Cross-Country Rates Relative Value:

A Systematic Regime-Conditional Strategy

Kairav Shaurya Makhija

Kiefer Ong Xian Yao

Mitchel Lee Jun Xian

Sakthidaran Avanindran

NUS Investment Society

Fixed Income Pod

2026

Abstract

This paper presents a systematic relative-value (RV) strategy targeting sovereign yield spreads across USD, EUR, JPY and AUD. We demonstrate that traditional unconditional pairs trading is insufficient in fixed income markets due to structural regime shifts and central bank interventions. To systematically capture transient dislocations, we implement a multi-stage pipeline involving a 4-state Gaussian Hidden Markov Model (HMM) and regime-conditional cointegration testing. Using a strict walk-forward framework with no lookahead bias, we show that an adaptive Kalman-filter signal achieves an annualised Sharpe ratio of 0.80 and captures 15,220 basis points of PnL over a 15-year out-of-sample window (2010–2025). We detail the methodology, parameter robustness and core failure modes, including central bank policy breaks and cross-currency funding stress.

Contents

1 Introduction and Research Question

Sovereign yield spreads between developed market economies are fundamentally anchored over long horizons by relative monetary policy expectations, inflation differentials and global risk premia. However, in the short run, these spreads dislocate due to order flow, funding stress or geopolitical shocks, before reverting as macro fundamentals reassert themselves.

The central research question of this paper is: **When and why do cross-country interest rate relative value relationships hold—and when do they break?**

Our motivation stems from the observation that unconditional relative value trading generally fails. Pairs that appear cointegrated in a full-sample analysis break down dramatically across shifting macroeconomic regimes. We hypothesise that a successful systematic RV strategy requires three conditions to hold simultaneously:

1. **Statistical:** The spread must be cointegrated specifically within the current macroeconomic regime.
2. **Macro:** The regime must represent stable mean-reversion, not an abrupt structural break (e.g., central bank divergence or yield curve control).
3. **Risk:** Positions must be constrained by objective hard stops to prevent tail losses before reversion occurs.

This paper's primary contribution is a fully lookahead-free framework. Every stage of our multi-gate pipeline: (1) macroeconomic feature analysis, (2) HMM state probability estimation, (3) conditional cointegration validation and (4) Kalman filter signal generation is strictly re-estimated within expanding training windows before being applied out-of-sample.

2 Dataset and Universe Characterization

2.1 Data Coverage

Our core universe consists of 8 instruments: 5Y and 10Y par sovereign yields for USD (US Treasury), EUR (German Bund), JPY (JGB) and AUD (ACGB). The primary data is sourced from Bloomberg and FRED, with the sample starting on January 6, 2005. The out-of-sample (OOS) backtest spans 2010 to 2025 across 15 expanding folds, utilising 2005–2009 purely as a burn-in period.

2.2 Stationarity Testing

Prior to modeling, we conducted rigorous stationarity testing across 41 distinct series using joint Augmented Dickey-Fuller (ADF) and KPSS tests. The joint decision rule classified a series as $I(0)$ if and only if ADF $p < 0.05$ and KPSS $p > 0.05$.

Results indicated that sovereign yields, policy rates and FX spot rates are $I(1)$ non-stationary processes. Because bond yields are $I(1)$, RV strategies must operate on spread levels, relying on cointegration frameworks rather than assuming simple mean-stationarity of the underlying assets.

Figure 1: Exploratory Data Analysis (US 2Y Treasury): Visual confirmation of $I(1)$ non-stationarity in raw levels (top), stationary daily differencing (middle), and severe volatility clustering (bottom) which necessitates regime-aware modeling.

2.3 Structural Blocs

Visual inspection of the universe revealed distinct regional blocs. The **Global Core** (USD, EUR, AUD) exhibits highly correlated trends and synchronised volatility spikes during crisis events (e.g., 2008 and 2020). Conversely, JPY acts as a **Deflationary Outlier**, characterised by a secular yield decline and severe decoupling post-2016 due to the Bank of Japan's Yield Curve Control (YCC) policy to curb deflation.

3 Empirical Foundations and Asset Selection

3.1 Principal Component Analysis

Full-sample Principal Component Analysis (PCA) on 5Y yield changes showed that PC1 explains approximately 35% of the variance, acting as a global rates level/duration factor with positive, uniform loadings across currencies.

