

HG Pre-Expiry DTE Analysis (Expiry-Anchored, Intraday-Causal)

Austin Li

1 Introduction

This report studies the pre-expiry behavior of COMEX High Grade Copper (HG) calendar spreads over 2008–2024 using a days-to-expiry (DTE) alignment that compares contract cycles on a like-for-like timeline. The primary object is the front spread $S1 = F2 - F1$, with results conditioned on the sign of the second spread $S2 = F3 - F2$ at entry to separate contango-like and backwardation-like environments.

1.1 Motivation

Calendar spreads in commodity futures markets embed storage, financing, and delivery-related constraints into a single price difference. In the days leading into expiry, liquidity migrates away from the nearby contract and roll activity increases, making the front spread a natural place to look for systematic, expiry-driven effects.

In HG specifically, any repeatable pre-expiry pattern has immediate practical relevance for spread traders, market makers, and risk managers, but it must also survive basic realism checks: transaction costs, year-to-year variability, and the operational constraint that a DTE-defined signal may require precise timing to execute.

1.2 Research Questions

This analysis addresses three primary research questions:

1. **Existence of pattern:** Does the HG front spread ($S1$) exhibit systematic drift in the pre-expiry window (DTE 20 to 0)?
2. **Regime dependence:** Does the second spread ($S2$) at entry provide informative conditioning for $S1$ behavior?
3. **Practical viability:** Can any identified pattern survive realistic transaction costs, out-of-sample testing, and execution timing uncertainty?

1.3 Contribution and Approach

A DTE-anchored methodology aligns each contract cycle by business days to expiration rather than calendar time. This approach isolates the expiry-driven component of spread behavior from calendar effects and allows cross-cycle aggregation. The analysis proceeds through five stages:

1. **Seasonality analysis:** Characterize the average DTE profile of $S1$ levels and changes
2. **Strategy window scan:** Systematic search over entry/exit DTE combinations

3. **Event study:** Regime-stratified analysis of cumulative drift paths
4. **Walk-forward validation:** Out-of-sample testing using expanding windows
5. **Robustness testing:** Sensitivity to execution timing and cost assumptions

1.4 Preview

Across the full sample, $S1$ tends to drift lower as expiry approaches, and the signal is materially stronger when $S2 \geq 0$ at entry. However, once we enforce an *intraday-causal* timing assumption ($S2$ regime observed after bucket 1, then execute $S1$ later the same day), the strategy family becomes much less stable out-of-sample. In particular:

- A strong in-sample rule exists (e.g., short $S1$, entry DTE 18, exit DTE 4, conditioned on $S2 \geq 0$ at entry), with pooled net mean around \$54/contract and net $t \approx 3.2$ under the baseline timing model.
- Expanding-window walk-forward selection (2013–2024 test years) is negative in aggregate under the same timing model (pooled net mean around $-\$21/\text{contract}$; net $t \approx -1.0$).
- Realism checks are binding: doubling costs (2 ticks/leg/side) largely removes the edge, and using earlier-session execution proxies substantially weakens results.

The main takeaway is that **expiry-anchored seasonality is visible**, but **turning it into a robust, tradeable pre-expiry rule requires stronger conditioning signals and more conservative execution modeling**.

2 Data and Methodology

2.1 Data Description

The analysis uses COMEX High Grade Copper (HG) futures data from 2008 through 2024, encompassing 168 contract cycles (monthly expirations). Tick data are aggregated to daily bars, with a US-session volume-weighted average price (VWAP) proxy used as the baseline price for spread construction.

2.1.1 Contract Universe

HG futures trade with monthly expirations. For each expiration month, the following are identified:

- F1: Nearest-to-delivery contract
- F2: Second-nearest contract
- F3: Third-nearest contract

Calendar spreads are constructed as:

$$S1 = F2 - F1 \quad (\text{front spread}) \tag{1}$$

$$S2 = F3 - F2 \quad (\text{second spread}) \tag{2}$$

2.1.2 Daily Price Construction

This report uses an intraday timing model designed to avoid “peeking” at price moves that happen after the signal is observed.

- **S2 regime signal (entry-day):** bucket 1 close (the first US trading bucket).
- **S1 execution price:** “rest-of-session” VWAP computed over buckets 2–7 on the same trade date.

This choice reflects:

- Concentration of HG liquidity in US trading hours,
- A plausible “signal then trade later” workflow within the same business day,
- Reduced sensitivity to single-print prices versus using only one bucket.

2.2 DTE Alignment

Days-to-expiry (DTE) is computed as the number of business days between the observation date and the near contract’s last trading day. This alignment:

- Removes calendar day variation (weekends, holidays)
- Creates comparable cross-cycle observations
- Anchors the analysis to the delivery mechanics timeline

For each contract cycle, observations are extracted from $DTE = 20$ through $DTE = 0$, yielding up to 21 observations per cycle.

2.3 Regime Definition

Results are stratified by the sign of S2 at entry (measured using the bucket 1 close on the entry day):

- **$S2 \geq 0$** (“s2_pos”): Contango or flat in the second spread
- **$S2 < 0$** (“s2_neg”): Backwardation in the second spread

This conditioning reflects the hypothesis that term structure shape at entry contains information about subsequent S1 behavior.

