Recap of Assignment 10 Implementation:

In assignment 10, the implementation began with sovereign yield curves to construct DV01-neutral 2s5s10s flies, and evaluated carry-and-roll-down (CRD). A known limitation was the fixed normalisation signal of yields that did not roll, and resulted in limited trades (because the mean was skewed from high IR era).

Part A:

Potential refinements and enhancements were evaluated in light of feedback from Prof, TAs, and some of the issues that we have identified ourselves.

1) Signal Generation Pipeline for Systematic Entries and Exits

For each country and potentially fly combinations (2s5s10s, 5s10s30s), a CRD was computed from the spot yield curve. These signals were then smoothed into a 5-day moving average (to dampen day to day noises), and it was standardised using a rolling 6-month z-score, with a 126 day lookback window. Previously a 10-year average lookback was used.

Now, we have pivoted to instead of a 1 month holding period, now, instead there are hard entry and exit rules, such that a long or short trade is entered such that:

- Enter a fly when |z| exceeds the entry threshold (currently 1.6 in absolute value).
- Exit when the signal mean-reverts inside a tighter exit band, currently 0.2

```
H_MONTH = 1
H = H_MONTH / 12.0

ROLL_WIN_6M = 126
ENTRY_Z = 1.6
EXIT_Z  = 0.2
```

A cap is also ensured that we cap the portfolio at four simultaneous flies, selecting those with the largest |z| signals on a given day (when there is capacity), and allocate to each fly to hit a fixed DV01 budget per position. In the case where there is fund capacity, like let's say there are only 3 fly positions, the remaining 25% is allocated to SOFR, which is basically a cash proxy.

2) Expansion of Universe to comprise of other sovereign yield curves (US, DE, UK, IT, JP, AU, CA)

Now, the sovereign curves are expanded for US, Germany, UK, Italy, Japan, Australia and Canada, constructing 2s5s10s and 5s10s30s flies for each country wherever tenors exist. As a baseline, we used a simple equal-weight cross-country allocation across active flies, combined with a per-fly DV01 cap and a total portfolio DV01 budget (25%).

This is more realistic than signal-weighted allocation (which tends to over-concentrate in the noisiest signals), and reduces transaction costs, as we have to rebalance every time there is an exit or entry of a new position. Risk-aware or volatility weighing scheme, can help control concentration risk, and can be considered for future enhancements. The reason why signal-weight was not chosen for PC is that there are underlying cost implications.

3) Extending to futures with TU (2y), FV (5y), TY (10y), US (30y),

Beyond the spot baseline implementation, the strategy is structured such that the signals are extended to execution through futures, which naturally has some slippages and transaction cost implications. The following were extracted for implementation:

- UK: Schatz (2y), Bobl (5y), Bund (10y), Buxl (30y)
- US: TU (2y), FV (5y), TY (10y), US (30y)
- UK Short/Medium/Long Gilt futures (approx. 2/5/10–30y equivalents)

4) Preliminary modelling reveals 0.9529 gross sharpe to 0.86 net sharpe. This enhancement was considered after some modelling of Enhancement #1, #2, #5, with transaction costs being quite substantial. Signal change to be based on cost-modelling. We had initially assumed the following:

```
COST_ROUNDTRIP_BP_COUNTRY = {
    "US": 0.12,   # UST fly via futures, very tight
    "DE": 0.12,   # Bund - Germany
    "UK": 0.18,   # Gilts a bit wider - wider than above, esp in recent markets - refer to marketwatch
    "JP": 0.08,   # JGBs - generally very very liquid
    "AU": 0.35,   # ACGBs - Aussies
    "CA": 0.25,   # Canadas - wider than US/DE/JP, generally not that liquid, but more liquid than AUS
    "IT": 0.50,   # BTPs widest - LOL
}
```

It does make more sense to account for the fact that costs for certain sovereigns to be substantially more expensive, than, let's say the US, in which it is far more costly.

5) IRS swap rates by Tenor in parallel to Yield Curves, and utilise in portfolio construction or a filter (if both Yield Curve and IRS swap agrees in sign → execute)

Finally, we introduce interest-rate swap (IRS) curves by tenor in parallel to the government yield curves as an additional layer of signal validation. The idea is to require that both the sovereign curve and the swap curve agree on the sign of the CRD signal (e.g. both suggest paying the belly and receiving wings) before executing a fly.

