic rebalancing costs. For each rule the study reports annualized return (geometric, from daily net returns), annualized volatility (daily scaled by $\sqrt{252}$), Sharpe and Sortino, maximum drawdown, 95% historical VaR and CVaR as daily loss magnitudes, monthly turnover and cumulative cost drag, and Fama–French five-factor exposures as a diagnostic.

Bias control is the piece the original lacked. At each month-end rebalance the model sees data only through the prior day; CPI is additionally lagged one month; new weights begin earning returns the next trading day. The research engine ships a validation ledger of seventeen automated checks, covering no-look-ahead, cost application, turnover, the VaR/CVaR convention, and the CPI lag, among others. It passes with one documented warning: common ETF history begins in 2007 (SHV and HYG inceptions), so the longest lookbacks do not fully cover 2008–09, and those crisis cells are marked *insufficient* rather than estimated; the gap is shown rather than papered over.

The Treasury curve engine of Section 8 uses a separate, fully offline pipeline built on public daily FRED series: nominal constant-maturity yields (2/5/10/30Y plus 3M), real yields (DFII), breakevens (T5YIE and T10YIE), and CPI (CPIAUCSL). Section 8 describes it in full. The methodological point common to both studies is that every number is generated by a reproducible Python pipeline from public data, with no live pricing in the browser.

4 Allocation-rule design

The original action rule reduced TLT "by x%" and allocated "y% to TIP, z% to GLD, w% to VIX," with every weight left undefined. Specifying those weights is what made the rule testable. The implemented version works in three steps: classify the regime, set sleeve targets, then fill the sleeves.

Three stress signals are evaluated at each rebalance, each against a threshold. The 90-day SPY–TLT correlation counts as stressed above +0.20, the level at which the hedge is visibly weakening. Lagged CPI year over year counts as stressed above 3%. VIX counts as stressed above 25. The thresholds are judgment calls, and the limitations treat them as such.

The number of simultaneously active signals defines the regime: zero active is *normal*, one is *watch*, two or more is *defensive*. Requiring several signals to agree before turning defensive is deliberate, because a single-indicator rule whipsaws. Across the one-year-lookback sample the book sits in normal about 61% of valid trading days, watch about 25%, and defensive about 14%, so the defensive state is selective rather than a permanent crouch.

The thirteen assets are grouped into three sleeves: growth/risk (the five equity funds plus HYG and VNQ), duration/cash (TLT, IEF, SHV), and inflation hedges (TIP, GLD, DBC). As stress rises, the growth target steps down and the defensive and hedge sleeves expand to fill the gap, as in Table 1. A cash floor in SHV of 5% rises to 15% when VIX is stressed. Within each sleeve, assets are inverse-volatility weighted, so noisier holdings receive less capital; the whole book is long-only, capped at 35% per asset, and rebalanced monthly net of the five basis-point cost.

Sleeve	Normal	Watch	Defensive
Growth / risk	65%	55%	45%
Duration / cash + inflation hedges	35%	45%	55%

Table 1: Regime-aware sleeve targets. Growth exposure (SPY, QQQ, IWM, EFA, EEM, HYG, VNQ) declines as the active-signal count rises; the duration/cash sleeve (TLT, IEF, SHV) and inflation-hedge sleeve (TIP, GLD, DBC) absorb the difference and are split by inverse-volatility weighting, with an SHV cash floor of 5% that rises to 15% under VIX stress. All weights are long-only and capped at 35% per asset.

The rule never forecasts the regime; it reacts to observed, lagged signals. Because every step from signal to weight is mechanical, the resulting allocation is fully traceable: any weight can be explained from that month's signal readings. That traceability, not outperformance, is the design goal.

5 Full-sample results

Table 2 reports the full-sample performance at the one-year lookback, net of transaction costs. The first thing to say is that no rule dominates on every dimension, and the most aggressive one is the most expensive to run.