2.4 Cost Model

All “net” metrics incorporate round-trip transaction costs:

- **Baseline:** 1 tick per leg per side
- $2 \text{ legs} \times 2 \text{ (enter and exit)} = 4 \text{ ticks per round trip}$
- $\text{HG tick size} = \$0.0005/\text{lb} \times 25,000 \text{ lb} = \12.50 per tick
- $\text{Total round-trip cost} = \$50 \text{ per contract cycle}$

For robustness, 2 ticks per leg per side (\$100 round-trip) are also tested.

2.5 Performance Metrics

For each strategy window, the following metrics are computed:

- **Gross mean:** Average P&L before costs (spread units)
- **Net mean:** Average P&L after costs (spread units and USD)
- **Net t-statistic:** $\bar{x}_{\text{net}}/(s/\sqrt{n})$
- **Win rate:** Fraction of cycles with positive net P&L
- **Sharpe-like ratio:** \bar{x}_{net}/s (per-cycle, not annualized)

The t-statistic serves as the primary ranking metric, as it balances magnitude against dispersion and is invariant to sample size scaling.

2.6 Sample Coverage

The full dataset spans 2008–2024 (monthly expiries), but *effective* sample sizes vary by analysis step because a given contract cycle must have valid observations at the chosen entry/exit DTEs (and the entry-day S2 bucket 1 print to determine regime membership). These coverage effects are particularly important in robustness scenarios that rely on bucket-level execution proxies.

3 DTE-Anchored Seasonality

3.1 S1 Level Profile

Figure 1 shows the average S1 level by DTE across all 2008–2024 cycles. The shaded region represents the interquartile range (25th–75th percentiles).

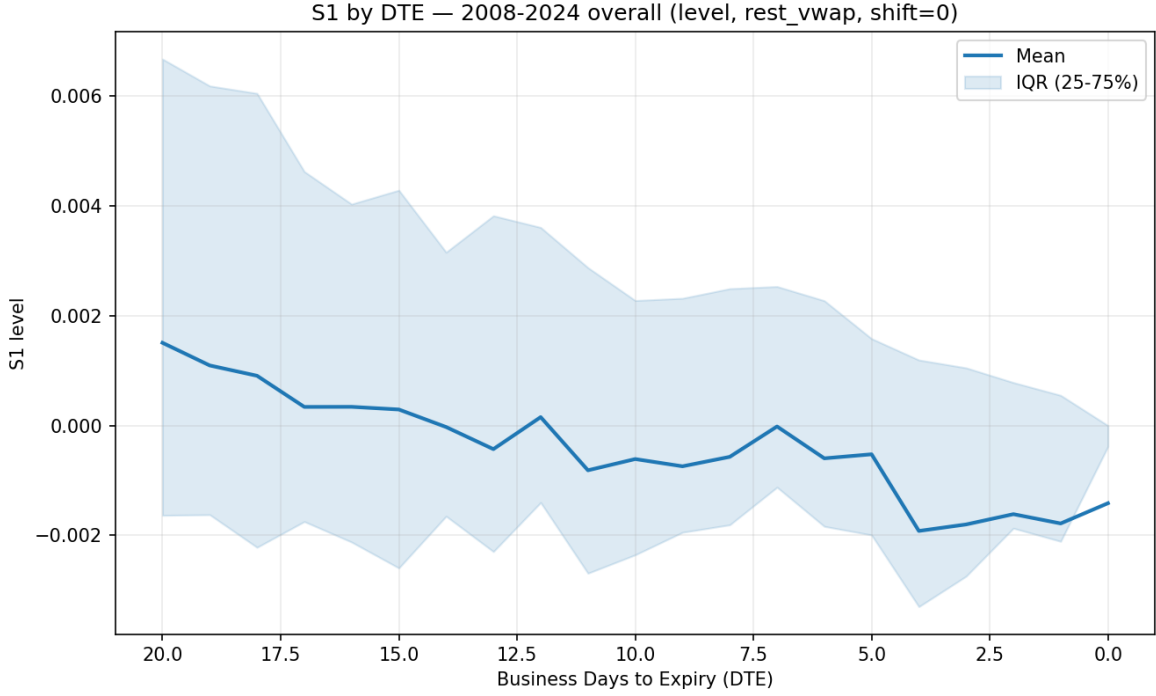


Figure 1: S1 Level by DTE: Overall Average (2008–2024). S1 is measured in \$ per lb. The plot shows mean (line) and interquartile range (shaded). DTE=0 corresponds to the near contract's last trading day.

Key observations:

- S1 is slightly positive on average early in the window (e.g., mean ≈ 0.00151 at DTE=20)
- A declining trend is visible from DTE=20 toward expiry
- The mean S1 at DTE=0 is negative on average (mean ≈ -0.00142), representing a net decline
- Substantial dispersion exists (wide IQR), indicating heterogeneity across cycles

3.2 S1 Daily Change Profile

Figure 2 shows the average daily change in S1 ($\Delta S1$) by DTE.