Crucially, PC1 dominance is highly regime-dependent. In low-volatility regimes, country loadings diverge, leaving idiosyncratic room for RV strategies to be viable. However, in high-volatility crisis regimes, PC1 accounts for $> 50\%$ of variance, and cross-country differentiation collapses. This difference in PC1 dominance between stress and normal regimes is statistically significant ($p = 5.3 \times 10^{-23}$), providing a quantitative mandate for regime-filtering.

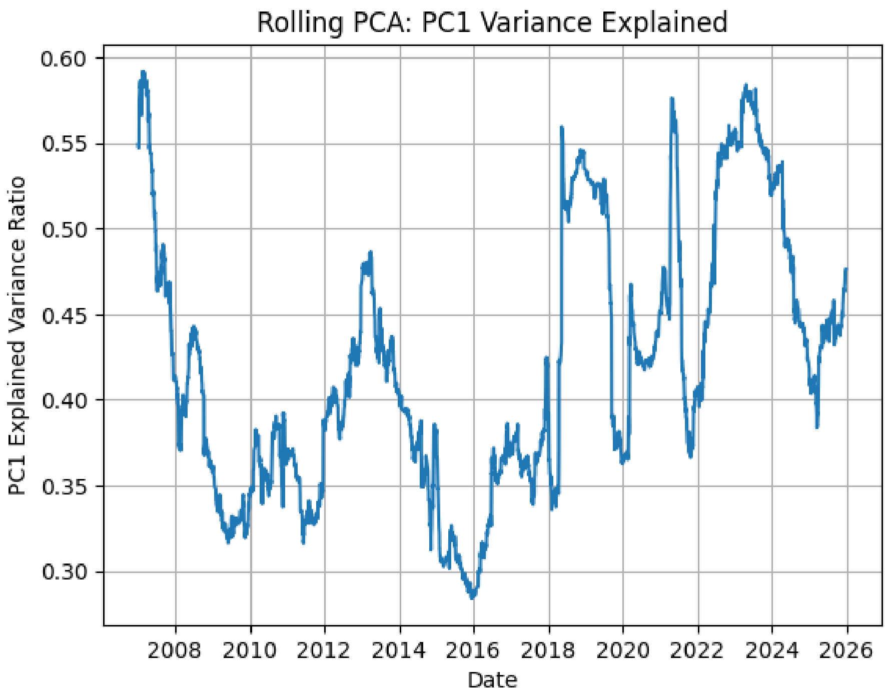

Figure 2: Rolling 252-day PCA Variance Explained by PC1, illustrating structural spikes during market stress regimes.

3.2 Rolling Correlation and Clustering

Rolling correlation analysis (252-day windows) demonstrated that most pairs fail to maintain consistent positive correlation, spending $< 50\%$ of the time above a 0.6 threshold. To identify persistent co-movement blocs, we applied 252-day rolling K-Means clustering ($k = 3$) on daily yield changes. Strong co-cluster persistence ($> 50\%$) was isolated almost exclusively to JPY-AUD 5Y and USD-EUR pairs. Consequently, only pairs with high historical persistence proceeded to formal cointegration testing.

4 Regime Identification

To prevent the strategy from executing trades during structural breaks and periods of fundamental decoupling, we implement a regime-switching framework. By integrating our feature engineering and state detection into a strict expanding-window architecture, we ensure that all regime inferences are free of lookahead bias, relying solely on data available prior to the time of execution.

4.1 Macro Feature Engineering

We engineered four macro-financial features to capture the global environment:

1. **MOVE Index:** 1-day change, acting as a proxy for US rates volatility.
2. **BBDXY Index:** 1-day change, proxying USD funding and cross-market stress.
3. **Macro Surprise:** PC1 of the Citi Economic Surprise Indices (CESI) for USD and EUR.
4. **Curve Signal:** PC1 of 2Y-10Y spreads across the 4 currencies.

4.2 Hidden Markov Model (HMM)

We specified a 4-state Gaussian HMM ($s_t \in \{0, 1, 2, 3\}$), ordered by the MOVE mean. State 0 represents a "Low-vol Carry" environment, while State 3 represents a "Crisis" state where volatility is $4\times$ the baseline. Parameters were estimated via the Baum-Welch algorithm (200 EM iterations).

