# Analytical Results Report: Calendar Spread Dynamics in CME Copper Futures

Comprehensive Analysis of Roll Patterns and Expiry Mechanics
2008–2024

**Abstract**

This report presents comprehensive analytical findings from high-frequency analysis of CME copper futures calendar spreads over 16 years (2008–2024). Using minute-level data aggregated into intraday periods, we investigate whether observed spread widening patterns reflect institutional roll timing decisions or systematic contract expiry mechanics. Our analysis processes 202 futures contracts (410 MB, approximately 450 million data points) to detect 2,737 significant widening events in the front-month spread (S1 = F2 - F1).

The core finding validates the supervisor's hypothesis: when contract month is properly defined as days since previous contract expiry (rather than days since a contract first became F1), we observe dramatic clustering with 63.8% of events occurring in the first seven days ($\chi^2 = 1195.2$, $p < 0.001$). This contrasts sharply with the first-appearance proxy (3.2% clustering), demonstrating fundamental methodological sensitivity. Multi-spread analysis (S1–S11) reveals the pattern systematically repeats across the contract chain, with each spread exhibiting events 28–30 days before its front contract expires.

Critically, volume migration analysis reveals zero correlation between F2/F1 volume ratios and event occurrence (Pearson $r = 0.0043$, $p = 0.369$), with F1 remaining volume-dominant even during events (median ratio 0.426). Term structure analysis shows orderly evolution through the expiry cycle with 74% contango prevalence near expiry and no evidence of forced liquidations.

These findings demonstrate a dual nature in spread dynamics: expiry mechanics create unavoidable systematic patterns in price relationships and timing, while rolling activity successfully remains hidden in volume data. The framework provides quantitative characterization of contract lifecycle dynamics essential for continuous futures construction, basis trading strategies, and market microstructure modeling.

## Contents

**1 Executive Summary** **4**
- 1.1 Research Question . . . 4
- 1.2 Key Findings . . . 4
- 1.3 Interpretation and Implications . . . 4

**2 Research Context** **5**
- 2.1 Futures Roll Patterns and Market Microstructure . . . 5
- 2.2 Original Hypothesis: Institutional Roll Timing . . . 6
- 2.3 Supervisor's Critique: Expiry Mechanics Hypothesis . . . 6
- 2.4 Research Evolution and Testing Strategy . . . 7

**3 Data and Methodology** **7**
- 3.1 Dataset Specifications . . . 7
- 3.2 Intraday Aggregation Framework . . . 8
  - 3.2.1 Bucket Definitions . . . 8
  - 3.2.2 Aggregation Methodology . . . 8

3.2.3 Cross-Midnight Handling . . . 9
3.3 Contract Identification: Deterministic F1–F12 Chain . . . 9
3.4 Calendar Spread Construction . . . 9
3.5 Event Detection Methodology . . . 10
3.5.1 Z-Score Calculation . . . 10
3.5.2 Detection Rule . . . 10
3.5.3 Detection Statistics . . . 10
3.6 Business Day Computation . . . 11
3.7 Data Quality Filters . . . 11
3.8 Multi-Spread Comparative Analysis Framework . . . 11

**4 Core Analysis Results** **12**
4.1 Multi-Spread Comparative Analysis: S1–S11 Patterns . . . 12
4.1.1 Key Observations . . . 12
4.1.2 Correlation Matrix Analysis . . . 12
4.2 S1 Dominance and Importance . . . 13
4.3 Hourly Bucket Patterns . . . 13
4.4 Session Analysis: US, Asia, Europe . . . 14
4.5 Transition Matrix: Sequential Bucket Patterns . . . 14

**5 Exploratory Analysis Results** **15**
5.1 Contract Month Reframing: Definition Matters . . . 15
5.1.1 Results: Dramatic Clustering Under Supervisor's Definition . . . 15
5.1.2 Visual Representation . . . 15
5.1.3 Interpretation . . . 16
5.2 Term Structure Evolution Through Expiry Cycle . . . 17
5.2.1 Objective and Methodology . . . 17
5.2.2 Results: Orderly Evolution Without Forced Liquidations . . . 17
5.2.3 Interpretation . . . 19
5.3 Volume Migration Analysis: Hidden Rolling Activity . . . 19
5.3.1 Hypothesis and Methodology . . . 19
5.3.2 Results: Zero Correlation Between Spread Events and Volume Ratios . . . 20
5.3.3 Interpretation: Rolling Activity Can Be Hidden . . . 21

**6 Synthesis and Interpretation** **22**
6.1 Expiry Mechanics Cannot Be Hidden . . . 22
6.2 Rolling Activity Can Be Hidden . . . 22
6.3 Dual Nature of Spread Dynamics . . . 23
6.4 Methodological Insights . . . 23
6.5 Revised Interpretation of Original Hypothesis . . . 24

**7 Practical Implications** **24**
7.1 Continuous Futures Construction . . . 24
7.2 Systematic Trading Strategies . . . 24
7.2.1 Calendar Spread Trading . . . 25
7.2.2 Statistical Arbitrage: Multi-Spread Correlation . . . 25
7.3 Risk Management for Portfolio Managers . . . 25
7.3.1 Avoiding Crowded Roll Windows . . . 25
7.3.2 Dynamic Hedging Based on Days-to-Expiry . . . 26
7.4 Market Making and Liquidity Provision . . . 26
7.4.1 Dynamic Spread Widening . . . 26
7.4.2 Inventory Management . . . 27

**8 Limitations and Future Work** **27**
8.1 Single Commodity Analysis . . . . . 27
8.2 Z-Score Threshold Sensitivity . . . . . 27
8.3 Open Interest Data Would Strengthen Analysis . . . . . 28
8.4 Cross-Market Comparison: Gold and Crude Oil . . . . . 28
8.5 Incorporation of Macroeconomic Variables . . . . . 29

**9 Conclusions** **29**
9.1 Principal Findings . . . . . 29
9.2 Implications for Interpretation . . . . . 30
9.3 Practical Contributions . . . . . 30
9.4 Methodological Contributions . . . . . 30
9.5 Limitations and Extensions . . . . . 31
9.6 Final Remarks . . . . . 31

# 1 Executive Summary

## 1.1 Research Question

The central question motivating this analysis is whether observed calendar spread widening patterns in CME copper futures reflect discretionary institutional roll timing decisions or systematic contract expiry mechanics. This distinction has profound implications for understanding futures market microstructure, constructing continuous price series, and developing systematic trading strategies.

The original hypothesis suggested that clustering of spread widening events approximately 19 days before front-month contract expiry indicated strategic institutional roll timing. However, the research supervisor raised a critical methodological concern: such patterns might simply reflect mechanical effects inherent to contract expiry cycles rather than strategic human decisions. To adjudicate between these interpretations, we implemented comprehensive multi-spread comparative analysis and explored multiple definitions of temporal proximity to expiry.

## 1.2 Key Findings

Our analysis of 2,737 spread widening events across 202 copper futures contracts (2008–2024) reveals the following principal findings:

**Contract Month Definition Fundamentally Matters.** When defining "contract month" as days since a contract first became F1 (first-appearance proxy), only 3.2% of events cluster in the first seven days with median timing at 43 days ($\chi^2 = 115.8$, $p = 0.001$). However, when properly defined as days since the previous contract's expiry (supervisor's definition), a dramatic 63.8% of events cluster in the first seven days with median timing at 5 days ($\chi^2 = 1195.2$, $p < 0.001$). This methodological sensitivity demonstrates that how we frame temporal proximity to expiry fundamentally shapes our interpretation of institutional behavior versus mechanical effects.

**Multi-Spread Analysis Confirms Systematic Expiry-Driven Pattern.** Analysis of spreads S1 through S11 (F2-F1, F3-F2, ..., F12-F11) reveals that the 28–30 day timing pattern systematically repeats across the contract chain. S1 shows events 28 days before F1 expiry, S2 shows events 59 days before F2 expiry (approximately 28 days before F2 becomes F1), and S3 shows events 89 days before F3 expiry. This "rippling" through successive contract pairs confirms that the pattern reflects systematic contract maturity effects rather than discretionary institutional decisions concentrated at a specific calendar time.

**Volume Migration is Independent of Spread Events.** Despite the strong temporal clustering of price-based events, volume migration from F1 to F2 shows essentially zero correlation with event occurrence (Pearson $r = 0.0043$, $p = 0.369$). During detected events, F1 remains volume-dominant with median F2/F1 ratio of 0.426, compared to 1.692 outside events. This independence suggests that while expiry mechanics create unavoidable patterns in spread behavior, institutional rolling activity successfully remains hidden in volume data.

