When Bonds Stop Hedging: An Exogenous Inflation State Selects the Equity Hedge Instrument

Maksim Aleksandrovich Andreev, Joseph Howden
Quark Quantitative Research

Abstract

The classic bond hedge for equity portfolios failed in 2022: stocks and Treasuries fell together, and static bond–gold hedge sleeves lost double digits precisely when protection was needed. Campbell, Pflueger and Viceira (2020) attribute the sign of the stock–bond correlation to the monetary–inflation regime: bonds hedge growth shocks but amplify inflation shocks. We first replicate the mechanism on entirely free data (FRED consumption-price and breakeven series; ETF total returns, 2005–2026, 244 monthly observations): a point-in-time composite inflation state predicts the forward 63-day realized SPY–IEF correlation with a Newey–West t of 2.46 in the presence of correlation-persistence, volatility, and yield-slope controls, exceeding a 2,000-draw shuffle null — to our knowledge the first replication of the mechanism at free-data resolution. We then ask whether the mechanism can be traded as a hedge-*instrument* selector: hold Treasuries as the equity hedge when the state predicts negative correlation, rotate the hedge sleeve to real assets (gold, broad commodity) when it predicts positive correlation. A naive median-threshold rotation *fails* our preregistered bars: it classifies 56% of history as inflationary, surrendering the bond hedge in the many ordinary growth-shock drawdowns, and slightly underperforms static hedges on the average equity down month. A rare-crisis calibration — rotate only when the state exceeds its expanding top quintile, with no fitted parameters — passes every bar: it improves the average down-month return over both a static bond hedge and a static diversified hedge net of costs, beats a random-timing shuffle at the 95th percentile, retains the decisive 2022 result (−0.9% versus −8.7% for the static bond hedge), and survives the exclusion of 2022 entirely, firing usefully across eight distinct years. A continuous exposure-weighted variant fails by re-creating the over-firing, isolating "fire rarely" as the operative principle. All bars, controls, and selftests were frozen before estimation; the failed variant is reported with the same prominence as the successful one. The verdict on the calibrated gate is **confirmed in sample (2005–2026) and forward-unvalidatable**: the measured down-month improvement is roughly 14 to 24 times smaller than the effect this design can detect, so no feasible forward window can settle it, and the paper makes no claim that validation is pending.

1. Introduction

An equity portfolio's hedge sleeve is usually a standing allocation: some mix of Treasuries and gold held statically alongside the equity book. The construction rests on the post-1990s stylized fact that the stock–bond correlation is negative — bonds rally when stocks fall. In 2022 that fact failed: an inflation shock drove stocks and bonds down together, and static bond-heavy hedge sleeves recorded losses in the teens at exactly the moment they existed to protect against.

The failure was not unforeseeable. A substantial literature documents that the stock–bond correlation is regime-dependent, and Campbell, Pflueger and Viceira (2020) supply the structural account: when macroeconomic risk is dominated by *growth* shocks, discount-rate news moves stocks and bonds in opposite directions and bonds hedge; when it is dominated by *inflation* shocks (or by monetary policy responding to them), nominal bonds are repriced with equities and the correlation turns positive. Ilmanen (2003) and Baele, Bekaert and Inghelbrecht (2010) document the empirical regime dependence; Neville, Draaisma, Funnell, Harvey and van Hemert (2021) document which assets actually defend in inflationary episodes.

This paper asks a narrower and more operational question than the literature usually poses: can a portfolio use an *exogenous, point-in-time observable* inflation state to select *which instrument* serves as the equity hedge — and does the selection survive a preregistered falsification protocol with honest

multiple-testing accounting? Three results follow. First, the mechanism replicates on free data. Second, the naive tradeable form fails, for an identifiable reason. Third, a rare-crisis calibration of the same signal passes every preregistered bar, with robustness that a single-episode artifact would not display.

> **Why exogeneity matters.** Our register of settled results contains a family of nulls in which signals derived from the realized price path (rolling correlations, realized volatility, crash-probability transforms) failed to *lead* regime shifts — they are contemporaneous at best, because they are causal functionals of the very path they try to anticipate. A consumption-price index and a market breakeven rate are not price-path transforms of the assets being hedged; they are external state variables. The hypothesis is therefore not subject to that structural objection, which is precisely why it was selected for testing.