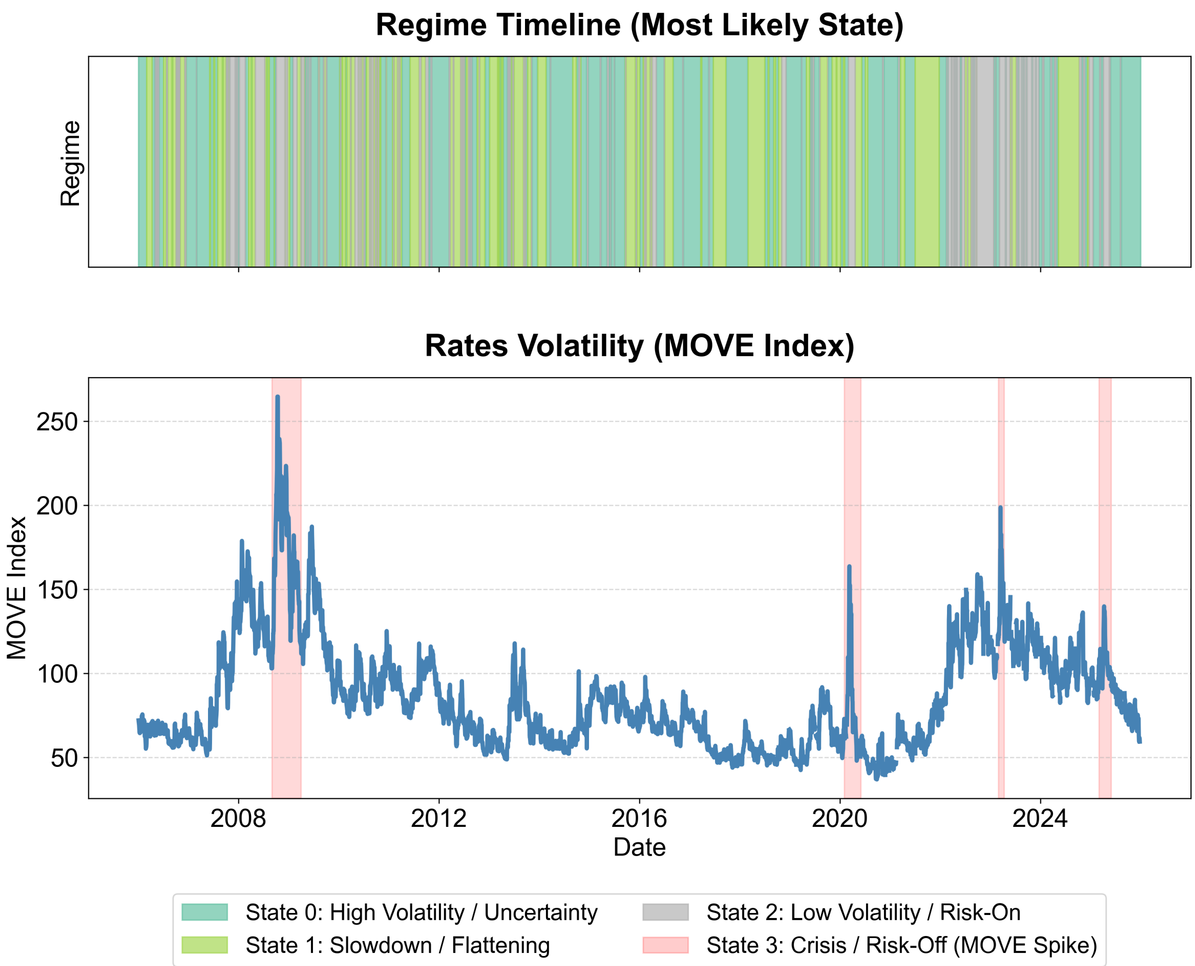

Figure 3: Historical Time Series shaded by identified HMM Regimes. Note the distinct clustering of the 'Risk-off' / Crisis states during 2008 and 2020.

In our walk-forward inference, the HMM was retrained monthly on an expanding window. Regime labels were smoothed using a 3-day rolling mode to eliminate 1-day false flickers. The Viterbi algorithm was used to infer the most likely state sequence.