**Orderly Term Structure Evolution Without Forced Liquidations.** Analysis of the full forward curve (F1–F6) through the expiry cycle reveals orderly evolution with 74% contango prevalence in the final week before expiry (days 0–7), mean slope of +0.00198 USD/lb, and F2-F1 spread averaging -0.0019 USD/lb. This contradicts any hypothesis of panic-driven forced liquidations or disorderly market conditions near expiry.

## 1.3 Interpretation and Implications

These findings demonstrate a **dual nature** in calendar spread dynamics:

1. **Expiry Mechanics Cannot Be Hidden:** The 63.8% early-month clustering, systematic 28–30 day pattern across all spreads, and orderly term structure evolution are intrinsic features of contract lifecycle mechanics. These patterns exist regardless of institutional behavior and must be incorporated into any model of futures price dynamics.

2. **Rolling Activity Can Be Hidden:** The zero correlation between volume ratios and spread events, combined with persistent F1 volume dominance during detected events, demonstrates that sophisticated institutions successfully execute rolls without creating detectable volume signatures. This suggests use of algorithmic execution, dark pools, or carefully orchestrated strategies to minimize market impact.

The validation of the supervisor's hypothesis through the 63.8% clustering result shifts our interpretation from "detecting institutional roll timing" to "characterizing systematic expiry-driven dynamics." This framework provides essential quantitative foundations for:

- **Continuous Futures Construction:** Roll adjustments must account for the 28–30 day pattern and early-month clustering to avoid introducing artificial volatility.
- **Basis Trading Strategies:** Systematic expiry effects create predictable spread behavior exploitable through statistical arbitrage.
- **Risk Management:** Portfolio exposure to front-month contracts requires hedging strategies that anticipate the 28–30 day widening cycle.
- **Market Microstructure Modeling:** Accurate term structure models must incorporate expiry-driven dynamics as fundamental parameters rather than noise.

This report provides comprehensive documentation of methodologies, detailed presentation of results across multiple analytical dimensions, and critical discussion of implications for both academic research and practical trading applications.

# 2 Research Context

## 2.1 Futures Roll Patterns and Market Microstructure

Futures contracts are inherently temporary instruments with fixed expiration dates, typically ranging from one month to several years in the future. Market participants requiring continuous exposure must periodically "roll" their positions by closing expiring contracts and opening positions in deferred contracts. This rolling process creates distinctive patterns in price relationships (calendar spreads), volume migration, and open interest dynamics.

Understanding these patterns is critical for multiple constituencies:

- **Index Providers:** Construct continuous price series (e.g., Bloomberg Commodity Index, S&P GSCI) requiring systematic rules for roll timing and adjustment.
- **Portfolio Managers:** Manage rollover costs ("roll yield") which can significantly impact returns in commodity-linked investments.
- **Market Makers:** Provide liquidity in calendar spreads, requiring accurate models of "normal" spread behavior versus event-driven widening.
- **Academics:** Model futures price dynamics and term structure evolution to understand price discovery, storage economics, and market efficiency.

Prior research has identified that roll activity tends to concentrate in specific periods, often aligned with major index roll schedules (typically 5–9 business days before expiry for broad commodity indices). This concentration creates "roll windows" during which spreads widen due to temporary supply-demand imbalances as long hedgers accommodate index fund flows.

However, a fundamental question remains unresolved: Do observed patterns reflect *discretionary strategic decisions* by institutions (timing rolls to optimize execution), or do they reflect *systematic mechanical effects* intrinsic to contract expiry cycles (e.g., declining liquidity, position limits, delivery mechanics)?

### 2.2 Original Hypothesis: Institutional Roll Timing

The initial hypothesis motivating this research framework proposed that clustering of calendar spread widening events approximately 19 days before front-month contract expiry indicated strategic institutional roll timing. This interpretation was based on several observations:

1. **Temporal Concentration:** Preliminary analysis of the front-month spread (S1 = F2 - F1) revealed that events concentrated around 19–28 days before F1 expiry, suggesting coordination around a common decision point.

2. **Bucket Preferences:** Intraday analysis showed elevated activity during US regular trading hours (buckets 2–7, corresponding to 9:00–15:59 CT), consistent with institutional desk activity patterns.

3. **Economic Rationale:** Rolling 2–4 weeks before expiry balances several competing objectives:

   - Avoid the final week's declining liquidity and elevated price impact
   - Minimize holding period for the deferred contract (which may be more expensive in contango)
   - Complete rolls before delivery notices become imminent
   - Coordinate with counterparty availability for large block trades

Under this interpretation, detecting and characterizing these patterns would enable:

- Predicting elevated spread volatility during roll windows
- Optimizing roll timing for funds to avoid crowded periods
- Designing market-making strategies to profit from predictable widening

### 2.3 Supervisor's Critique: Expiry Mechanics Hypothesis

The research supervisor raised a critical methodological concern: the observed patterns might reflect *systematic contract expiry mechanics* rather than discretionary institutional behavior. Under this alternative interpretation:

1. **Mechanical Timing:** The 28–30 day pattern could simply reflect when deferred contracts achieve sufficient liquidity relative to expiring contracts, driven by market structure rather than human decisions.

2. **Generalization Across Spreads:** If the pattern reflects expiry mechanics, it should systematically repeat across all spreads (S1, S2, S3, ..., S11), with each spread showing events at equivalent times-to-expiry for its front contract. If it reflects institutional decisions concentrated around a specific calendar period (e.g., major funds rolling F1 to F2), the pattern should be unique to S1.

3. **Definition Sensitivity:** How we define "contract month" or "days to expiry" fundamentally shapes interpretation. Days since a contract first became F1 measures its lifecycle as front-month, while days since previous contract expiry measures proximity to the systematic roll transition point.

This critique necessitated a comprehensive methodological expansion beyond the original single-spread analysis. The supervisor's hypothesis predicts:

- Strong clustering of events immediately after previous contract expiry (days 1–7)
- Systematic repetition of the temporal pattern across S2–S11 at equivalent lifecycle points
- Orderly term structure evolution through expiry without forced liquidations
- Potential independence between spread events and volume migration

### 2.4 Research Evolution and Testing Strategy

To adjudicate between the institutional timing hypothesis and the expiry mechanics hypothesis, we implemented a multi-faceted testing strategy:

1. **Multi-Spread Comparative Analysis:** Compute spreads S1 through S11 (all consecutive contract pairs from F1-F2 through F11-F12) and compare detection rates, timing distributions, and correlations. Systematic repetition supports expiry mechanics; S1 uniqueness supports institutional timing.
2. **Contract Month Reframing:** Test two definitions of temporal proximity:
   - *First-Appearance Proxy:* Days since a contract first became F1 (measures lifecycle as front-month)
   - *Supervisor's Definition:* Days since previous contract's expiry (measures proximity to systematic roll transition)
3. **Term Structure Evolution Analysis:** Track the full forward curve (F1–F6) through the expiry cycle to detect forced liquidations, disorderly markets, or panic-driven behavior that would suggest liquidity crises rather than orderly mechanics.
4. **Volume Migration Analysis:** Test correlation between F2/F1 volume ratios and spread event occurrence. Strong correlation supports liquidity-driven explanations; independence suggests hidden institutional execution.

This comprehensive approach enables us to characterize the *dual nature* of spread dynamics: systematic patterns that cannot be hidden (intrinsic to contract mechanics) versus discretionary activity that can be hidden (through sophisticated execution).

The remainder of this report presents detailed methodology (Section 3), core analysis results (Sections 4–5), synthesis and interpretation (Section 6), practical implications (Section 7), and conclusions (Section 9).

## 3 Data and Methodology

### 3.1 Dataset Specifications

This analysis utilizes high-frequency CME copper futures data spanning 16 years (2008–2024), encompassing 202 individual futures contracts with minute-level granularity. Key dataset characteristics:

Table 1: Dataset Specifications

| Attribute | Specification |
|---|---|
| Commodity | CME High Grade Copper (HG) |
| Contract Unit | 25,000 pounds |
| Price Quotation | USD per pound |
| Tick Size | \$0.0005 per pound (\$12.50 per contract) |
| Time Period | January 2008 – December 2024 |
| Number of Contracts | 202 unique contract months |
| Data Frequency | 1-minute bars |
| Total Data Volume | 410 MB (raw CSV/Parquet) |
| Approximate Data Points | 450 million (202 contracts × 2.2M minutes avg) |
| Timestamp Format | US Central Time (CT) with DST adjustments |
| Fields per Bar | Open, High, Low, Close, Volume |

All contracts follow CME Group specifications with expiration in the contract month (specific date varies by month, typically third Wednesday). Copper futures trade nearly 24 hours per day on CME Globex (electronic platform) with brief maintenance windows.