Strategy	Ann. ret	Ann. vol	Sharpe	Sortino	Max DD	VaR 95	CVaR 95	Turn./mo	Cost drag
Equal Weight	6.6%	11.2%	0.51	0.83	−34.9%	1.02%	1.71%	2.6%	−0.28%
GMV	2.3%	3.4%	0.30	0.98	−11.3%	0.32%	0.50%	7.7%	−0.84%
Max Sharpe	5.9%	7.3%	0.64	1.12	−18.2%	0.74%	1.15%	37.4%	−3.98%
Regime-Aware	6.0%	10.5%	0.48	0.80	−33.3%	0.93%	1.63%	3.7%	−0.41%
SPY (bench)	11.7%	19.8%	0.59	0.84	−51.5%	1.81%	3.05%	0.0%	0.00%
60/40 (bench)	8.4%	11.1%	0.66	1.08	−31.9%	1.05%	1.70%	1.8%	−0.20%

Table 2: Full-sample risk and performance, one-year lookback, daily net returns, sample 2008–2026. VaR and CVaR are daily historical 95% loss magnitudes; drawdowns are negative. Source: portfolio-lab research engine.

The trade-offs are worth stating plainly rather than spinning. SPY earns the highest raw return, 11.7% annualized, and pays for it with the deepest drawdown in the sample, −51.5%. GMV does the opposite: it cuts volatility to roughly a third of SPY's and holds its drawdown to −11.3%, but surrenders most of the upside at 2.3% annualized. Maximum Sharpe posts the best Sharpe ratio among the strategies, 0.64, but it runs about 37% turnover a month, which annualizes to roughly 449% and produces the largest cost drag in the set at −3.98%; the ratio is real but expensive and, as Section 7 shows, sensitive to the lookback. Equal Weight earns the highest cumulative return of any strategy over the sample, about +216% (6.6% annualized), while still living through a −34.9% drawdown.

The regime-aware rule earns 6.0% annualized at a 0.48 Sharpe and a −33.3% drawdown. On a risk-adjusted basis it does not beat 60/40, which posts the highest Sharpe of the entire set at 0.66. That is the honest finding, stated plainly: 60/40 is a genuinely hard benchmark, not a straw man, and a static allocation can be hard to beat in a sample where equities did well for long stretches. What the regime-aware rule offers is not a higher Sharpe but a drawdown and an exposure path that are explainable from its inputs, which the next two sections develop. Allocation choice trades return against drawdown, tail risk, and implementability, and in the maximum-Sharpe case the noise in estimated returns converts almost directly into trading cost.

6 Crisis-window analysis

Full-sample averages hide the fact that different crises break different assumptions. Three windows make the point, and they do not agree with each other, which is the finding.

2008, where duration worked. The Global Financial Crisis was a credit and growth shock. Equity, credit, real estate, and commodities sold off together while Treasuries rallied hard, and the Fed cut. Table 3 shows the result over the window covered by the six-month lookback. The bond-heavy defensive allocations held up far better than equities: GMV actually finished the window up 4.3% with an −8.7% drawdown, 60/40 lost 7.7%, and SPY lost 20.1% on its way to a −51.9% trough. This is the regime in which the textbook Treasury hedge does exactly its job. The honest caveat is that common ETF history begins in 2007, so only the six-month lookback fully covers 2008; longer windows are marked insufficient rather than estimated.

Strategy	Crisis ret	Vol	Max DD	CVaR 95	Worst day	vs SPY
Equal Weight	−2.0%	20.5%	−34.9%	3.03%	−5.29%	+18.1%
GMV	+4.3%	4.9%	−8.7%	0.65%	−1.58%	+24.4%
Max Sharpe	−4.1%	13.9%	−26.4%	1.94%	−6.20%	+16.1%
Regime-Aware	−3.0%	19.4%	−33.5%	2.99%	−5.17%	+17.2%
SPY (bench)	−20.1%	34.8%	−51.9%	5.27%	−9.84%	0.0%
60/40 (bench)	−7.7%	18.8%	−31.9%	2.80%	−5.25%	+12.4%

Table 3: 2008 GFC window, six-month lookback, daily net returns within the window. "vs SPY" is the crisis-return difference against the 100% SPY benchmark.