Practically, this acts as a filter such that:

- If both govt and IRS signals have the same sign, the fly is traded
- If they contradict, the trade is skipped.

The idea is that this refinement should reduce the number of false positives where the government curve is distorted by idiosyncratic factors like Liberation day impacting demand and supply, but the curve swap is not. It therefore pushes the strategy towards a more robust signal that offers cross-instrument consensus, but with the cost of potentially lowering trading frequency.

Part B:

Following the baseline results, and guided by lecturer feedback, we selected three refinements that directly addressed the main issues observed in the strategy:

1) Signal efficacy & Cost-Sensitive signal Thresholding (A1, A4) - Simple + Simple Refinement

2) Dual-Curve Confirmation (IRS & Govt Yield) (A5) - Major Refinement

3) Expansion into other sovereigns + Futures - Expansion refinement (A2, A3)

B1)

Preliminary backtests showed that the baseline strategy suffered from excessive turnover once realistic country-dependent cost assumptions were applied (especially for UK, AU, CA, IT). To address this, we refine the signal generation by tightening entry/exit thresholds in a way that is explicitly cost-sensitive, ensuring trades are only taken when the expected CRD edge is large enough to overcome estimated round-trip costs. Notably for IT trades, volatility in the CRD meant some trades short holding periods resulting in losses net of transaction costs.

Specifically, we raise the absolute z-entry threshold (currently 1.6) and narrow the exit threshold (currently 0.2), reducing unnecessary churn while retaining the most profitable dislocations. This is so that the exit threshold is slightly harder to reach.

B2)

Following, the second refinement introduces the swap-curve confirmation as an additional layer, such that a trade is only commenced with both signals in agreement (IRS and Govt curve), such that there is cross-instrument consensus.

This can remove many false positives curve signals by liquidity events, flight to quality, or regulatory or seasonal flows imbalances. Although this lowers trade frequency, it increase signal reliability and reduces the number of lower conviction positions entering and exiting the portfolio. It will be constructed in such a way that if we have low confidence entry, we will not enter. If we have either signal indicating exit, we will exit.

Because IRS markets often reflect a “cleaner” macro-rate expectation free of sovereign-specific distortions, this refinement materially improves the logic in multi-country curve trading, and to a certain

extent the reliability of the trade in terms of signal generation. The four-fly DV01 cap and the 25% unused-capital allocation rule remain unchanged under this refinement. However, do note that IRS data is only available on BBG from 2009 onwards, so previous years relied on yield curve signal, and following years rely on this implementation.

B3)

The addition of futures implementation, on top of the spot-curve flies, forms a parallel execution option. All of these sleeves (spot bond and futures) rely on the yield curves as the generator of CRD and z-score signals. We include futures only for countries where deep and reliable contracts exist (US, Germany, UK). Other sovereign markets either lack liquid long-end futures (e.g., AU 30-year does not exist), or the contract depth is insufficient for systematic fly trading (e.g., Italy, Canada - B/A spreads are way too large).

The futures positions are layered on top, by mapping each specific sleeve's DV01 exposure to the nearest-maturity bond future (e.g., TU/FV/TY/US for the US curve, Schatz/Bobl/Bund/Buxl for Germany, and short/medium/long gilt futures for the UK).

Like before, PC remains unchanged across each position, such that:

- No fly has allocation of greater than 25% of the total portfolio
- Unused DV01 capacity allocated to SOFR
- Both spot and futures implementations obey the same exit and entry logic, and is sorted by the Z-score. Note that transaction costs from B1) should also be accounted for.

Part C

Refer to Github and codebase for different implementations.