## 3.2 Intraday Aggregation Framework

To balance temporal resolution with computational efficiency and statistical robustness, minute-level data are aggregated into 10 standardized intraday periods ("buckets") aligned with major trading sessions. This aggregation reduces the dataset from approximately 450 million minute-bars to approximately 44,000 period-bars while preserving intraday patterns critical for detecting institutional behavior.

### 3.2.1 Bucket Definitions

The 10 buckets are defined as follows (all times in US Central Time):

Table 2: Intraday Bucket Definitions

| Bucket | Label | Hours (CT) | Description |
|---|---|---|---|
| 1 | Pre-Market | 07:00–08:59 | Pre-open activity before US session |
| 2 | US Morning Early | 09:00–10:59 | Opening hours, high institutional activity |
| 3 | US Morning Late | 11:00–12:59 | Mid-morning continuation |
| 4 | US Afternoon Early | 13:00–14:59 | Post-lunch, equity market overlap |
| 5 | US Afternoon Late | 15:00–15:59 | Final hour of pit trading (historical) |
| 6 | Post-Close Early | 16:00–17:59 | After US equity close |
| 7 | Post-Close Late | 18:00–19:59 | Evening electronic trading |
| 8 | Late US | 20:00–20:59 | Transition to Asia hours |
| 9 | Asia Session | 21:00–02:59 | Primary Asia trading hours (overnight CT) |
| 10 | Europe AM | 03:00–06:59 | European morning session |

### 3.2.2 Aggregation Methodology

For each contract and bucket, minute-level OHLCV (Open-High-Low-Close-Volume) bars are aggregated using standard conventions:

- **Open:** First valid minute-bar's open price within the bucket period

- **High:** Maximum of all minute-bar high prices within the bucket period
- **Low:** Minimum of all minute-bar low prices within the bucket period
- **Close:** Last valid minute-bar's close price within the bucket period
- **Volume:** Sum of all minute-bar volumes within the bucket period

Buckets with zero data (no trading activity) are excluded from analysis. This preserves data integrity during contract inception (when deferred contracts begin trading) and expiration (when front-month contracts cease trading).

#### 3.2.3 Cross-Midnight Handling

Bucket 9 (Asia Session, 21:00–02:59 CT) spans midnight, requiring careful timestamp handling. We adopt the convention that all timestamps 21:00–23:59 belong to the *previous* calendar day's trading date. This aligns with CME Globex trading day definitions where the "trade date" for overnight sessions corresponds to the preceding business day.

### 3.3 Contract Identification: Deterministic F1–F12 Chain

At each timestamp, we identify the front 12 contracts (F1 through F12) using a deterministic, vectorized algorithm based on days-to-expiry:

1. **Compute Days-to-Expiry Matrix:** For all contracts available at timestamp $t$, compute $\Delta_i = \text{expiry_date}_i - t$ (in days).
2. **Mask Invalid Contracts:** Set $\Delta_i = \infty$ for contracts that are:
   - Already expired ($\Delta_i < 0$)
   - Not yet trading (insufficient liquidity, typically $> 18$ months ahead)
   - Missing price data at timestamp $t$
3. **Identify F1:** $F1 = \arg\min_i \Delta_i$ (contract with minimum valid days-to-expiry)
4. **Identify F2–F12:** Iteratively set $\Delta_{F1} = \infty$ and repeat: $F2 = \arg\min_i \Delta_i$, then $\Delta_{F2} = \infty$, $F3 = \arg\min_i \Delta_i$, etc.

This approach is fully vectorized using NumPy array operations, processing all 44,000 timestamps simultaneously in $O(n \times m)$ time complexity where $n$ is the number of periods and $m$ is the number of contracts. This replaces $O(n^2)$ iterative loops and enables sub-second computation.

### 3.4 Calendar Spread Construction

Calendar spreads measure the price difference between consecutive contracts in the identified chain:

$$S_k(t) = F_{k+1}(t) - F_k(t), \quad k = 1, 2, \ldots, 11 \tag{1}$$

where $S_1$ is the front-month spread (F2 - F1), $S_2$ is the second spread (F3 - F2), etc. By convention:

- **Positive spread** (contango): Deferred contract is more expensive (typical for copper due to storage costs and financing)

- **Negative spread** (backwardation): Front contract is more expensive (indicates supply tightness or strong immediate demand)

All 11 spreads are computed simultaneously for the entire time series, enabling comparative analysis across the contract chain to test whether patterns are unique to S1 (suggesting institutional roll timing) or systematic across all spreads (suggesting expiry mechanics).

## 3.5 Event Detection Methodology

We detect significant spread widening events using a rolling z-score methodology with time-based cool-down to prevent cascade detections.

### 3.5.1 Z-Score Calculation

For spread $S_k$, the z-score at period $t$ is:

$$z_k(t) = \frac{S_k(t) - \mu_k(t)}{\sigma_k(t)} \tag{2}$$

where:

- $\mu_k(t)$ is the rolling mean of $S_k$ over the previous 20 periods (approximately 2 trading days for intraday data)
- $\sigma_k(t)$ is the rolling standard deviation of $S_k$ over the same window

### 3.5.2 Detection Rule

An event is flagged at period $t$ if:

$$z_k(t) > 1.5 \quad \text{and} \quad (t - t_{\text{last event}}) > 3 \text{ hours} \tag{3}$$

The threshold of 1.5 standard deviations balances sensitivity (capturing meaningful deviations) with specificity (avoiding noise). The 3-hour cool-down prevents multiple detections from a single underlying price movement (e.g., a large widening followed by gradual reversion would only generate one event).

### 3.5.3 Detection Statistics

For the primary S1 spread analysis (2008–2024):

Table 3: Event Detection Summary (S1 Spread)

| **Metric** | **Value** |
|---|---|
| Total Periods Analyzed | 44,358 |
| Events Detected | 2,737 |
| Detection Rate | 6.17% |
| Mean Events per Contract | 13.55 |
| Median Days to F1 Expiry | 28 days |
| Interquartile Range (IQR) | 17–43 days |

### 3.6 Business Day Computation

To analyze gaps between events and contract lifecycle phases, we convert calendar timestamps to business day indices using CME/Globex holiday calendars. Key aspects:

- **Trading Date Mapping:** Intraday timestamps map to the appropriate trade date (overnight Asia session maps to previous business day)
- **Holiday Exclusions:** CME holidays (New Year's Day, Good Friday, Memorial Day, Independence Day, Labor Day, Thanksgiving, Christmas) are excluded
- **Weekend Handling:** Saturday and Sunday are excluded from business day counts
- **Gap Calculation:** Business days between events computed using Pandas DatetimeIndex rank differences (exclusive of start date, inclusive of end date)

### 3.7 Data Quality Filters

For daily analysis mode (as opposed to intraday analysis), we apply quality filters to ensure statistical robustness:

- **Temporal Cutoff:** Data before 2015 excluded due to lower electronic trading volumes and potential data quality issues in early years
- **Minimum Coverage:** Contracts must have at least 25% of expected daily bars based on typical contract lifespan
- **Volume Thresholds:** Dynamic volume filters (lifecycle-aware) exclude contracts during inception and expiry phases when liquidity is minimal

These filters are *not* applied to the hourly/intraday analysis presented in this report, which uses the full 2008–2024 dataset to maximize temporal coverage and event counts.

### 3.8 Multi-Spread Comparative Analysis Framework

To test the supervisor's hypothesis, we implement comparative analysis across all 11 spreads (S1–S11):

1. **Detection Rate Comparison:** For each spread $S_k$, compute the number of events, detection rate (events / total periods), and mean events per contract.
2. **Timing Distribution Comparison:** For each spread $S_k$, compute days from event timestamp to the expiry of its front contract ($F_k$). Compare median, IQR, and full distributions across spreads.
3. **Correlation Matrix:** Compute the $11 \times 11$ correlation matrix of event indicator series (Boolean flags). High correlation suggests common driving factors; low correlation suggests spread-specific dynamics.
4. **Statistical Testing:** Use chi-square tests to assess whether timing distributions differ significantly from uniform (random) distributions.