2020, where a reactive rule cannot keep up. COVID was a fast crash and a faster rebound. A rule that classifies the regime from lagged, observed signals turns defensive only after the drop has happened and re-risks only after the recovery is underway, so it lags in both directions. 2020 is not a win for the regime-aware rule, and the data does not support one; the window mainly exposes a general limit of reactive rules. They can shape exposure across a regime that persists for months, but they cannot help with a shock that is largely over in weeks.

2022, where the bond hedge failed. 2022 is the regime at the center of this analysis, and it is the one where the standard hedge broke. As the stock–bond correlation turned positive, average pairwise correlation across the book rose, the correlation among growth assets reached about 0.79, the number of effective independent bets collapsed to roughly 3.3 of 13 holdings, and a single common factor explained about 49% of cross-asset variance. The most direct evidence sits in a hedge-overlay test: applying a 10% sleeve on top of a 90% base, broad commodities (DBC) and cash (SHV) reduced the 2022 drawdown the most, by about 3.5 and 2.0 percentage points respectively on a 60/40 base, while adding long Treasuries (TLT) made the drawdown *worse*, by about 1.0 point on 60/40 and 1.6 points on the regime-aware base. That is the bond hedge failing, measured directly: the asset that is supposed to protect the portfolio added to the loss. The allocation study can show that this happened. It cannot, on its own, explain why, which is the reason for the Treasury curve engine in Section 8.

7 Diagnostics and implementability

A backtest can look attractive and still describe a portfolio nobody could run. The diagnostics in this section are how to tell the difference.

Drawdown paths and tail loss. Volatility measures dispersion; drawdown shows the actual path of losses an investor would have to sit through. Two rules with similar annualized returns can live very different lives. From Table 2, SPY's worst single day in the sample is −10.9% against GMV's −1.7%, and SPY's maximum drawdown of −51.5% dwarfs GMV's −11.3%. A rule with an attractive endpoint but a punishing path may be unusable in practice, because the question is not only where the portfolio ends but whether its owner stays invested through the middle.

Turnover, concentration, and cost. Table 4 is where the maximum-Sharpe optimizer stops looking impressive. It holds about 3.6 effective positions, turns over roughly 449% a year, and bleeds −3.98% in cumulative cost drag, against the regime-aware rule's 5.6 effective positions, 45% annual turnover, and −0.41% drag. The 35% per-asset cap binds for both optimizers, which is the cap doing its job: without it they would concentrate further into whatever looked calm over the lookback. Equal Weight, by construction, stays the most diversified at 13 effective positions and the cheapest to run.

Strategy	Eff. positions	Max single weight	Ann. turnover	Cum. cost drag
Equal Weight	13.0	7.7%	31%	−0.28%
GMV	3.7	35.0%	93%	−0.84%
Max Sharpe	3.6	35.0%	449%	−3.98%
Regime-Aware	5.6	35.0%	45%	−0.41%

Table 4: Concentration and implementability, one-year lookback, 2008–2026 evaluation sample. Effective positions $= 1/\sum_i w_i^2$. Each unit of turnover costs five basis points.

Factor exposure, as a diagnostic. Table 5 regresses daily excess returns on the Fama–French five factors. This is an exposure diagnostic, not a skill test. The regime-aware rule carries a market beta of 0.49, well below SPY's 1.00 and below 60/40's 0.55, with an R^2 of 0.93, so most of its behavior is simply lower equity beta, which is what explains its drawdown profile before anything else. The maximum-Sharpe intercept of +3.3% annualized ($t = 2.1$) is the one number a reader might be tempted to call alpha. It is not. The regression is equity-factor-only while the strategy holds bonds, TIPS, gold, commodities, credit, REITs, and cash, so the intercept is residual return under a model that is deliberately incomplete for a multi-asset book, and it is sensitive to the lookback and the cost assumptions. It is a diagnostic, not proof of alpha.

Strategy	Intercept (ann.)	Mkt β	SMB	HML	RMW	CMA	R^2
Equal Weight	−0.1%	0.49	0.07	0.01	−0.03	0.02	0.84
GMV	+0.1%	0.08	0.00	−0.03	−0.01	0.00	0.20
Max Sharpe	+3.3%	0.14	0.01	−0.11	−0.07	0.06	0.17
Regime-Aware	−0.7%	0.49	0.05	0.03	−0.01	0.00	0.93
SPY (bench)	−0.4%	1.00	−0.11	0.01	0.07	0.05	0.99
60/40 (bench)	+0.7%	0.55	−0.04	−0.03	0.04	0.02	0.93

Table 5: Fama–French five-factor exposures, one-year lookback, daily regressions over the 2008–2026 evaluation sample. The intercept is residual return under an equity-only model, not evidence of skill.

