

Stock–Bond Diversification in an Inflation-Driven Rates Regime

AVTORJI: Vatsal Maniar

POVZETEK: Raziskava obravnava, kako se odnose naštevanih akcij in zavodov zavira v okviru inflacije vpliva na cene zavodov. V 2022 je odnosa med akcijami in dolgoročnimi zavodi spremenila, zato je 60/40 poročilo enega od najslabših let. Raziskava opisuje, da je ta odnos regime-zavisen in da je pravila-odvisna prilagajalna strategija, ki je testirana od leta 2007 do 2026, ne izpostavljena standardnemu 60/40 poročilu. Ugotovljeni razlog je, da je glavni vpliv na odnos realne stopnje oprostita, nisem inflacijskih breakevens.

METODA: Raziskava je testirala pravila-odvisno prilagajalno poročilo, ki je uporabljala lagirane, vidljive signalje, podpisane z transakcijskimi stroški in omejitvami. Poročilo je preverjeno proti standardnim merilom od leta 2007 do 2026.

KLJUČNE UGOTOVITVE:

- Odnos med akcijami in zavodoma je regime-zavisen.
- Pravila-odvisna prilagajalna strategija ne je izpostavljena standardnemu 60/40 poročilu.
- Glavni vpliv na odnos je realna stopnja oprostita, nisem inflacijskih breakevens.

POMEN ZA TRGOVANJE: Raziskava pokazuje, da je pravila-odvisna prilagajalna strategija težko izpostavljena standardnemu 60/40 poročilu, če je uporabljena z lagiranimi, vidljivimi signalji in omejitvami. To pomeni, da je ključno razumeti, kako se odnosi med akcijami in zavodoma spreminja v različnih ekonomskih okolijih.

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

Balanced portfolios rely on Treasuries rising when equities fall. In 2022 that relationship broke down: stocks and long-duration Treasuries fell together, and a 60/40 portfolio had one of its worst years on record. That stock–bond diversification is regime-dependent is well established (Campbell et al., 2026), and this memo does not revisit it. Instead it asks whether acting on that fact helps in practice. Restricted to lagged, observable signals and subject to realistic transaction costs and position limits, a rule-based regime-aware allocation is tested over 2007–2026 against standard benchmarks; it does not outperform a static 60/40 on a risk-adjusted basis (Sharpe 0.48 versus 0.66). The memo then examines why the hedge fails: a Treasury curve model reprices the 2022 move and attributes a long-duration book’s roughly 19% loss primarily to rising real yields rather than inflation breakevens. The emphasis is measurement and attribution rather than 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, nor that inflation alone determines the hedge. It is that the property investors pay bonds to provide, a hedge against equity drawdowns, is conditional on the macroeconomic environment: on whether the dominant shock is to growth or to inflation, on the monetary-policy response, and on how real rates, inflation compensation, and risk premia reprice. An inflation-driven rate shock is one state in which the hedge is weak, but the map from “inflation” to “hedge failure” is not mechanical.

This is not a new observation. A substantial literature establishes that stock–bond comovement varies with the macroeconomic environment, and recent work decomposes nominal bond–stock comovement into real and inflation-related components whose relative importance shifts over history (Campbell et al., 2026, 2020; AQR Capital Management, 2023). The intuition is familiar: 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, a common rising discount rate can push stocks and bonds down together. The correlation is not a constant of nature, and treating it as fixed misjudges the hedge in exactly the states where the misjudgment is most expensive. Figure 1 shows the pattern directly: the 90-day rolling SPY–TLT return correlation is strongly negative through the 2010s, then rises into positive territory during the 2022–23 inflation episode, peaking near +0.41.

As of mid-June 2026 the question is live again rather than historical. 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% (U.S. Bureau of Labor Statistics, 2026). The FOMC held the funds rate at 3.50–3.75% at its June 17, 2026 meeting, attributing elevated inflation in part to energy supply shocks and “the conflict in the Middle East” while restating a commitment to price stability (Federal Open Market Committee, 2026). The ten-year Treasury yield sat near 4.45% (Federal Reserve Bank of St. Louis, 2026b). 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 (Federal Reserve Bank of New York, 2026; Adrian et al., 2013) – a model-dependent figure, but one corroborated in direction – and large fiscal deficits feature in most 2026 fixed-income outlooks as a reason the long end stays under pressure (J.P. Morgan Asset Management, 2026).