5 Cointegration and Signal Generation

5.1 Regime-Conditional Cointegration

We demand that a pair is cointegrated *conditional* on the current regime. This involves identifying contiguous spells of a specific regime with ≥ 40 observations. We run an ADF test within each spell and combine the p-values via Fisher's method:

$$\chi^2 = -2 \sum_{i=1}^{k} \ln(p_i)$$

where the statistic follows a $\chi^2(2k)$ distribution.

We also fit an AR(1) model on the pooled, de-meaned spread to calculate the reversion speed:

$$\Delta s_t = c + \alpha s_{t-1} + u_t$$

The tradability gate requires the AR(1) coefficient $\alpha < 0$ (with $p < 0.05$), and a probability $> 60\%$ that the regime spell duration will outlast the calculated half-life.

5.2 Kalman Filter Dynamic Hedge Ratio

To avoid stale-beta risk inherent in static OLS, we utilise a Kalman filter to dynamically update the hedge ratio (β_t) and intercept (α_t). The state-space system is defined as:

$$y_t = \beta_t x_t + \alpha_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, R)$$

$$[\beta_t, \alpha_t]^T = [\beta_{t-1}, \alpha_{t-1}]^T + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q)$$

The process noise covariance Q is kept intentionally small to allow the hedge ratio to drift smoothly without overfitting to daily noise. The Kalman innovation, standardised by its predicted variance, provides our trading signal: Z_t.

6 Backtesting Framework

6.1 Execution Logic and Hard Stops

Signals are generated based on the Kalman Innovation Z-score:

- **Entry:** Enter Long if $Z_t \leq -Z_{\text{entry}}$; Short if $Z_t \geq +Z_{\text{entry}}$.
- **Exit:** Close when $|Z_t| < Z_{\text{exit}}$, or if the regime changes to an untradable state.
- **Execution:** $T + 1$. A signal generated at the close of day t is filled at the open of $t + 1$, enforcing a strict causal barrier.
- **Hard Stop:** We calibrated a fixed hard stop of $|Z_t| \geq 3.5$ based on the MAE distribution (between p90 and p95 of MAE on historically winning trades). This prevents catastrophic tail losses.

6.2 Walk-Forward Optimisation Engines

We deployed three backtest engines across 2010–2025, assuming a 5 bps flat round-trip transaction cost:

1. **Static:** Fixed parameters ($Z_{\text{entry}} = 2.0$, $Z_{\text{exit}} = 0.5$).

2. **WF PnL-Optimised:** Grid search maximising cumulative training PnL per expanding fold.

3. **WF Sharpe-Optimised:** Grid search maximising annualised training Sharpe.

7 Results and Performance

7.1 Headline Metrics

Engine	Trades	Win Rate	PF	Total PnL	Sharpe	Max DD
Static	56	64.3%	3.13	8,955 bps	0.63	-1,849 bps
WF PnL	121	65.3%	2.37	15,220 bps	0.80	-3,453 bps
WF Sharpe	121	64.5%	2.30	14,487 bps	0.77	-3,453 bps

Table 1: Headline Performance (2010-2025). Profit Factor (PF) indicates the ratio of gross profits to gross losses. Drawdowns are reported in basis points.