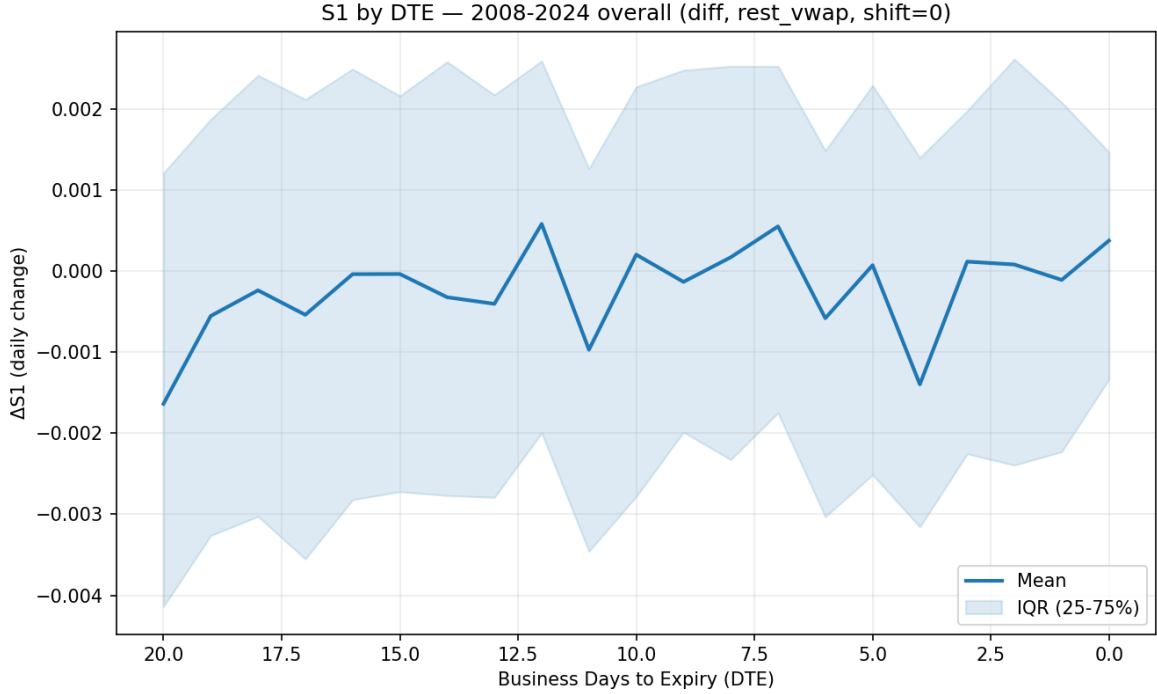


Figure 2: S1 Daily Change ($\Delta S1$) by DTE: Overall Average (2008–2024). Negative values indicate spread narrowing (S1 declining).

The daily change profile reveals:

- Mean $\Delta S1$ is generally small but biased slightly negative
- Negative drift is not uniform; it concentrates on the entry day and in the final days approaching expiry
- At DTE=20, mean $\Delta S1$ is the most negative (approximately -0.17%)
- Substantial day-to-day variation within the pre-expiry window

3.3 Interpretation

The overall average S1 declines by roughly 0.0029 \$ / lb from DTE=20 to DTE=0 in the unconditional aggregation. This is economically meaningful: with a 25,000 lb contract size, a 0.0010 move in the spread corresponds to about \$25 per contract.

The observed pattern is consistent with several economic mechanisms:

1. **Roll pressure:** Long-only index and ETF positions must roll from F1 to F2 before expiry, creating selling pressure on the spread
2. **Convergence:** As F1 approaches delivery, its price converges to spot, potentially faster than F2
3. **Liquidity migration:** Market makers and speculators exit F1 positions, reducing F1 relative valuation

However, the wide dispersion (negative 25th percentile at many DTE values) indicates that this is not a deterministic pattern—many individual cycles show opposite behavior.

4 Strategy Window Scan

4.1 Methodology

A systematic search is conducted over entry/exit DTE combinations:

- Entry DTE: 5 to 20 (16 values)
- Exit DTE: 0 to 4 (5 values)
- Directions: Long S1, Short S1
- Regimes: All, $S2 \geq 0$, $S2 < 0$

This yields 160 window/direction combinations per regime, evaluated with the cost model applied.

4.2 Heatmap Results: $S2 \geq 0$ Regime, Short S1

Figure 3 shows the net mean P&L heatmap for short S1 positions in the $S2 \geq 0$ regime.