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 (State Street Global Advisors, 2025). That figure is one data-provider’s estimate and, like the 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.

¹Calendar-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%.

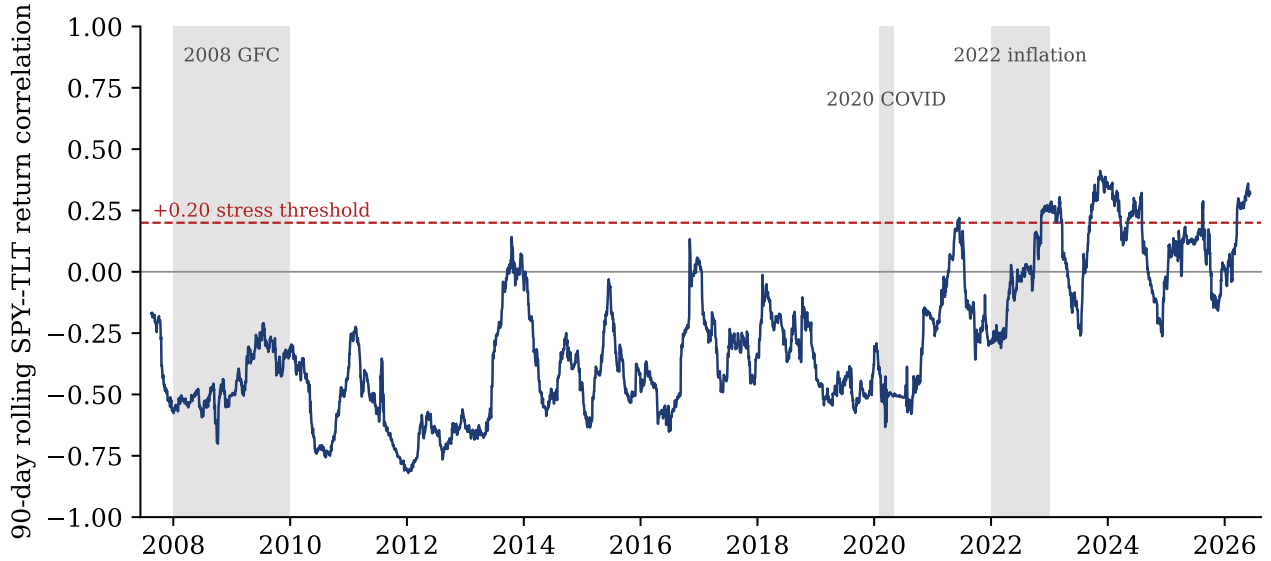


Figure 1: Rolling 90-day SPY–TLT return correlation, 2007–2026, with the three crisis windows shaded and the +0.20 regime-stress threshold marked. The hedge relationship is strongly negative through the 2010s (a 2010–2019 average near -0.43), then rises during the 2022–23 inflation episode, breaching the stress threshold and peaking near $+0.41$ in late 2023. It is the direct motivation for the paper: the diversification a 60/40 portfolio relies on is regime-dependent, not constant. Generated from the portfolio-lab signal series (`signal_data.csv`).

The question this memo takes up is therefore not *whether* stock–bond diversification is regime-dependent – the literature settles that – but how that established economics manifests once an investor is constrained to observable, lagged information, realistic trading frictions, and the drawdown path and portfolio-level Treasury hedge-book P&L implied by the realized curve move. 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 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 weakens can be measured, and the realized 2022 loss 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, how does the established regime dependence of stock–bond diversification manifest in an implementable multi-asset portfolio when allocation decisions use only observable, lagged signals – inflation, correlation, and volatility – and are subject to realistic costs and constraints? This is the *when*: when, and how visibly, does the hedge weaken in a portfolio an investor could actually run.

Second, does reacting to observable indicators of deteriorating stock–bond diversification improve portfolio outcomes relative to Equal Weight, global minimum variance, maximum Sharpe, SPY, and static 60/40, once transaction costs, constraints, drawdown paths, and tail loss are all accounted for? This is the test of whether the added complexity earns its keep; the answer reported below is that it

does not clearly, and in particular the regime-aware rule does not beat 60/40 on full-sample Sharpe.