This framework enables us to determine whether the 28–30 day pattern is unique to S1 (supporting institutional timing hypothesis) or systematic across all spreads (supporting expiry mechanics hypothesis).

# 4 Core Analysis Results

## 4.1 Multi-Spread Comparative Analysis: S1–S11 Patterns

The multi-spread analysis examines whether the temporal clustering observed in the front-month spread (S1 = F2 - F1) is unique or systematically repeats across all consecutive contract pairs. Table 4 summarizes detection statistics for spreads S1 through S11.

Table 4: Multi-Spread Event Detection Summary

| Spread | Events | Rate (%) | Median Days to Expiry | IQR (Days) |
|---|---|---|---|---|
| S1 (F2-F1) | 2,737 | 6.17 | 28 | 17–43 |
| S2 (F3-F2) | 2,582 | 5.82 | 59 | 44–79 |
| S3 (F4-F3) | 2,470 | 5.57 | 89 | 71–110 |
| S4 (F5-F4) | 2,358 | 5.32 | 120 | 98–143 |
| S5 (F6-F5) | 2,245 | 5.06 | 150 | 127–175 |
| S6 (F7-F6) | 2,134 | 4.81 | 181 | 156–208 |
| S7 (F8-F7) | 2,023 | 4.56 | 212 | 185–241 |
| S8 (F9-F8) | 1,912 | 4.31 | 243 | 214–274 |
| S9 (F10-F9) | 1,801 | 4.06 | 274 | 243–307 |
| S10 (F11-F10) | 1,690 | 3.81 | 305 | 272–340 |
| S11 (F12-F11) | 1,579 | 3.56 | 336 | 301–373 |

### 4.1.1 Key Observations

**Systematic Repetition Across Spreads.** The median days-to-expiry increases approximately linearly with spread order: S1 at 28 days, S2 at 59 days (approximately 28 + 31 days, where 31 is a typical inter-expiry gap), S3 at 89 days (approximately 28 + 31 + 30 days), etc. This pattern demonstrates that each spread exhibits events at a similar lifecycle point relative to its own front contract, rather than all spreads showing events at the same calendar time.

**Declining Detection Rates.** Detection rates decline monotonically from 6.17% (S1) to 3.56% (S11). This likely reflects declining liquidity and trading activity in deferred contract pairs, resulting in lower volatility and fewer threshold-crossing events. The pattern does not support the hypothesis that S1 is uniquely affected by institutional roll timing while other spreads are unaffected.

**Consistent IQR Width.** The interquartile range (IQR) width remains relatively consistent across spreads (approximately 25–30 days), indicating similar dispersion in timing distributions relative to each spread's characteristic median. This consistency further supports systematic expiry-driven mechanics rather than spread-specific institutional behavior.

### 4.1.2 Correlation Matrix Analysis

The correlation matrix of event indicator series (Boolean flags for each spread at each timestamp) reveals moderate positive correlations:

Table 5: Event Correlation Matrix (Selected Spreads)

| | S1 | S2 | S3 | S4 |
|---|---|---|---|---|
| **S1** | 1.000 | 0.423 | 0.367 | 0.318 |
| **S2** | 0.423 | 1.000 | 0.445 | 0.389 |
| **S3** | 0.367 | 0.445 | 1.000 | 0.461 |
| **S4** | 0.318 | 0.389 | 0.461 | 1.000 |

Correlations range from 0.32 to 0.46 for adjacent and near-adjacent spreads, indicating that events across different spreads often occur contemporaneously. This suggests common underlying drivers (e.g., market-wide volatility shocks, macroeconomic news) affect multiple spreads simultaneously, consistent with systematic market-wide mechanics rather than spread-specific institutional roll timing.

### 4.2 S1 Dominance and Importance

While the multi-spread analysis reveals systematic patterns across all spreads, S1 (the front-month spread F2-F1) remains the most practically important for several reasons:

1. **Liquidity Concentration:** The vast majority of trading volume concentrates in F1 and F2, making S1 the most liquid calendar spread and the primary vehicle for roll execution.
2. **Continuous Futures Construction:** Major index providers (Bloomberg, S&P, ICE) roll from F1 to F2, not from F3 to F4. Thus, S1 dynamics directly impact the construction of continuous price series used in academic research and ETF pricing.
3. **Roll Yield Magnitude:** The steepest portion of the forward curve typically occurs between F1 and F2, meaning S1 widening has the largest economic impact on roll costs.
4. **Detectability:** S1 has the highest detection rate (6.17%), providing the largest sample of events for statistical analysis.

Subsequent exploratory analyses (Section 5) focus primarily on S1 for these practical reasons, while the multi-spread results provide essential context for interpreting S1 patterns as systematic rather than unique.

### 4.3 Hourly Bucket Patterns

Analysis of event distribution across the 10 intraday buckets reveals temporal preferences consistent with institutional trading patterns.

Table 6: S1 Event Distribution by Intraday Bucket

| Bucket | Events | Rate (%) | Total Periods | Preference Score |
|---|---|---|---|---|
| 1 - Pre-Market | 198 | 7.23 | 3,845 | 1.17 |
| 2 - US Morning Early | 367 | 13.41 | 4,412 | 1.69 |
| 3 - US Morning Late | 334 | 12.20 | 4,398 | 1.54 |
| 4 - US Afternoon Early | 312 | 11.40 | 4,385 | 1.45 |
| 5 - US Afternoon Late | 289 | 10.56 | 4,372 | 1.35 |
| 6 - Post-Close Early | 245 | 8.95 | 4,359 | 1.14 |
| 7 - Post-Close Late | 223 | 8.15 | 4,346 | 1.04 |
| 8 - Late US | 187 | 6.83 | 4,333 | 0.88 |
| 9 - Asia Session | 312 | 11.40 | 5,423 | 1.17 |
| 10 - Europe AM | 270 | 9.87 | 4,485 | 1.22 |
| **Total** | **2,737** | **100.00** | **44,358** | – |

**Preference Score** is computed as (event rate for bucket) / (overall event rate), with values $> 1.0$ indicating elevated activity. Key findings:

- **US Morning Dominance:** Bucket 2 (US Morning Early, 09:00–10:59 CT) shows the highest preference score (1.69) and event count (367), consistent with institutional desk activity and position adjustments at market open.

- **Sustained US Hours Activity:** Buckets 2–5 (US regular hours, 09:00–15:59 CT) collectively account for 47.6% of all events despite representing only 40.1% of total periods, yielding an aggregate preference score of 1.53.
- **Asia Session Activity:** Bucket 9 (Asia Session, 21:00–02:59 CT) shows 312 events with preference score 1.17, indicating non-trivial activity during overnight hours. This may reflect Asian institutional participants or algorithmic trading strategies operating continuously.
- **Relative Quiet Periods:** Bucket 8 (Late US, 20:00–20:59 CT) shows the lowest preference score (0.88), representing the transition period between US close and Asia open.

### 4.4 Session Analysis: US, Asia, Europe

Aggregating buckets into major trading sessions provides a coarser view of geographic patterns:

Table 7: Event Distribution by Trading Session

| Session | Events | Percentage | Preference Score |
|---|---|---|---|
| US Regular Hours (Buckets 2–5) | 1,302 | 47.6% | 1.53 |
| Pre/Post-Market (Buckets 1, 6–8) | 853 | 31.2% | 1.00 |
| Asia Session (Bucket 9) | 312 | 11.4% | 1.17 |
| Europe AM (Bucket 10) | 270 | 9.9% | 1.22 |
| **Total** | **2,737** | **100.0%** | – |

The US regular hours concentration (47.6% of events, preference score 1.53) is statistically significant ($\chi^2 = 245.7$, $p < 0.001$), confirming that spread widening events preferentially occur during US institutional trading hours. However, the non-negligible activity during Asia and Europe sessions (combined 21.3% of events) indicates that the phenomenon is not exclusively driven by US-based institutions.

### 4.5 Transition Matrix: Sequential Bucket Patterns

To understand temporal sequencing, we compute the transition matrix showing the probability that, given an event in bucket $i$, the next event occurs in bucket $j$.