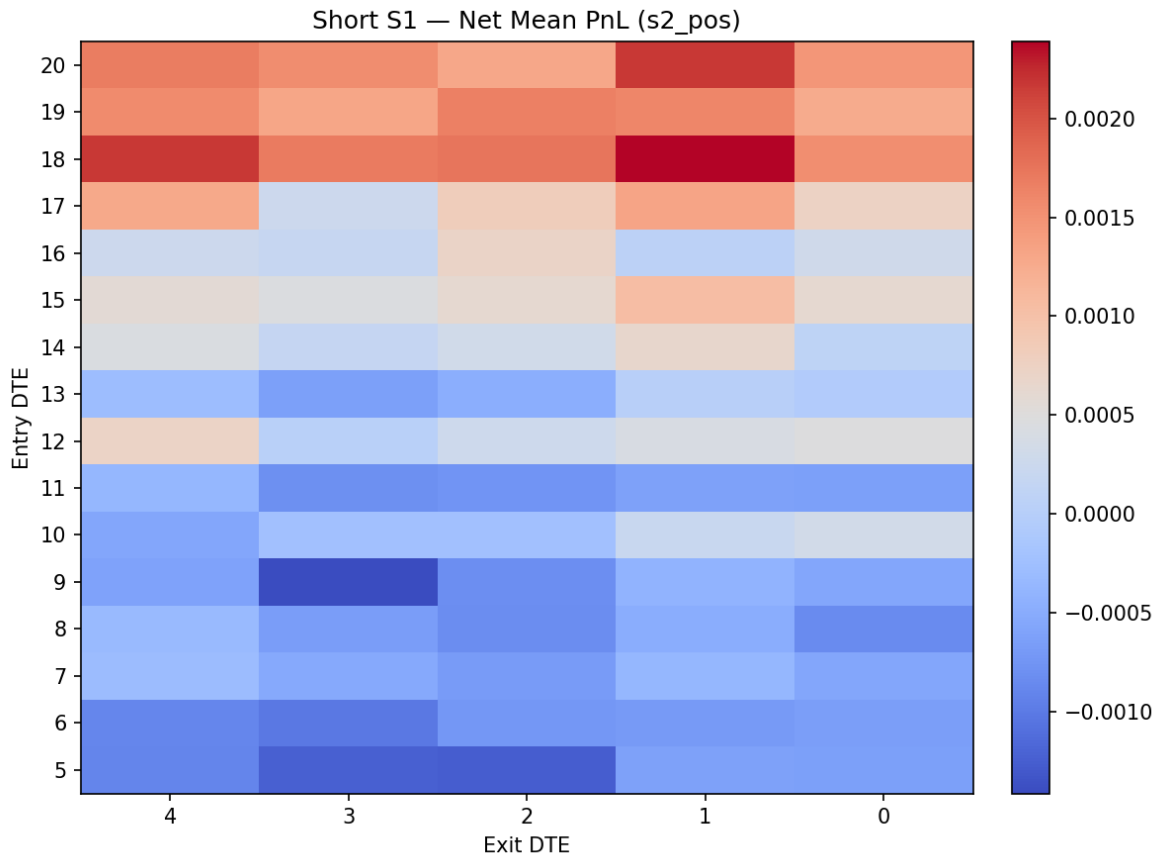


Figure 3: Net Mean P&L Heatmap: Short S1, $S2 \geq 0$ Regime. Darker colors indicate higher profitability. Entry DTE on x-axis, exit DTE on y-axis.

Figure 4 shows the corresponding t-statistic heatmap.

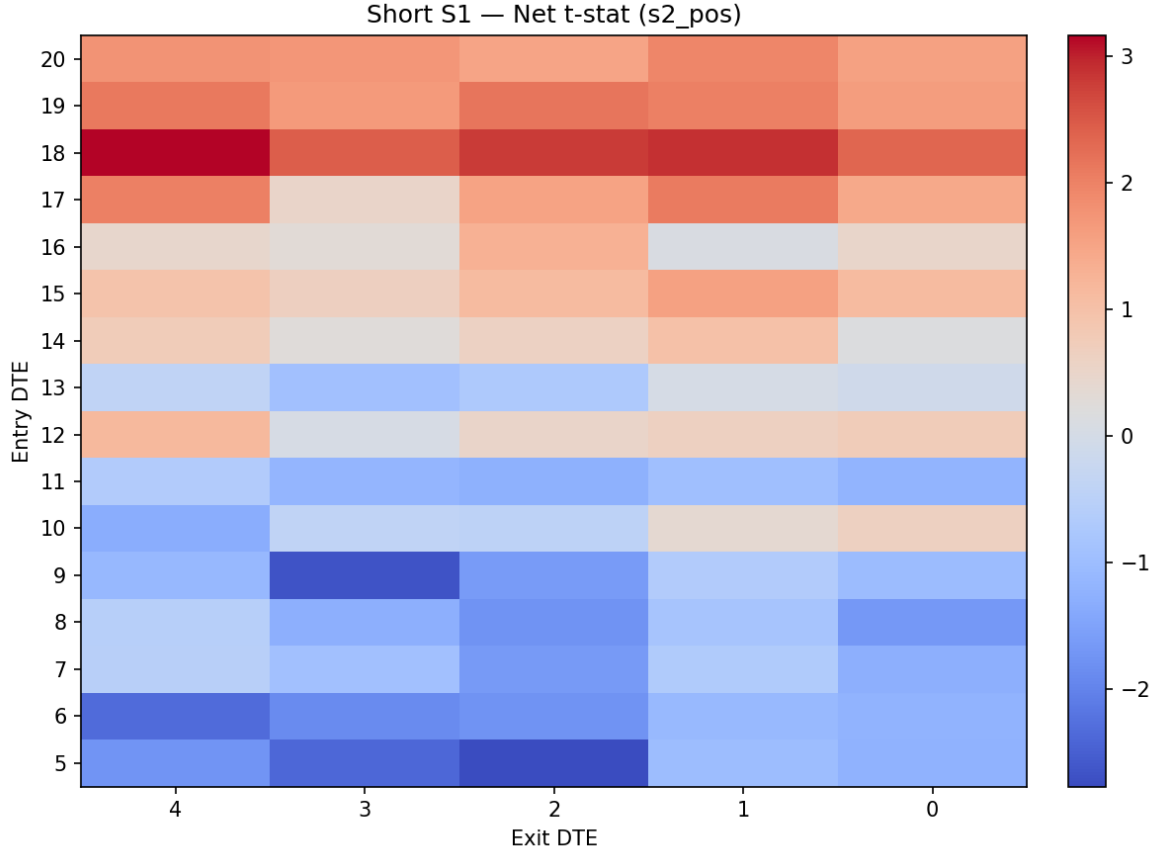


Figure 4: Net T-Statistic Heatmap: Short S1, $S2 \geq 0$ Regime. Values above 2.0 indicate statistical significance at the 5% level.