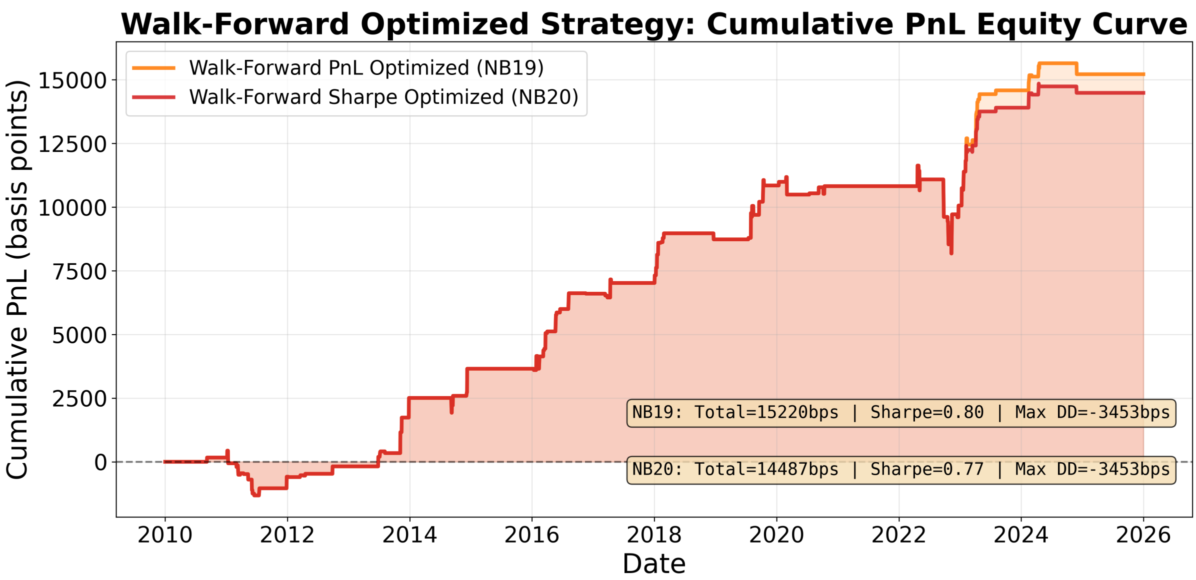

Figure 4: Cumulative PnL (in basis points) of the Walk-Forward Optimized strategy.

The WF optimisation resulted in a massive +70% PnL boost and a +0.17 Sharpe improvement compared to the static baseline. Crucially, robustness sweeps confirm that the strategy's statistical edge is not overly sensitive to the conservative 5 bps transaction cost assumption. The walk-forward engines remain strictly profitable (Sharpe $>$ 0) at round-trip execution costs of up to 10–12 bps, providing a substantial margin of safety for real-world slippage and market impact.

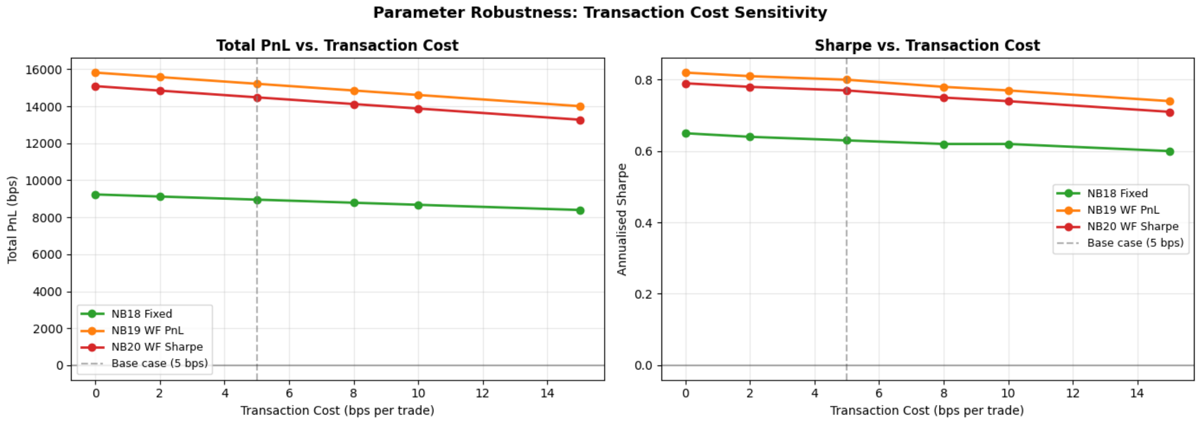

Figure 5: Transaction Cost Sensitivity Analysis: Total PnL and Annualised Sharpe Ratio demonstrate strategy robustness up to 12 bps of execution drag.

7.2 Optimal Parameter Convergence and Trading Frequency

A primary observation from the WF optimisation process was the consistent convergence of the entry threshold toward $Z_{\text{entry}} = 1.5$ across the majority of annual folds. This preference was invariant to the choice of objective function, appearing in both the PnL and Sharpe-optimised engines.

By recalibrating the model to this more sensitive entry threshold, the strategy successfully captured higher-frequency, lower-amplitude mean-reversion oscillations. This shift in parameterisation resulted in effectively doubling the total trade count from 56 to 121, while maintaining a robust win rate of approximately 65%. Furthermore, the near-identical outcomes observed between the PnL and Sharpe optimisations indicate a high degree of collinearity between raw returns and risk-adjusted performance within our defined parameter grid. This suggests that the captured spread wiggles provide sufficient risk-adjusted alpha to dominate the objective function regardless of the specific reward-to-risk weighting.