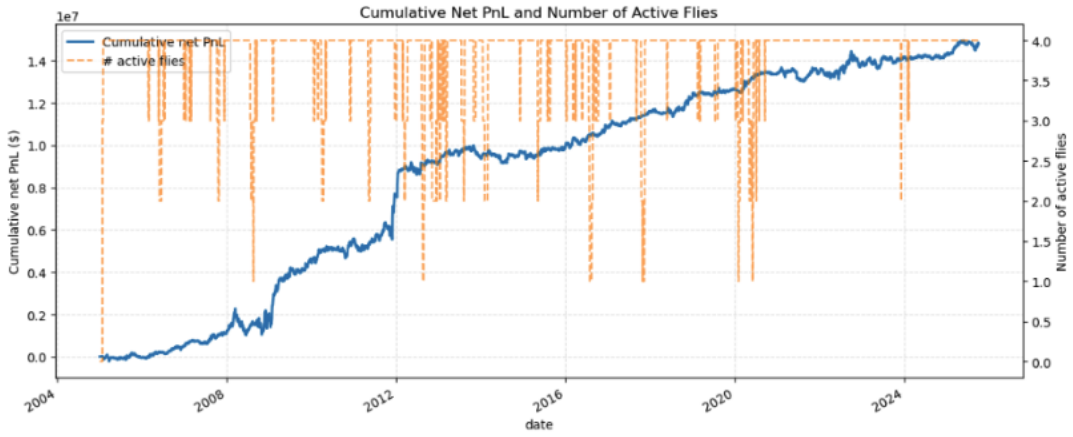
Part D

For Part D, our key intermediate results are the trade-level and portfolio refined signal as we move on from the intermediate refinements with:

1) Portfolio construction sanity: fly cap and DV01 logic

In the above logic, we mentioned that we should only have active position-sizing and fly-cap constraints (max 4 flies, unused DV01 parked in SOFR) are actually being enforced in the code and that we are not accidentally exceeding the intended risk budget.

Using the cumulative net Pnl vs active flies plot, for the finalised version, this has been achieved:



On this, we can see that the average number of active position is less than 4, as per design.

2) Risk/return and horizon sanity - are the stats generally consistent with signal design?

On this front, referring to the below results that are presented from B1 for the tighter, cost-aware entry/exit thresholds (B1), we expect few but longer-lived trades (order of 1–3 months). The final implementation shows average holding period of about 69 days, but the turnover of 13.8 trips per year, is in line with the 1-3 mean reversion horizon.

3) Var vs Volatility

With annual volatility around 6–6.5%, the implied daily volatility is roughly 0.4%, meaning that a 99% normal VaR would then be about 2.3 × 0.4% be about 0.9%. On the other hand, the reported VaR(99%) (1.0%) and ES(99%) (1.7%), which is larger but not implausible.

4) Versioning Sanity checks with Baseline implementation

Comparing the US-only (net Sharpe = 0.49, max DD = −27.7%) with the multi-country spot portfolio (net Sharpe = 0.83, max DD = −43%). This indicates that Sharpe improves meaningfully, which is consistent with having more independent curve opportunities. On the other hand, Volatility and drawdown also

increase (vol 6.1% to 7.1%, DD 27.7% to\ 42.9%), which is what we expect when we run up to four flies simultaneously instead of one: more risk is deployed even if each fly is DV01-neutral. Resultantly, we have higher risk and higher returns with better Sharpe, which is directionally consistent.

5) Gross performance vs net profmrance (transaction cost sanity)

In the multi-country version, gross Sharpe of 0.95 falls to net Sharpe to 0.83, and total costs are about 1.9m with an average daily cost of approx 358. This magnitude is reasonable given about 14 round trips (turnover) per year in our multi-country futures / bond trading. If the drop.

6) Trade Level PnL structure reasonability.

Looking at hteh final (multi-country + futures + B1–B3) trade stats:

- Hit rate of 72.7% per trade
- Avg winning trade of 103k
- Average losing trade of -71k

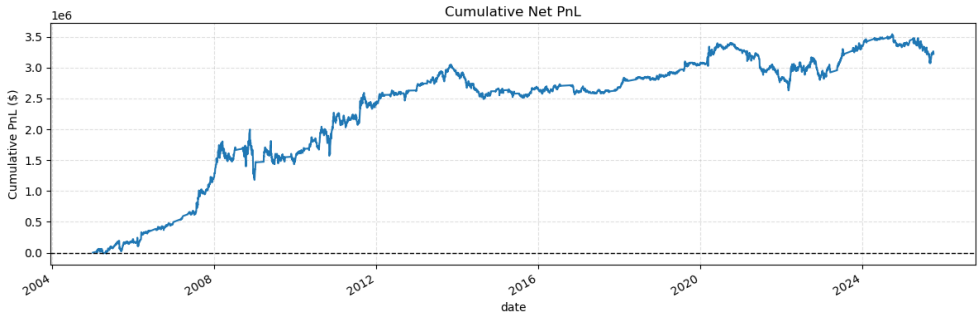
This indicates that adding cost-aware thresholds and the IRS confirmation, we expect fewer higher conviction trades, so win-rate and average win should both improve relative to early baselines.