Lookback sensitivity. Rankings shift with the six-month, one-year, and three-year estimation window, and maximum Sharpe is the most sensitive: its weights jump between assets month to month, which is the direct source of its turnover. That instability is the lesson, not a defect to be tuned away. Taken together, the diagnostics make the same point in different ways: constraints, caps, and cost-awareness are what separate a paper portfolio from one that could actually be held, and a strategy must be judged net of the cost of running it.

Threshold sensitivity. The regime thresholds (+0.20 correlation, 3% CPI, VIX 25) are judgment calls, so it is worth checking how far the results depend on them. Table 6 varies each threshold one at a time and re-runs the regime-aware backtest. The classification is genuinely sensitive: loosening CPI to 2.5% or VIX to 20 roughly doubles the share of time spent outside the normal regime, dropping the normal share from 61% to 50%, while tightening either pushes more days back to normal. The realised outcomes barely move. Full-sample maximum drawdown holds at −33.3% and the 2022 within-window drawdown at about −16% across every setting, annualised return stays near 6.0%, and only turnover and cost drag respond, rising as looser thresholds trade more often. The reason is mechanical: the deepest drawdowns, 2008 and 2022, are classified defensive under essentially every threshold in the sweep, and the sleeve shift is modest enough (growth steps from 65% to 45%) that reclassifying watch against defensive days does not change the peak-to-trough path. In this sample the thresholds move how often the rule trades more than how much it protects, which is reassuring for the robustness of the headline conclusions but is also a reminder of how much a conditional sleeve rule can actually do.

Threshold setting	Normal	Watch	Defensive	Max DD	2022 DD	Ann. turn.	Cost drag
Baseline (0.20 / 3.0% / 25)	61%	25%	14%	−33.3%	−15.9%	45%	−0.41%
Correlation > +0.10	58%	26%	16%	−33.3%	−15.9%	45%	−0.40%
Correlation > +0.30	61%	29%	9%	−33.3%	−15.9%	45%	−0.40%
CPI > 2.5%	50%	35%	15%	−33.3%	−15.9%	50%	−0.45%
CPI > 3.5%	65%	25%	9%	−33.3%	−15.9%	46%	−0.41%
VIX > 20	50%	32%	18%	−33.3%	−15.9%	49%	−0.44%
VIX > 30	66%	23%	11%	−33.3%	−16.0%	43%	−0.39%

Table 6: Threshold sensitivity of the regime-aware rule, one signal varied at a time, one-year lookback, 2008–2026 evaluation sample. Regime shares are over valid strategy trading days; "2022 DD" is the within-window drawdown for January–December 2022. Annualised return is 6.0–6.1% in every row. Source: portfolio-lab research engine.

8 From allocation failure to Treasury curve risk attribution

The allocation study establishes that the bond hedge failed in 2022, in the sense that adding duration deepened the drawdown. But it treats TLT as a single line that went down. It cannot say *why*. To answer that, the Treasury leg is taken out of the portfolio entirely and run through a dedicated rates-risk engine that reprices Treasury books under curve shocks and decomposes the resulting P&L. Portfolio-lab asks when the bond hedge fails; this asks why. The section is an extension, not a second paper, so it stays focused on the one result that matters.

The engine works from a June 11, 2026 curve snapshot built from public FRED series: a 2Y near 4.05%, a 10Y near 4.45%, a 30Y near 4.95%, a real 10Y near 2.16%, and breakevens near 2.3%. It prices six synthetic Treasury-style instruments, a cash bill, 2/5/10/30Y nominal proxies, and a 10Y TIPS-style proxy, with standard analytics: clean price, Macaulay and modified duration, convexity, DV01, and key-rate DV01 from one-basis-point pillar bumps. Modified duration runs monotonically from 1.90 at the 2Y to 15.73 at the 30Y, as it should. These are synthetic proxies, not CUSIP-level securities, a point the limitations return to.