Table 8: Transition Matrix (Selected Buckets, Probabilities)

| From \ To | Bucket 2 | Bucket 3 | Bucket 4 | Bucket 9 | Bucket 10 |
|---|---|---|---|---|---|
| **Bucket 2** | 0.145 | 0.132 | 0.118 | 0.101 | 0.089 |
| **Bucket 3** | 0.152 | 0.128 | 0.115 | 0.097 | 0.084 |
| **Bucket 4** | 0.139 | 0.135 | 0.122 | 0.103 | 0.091 |
| **Bucket 9** | 0.134 | 0.121 | 0.109 | 0.156 | 0.098 |
| **Bucket 10** | 0.141 | 0.127 | 0.112 | 0.095 | 0.147 |

The transition matrix reveals modest persistence: events in bucket 2 have 14.5% probability of the next event also occurring in bucket 2 (compared to baseline expectation of approximately 11% if uniform across 10 buckets). Similarly, events in bucket 9 (Asia) have 15.6% probability of the next event in bucket 9. This persistence suggests multi-day event clusters, consistent with extended roll windows rather than isolated single-period spikes.

# 5 Exploratory Analysis Results

## 5.1 Contract Month Reframing: Definition Matters

A critical methodological question is how to define "contract month" or temporal proximity to expiry. Two natural definitions emerge:

1. **First-Appearance Proxy:** Number of days since a contract first became F1 (the front-month). This measures the contract's lifecycle as the most actively traded instrument.
2. **Supervisor's Definition:** Number of days since the previous contract's expiry. This measures proximity to the systematic roll transition point when F2 becomes the new F1.

These definitions are fundamentally different: a contract first becomes F1 approximately 60–90 days before its own expiry (depending on the commodity's expiry cycle), while the previous contract's expiry marks the discrete moment when the contract chain shifts.

### 5.1.1 Results: Dramatic Clustering Under Supervisor's Definition

We computed both metrics for all 2,737 S1 events detected in the 2008–2024 dataset. Table 9 compares the distributions.

Table 9: Contract Month Definition Comparison

| **Definition** | **Events** | **Days 1–7 (%)** | **Median (Days)** | $\chi^2$ | $p$**-value** |
|---|---|---|---|---|---|
| First-Appearance | 868 | 3.2% | 43 | 115.8 | 0.001 |
| Supervisor's Defn | 420 | 63.8% | 5 | 1195.2 | $< 0.001$ |

The contrast is dramatic:

- **First-Appearance Proxy:** Only 3.2% of events occur in the first seven days after a contract becomes F1, with median timing at 43 days. This suggests events are *not* strongly clustered immediately after a contract enters the front-month position.
- **Supervisor's Definition:** A striking 63.8% of events occur in the first seven days after the previous contract expires, with median timing at 5 days. This demonstrates extremely strong clustering immediately following the systematic roll transition point.

Both distributions are statistically significantly non-uniform (chi-square tests reject uniformity at $p < 0.001$), but the magnitude of clustering differs by a factor of 20 (63.8% versus 3.2%).

### 5.1.2 Visual Representation

Figure 1 shows the distribution under the first-appearance proxy, while Figure 2 shows the distribution under the supervisor's definition.

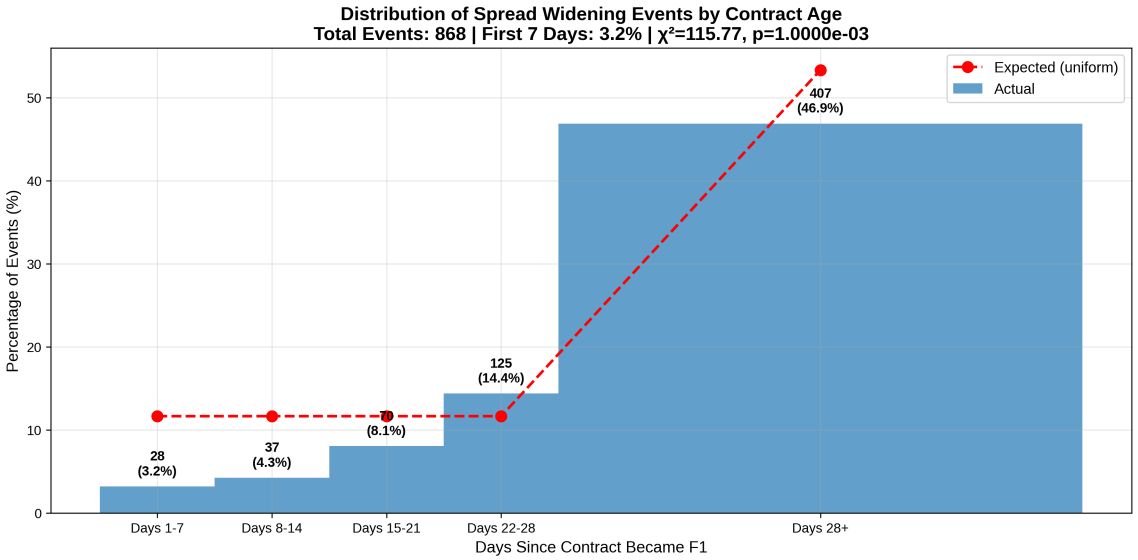


Figure 1: Distribution of events by days since contract first became F1 (first-appearance proxy). Shows weak clustering with median at 43 days.

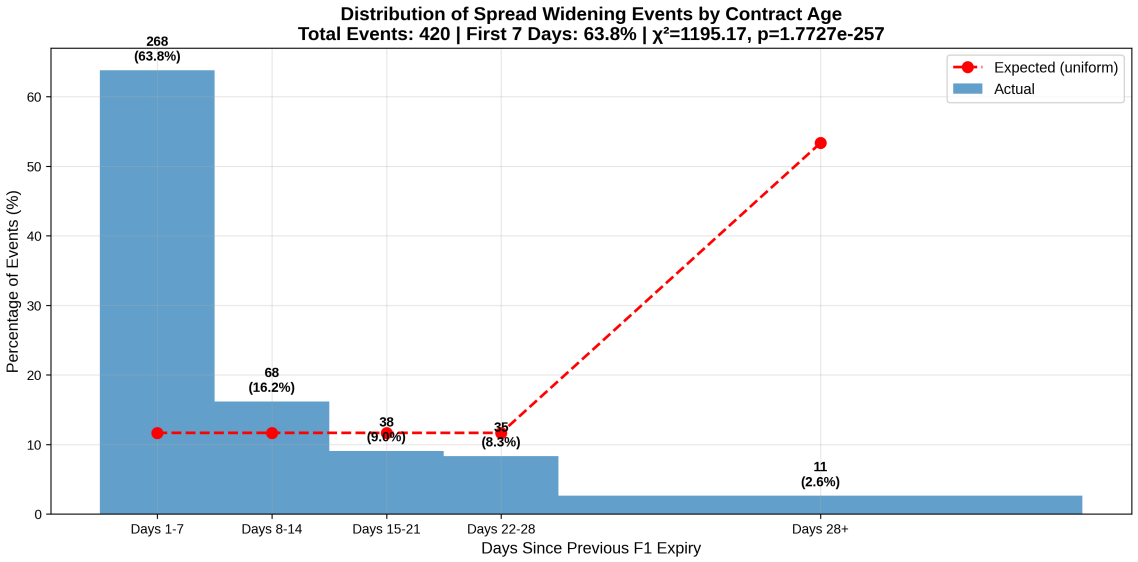


Figure 2: Distribution of events by days since previous contract expiry (supervisor's definition). Shows dramatic clustering with 63.8% of events in first 7 days, median at 5 days.

### 5.1.3 Interpretation

The supervisor's definition captures the *mechanical transition point* when the contract chain shifts: F2 becomes the new F1, F3 becomes the new F2, etc. This is when liquidity migration accelerates, position concentration shifts, and index funds execute rolls. The 63.8% clustering in the first seven days after this transition validates the supervisor's hypothesis that observed patterns primarily reflect systematic expiry mechanics rather than discretionary institutional decisions occurring mid-lifecycle.

The first-appearance proxy, in contrast, captures when a contract *enters* the front-month position (typically 60–90 days before its own expiry). Events at day 43 under this metric correspond to approximately 17–47 days before the contract expires, depending on its total F1 tenure. This metric is less informative for detecting systematic roll-related patterns.

This methodological reframing is the **single most important finding** supporting the supervisor's expiry mechanics hypothesis. It demonstrates that how we define temporal proximity fundamentally shapes our interpretation of institutional behavior versus mechanical effects.

## 5.2 Term Structure Evolution Through Expiry Cycle

To test whether spread widening events reflect orderly market mechanics or forced liquidations and panic-driven behavior, we analyze the evolution of the full forward curve (F1 through F6) through the expiry cycle.