7) Crisis-Regime behaviour (time-series sanities)

Inspecting PnL paths around 2008, 2013 and 2020, all versions suffer drawdowns in major rate shocks, but do not completely blow up. Looking specifically at the finalised version, we have smaller percentage drawdowns.

Part E)

1st version with the only systematic signal trading. Only US curve (B1 Refinement)



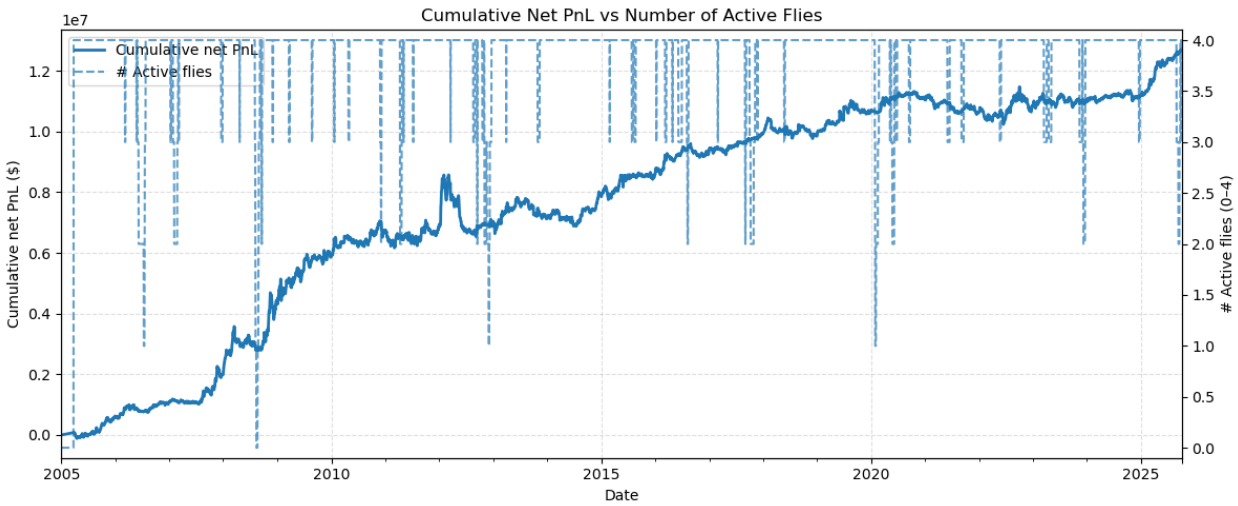

```
=== Core performance stats (US-only, net) ===

Ann. return (net, % per year)          3.0015
Ann. volatility (net, % per year)      6.1394
Sharpe (net, unitless)                 0.4889
Max drawdown (capital, %)            -27.7494
Calmar ratio (unitless)                0.1082
Sortino ratio (unitless)               0.5563
Hit rate (% of days profitable)       59.5564
Avg. daily win (capital %)             0.2022
Avg. daily loss (capital %)           -0.2683
VaR(99%, daily, % cap)                -1.0907
ES(99%, daily, % cap)                 -1.6980
VaR(95%, daily, % cap)                -0.5347
ES(95%, daily, % cap)                 -0.9392
dtype: float64
```

The first implementation evaluates the behaviour of the pure signal engine applied only to the US 2s–5s–10s fly, executed via spot-curve DV01-neutral positions. This version uses the refined signal pipeline introduced in Part B1. In general the cumulative PnL from 2004-24 shows a long and relatively consistent upward trajectory, particularly strong from 2008–2015, followed by a slower but still positive growth phase. Performance flattens post 2020 with US curve volatility becoming more macro driven and the mean reversion in the belly becomes less stable. Sharpe is still solid, and indicates that this is still solid, and the 59.6 hit rate indicates that this z-score logic remains solid with solid dislocations.