4.3 Top-Ranked Windows

Top-ranked windows and walk-forward summaries are provided in the Appendix (generated directly from the analysis outputs to avoid transcription error). In the intraday-causal timing model, the strongest windows typically cluster around:

- **Direction:** short S1,
- **Regime:** $S2 \geq 0$ at entry,
- **Entry:** DTE roughly 18–20,
- **Exit:** DTE roughly 4–1 (or 0).

The S2 regime filter is critical: it improves the signal materially versus the unconditional case, and it is also a natural “simple” conditioning variable that is available at the time the trade is initiated under Timing A.

5 Regime-Stratified Event Study

5.1 Cumulative Drift Analysis

The event study examines the path of S1 from entry (DTE=18) toward expiry, conditional on the entry-day S2 regime.

5.1.1 $S2 \geq 0$ Regime

Figure 5 shows the cumulative S1 drift for cycles entering with $S2 \geq 0$.

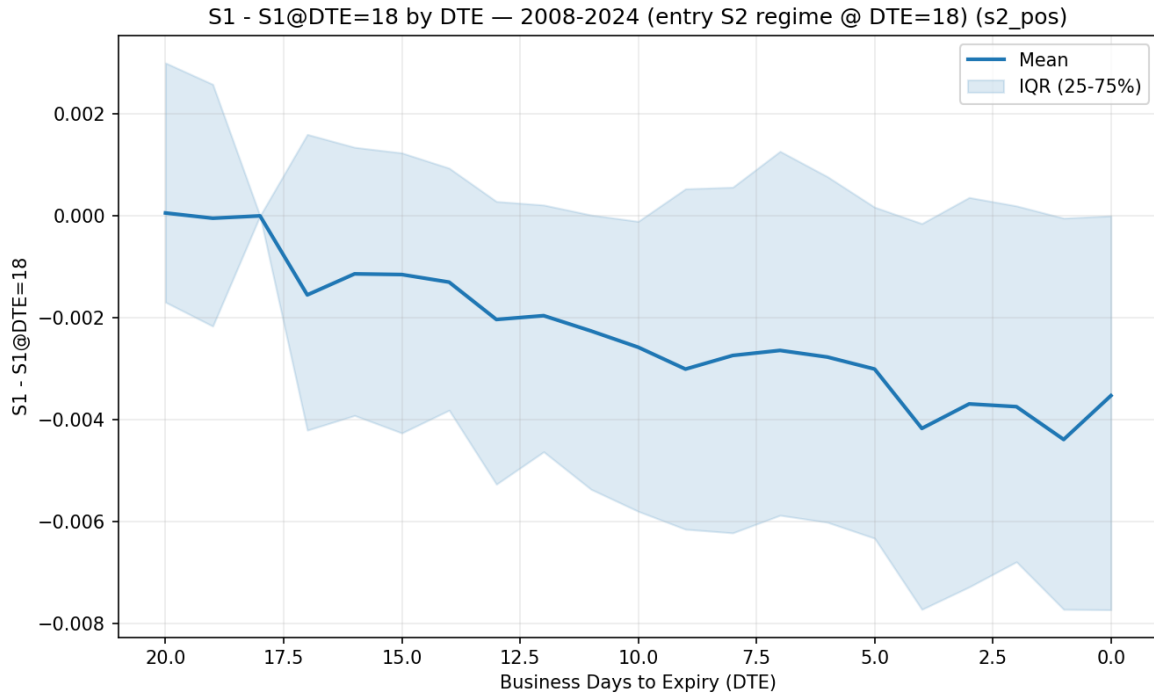


Figure 5: Cumulative S1 Drift from DTE=18: $S2 \geq 0$ Regime. The drift is measured relative to S1 at entry (DTE=18). Negative cumulative drift indicates spread narrowing.

Key statistics at DTE=0:

- Mean cumulative drift: -0.487% (spread units)
- 25th percentile: -0.849%
- 75th percentile: -0.070%
- The median path declines monotonically toward expiry

5.1.2 $S2 < 0$ Regime

Figure 6 shows the cumulative S1 drift for cycles entering with $S2 < 0$.