Third, for the 2022 episode specifically, what does a portfolio-level Treasury curve attribution imply about the sources of the realized hedge loss – how much traces to real-rate repricing, how much to inflation compensation and breakevens, and how much to curve shape? This is the *why*, and it is deliberately a historical case attribution rather than a claim about a universal Treasury-hedge failure mechanism.

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 Related literature

This memo sits downstream of an established literature on stock–bond comovement, and it does not claim any of that economics as new. The first result it relies on is that the stock–bond correlation is not structural. Campbell et al. (2017) show the bond–stock covariance moved from positive in the early 1980s to negative in the 2000s, Campbell et al. (2020) tie that switch to the changing correlation between inflation and the output gap – whether the dominant shock is to growth or to inflation – and AQR Capital Management (2023) reach the same practitioner conclusion, that the diversification benefit of Treasuries is regime-dependent and weakened as inflation returned in 2021–2022. The allocation study here takes that regime dependence as given and as the object of *measurement*, not as a hypothesis to establish.

A more recent strand decomposes nominal bond–stock comovement into its real and inflation-related components and studies how their relative importance moves over time. Campbell et al. (2026) document that nominal government bonds and stocks comoved positively in the late twentieth century and negatively in the first quarter of the twenty-first, that a similar sign switch appears in the comovement of real bonds and breakeven inflation, and that periods of high risk premia are periods in which bond–stock comovements are large in absolute value. That work bears directly on this memo, because the distinction between real-rate and inflation-compensation channels – and the historical variation in which one dominates – is exactly what the Treasury attribution below measures for a single episode. The memo therefore does not introduce that distinction; it borrows it and applies it at the level of a hedge book’s realized P&L. The term-premium reading used in the introduction follows the Adrian et al. (2013) decomposition.

On the mechanics, the Treasury attribution uses standard fixed-income analytics – duration, convexity, and key-rate measures follow Tuckman and Serrat (2022), and the decomposition of curve moves into a few factors traces to Litterman and Scheinkman (1991) – and the short-rate simulations use the canonical Vasicek (1977) and Cox et al. (1985) models as stylized rates-risk distributions, not forecasts. On the allocation side, Ang and Bekaert (2002) solve dynamic allocation when correlations switch with a hidden state and find that regime structure matters most when correlations rise in bad times; the rule tested here is a deliberately transparent, threshold-based analogue that reacts to observed, lagged signals rather than estimating a latent state. Factor diagnostics use Fama and French (2015), strictly as an exposure check.

Against that backdrop the contribution is explicitly applied rather than structural. The memo does not claim to introduce the regime dependence of stock–bond covariance, the decomposition of nominal bond risk into real and inflation components, or a new theory of stock–bond comovement; those mechanisms are established, most recently by Campbell et al. (2026). It instead asks how those insights appear in an implementable portfolio-risk setting, by (i) translating the established regime-dependent

economics into a cost- and constraint-aware multi-asset allocation experiment, (ii) testing whether reacting to observable regime signals actually improves outcomes after costs and constraints, (iii) reporting the negative benchmark result rather than optimizing it away, (iv) taking the realized 2022 Treasury hedge failure into a portfolio-level curve-repricing framework, and (v) expressing that failure as portfolio-level hedge-book P&L attribution. The intended contribution is applied measurement and diagnostic integration, not a new structural account of stock–bond comovement.

4 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 (Cboe Global Markets, 2024). 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 (Fama and French, 2015; French, 2026) 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 inception), 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 9 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) (Federal Reserve Bank of St. Louis, 2026a). Section 9 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.

5 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.

6 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 8 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.

7 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.

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

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.

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 9.

8 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 path of losses an investor must sit through, and two rules with similar annualized returns can live very different lives. From Table 2, SPY’s worst day is −10.9% against GMV’s −1.7%, and SPY’s −51.5% maximum drawdown dwarfs GMV’s −11.3%. An attractive endpoint reached through a punishing path may be unusable, because what matters is whether the 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% ($t = 2.1$) is the one number a reader might 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 deliberately incomplete model and is sensitive to lookback and costs. 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 window, and maximum Sharpe is the most sensitive: its weights jump between assets month to month, the direct source of its turnover. That instability is the lesson, not a defect to be tuned away.