7.3 Trade Diagnostics and Exit Logic

To further validate the strategy's mechanics, we analysed the underlying exit conditions for the 121 generated trades. As illustrated in Figure 6, the strategy operates structurally as intended, with natural mean-reversion serving as the primary baseline exit mechanism (accounting for nearly 82% of all trades).

However, the performance breakdown reveals a crucial secondary benefit of the HMM framework: when macro conditions transition and force a 'Regime Shift' exit, these trades act as highly profitable outliers. Regime-forced exits captured an average of 303 bps with a 100% win rate over the sample. This indicates that the HMM not only filters out toxic entries but effectively locks in highly profitable dislocations the moment the macro environment shifts. Furthermore, the hard-stop circuit breaker was only triggered 7 times across the entire 15-year OOS period, confirming the initial regime gates successfully avoid the vast majority of dangerous, non-reverting environments.

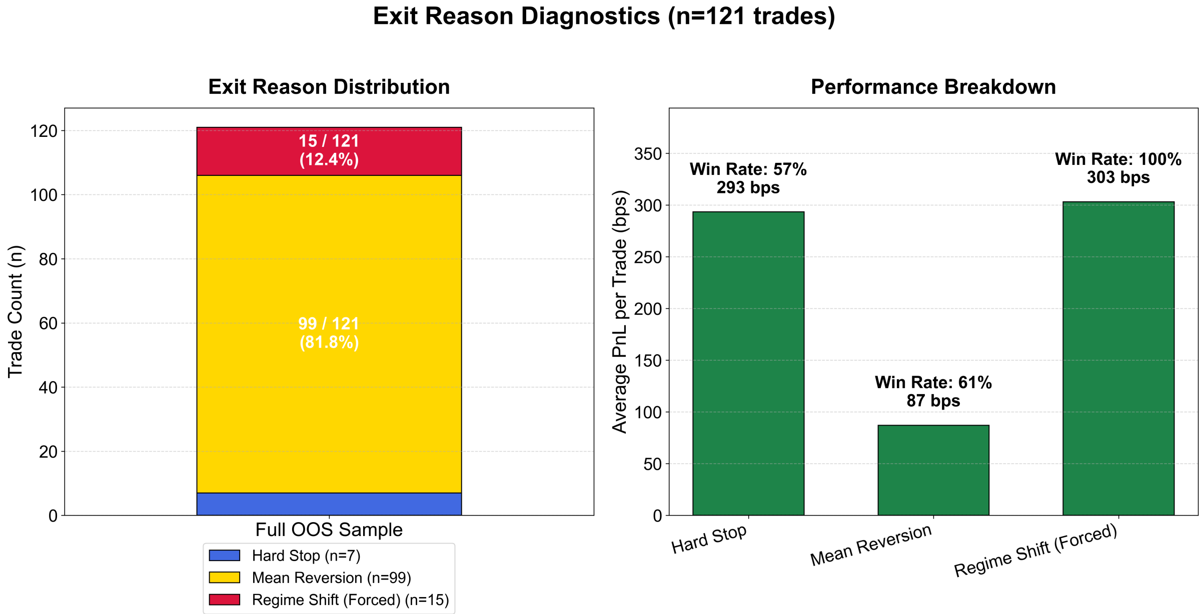

Figure 6: Trade Diagnostics: Exit Reason Breakdown. While natural Mean Reversion drives the highest volume of trade exits (81.8%), HMM-forced Regime Shifts capture the highest average profitability (303 bps) with a 100% win rate.

8 Pair-Level Analysis

Breaking down the aggregate performance reveals that returns were heavily clustered in specific macro-aligned states and distinct currency pairs.