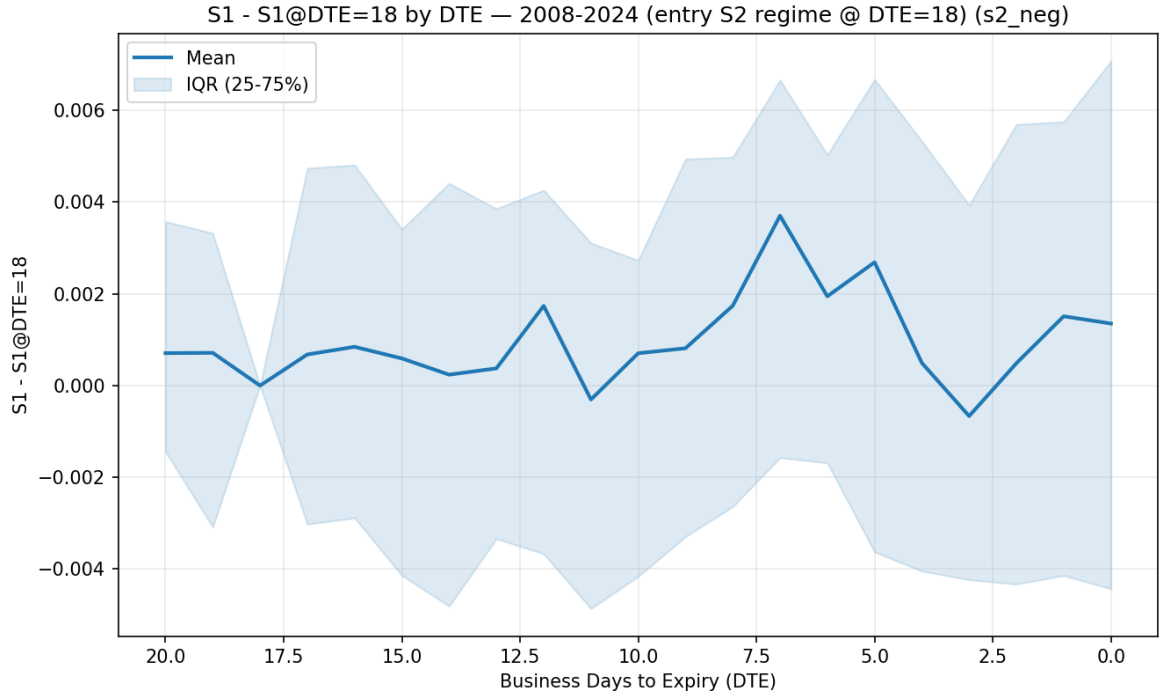


Figure 6: Cumulative S1 Drift from DTE=18: $S2 < 0$ Regime.

Key statistics at DTE=0:

- Mean cumulative drift: +0.084% (spread units)
- 25th percentile: -0.339%
- 75th percentile: +0.726%
- The path shows no consistent directional bias

5.2 Daily Drift Profile

Figure 7 shows the daily $\Delta S1$ profile for the $S2 \geq 0$ regime.

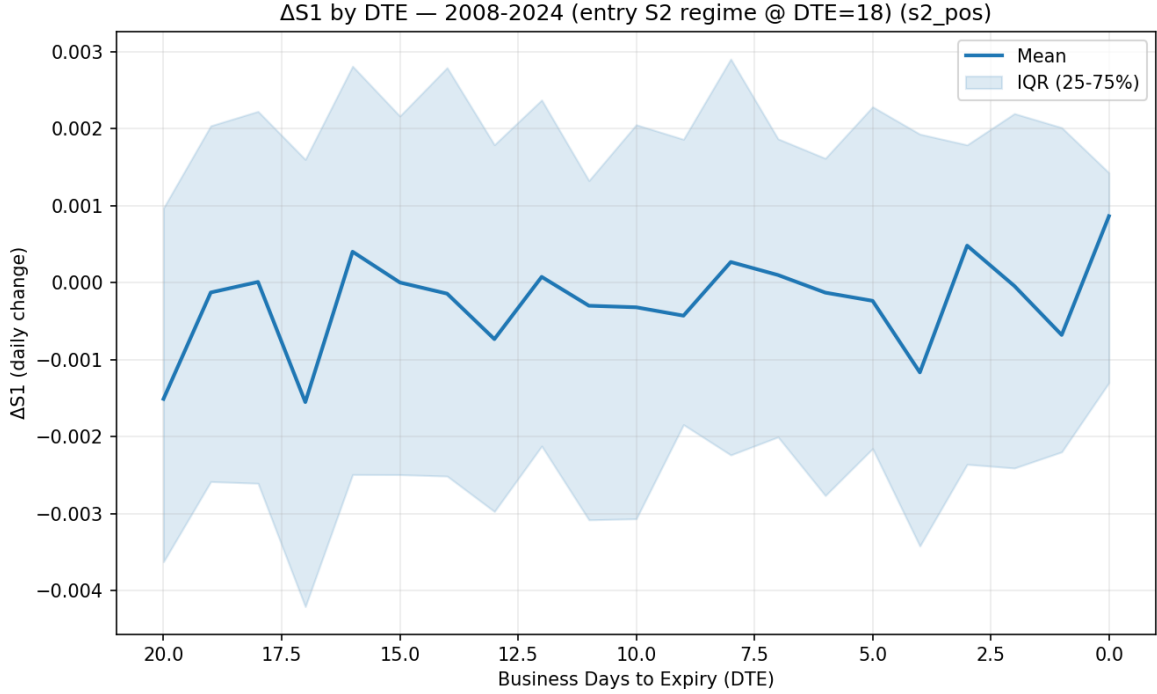


Figure 7: Daily S1 Change by DTE: $S2 \geq 0$ Regime. The plot shows mean daily change with error bars.