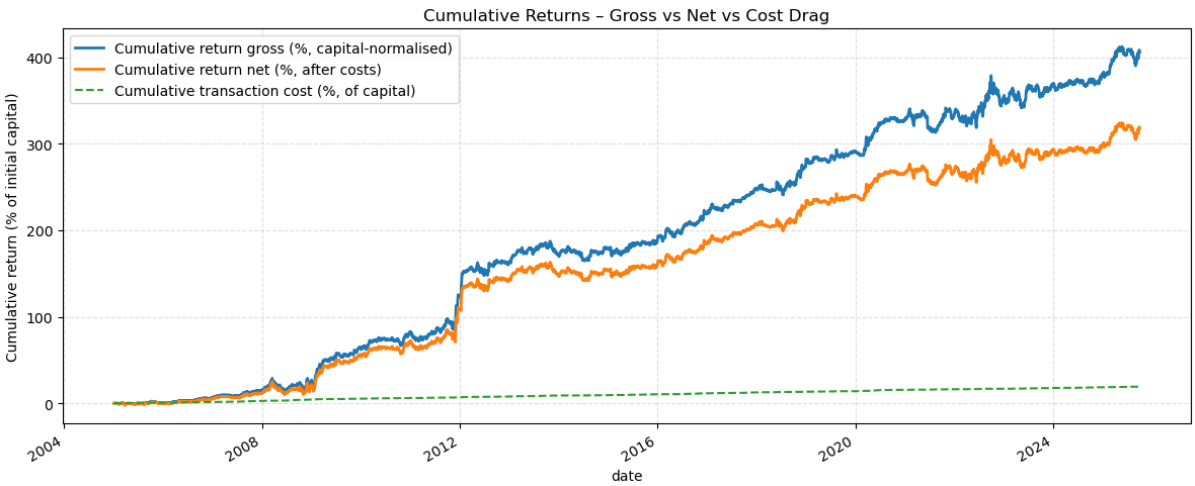
2nd finalised version (No futures) (B1,B2, B3.5)

```
=== Core performance stats (net, after costs) ===
ann_ret_capital_gross    0.0680
ann_vol_capital_gross    0.0714
sharpe_gross             0.9529
ann_ret_capital_net      0.0591
ann_vol_capital_net      0.0715
sharpe_net               0.8264
max_drawdown_net_pct    -0.4297
calmar_net               0.1375
sortino_net              1.1418
hit_rate_net             0.5284
avg_win_net              0.0029
avg_loss_net            -0.0028
VaR_95_net               0.0061
ES_95_net                0.0100
VaR_99_net               0.0118
ES_99_net                0.0174
dtype: float64
```



The performance of the refined strategy shows a balanced but still imperfect risk-return profile: a net annualized return of 5.9% against a 7.15% volatility yields a solid Sharpe of 0.83, and the Sortino of 1.14 indicates relatively well-controlled downside variation. However, the strategy continues to exhibit a large maximum drawdown of roughly 43%, reflecting long periods of slow recovery despite generally stable month-to-month PnL. The hit rate of 52.8% combined with a slightly favorable win–loss asymmetry suggests that the signal has genuine directional power, but remains sensitive to noise and regime shifts. Tail-risk metrics (VaR/ES) remain modest, highlighting that losses come from persistence rather than shocks. The cumulative net PnL curve confirms this interpretation which is steady compounding over the long run, punctuated by extended flat periods then indicating that while refinements have improved stability, the strategy still relies on further diversification and confirmation layers to fully control drawdowns.