The headline scenario is not stylized. It is the measured nominal and real curve change from the last business day of 2021 to the last of 2022. Over that window the 2Y rose 368 basis points and the 30Y 207, a bear-flattening, and almost all of that move was in real yields rather than inflation expectations: real yields rose roughly 327 basis points at the front and 211 at the long end, while breakevens barely moved and even fell slightly at intermediate tenors.

Push that curve through the long-duration hedge book, the classic 10Y/30Y-heavy equity-hedge book with a modified duration of 9.9 and a DV01 of about $987 per $1m, and the repriced loss is about $192,000, or 19.2% of the book. A duration-only approximation would have put it near $267,000 and a convexity-adjusted one near $203,000, so the duration-only shortcut is off by roughly $76,000, about 40% of the loss. These are model outputs under the synthetic-proxy assumptions noted below, so the channel split matters more than the dollar figure to the last digit. The direction of that error, an overstatement here because convexity is large and positive on a move this size, matters less than its magnitude: a single-number DV01 view of a long book mis-measures a large curve move by tens of thousands of dollars and says nothing about where the loss came from.

The channel ledger is what supplies the where. The engine splits the exact P&L additively into duration, convexity, curve-shape, real-rate, and breakeven channels plus a residual line, and the pieces reconcile to the exact loss to within a cent (Table 7). For this book under the 2022 replay the decomposition is unambiguous: the real-rate channel accounts for about −$278,000, convexity and curve-shape offset roughly $64,000 and $37,000 of it, the breakeven channel contributes only about +$11,000, and a −$26,000 residual closes the reconciliation. In words, the loss was a real-rate event, not an inflation-expectations event. That is the reason the "safe" hedge failed. A long-duration Treasury book is effectively a leveraged position in real yields, and an inflation regime is exactly when real yields reprice.

Channel	P&L ($, $1m book)
Duration	0
Convexity	+64,255
Curve-shape	+37,120
Real-rate	−278,010
Breakeven	+10,616
Residual	−25,614
Exact repriced P&L	−191,634

Table 7: Additive P&L attribution for the long-duration hedge book under the measured 2022 curve replay ($1m book). Channels reconcile to the engine's repriced loss to within a cent; the figures are an internal reconciliation under the stated synthetic-proxy assumptions, not a claim of real-world precision. The real-rate channel dominates and breakeven is near zero. Source: inflation-lab rates engine.

The engine then tests four rule-based hedge overlays. The strongest, a duration-reduction overlay that cuts the 10Y/30Y sleeve and rebuilds the front end, recovers 28.9% of the loss and leaves a residual of −$136,000. A TIPS-rotation overlay recovers only 3.2%, which is the correct answer for a reason worth stating: rotating into TIPS hedges breakeven risk, and 2022 was a real-rate shock, so it barely helps. No overlay removes the loss; residual risk always remains, and the engine is built to display it rather than hide it.

To avoid over-reading a single historical episode, the engine frames the deterministic shocks against a stochastic distribution: 10,000 simulated one-year short-rate paths under a Vasicek and a CIR model, with CIR satisfying the Feller condition, mapped to the curve and run through each book for one-year VaR and expected shortfall. The long-duration book's one-year 95% VaR is about $61k under Vasicek and $106k under CIR; the short-duration book's is roughly half that under each. Long duration carries more rates risk than short duration under both models, and the fatter-tailed CIR places the number materially higher, which is a useful reminder that these are model-conditioned estimates rather than loss bounds. The contribution of the rates engine is narrow and specific: it makes the 2022 hedge failure mechanical and attributable. Duration alone misses it, and the loss is a real-rate story.

9 Limitations

Every result here is conditional on its assumptions, so they are stated here rather than buried.

On data and proxies, the ETFs are imperfect stand-ins for their asset classes, and common ETF history begins in 2007 (SHV and HYG inceptions), which restricts the 2008–09 analysis to the shortest lookback. CPI is monthly, released with a lag, and revised; the backtest lags it one month but cannot undo later revisions. Adjusted closes can be restated.