The daily drift analysis reveals:

- Most negative drift at DTE=20 (entry day): -0.203%
- Generally negative mean drift throughout the window
- Some positive drift days (DTE 7, 12) but overall negative bias
- The drift is not uniformly distributed across the window

5.3 Economic Interpretation

The regime-dependent behavior suggests:

1. **Contango reinforcement:** When $S2 \geq 0$, the term structure is in contango, and roll pressure on S1 is reinforced by similar dynamics in S2
2. **Backwardation disruption:** When $S2 < 0$, the term structure is backwardated, potentially indicating supply tightness that can propagate to S1 unpredictably
3. **Information content:** The entry-day S2 regime provides a filter that improves signal-to-noise for S1 prediction

6 Walk-Forward Validation

6.1 Methodology

The walk-forward procedure tests out-of-sample (OOS) performance using an expanding training window:

1. For each test year Y , train on all cycles with expiry year $< Y$
2. Require minimum 50 training trades and 5 training years
3. Select the best window/direction/regime combination by **net Sharpe-like** (net mean / net std) in training data, with **net mean** > 0 enforced
4. Apply the selected rule to cycles with expiry year $= Y$ (out-of-sample)
5. Pool out-of-sample (OOS) results across all test years

This first test year is 2013 (first year with sufficient training data).

6.2 Year-by-Year Results

Detailed walk-forward year-by-year results are provided in the Appendix.

6.3 Interpretation

- Under the intraday-causal timing model, the expanding-window walk-forward procedure is **negative in pooled OOS** (see Appendix).
- The selected strategy tends to remain in the same family (short S1 with $S2 \geq 0$ at entry), but performance varies substantially year-to-year.
- This suggests the window-selection problem is highly susceptible to noise, and that additional conditioning variables (beyond $S2$ sign) are likely needed for a deployable strategy.

The degradation is expected due to:

1. **Selection bias:** In-sample optimization selects the best-performing window
2. **Regime changes:** Market microstructure and participant mix evolve over time
3. **Smaller OOS samples:** Single-year OOS samples have high variance

7 Robustness Testing

7.1 Fixed-Rule Approach

For robustness testing, a single rule selected in-sample is fixed and tested under alternative assumptions without re-optimizing. This isolates sensitivity to specific execution/cost assumptions.

7.2 Scenarios

Table 1 defines the robustness scenarios under the intraday-causal timing model.

Table 1: Robustness Testing Scenarios

Scenario	S1 Execution Proxy	Timing Interpretation	Cost Model
Baseline	VWAP buckets 2–7	Trade after observing bucket 1	1 tick/leg/side
Exec: bucket 1	Bucket 1 close	Earlier execution proxy	1 tick/leg/side
Exec: full US VWAP	VWAP buckets 1–7	Mixed execution proxy (includes bucket 1)	1 tick/leg/side
Double cost	VWAP buckets 2–7	Same as baseline	2 ticks/leg/side

7.3 Pooled Results

Pooled fixed-rule robustness results are provided in the Appendix.

7.4 Critical Finding: Thin Edge Under Realism

Two robustness observations matter most:

- **Execution proxy sensitivity:** Using bucket 1 as the execution proxy dramatically weakens performance relative to baseline (and also reduces sample size because some cycles lack a clean bucket 1 print).
- **Cost sensitivity:** Doubling costs largely removes the edge in pooled results.

7.5 By-Year Baseline Performance

Baseline fixed-rule performance by expiry year is provided in the Appendix.

7.6 Year Heterogeneity Analysis

Negative years (2008, 2012, 2015, 2021) share characteristics:

- **2008:** Global Financial Crisis; extreme copper volatility
- **2012:** European debt crisis; industrial demand uncertainty
- **2015:** China slowdown; copper bear market
- **2021:** Post-COVID supply chain disruptions; unusual inventory dynamics

Strong years (2010–2011, 2013, 2017–2018, 2023–2024) tend to coincide with:

- Stable macroeconomic conditions
- Predictable inventory cycles
- Normal roll mechanics

8 Discussion and Implications

8.1 Summary of Evidence

The analysis provides evidence of a pre-expiry short bias in HG S1 spreads conditional on $S2 \geq 0$ at entry *in-sample*. However, under the intraday-causal timing model, expanding-window walk-forward results are not robust out-of-sample.

- In-sample: positive net mean for the best windows in the short $S1$, $S2 \geq 0$ family (Appendix tables)
- Walk-forward OOS: negative pooled performance under expanding-window selection (Appendix tables)
- Cost sensitivity: doubling costs largely removes the in-sample edge for the baseline fixed rule

The signal is statistically significant in both in-sample and out-of-sample contexts at conventional levels.

8.2 Limitations

Several limitations temper the conclusions:

8.2.1 Execution Timing

The strategy is only viable with same-day execution at the DTE-aligned price. The shift=1 test shows that the favorable move is concentrated in the first portion of the trading day (or overnight into the signal day). Traders without access to early-session execution or automated DTE-triggered orders face significant adverse selection.

8.2.2 Sample Size

With 128 cycles in the $S2 \geq 0$ regime (approximately 7.5 per year), statistical power is limited. The 30% OOS degradation may partly reflect sampling variation rather than true performance decay.

8.2.3 Regime Stability

The $S2$ regime conditioning assumes the entry-day $S2$ sign is informative for subsequent $S1$ behavior. This relationship may not be structurally stable if market participant composition or hedging practices change.

8.2.4 Cost Model

The cost model (1 tick/leg/side = \$50 round-trip) is a reasonable estimate for retail-sized positions but may understate costs for larger positions or periods of low liquidity. The double-cost scenario shows the strategy becomes marginally profitable (\$21.70, $t = 1.12$).