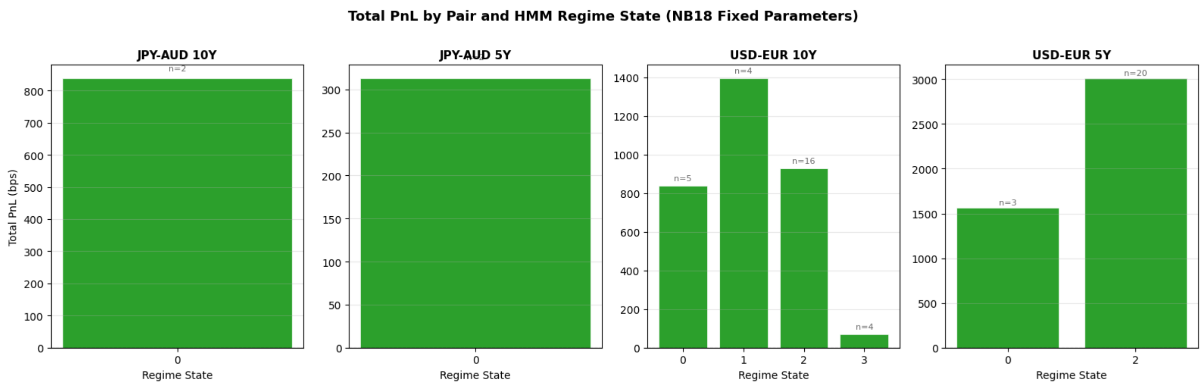

Figure 7: Strategy PnL decomposed by currency pair and identified macroeconomic regime state.

- **USD-EUR 10Y & 5Y:** These were the most reliably activated pairs. They performed exceptionally well in low-volatility regimes where US Federal Reserve and ECB monetary policies were directionally aligned, allowing fundamental inflation differentials to act as a strong anchor for the spread.

- **JPY-AUD:** The HMM effectively sidelined this pair for most of the decade. The Bank of Japan's Yield Curve Control since 2016 structurally broke the natural cointegration. The regime gate correctly identified the spread as a dangerous trending asset, saving the portfolio from significant losses that an unconditional model would have incurred.

9 Critical Assessment and Failure Modes

Understanding the structural vulnerabilities of the strategy is as essential as analysing its returns. We identify four primary failure modes that currently constrain performance:

1. **HMM Detection Lag:** While the 1-day lag and 3-day smoothing protocols strictly prevent lookahead bias, the real-time regime assignment inherently lags fast market dislocations (e.g., March 2020). While the hard stop circuit-breaker limits the severity of these entries, it cannot entirely preempt the initial drawdown.

2. **Funding Stress Dislocations:** The current model simulates local yield spreads but does not account for the cross-currency (XCCY) basis swap. During episodes of acute funding stress, XCCY basis blowouts can erase spread PnL, rendering seemingly attractive relative-value trades economically unviable, particularly for AUD-denominated carry legs.

3. **Structural Breaks:** Macroeconomic policies, such as the BoJ's YCC, permanently relocate yield anchors outside of historical cointegrating relationships. While the WF retraining eventually adapts to these shifts, the transition period frequently produces false entries before the Kalman hedge ratio and HMM parameters fully adjust to the new policy.

4. **Grid Search Resolution:** The current 3×3 parameter grid lacks the resolution to significantly distinguish between PnL and Sharpe ratio objectives, often resulting in identical parameter selection across different optimisations. A finer grid or Bayesian optimisation is required to isolate the distinct optimal points for each specific reward-to-risk profile.

Furthermore, the generation of only 121 trades over a 15-year horizon results in relatively wide confidence intervals for our performance estimates. From a risk-management perspective, the assumption of equal notional sizing ignores the potential benefits of regime-aware scaling, which may lead to unoptimised portfolio variance during periods of shifting volatility.

10 Conclusion and Future Work

We successfully answer our research question: Cross-country RV relationships are exclusively tradeable when the macro regime is stable, monetary policy cycles are correlated and funding conditions are benign. Unconditional RV will simply fail outright.

By implementing a 4-state HMM model combined with a dynamic Kalman Filter, we successfully filtered out broken, trending markets where mean-reversion is unlikely. The WF Optimised engine proved that adapting Z-score thresholds to historical volatility (instead of sticking to fixed Z-score thresholds) significantly improves performance, capturing over 15,000 bps of out-of-sample profit across 15 years.

Future Work: To transition this framework from a robust research pipeline into a deployable portfolio strategy, future enhancements must directly address our identified failure modes. Structurally, we will focus on: (1) introducing an XCCY basis filter as a secondary gate to protect against funding squeezes, (2) applying a finer parameter grid via Bayesian optimisation to better separate PnL and Sharpe objectives, and (3) implementing volatility-scaled position sizing to optimise portfolio variance across shifting market environments.

References

[1] Huggins, D. and Schaller, C. (2013). *Fixed Income Relative Value Analysis*. Bloomberg Press / Wiley, Hoboken, NJ.

Disclaimer