Finalised Version (with futures and B1, B2, B3)

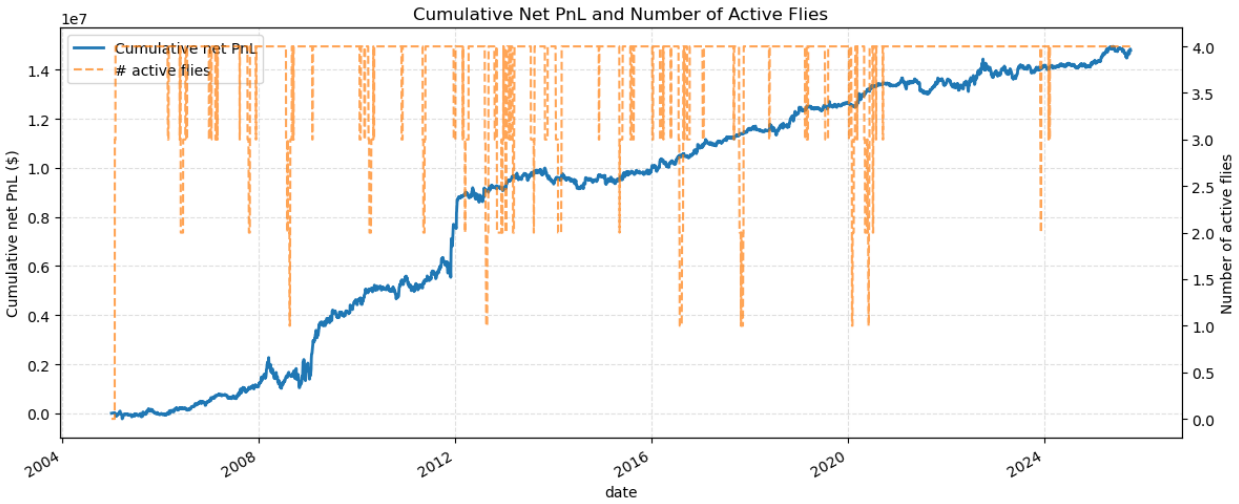


The finalised version of the strategy(combining futures implementation with the B1, B2, and B3 refinements) produces a much more stable and scalable PnL profile, with smooth long-term compounding across multiple regimes and a notably tighter gap between gross and net returns compared to earlier versions. The inclusion of cross-sovereign flies and futures reduces dependence on any single curve, while cost-sensitive thresholds and dual-curve confirmation filters significantly cut down on noise-driven entries and unnecessary turnover. This is reflected in the high hit rate (73%), long average holding periods (~70 days), and relatively modest daily VaR, indicating that returns are driven by persistent macro dislocations rather than short-term volatility. Although drawdowns still occur, they are shallower and recover more quickly thanks to diversified exposure and fewer false positives.