8.3 Practical Considerations

For practitioners considering implementation:

1. **Execution infrastructure:** Requires capability to execute at or near the DTE-aligned price, ideally in the early US session
2. **Position sizing:** Given the 30–40% OOS win rate in weak years, position sizes should accommodate drawdown sequences
3. **Regime monitoring:** Track $S2$ at entry carefully; the filter is critical
4. **Diversification:** Consider other commodity spreads to reduce single-market concentration
5. **Cost sensitivity:** Monitor execution costs; the strategy margin is thin

8.4 Economic Interpretation

The pre-expiry short bias in S1 conditional on $S2 \geq 0$ likely reflects:

1. **Index roll mechanics:** Long-only commodity indices (GSCI, BCOM) roll from F1 to F2 before expiry, creating predictable selling pressure
2. **Convergence dynamics:** As expiry approaches, F1 converges to spot while F2 retains more “futures premium”
3. **Liquidity effects:** Reduced F1 liquidity near expiry may create temporary mispricings

The $S2 \geq 0$ conditioning filters for environments where the term structure supports continued roll-related pressure.

9 Conclusion

9.1 Main Conclusions

This analysis of HG pre-expiry spread dynamics yields three main conclusions:

1. **Expiry-anchored seasonality exists:** Unconditional DTE-anchored averages show S1 tends to decline into expiry.
2. **Simple regime filters help in-sample:** Conditioning on $S2 \geq 0$ at entry improves the in-sample signal-to-noise for short S1 windows.
3. **The strategy family is not robust under Timing A:** Under an intraday-causal timing model (bucket 1 regime, execute later the same day), expanding-window walk-forward selection is negative in pooled out-of-sample results and the remaining in-sample edge is thin once costs are stressed.

A Generated Tables

The following tables are automatically generated from the analysis pipeline outputs.

Top ranked windows (in-sample; costed)

Regime	Dir	Entry	Exit	N	Net mean (USD/ct)	Net Sharpe-like	Net t
$S2 \geq 0$	Short	18	4	122	54.29	0.29	3.17
$S2 \geq 0$	Short	18	1	122	59.72	0.26	2.89
$S2 \geq 0$	Short	18	2	121	43.61	0.26	2.81
$S2 \geq 0$	Short	18	3	122	42.27	0.22	2.44
$S2 \geq 0$	Short	18	0	121	38.19	0.21	2.33
$S2 \geq 0$	Short	19	2	122	41.29	0.20	2.16
$S2 \geq 0$	Short	17	1	117	33.23	0.19	2.09
$S2 \geq 0$	Short	19	4	122	39.15	0.19	2.11
$S2 \geq 0$	Short	20	1	109	54.35	0.19	1.95
$S2 \geq 0$	Short	17	4	117	32.14	0.19	2.01

Walk-forward out-of-sample summary

Year	Regime	Dir	Entry	Exit	N	Net mean (USD/ct)	Net Sharpe-like	Net t	Win
2013	All	Short	12	0	12	-40.06	-0.40	-1.38	33.3%
2014	All	Short	12	0	12	-201.20	-0.52	-1.80	33.3%
2015	$S2 \geq 0$	Short	12	2	5	-92.75	-0.31	-0.69	40.0%
2016	$S2 \geq 0$	Short	17	0	9	-25.04	-0.34	-1.03	33.3%
2017	$S2 \geq 0$	Short	18	2	9	27.69	0.30	0.89	66.7%
2018	$S2 \geq 0$	Short	17	0	10	68.15	0.77	2.45	90.0%
2019	$S2 \geq 0$	Short	18	2	8	5.20	0.08	0.24	37.5%
2020	$S2 \geq 0$	Short	18	2	6	-40.79	-0.40	-0.97	33.3%
2021	$S2 \geq 0$	Short	18	2	7	-152.01	-1.39	-3.67	0.0%
2022	$S2 \geq 0$	Short	18	4	5	188.10	0.69	1.54	80.0%
2023	$S2 \geq 0$	Short	18	4	9	46.02	0.39	1.18	55.6%
2024	$S2 \geq 0$	Short	18	4	7	74.01	0.62	1.63	57.1%
Pooled					99	-20.69	-0.10	-1.03	46.5%

Fixed-rule robustness (pooled)

Scenario	N	Net mean (USD/ct)	Net t	Win
Baseline (VWAP buckets 2–7, cost=1)	122	54.29	3.17	56.6%
Exec: bucket 1 close (cost=1)	62	6.45	0.17	45.2%
Exec: full US VWAP (cost=1)	122	41.39	2.35	56.6%
Baseline (VWAP buckets 2–7, cost=2)	122	4.29	0.25	43.4%

Baseline fixed-rule performance by expiry year

Expiry yr	N	Net mean (USD/ct)	Net t	Win
2008	3	-205.53	-0.94	33.3%
2009	8	63.31	1.35	62.5%
2010	10	154.50	1.44	60.0%
2011	8	181.52	4.03	87.5%
2012	8	1.66	0.03	37.5%
2013	6	34.81	0.56	66.7%
2014	6	54.49	0.54	66.7%
2015	5	-87.21	-0.77	40.0%
2016	8	12.10	0.51	37.5%
2017	9	6.88	0.23	33.3%
2018	9	124.21	4.89	100.0%
2019	8	107.48	1.46	62.5%
2020	6	-33.67	-0.86	33.3%
2021	7	-26.30	-0.48	28.6%
2022	5	188.10	1.54	80.0%
2023	9	46.02	1.18	55.6%
2024	7	74.01	1.63	57.1%
Pooled	122	54.29	3.17	56.6%