2. Data and the inflation state

All data are free and public. Monthly ETF total returns for SPY (equity), IEF and TLT (Treasuries), GLD (gold) and DBC (broad commodity) from 2004–2026; daily returns of SPY and IEF for realized-correlation construction; the CBOE volatility index and the 10-year minus 3-month Treasury slope as controls. The inflation state is a composite of three FRED series, each transformed to an expanding-window z-score so that no future information enters the standardization:

> z_t = mean of three expanding z-scores: trailing 12-month CPI inflation (lagged one month for publication delay), the 10-year breakeven inflation rate (real-time market data), and the trailing 12-month realized volatility of monthly CPI changes (lagged one month).

The lag convention makes the state point-in-time: at month-end t the latest available CPI print refers to month t−1, while breakevens are live market prices. The composite is deliberately simple — no weights are estimated anywhere in the paper.

3. The mechanism: forward stock–bond correlation

For each month-end t we compute the realized correlation of daily SPY and IEF returns over the following 63 trading days and regress it on the inflation state with three controls: the trailing realized correlation (persistence), the volatility index level, and the yield-curve slope. Standard errors are Newey–West with automatic bandwidth. Significance is assessed against both the conventional threshold and a shuffle null in which the inflation state is permuted in time (2,000 draws) and the regression re-estimated.

> **Result (244 months, 2005–2026).** The inflation-state coefficient on forward realized correlation is +0.089 with Newey–West t = 2.46, against a shuffle-null 95th-percentile |t| of 2.11. The sign matches the Campbell–Pflueger–Viceira prediction: a higher inflation state forecasts a higher (less negative, eventually positive) stock–bond correlation. The controls confirm the state is not a disguised volatility or curve effect. We believe this is the first replication of the mechanism using only free public data.

4. The naive gate fails — and why

The natural tradeable form holds a hedged equity book — 60% SPY plus a 40% hedge sleeve — and rotates the sleeve between Treasuries (when the state predicts negative correlation) and a 50/50 gold–commodity mix (when it predicts positive correlation). The naive trigger is a median split: rotate whenever the state is above its expanding median. Under our preregistered bars this fails. Over the 81 equity down months in the sample the median-gated book averages −2.23% per month against −2.12% for a static bond hedge and −2.06% for a static diversified hedge, and it does not beat a random-timing shuffle.

The mechanism of the failure is instructive and diagnosable: the median gate labels *56% of all months* as inflationary, but genuine inflation-shock regimes are rare. The gate therefore surrenders the bond hedge in the many ordinary growth-shock drawdowns — where bonds do hedge — collecting a large 2022 win but paying a small penalty in dozens of other down

months. Gross and net results are nearly identical, so transaction costs are not the cause: the gate genuinely selects the wrong instrument on the average down month.

5. The rare-crisis calibration

The diagnosis dictates the fix, which was pre-declared in the first study's frozen protocol before its outcome was known: fire only in *extreme* inflation states. The calibrated gate rotates the hedge sleeve to real assets only when the state exceeds its expanding-window 80th percentile — again with no fitted parameters; the threshold is a fixed quantile of past data available at each date. The gate fires in 23% of months.

> **Result.** All three preregistered bars pass. (i) Down-month performance: the gated book averages −1.92% per equity down month versus −2.12% (static bonds) and −2.06% (static diversified), net of costs. (ii) Timing content: the gross down-month mean exceeds the 95th percentile of a 2,000-draw random-timing shuffle that preserves the firing frequency — the improvement is when the gate fires, not that it fires. (iii) The 2022 discriminator: the gated book returned −0.9% over 2022 against −8.7% for the static bond hedge and −8.3% for the static diversified hedge, because the gate had rotated the sleeve into gold and commodity while bonds fell with equities.

Three robustness results argue against a single-episode artifact. First, a stricter top-decile gate (firing 12% of months) also passes all three bars. Second, excluding 2022 entirely, the gated book still beats both static hedges over the remaining 75 down months (−1.75% versus −1.89% and −1.83%), and the months in which the gate fires during equity drawdowns span eight distinct years (2011, 2018, 2020–2023, 2025–2026). Third, a continuous variant that scales the real-asset weight by the state's expanding percentile rank *fails*: its average real-asset weight of 0.53 reproduces the median gate's over-firing and its down-month average reverts to −2.24%. The contrast isolates the operative principle: the signal earns its keep only when consulted rarely, at the extremes it was built to identify.