Threshold sensitivity. The +0.20 correlation, 3% CPI, and VIX-25 thresholds are judgment calls, but varying each one at a time moves the regime classification far more than the conclusions: loosening CPI or VIX roughly doubles the share of days outside the normal regime, so the normal share ranges 50–66%, and turnover and cost rise, yet the full-sample maximum drawdown holds at −33.3% and the 2022 drawdown near −16% in every setting. In this sample the thresholds change how often the rule trades more than how much it protects. The full one-at-a-time sweep is reported in the replication package (`scripts/threshold_sweep_results.csv`) and in the full version of this paper.

Taken together, the diagnostics make one point: 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.

9 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, modified duration 9.9, DV01 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. The dollar precision is not the point (these are model outputs under the synthetic-proxy assumptions below); the point is that 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 6). For this book under the 2022 replay the decomposition is dominated by the real-rate channel: it 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. Within this portfolio-level attribution specification, and for this synthetic long-duration book under the stated curve and attribution assumptions, the measured 2022 replay is dominated by real-yield repricing rather than breakeven inflation. That is a statement about the decomposition of this book’s realized P&L, not a causal identification of why the hedge failed in general: a long-duration Treasury book behaves like a levered position in real yields, so a curve move concentrated in real yields necessarily registers here as a real-rate loss. Whether real-rate dominance generalizes beyond 2022 is a question a single-episode attribution cannot settle, and the research extension below takes it up.

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 primarily addresses inflation-compensation exposure, while the measured 2022 replay is dominated by real-rate repricing under this attribution, so the overlay provides little protection. No overlay removes the loss; residual risk always remains, and the engine is built to display it rather than hide it.

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 6: 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.

To avoid over-reading one episode, the engine frames the deterministic shocks against a stochastic distribution: 10,000 simulated one-year short-rate paths under Vasicek and CIR (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, roughly double the short-duration book’s under each; the fatter-tailed CIR sits materially higher, a reminder that these are model-conditioned estimates, not loss bounds. The rates engine’s contribution is narrow: it makes the 2022 hedge failure mechanical and attributable – duration alone misses it, and under this attribution specification the loss is dominated by the real-rate channel.

10 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.