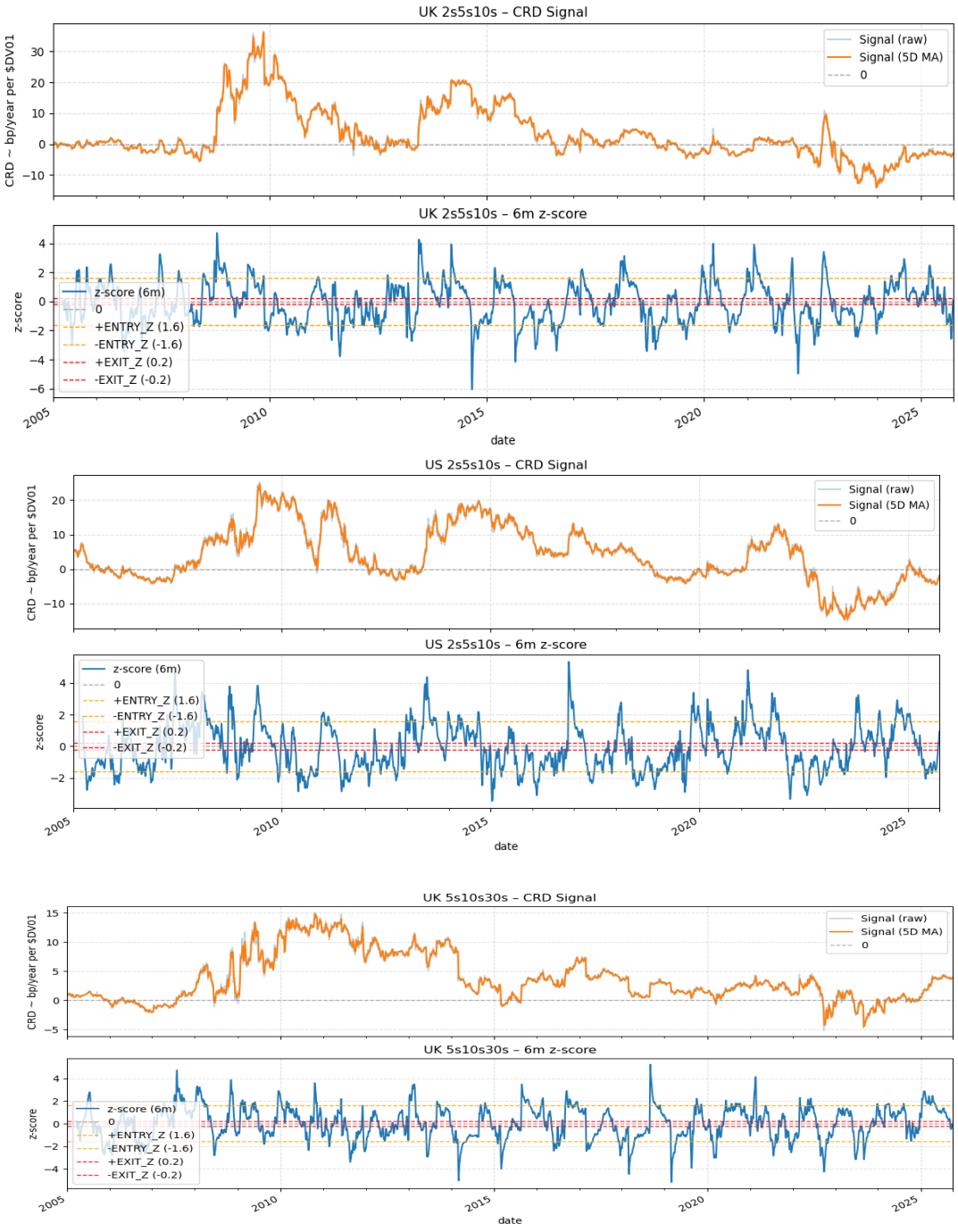
Ann. return (net, % per year)	6.78
Ann. volatility (net, % per year)	6.38
Sharpe (net, unitless)	1.06
Max drawdown (capital, %)	-8.50
Calmar ratio (unitless)	0.80
Sortino ratio (unitless)	1.64
Hit rate (% of trades profitable)	72.70
Avg. PnL per winning trade ($)	103394.03
Avg. PnL per losing trade ($)	-70398.03
Avg. holding period (days)	69.48
Round trips per year (turnover)	13.76
Total transaction cost ($)	1918562.50
Avg. daily transaction cost ($/day)	357.61
VaR(99%, daily, % of capital)	-1.04
ES(99%, daily, % of capital)	-1.37
VaR(95%, daily, % of capital)	-0.50
ES(95%, daily, % of capital)	-0.83
Number of round-trip trades	293.00
dtype: float64	

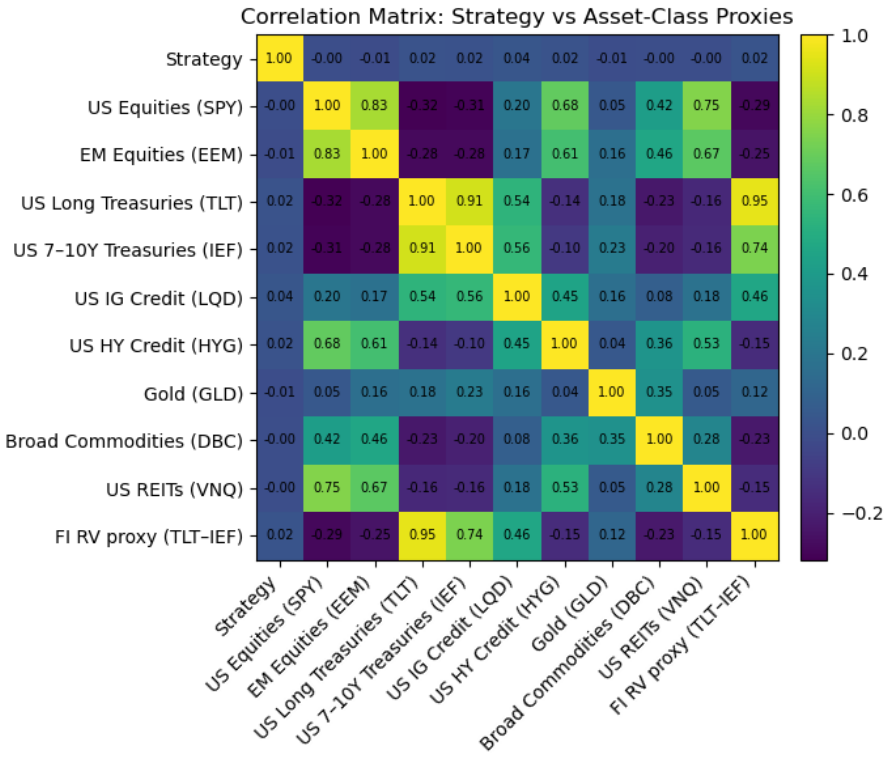
Overall, the combined refinements transform the initial directional signal into a robust, multi-curve trading framework with more consistent risk-adjusted performance and materially improved net profitability.