6. Preregistration and honesty accounting

Both studies ran under a frozen protocol: hypotheses, exact statistics, pass/fail bars, controls, and synthetic-data selftests were committed before any market metric was observed. The selftests require the harness to detect a planted regime effect, to produce no systematic false positives on independent-noise worlds (assessed as a rate across ten panels, since any 95th-percentile test admits occasional flukes by construction), and to discriminate a planted 2022-style both-down regime. Two methodological amendments were made before outcomes were known and are recorded in the protocol: the random-timing shuffle is evaluated gross-of-costs on both sides, because a persistent gate mechanically trades less than its time-shuffled counterfeits and charging costs inside the null would flatter the real book; and the null-world check is rate-based for the same reason.

Multiple testing is charged explicitly: four gate forms were examined across the two studies (median, quintile, decile, continuous), and the deflated-Sharpe accounting uses that count. The primary quintile form was pre-declared rather than selected after the fact, and the decile corroboration plus the ex-2022 stability argue the result is not selection residue. Because the significance object is a hedge overlay on a beta-dominated book, the preregistered test is the down-month timing comparison, not a full-sample Sharpe ratio — the latter is dominated by the 60% equity leg and would be uninformative about the hedge.

Limitations. The average down-month improvement is modest (roughly 0.2 percentage points per month); the economic case rests on the crisis asymmetry, where the improvement is large. Data are monthly and free; higher-frequency inflation states are not tested. The calibrated study is a second look at the same sample as the mechanism study, which is why the verdict is recorded as **confirmed in sample (2005–2026) and forward-unvalidatable**. That second label is permanent and is not a statement about effort or priority. The average down-month improvement of roughly 0.2 percentage points sits against a minimum detectable effect of 2.85 to 4.74 percentage points for this design, so the bar is 14 to 24 times the quantity being measured, and a forward record long enough to reach a Harvey threshold of 3 would run to somewhere between 2,100 and 5,900 years. No feasible forward window can settle this result either way, and the paper therefore withdraws any implication that one is pending. What is withdrawn is a promise of future validation the arithmetic cannot keep; the in-sample evidence and the conclusion drawn from it are unchanged. Any citation of this result carries both labels. Implementation uses liquid ETFs; futures implementations may differ. Nothing here constitutes investment advice.

7. Conclusion

The stock–bond correlation regime is forecastable from an exogenous, point-in-time inflation state at free-data resolution, replicating Campbell, Pflueger and Viceira (2020). The mechanism supports a tradeable hedge-instrument selection rule only in rare-crisis form: rotate the equity hedge from Treasuries to real assets when, and only when, the inflation state is in the top quintile of its own history. The naive form of the same idea fails by over-firing, and reporting that failure alongside the success is, we would argue, the more important half of the paper: the difference between a mechanism and an edge is calibration, and the difference between an edge and an artifact is a protocol that was written down before the data were consulted.

Reproducibility. Both studies, their frozen preregistrations, and full numerical results are versioned in the research repository (mechanism and median-gate study; rare-crisis calibration study). All data sources are free and public (FRED; exchange-traded fund total returns).

References

Baele, L., Bekaert, G. & Inghelbrecht, K. (2010). The determinants of stock and bond return comovements. *Review of Financial Studies*, 23(6), 2374–2428.

Bailey, D. H. & López de Prado, M. (2014). The deflated Sharpe ratio: correcting for selection bias, backtest overfitting, and non-normality. *Journal of Portfolio Management*, 40(5), 94–107.

Campbell, J. Y., Pflueger, C. & Viceira, L. M. (2020). Macroeconomic drivers of bond and equity risks. *Journal of Political Economy*, 128(8), 3148–3185.

Campbell, J. Y., Sunderam, A. & Viceira, L. M. (2017). Inflation bets or deflation hedges? The changing risks of nominal bonds. *Critical Finance Review*, 6(2), 263–301.

Harvey, C. R., Liu, Y. & Zhu, H. (2016). …and the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5–68.

Ilmanen, A. (2003). Stock–bond correlations. *Journal of Fixed Income*, 13(2), 55–66.

Neville, H., Draaisma, T., Funnell, B., Harvey, C. R. & van Hemert, O. (2021). The best strategies for inflationary times. *Journal of Portfolio Management*, 47(8), 8–37.

Newey, W. K. & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.

Newey, W. K. & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4), 631–653.

Andreev, M. A. & Howden, J. (2026a). Stochastic mechanical methods for quantitative portfolio management. *Quark Quantitative Research Working Paper*.

Andreev, M. A. & Howden, J. (2026e). Ontological limits of historical validation for wavefunction-calibrated financial systems. *Quark Quantitative Research Working Paper*.