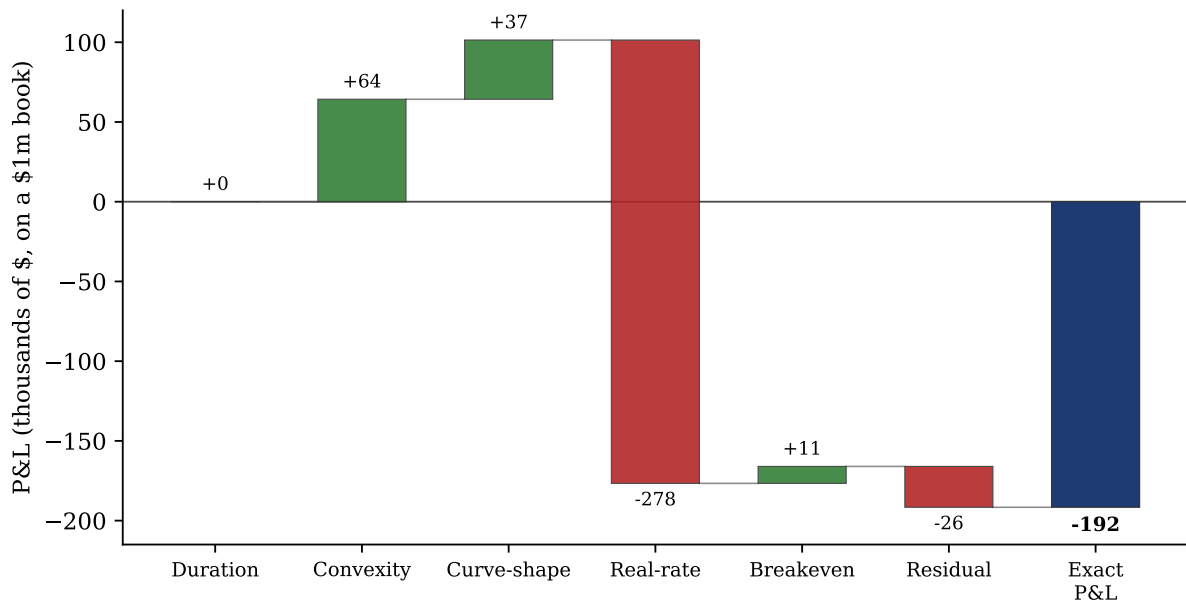


Figure 2: Additive P&L attribution for the long-duration hedge book under the measured 2022 curve replay, on a \$1m book. Green bars add to P&L and red bars subtract; the channels reconcile exactly to the repriced loss of about $-\$192\text{k}$ (-19.2%). The loss is dominated by the real-rate channel ($-\$278\text{k}$), while the breakeven (inflation-expectations) channel is small and slightly positive ($+\$11\text{k}$). Within this attribution specification, the measured 2022 replay is dominated by real-rate repricing rather than the breakeven channel. Generated from the inflation-lab engine (`scenario_results.json`).

The work measures and attributes; it does not predict.

11 Research extension: path to a thesis

The memo is deliberately bounded, and its limits point to a research programme rather than a single follow-up. Several concrete improvements come first and are largely mechanical: extending the sample with longer index and Treasury return series beyond the 2007 ETF-inception cap, so the regime tests span more than one full inflation cycle; replacing the synthetic constant-maturity proxies with a fitted curve and CUSIP-level, security-level instruments, particularly at the front end where the 2Y real/breakeven split is currently borrowed from the 5Y; a fuller TIPS treatment; genuine out-of-sample evaluation; and formalizing the regime rule as a change-point or state-switching model in the spirit of Ang and Bekaert (2002) rather than fixed thresholds, at the cost of the transparency the current rule is built around. Beyond those, the more interesting direction organizes into four stages.

Stage 1 – cross-episode portfolio P&L attribution. Apply the 2022 attribution framework to a set of historical Treasury hedge failures and test whether the realized losses decompose consistently into channels such as real rates, inflation compensation, term premium where identifiable, curve level, slope and curvature, convexity, carry and roll-down, and a residual. The question is whether the single-episode decomposition here generalizes at all.

Stage 2 – macro-regime association. Test whether particular macroeconomic environments are systematically associated with different dominant loss channels, without collapsing regimes to “high” versus “low” inflation. The relevant taxonomy is richer: supply and inflation shocks, demand and growth

shocks, monetary-policy shocks, recession and growth scares, fiscal and term-premium shocks, and inflation-expectations shocks may each leave a different attribution signature.

Stage 3 – out-of-sample identification. Ask whether, using only information available at the time, the emerging dominant Treasury-risk channel can be identified reliably – as distinct from explaining an episode after the fact. This stage lives or dies on avoiding hindsight: no thresholds defined by the outcome, and no using the same episode to both design and validate the mechanism.

Stage 4 – decision value. The strongest eventual question is normative: does identifying the dominant risk channel improve hedge selection or portfolio risk management relative to static duration reduction, conventional 60/40, or other simple defensive rules? This moves the enquiry from “can we explain what happened” toward “can mechanism-aware diagnosis deliver measurable out-of-sample decision value.” The present memo does not answer it; what it contributes is the attribution machinery and the honest negative allocation result that make the question worth asking.

12 Conclusion

The question this work started from was not whether stock–bond diversification is conditional – the literature settles that, most recently Campbell et al. (2026) – but how that conditionality shows up in a portfolio an investor could actually run. The analysis operationalizes the established regime dependence in an implementable setting and reports what it finds. 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; what the rule adds is that a portfolio’s hedge behavior becomes 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 realized loss is not mysterious: under this portfolio-level attribution the long-duration book’s roughly 19% loss is dominated by real-yield repricing, and a single-number DV01 view both mis-sizes it 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 weakens, which produced a negative, conditional result and a sharper question – what mechanically happened to the Treasury hedge in 2022 – which in turn became a portfolio-level curve attribution engine. None of it forecasts the next regime, and none of it claims to have identified a universal failure mechanism from one episode. The contribution is applied measurement: making hedge deterioration observable, benchmarking whether reacting to it helps, and expressing the 2022 failure as decomposable hedge-book P&L. Whether that attribution generalizes across episodes, whether the dominant channel can be identified in real time, and whether doing so has decision value are the open questions the four-stage extension above frames.