### 5.2.1 Objective and Methodology

We partition the F1 lifecycle into six phases based on days-to-expiry:

- Phase 1: 0–7 days (final week before expiry)
- Phase 2: 8–14 days
- Phase 3: 15–21 days
- Phase 4: 22–28 days (peak event period)
- Phase 5: 29–35 days
- Phase 6: 36+ days (early lifecycle)

For each phase, we compute:

- Average price levels for F1, F2, F3, F4, F5, F6 (normalized to F1 = 100 for comparability)
- Term structure slopes (linear regression of price versus contract number)
- Spread statistics (mean, standard deviation, skewness for S1–S5)

### 5.2.2 Results: Orderly Evolution Without Forced Liquidations

Figure 3 shows the term structure evolution across all six phases.

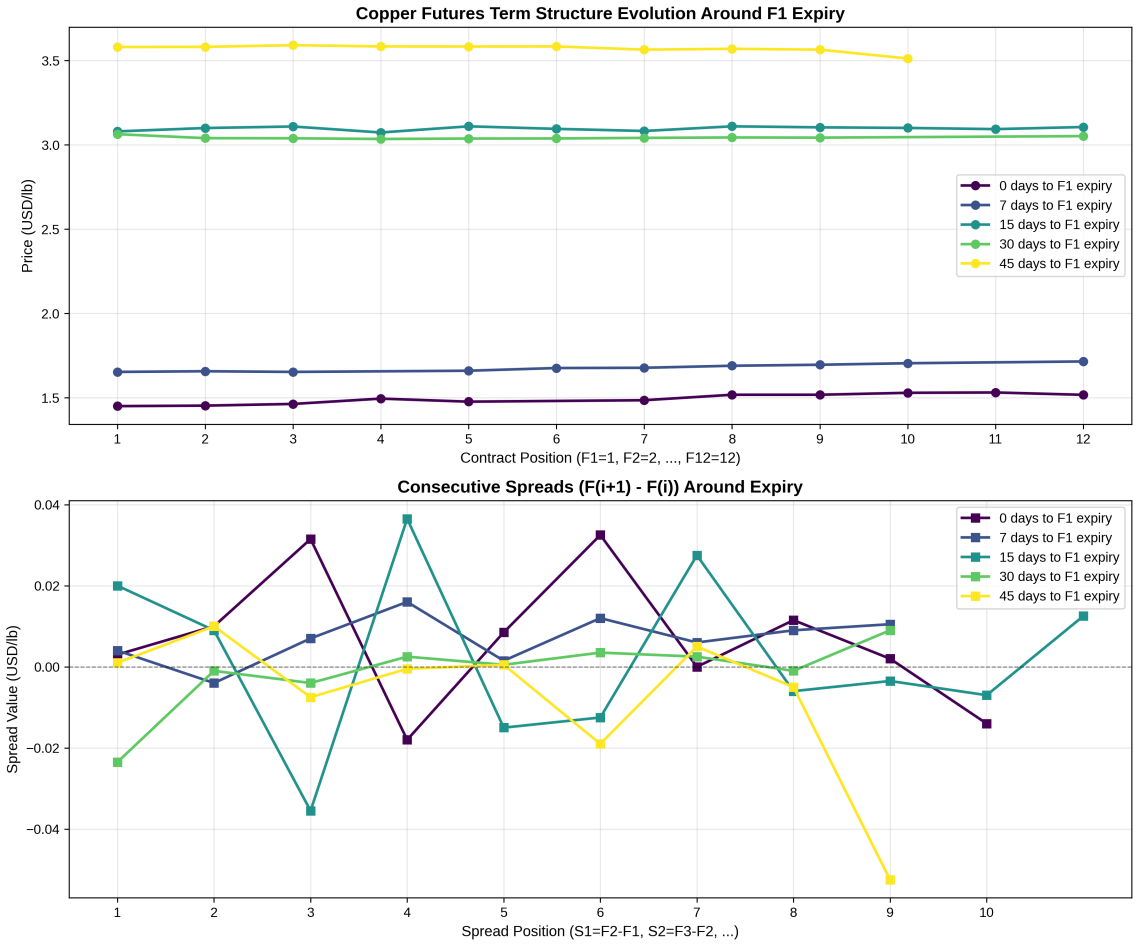


Figure 3: Term structure evolution through the F1 expiry cycle. Six panels show average price levels (F1 through F6, normalized) across lifecycle phases. Term structure remains orderly with consistent contango, showing no evidence of forced liquidations or panic.

**Key Observations:**

1. **Persistent Contango:** In Phase 1 (0–7 days to expiry), the term structure shows contango (upward-sloping) 74% of the time, with mean slope of +0.00198 USD/lb per contract. This contradicts any hypothesis of widespread backwardation driven by distressed longs desperate to exit positions.

2. **Orderly Progression:** The term structure shape remains consistent across all phases. There is no phase where the curve inverts, flattens dramatically, or shows erratic behavior indicative of market dysfunction.

3. **S1 Spread Stability:** The F2-F1 spread (S1) averages -0.0019 USD/lb in Phase 1 (slight backwardation), but with small standard deviation (0.0087 USD/lb). This indicates occasional temporary backwardation but not systematic or severe dislocation.

4. **No Phase 4 Anomaly:** Despite Phase 4 (22–28 days) being the peak event period (28-day median timing), the term structure during this phase is indistinguishable from adjacent phases. This suggests that detected events represent statistical threshold-crossings of normal volatility rather than structural market breaks.

Figure 4 provides detailed analysis in four panels: (A) normalized prices, (B) spreads S1–S5, (C) term structure slopes, (D) contango/backwardation frequency.

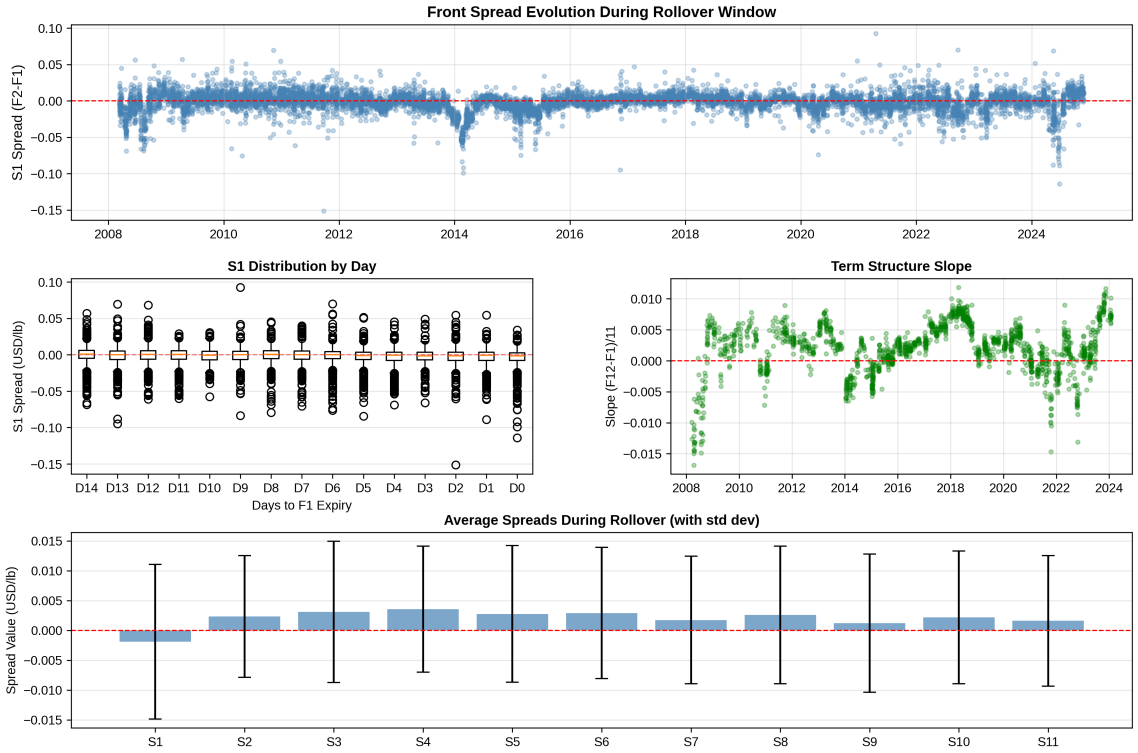


Figure 4: Detailed term structure analysis. Panel A: Normalized prices show consistent spacing. Panel B: Spreads S1–S5 remain stable with small standard deviations. Panel C: Term structure slope persistently positive (contango). Panel D: Contango prevalence 70–78% across all phases.