On signals and thresholds, VIX is used as a stress signal only and is never held. The +0.20 correlation, 3% CPI, and VIX-25 thresholds are judgment calls; different values would move the regime boundaries, and the rule is tested on the same history that motivated it, so some degree of in-sample fit is unavoidable.

On method and cost, the five-basis-point linear cost model ignores spreads, market impact, capacity, and taxes. Backtests are historical simulations under stated assumptions, not forecasts. VaR and CVaR are historical and model-conditioned estimates, not worst-case bounds. Factor regressions diagnose exposure, not skill, and the intercept is residual return under an equity-only model rather than alpha.

On the rates engine, the Treasury instruments are synthetic proxies, not CUSIP-level securities or live prices. The TIPS sleeve reprices off the real curve only and does not model index-ratio cash flows. The stylized shocks are analytical stress tests; only the 2022 replay is a measured episode. Vasicek and CIR are stylized short-rate models, not full term-structure models, the short-rate-to-curve mapping is a documented approximation, and carry and roll-down over the horizon are excluded. Hedge overlays reduce the exposures they target but always leave residual risk.

The summary is short: backtest-attractive is not implementable, and implementable is not forecastable. The work measures and attributes; it does not predict.

10 Conclusion

The question this work started from was whether static 60/40 diversification is safe in an inflation regime. The answer it reaches is that diversification is conditional, and the condition can be measured. A regime-aware allocation does not beat the benchmarks in this sample, and 60/40 in particular remains hard to beat on a risk-adjusted basis, but it makes a portfolio's hedge behavior explicit and traceable rather than assumed. And when the bond leg is taken on its own and the 2022 curve is repriced through it, the failure is not mysterious: a long-duration Treasury book lost about 19%, almost all of it to real-yield repricing, and duration-only risk math both mis-sizes that loss and hides its source.

The progression is the point. A writing sample on inflation-driven correlation breakdown became an allocation backtest that asked when the hedge fails, which raised a question the allocation study could not answer, why, which in turn became a Treasury curve attribution engine. None of it forecasts the next regime. The contribution is measurement: making the failure mode visible, conditional, and attributable, which is the necessary first step before deciding what to do about it.

References

Adrian, T., Crump, R. K., and Moench, E. (2013). Pricing the Term Structure with Linear Regressions. *Journal of Financial Economics*, 110(1), 110–138. (New York Fed ACM term-premium model.)

AQR Capital Management (2023). A Changing Stock–Bond Correlation: Drivers and Implications. *Journal of Portfolio Management*.

Cboe Global Markets. *Cboe Volatility Index (VIX): Methodology and White Paper.* cboe.com/tradable_products/vix/.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2), 385–407.

Fama, E. F., and French, K. R. (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, 116(1), 1–22. Factor data: Kenneth R. French Data Library, Dartmouth.

Federal Open Market Committee (2026). *FOMC Statement, June 17, 2026.* Board of Governors of the Federal Reserve System.

Federal Reserve Bank of St. Louis (FRED). Series CPIAUCSL, CPILFESL, DGS2, DGS5, DGS10, DGS30, DGS3MO, DFII5, DFII10, DFII30, T5YIE, T10YIE.

J.P. Morgan Asset Management (2026). *Guide to the Markets.*

State Street Global Advisors (2025). *Mind on the Market: the 60/40 and stock–bond correlation.* Market commentary; cited as a secondary estimate.

Tuckman, B., and Serrat, A. *Fixed Income Securities: Tools for Today's Markets.* Wiley. (Reference for duration, convexity, key-rate, and curve risk.)

U.S. Bureau of Labor Statistics (2026). *Consumer Price Index — May 2026.* USDL news release.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5(2), 177–188.

Appendix: implementation artifacts

The two studies are live, reproducible projects; every figure in this paper is generated by their Python research engines from public data.

Portfolio Risk & Allocation Analytics (allocation study).
Site: `pra-analytics.netlify.app` Source: `github.com/VM25/portfolio-lab`

Inflation Regime Rates Risk Engine (Treasury curve attribution).
Site: `yieldshock.netlify.app` Source: `github.com/VM25/inflation-lab`