Looking at the below, the trade is commonly active, except periods of relatively flat bellies, as shown. If not, we are invested in the SOFR.



Looking at the signals of the two most commonly executed trades, it is reasonable with the US and UK being the biggest profit drivers from market dislocations, where we benefit from this mean-reversion.





Evidently, the finalised version does not correlate with other major asset classes, represented by the ETF proxies.

=====Appendix====
Other possible refinements

Alt 1:
To further reduce false-positive entries triggered by noisy spikes in the CRD signal, we introduce a simple but effective "signal slope check." Instead of entering a position merely because the smoothed CRD signal crosses the threshold, we require the direction of the signal to be consistent. Specifically, we compute a short-term slope indicator， e.g., a 5-day moving average of the signal's daily change and only enter a long (short) fly when both the signal level and the slope are positive (negative). This filters out transient, one-day jumps that would otherwise generate unprofitable trades and ensures that we participate only when the curve carry/roll dynamic is trending in the expected direction. In practice, this slope filter substantially reduces noise-driven trades, improves selectivity, and strengthens risk-adjusted performance with minimal complexity. **This was not implemented as it can easily "double gate" the signal, such that sharp, fast mean-reversion trades are missed. It is also fragile to how we selected the lookback, which makes behaviour quite sensitive.**

Alt 2:
In addition to the standard 2s–5s–10s fly, we generalize the CRD signal to multiple tenor triplets to identify which "belly" offers the strongest relative-value opportunity at each point in time.We compute parallel CRD signals for combinations such as 3s–7s–20s, 5s–10s–30s, 2s–3s–5s, and 10s–20s–30s. Each fly has its own DV01-neutral weights and its own carry/roll-down profile, capturing different parts of the curve slope and curvature.On each trading day, we select the fly whose signal has the highest absolute CRD value (or the best Sharpe over a short rolling window). This cross-sectional selection helps avoid relying on a single belly structure and allows the strategy to dynamically rotate into whichever part of the curve is currently most mis-priced.This refinement is simple to implement yet meaningfully increases robustness, diversifies curve exposure, and reduces reliance on any single fly structure. It also naturally captures structural shifts in the yield curve where different bellies become rich or cheap over time. **We have a huge degree of freedom problem, with lots of overlapping flies sharing the same tenors, which makes the backtest very curve fitted. We need more curves if we were to implement tjis.**

Alt 3:
Introduces a risk-management overlay through volatility-scaled position sizing.Instead of holding a fixed DV01, the strategy adjusts its fly exposure based on a 20-day rolling volatility estimate of fly P&L.During high-volatility environments (e.g., MOVE spikes), the position size is automatically reduced, while in stable carry-rich periods it is increased. This keeps daily risk approximately constant, reduces tail losses, and results in materially better drawdown control and Sharpe ratio without changing the underlying CRD signal. **We need a stable estimate of Pnl vol, which is noisy for a small number of trades and short history. We also need a position sizing feedback loop - as the shiraz shrinking after losses or voll spikes can make the realised Pnl path very sensitive to the exact vol window.**