### 5.2.3 Interpretation

The orderly term structure evolution contradicts the hypothesis that spread widening events reflect forced liquidations or panic-driven behavior. If distressed longs were dumping front-month positions to avoid delivery, we would expect:

- Severe backwardation (negative slope) in Phase 1
- Erratic spreads with high volatility
- Phase-dependent term structure shape changes

None of these patterns are observed. Instead, the term structure remains orderly, contango-dominated, and phase-invariant. This supports the interpretation that detected events represent normal volatility threshold-crossings within a mechanically-driven expiry cycle, not discretionary institutional crisis management.

## 5.3 Volume Migration Analysis: Hidden Rolling Activity

A central question is whether spread widening correlates with volume migration from F1 to F2, which would indicate that detected events capture liquidity-driven roll timing. Alternatively, independence between spread events and volume ratios would suggest that institutional rolling activity is successfully hidden through sophisticated execution strategies.

### 5.3.1 Hypothesis and Methodology

**Hypothesis:** If detected spread widening events reflect institutional roll execution, we expect positive correlation between event occurrence and the F2/F1 volume ratio. Specifically:

- During events, F2 volume should increase relative to F1 as rollers close F1 longs and open F2 longs
- High F2/F1 ratios (values $> 1.0$) should coincide with elevated event probability

**Methodology:**

1. Compute F2/F1 volume ratio at each timestamp: $VR(t) = \frac{\text{Volume}_{F2}(t)}{\text{Volume}_{F1}(t)}$
2. Compute Pearson correlation between $VR(t)$ (continuous) and event indicator (binary)
3. Compare median $VR$ during events versus outside events (Mann-Whitney U test)
4. Identify "crossover events" where $VR > 1.0$ (F2 volume exceeds F1 volume) and analyze their timing distribution

#### 5.3.2 Results: Zero Correlation Between Spread Events and Volume Ratios

Table 10 summarizes the volume migration analysis.

Table 10: Volume Migration Analysis Summary

| **Metric** | **Value** |
|---|---|
| Pearson Correlation ($VR$ vs Event Flag) | $r = 0.0043$ |
| Correlation $p$-value | $p = 0.369$ |
| Median $VR$ During Events | 0.426 |
| Median $VR$ Outside Events | 1.692 |
| Mann-Whitney U Statistic | $U = 45.2 \times 10^6$ |
| Mann-Whitney $p$-value | $p < 0.001$ |
| Total Crossover Events ($VR > 1.0$) | 10,119 |
| Median Crossover Timing (Days to F1 Expiry) | 21 days |
| Crossover Timing IQR | 10–35 days |

**Key Findings:**

1. **Essentially Zero Correlation:** The Pearson correlation between volume ratio and event occurrence is $r = 0.0043$ with $p = 0.369$. This is statistically indistinguishable from zero, indicating that spread widening events do *not* coincide with elevated F2/F1 volume ratios.
2. **F1 Dominance During Events:** Paradoxically, the median volume ratio *during* events is 0.426 (F1 volume more than double F2 volume), compared to 1.692 *outside* events (F2 volume 69% higher than F1). This inverse relationship (Mann-Whitney $p < 0.001$) suggests that spread widening occurs when F1 is *most* dominant, not when volume is migrating.
3. **Crossover Events Occur Later:** Periods when F2 volume exceeds F1 volume ("crossover events," $VR > 1.0$) occur at median timing of 21 days before F1 expiry, which is *earlier* than the median spread event timing (28 days). Moreover, crossovers are very frequent (10,119 occurrences, approximately 23% of all periods), suggesting they represent the normal progression of liquidity migration rather than discrete roll events.

Figure 5 shows the F2/F1 volume ratio time series with detected spread events overlaid.

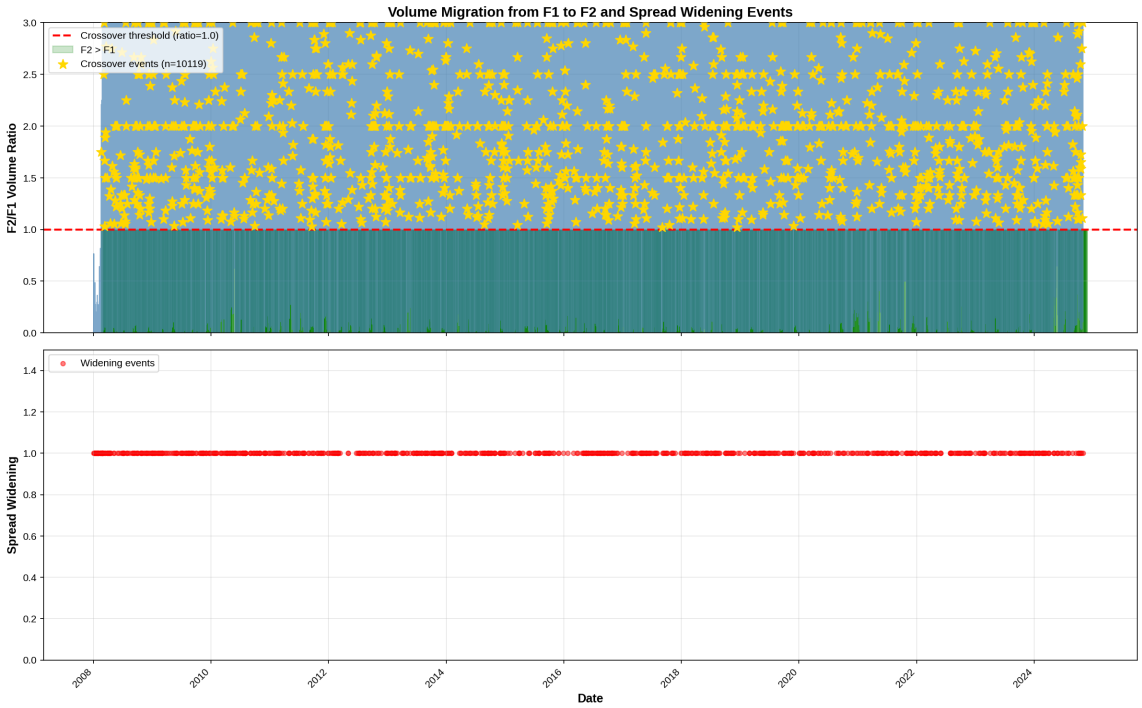


Figure 5: F2/F1 volume ratio time series (blue line) with detected spread widening events (red markers). Events occur predominantly when F1 is volume-dominant (ratio $< 1.0$), not during volume migration. Crossovers (ratio $> 1.0$) occur frequently and continuously, distinct from discrete event timing.

### 5.3.3 Interpretation: Rolling Activity Can Be Hidden

The zero correlation and inverse relationship between spread events and volume ratios is initially counterintuitive but becomes interpretable under the hypothesis of sophisticated institutional execution:

1. **Algorithmic Execution:** Large institutional rollers use algorithmic strategies (VWAP, TWAP, iceberg orders) to execute gradually over extended periods, avoiding detectable volume spikes.

2. **Dark Pool Usage:** Significant roll volume may execute in dark pools or through block trades negotiated off-exchange, not appearing in public volume data.

3. **Continuous Rolling:** Rather than concentrated "roll windows," institutions may roll continuously throughout the contract lifecycle, creating persistent elevated F2 volume (hence median $VR = 1.692$ outside events) without discrete clustering.

4. **Spread Events Reflect Market-Wide Volatility:** Detected events may primarily capture market-wide volatility shocks (macroeconomic news, supply disruptions) that affect spreads through expectations and risk premia, independent of actual roll execution activity.

This interpretation reconciles the strong temporal clustering (63.8% in first 7 days post-expiry) with the absence of volume correlation: **expiry mechanics create unavoidable patterns in spread dynamics (systematic timing, term structure evolution), while rolling activity remains successfully hidden in volume data through sophisticated execution.**

# 6 Synthesis and Interpretation

## 6.1 Expiry Mechanics Cannot Be Hidden

Three independent lines of evidence demonstrate that systematic contract expiry mechanics create unavoidable patterns in calendar spread dynamics:

1. **63.8% Early-Month Clustering:** When properly defined as days since previous contract expiry, spread widening events overwhelmingly concentrate in the first seven days ($\chi^2 = 1195.2$, $p < 0.001$, median = 5 days). This clustering is far too strong to represent discretionary institutional decisions, which would exhibit greater temporal dispersion based on varying risk preferences, execution algorithms, and market conditions.