References

- Tobias Adrian, Richard K. Crump, and Emanuel Moench. Pricing the term structure with linear regressions. *Journal of Financial Economics*, 110(1):110–138, 2013.
- Andrew Ang and Geert Bekaert. International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187, 2002.
- AQR Capital Management. A changing stock–bond correlation: Drivers and implications. *Journal of Portfolio Management*, 2023.
- John Y. Campbell, Adi Sunderam, and Luis M. Viceira. Inflation bets or deflation hedges? the changing risks of nominal bonds. *Critical Finance Review*, 6(2):263–301, 2017.
- John Y. Campbell, Carolin Pflueger, and Luis M. Viceira. Macroeconomic drivers of bond and equity risks. *Journal of Political Economy*, 128(8):3148–3185, 2020.

- John Y. Campbell, Carolin E. Pflueger, and Luis M. Viceira. Bond-stock comovements. *Annual Review of Financial Economics*, 18, 2026. Forthcoming; NBER Working Paper 34323.
- Cboe Global Markets. Cboe volatility index (vix): Methodology and white paper, 2024. https://www.cboe.com/tradable_products/vix/.
- John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.
- Eugene F. Fama and Kenneth R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.
- Federal Open Market Committee. FOMC statement, june 17, 2026. Board of Governors of the Federal Reserve System, 2026. <https://www.federalreserve.gov/newsevents/pressreleases/monetary20260617a.htm>.
- Federal Reserve Bank of New York. Treasury term premia (acm model), 2026. https://www.newyorkfed.org/research/data_indicators/term-premia-tabs.
- Federal Reserve Bank of St. Louis. FRED economic data: Cpiaucsl, cpilfsl, dgs2/5/10/30, dgs3mo, dfii5/10/30, t5yie, t10yie, 2026a. <https://fred.stlouisfed.org>.
- Federal Reserve Bank of St. Louis. Market yield on u.s. treasury securities at 10-year constant maturity (dgs10). FRED Economic Data, 2026b. <https://fred.stlouisfed.org/series/DGS10>.
- Kenneth R. French. Data library (fama–french research factors). Tuck School of Business, Dartmouth College, 2026. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
- J.P. Morgan Asset Management. Guide to the markets. Market Insights, 2026.
- Robert Litterman and José Scheinkman. Common factors affecting bond returns. *Journal of Fixed Income*, 1(1): 54–61, 1991.
- State Street Global Advisors. Mind on the market: the 60/40 and the stock–bond correlation. Market commentary (secondary, single-provider estimate), 2025. <https://www.ssga.com/us/en/institutional/insights/mind-on-the-market-03-october-2025>.
- Bruce Tuckman and Angel Serrat. *Fixed Income Securities: Tools for Today’s Markets*. Wiley, 4th edition, 2022.
- U.S. Bureau of Labor Statistics. Consumer price index — may 2026. USDL news release, 2026. <https://www.bls.gov/news.release/cpi.htm>.
- Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2): 177–188, 1977.

Appendix: Replication Package

All empirical results in this paper are generated by two reproducible Python research pipelines from public data. The browser applications listed below are presentation layers only; they are not the source of any empirical calculation. Every table and figure is produced by these pipelines and by the figure scripts documented in the replication map that accompanies this paper.

Allocation study — Portfolio Risk & Allocation Analytics. Rolling monthly backtests of four allocation rules and two benchmarks on thirteen ETF proxies, with no-look-ahead controls, transaction costs, crisis-window diagnostics, factor regressions, and the threshold-sensitivity sweep. Source code: github.com/VM25/portfolio-lab; presentation site: pra-analytics.netlify.app.

Treasury curve attribution — Inflation Regime Rates Risk Engine. Repricing of synthetic Treasury books under measured and stylized curve shocks, with additive P&L attribution, rule-based hedge overlays, and Vasicek/CIR rates-risk simulation. Source code: github.com/VM25/inflation-lab; presentation site: yieldshock.netlify.app.

Data. ETF adjusted closes (Yahoo Finance); CPI, Treasury nominal and real yields, and breakeven inflation (FRED); the VIX index (Cboe); and the Fama–French research factors (Kenneth French Data Library). All inputs are public and processed offline; neither pipeline makes live API calls or prices instruments in the browser.