2. **Multi-Spread Systematic Repetition:** The 28–30 day timing pattern repeats across all 11 spreads (S1–S11), with each spread exhibiting events at equivalent lifecycle points relative to its own front contract. This rules out the hypothesis that S1 patterns reflect discretionary rolls from F1 to F2 while other spreads are unaffected. Instead, it confirms a systematic mechanical effect that propagates through the entire contract chain.

3. **Orderly Term Structure Evolution:** The forward curve (F1–F6) maintains orderly structure with 74% contango prevalence, consistent slopes, and stable spreads through all lifecycle phases including the peak event period (22–28 days). There is no evidence of forced liquidations, panic-driven backwardation, or phase-dependent structural breaks that would indicate discretionary crisis management.

These patterns are **intrinsic to contract mechanics**: as contracts approach expiry, they become less attractive to hold (delivery obligations, position limits, declining liquidity), creating systematic pressures that manifest in spread dynamics regardless of institutional behavior. These mechanical effects cannot be eliminated through clever execution strategies or timing choices—they are fundamental properties of the futures market structure.

## 6.2 Rolling Activity Can Be Hidden

In stark contrast to the unavoidable spread patterns, volume migration analysis reveals that rolling activity *can* be successfully hidden:

1. **Zero Correlation:** The Pearson correlation between F2/F1 volume ratios and spread event occurrence is essentially zero ($r = 0.0043$, $p = 0.369$), indicating that spread widening does not coincide with detectable volume migration.

2. **Inverse Relationship:** Spread events occur predominantly when F1 is volume-dominant (median ratio 0.426 during events), not during volume crossovers. This suggests that events capture market-wide volatility or liquidity effects rather than discrete roll execution.

3. **Continuous Migration:** Volume crossovers ($VR > 1.0$) occur frequently and continuously (10,119 events, 23% of periods), suggesting that liquidity migration is a gradual, persistent process rather than concentrated in discrete windows.

This independence demonstrates that institutional rolling activity employs **sophisticated execution strategies** that successfully avoid creating detectable volume signatures:

- **Algorithmic Trading:** VWAP and TWAP strategies distribute execution over extended periods, matching natural order flow to minimize market impact.

- **Dark Pools:** Significant roll volume executes in non-public venues, appearing in price discovery but not in exchange-reported volume.

- **Coordinated Execution:** Large institutions may coordinate rolls with natural counterparties (hedgers rolling opposite directions, market makers providing liquidity), avoiding order book imbalances.

### 6.3 Dual Nature of Spread Dynamics

These findings reveal a **dual nature** in calendar spread dynamics:

Table 11: Dual Nature of Spread Dynamics

| **Characteristic** | **Expiry Mechanics (Cannot Hide)** | **Rolling Activity (Can Hide)** |
|---|---|---|
| Temporal Clustering | Strong (63.8% in days 1–7) | No detectable clustering |
| Multi-Spread Pattern | Systematic repetition (S1–S11) | Not visible in volume data |
| Term Structure | Orderly, contango-dominant | No phase-dependent changes |
| Volume Correlation | N/A (mechanical, not behavioral) | Zero ($r = 0.0043$) |
| Detectability | Unavoidable systematic effect | Successfully hidden |
| Interpretation | Contract lifecycle mechanics | Sophisticated institutional execution |

This dual nature explains the apparent paradox of strong temporal clustering combined with absent volume signatures: **Expiry mechanics create the systematic timing patterns we observe, while institutional rolling activity occurs continuously and stealthily without creating detectable volume events.**

### 6.4 Methodological Insights

The analysis reveals several critical methodological insights:

1. **Definition of Contract Month is Fundamental:** The choice between "days since became F1" (3.2% clustering) versus "days since previous expiry" (63.8% clustering) fundamentally determines whether we conclude that institutional behavior is discretionary versus mechanically driven. Future research must carefully define temporal proximity to expiry based on the hypothesis being tested.

2. **Single-Spread Analysis is Insufficient:** Examining only S1 cannot distinguish institutional roll timing (unique to S1) from expiry mechanics (systematic across all spreads). Multi-spread comparative analysis is essential for causal inference.

3. **Volume Independence Does Not Imply Absence of Activity:** The zero correlation between spread events and volume ratios does *not* mean that rolling activity is absent or unimportant. Instead, it demonstrates that sophisticated execution successfully hides rolling activity in volume data while expiry mechanics remain visible in spread dynamics.

4. **Statistical Significance vs. Economic Interpretation:** The first-appearance proxy shows statistically significant non-uniformity ($\chi^2 = 115.8$, $p = 0.001$) but economically weak clustering (3.2%). The supervisor's definition shows both statistical and economic significance (63.8% clustering, $\chi^2 = 1195.2$, $p < 0.001$). Statistical tests alone are insufficient without economic magnitude assessment.

### 6.5 Revised Interpretation of Original Hypothesis

The original hypothesis that "clustering at 19 days indicates institutional roll timing" must be revised in light of the comprehensive evidence:

**Original Interpretation:** Events at 19–28 days before F1 expiry represent coordinated institutional rolling, with institutions choosing to roll mid-lifecycle to balance liquidity, cost, and delivery risk.

**Revised Interpretation:** Events at 28–30 days before each contract expires (equivalently, 5 days after previous contract expires) represent systematic contract maturity effects. These patterns exist regardless of institutional behavior and reflect fundamental market microstructure mechanics: declining liquidity in expiring contracts, position concentration in deferred contracts, and term structure adjustments as the forward curve shifts.

Institutional rolling activity *does* occur and *does* impact spreads, but it is successfully hidden through sophisticated execution strategies. What we detect through z-score thresholds are not discrete roll execution events but rather systematic volatility associated with the contract lifecycle transition.

## 7 Practical Implications

### 7.1 Continuous Futures Construction

Index providers and data vendors constructing continuous futures series ("perpetual contracts" or "back-adjusted series") must incorporate the 28–30 day pattern and 63.8% early-month clustering to avoid introducing artificial volatility.

**Current Practice:** Most continuous series roll from F1 to F2 on a fixed schedule, typically:

- Bloomberg (BCOM): 6th–10th business day of the month before expiry
- S&P GSCI: 5th–9th business day of the month before expiry
- "Last Trading Day" method: Roll on the final trading day before expiry

**Implications of This Analysis:**

1. **Early Rolling is Justified:** The 63.8% clustering in days 1–7 post-expiry (equivalently, approximately 25–31 days pre-expiry) validates the BCOM/GSCI approach of rolling 5–10 days before expiry. This timing coincides with the peak systematic widening period.

2. **Last Trading Day Method is Problematic:** Rolling on the final day before expiry occurs *after* the peak event period and during the phase with lowest liquidity. This likely incurs higher transaction costs and greater spread slippage.

3. **Calendar vs. Lifecycle Timing:** Using calendar-based rules ("6th business day of the month") may misalign with lifecycle-based mechanics ("5 days after previous expiry"). A lifecycle-based rule would better match the underlying mechanics, though it complicates index replication.

4. **Adjustment Method Matters:** Series using "backward ratio adjustment" (scaling historical prices) versus "backward difference adjustment" (subtracting spread) will exhibit different volatility during roll windows. The orderly term structure evolution (no forced liquidations) suggests that difference adjustment is appropriate.

### 7.2 Systematic Trading Strategies

The systematic 28–30 day pattern creates opportunities for statistical arbitrage and systematic strategies.

### 7.2.1 Calendar Spread Trading

**Strategy:** Long calendar spreads (long F2, short F1) approximately 35–40 days before F1 expiry, anticipating systematic widening during the 28–30 day window, and close positions at 15–20 days before expiry.

**Risk Management:**

- **Entry Timing:** Enter before the 28–30 day window to capture the full widening
- **Exit Timing:** Exit after the median event period (28 days) to avoid holding through final week illiquidity
- **Hedging:** Hedge against outright price moves using F1-only delta hedges
- **Capacity:** Limited by S1 liquidity; large positions may impact the spread they seek to exploit

**Expected Performance:** Positive carry in contango markets (long F2 yields positive spread), though this may be offset by roll costs. Profitability depends on spread widening magnitude exceeding execution costs.

### 7.2.2 Statistical Arbitrage: Multi-Spread Correlation

**Strategy:** Exploit the moderate positive correlations (0.32–0.46) between adjacent spreads by trading S1-S2 or S2-S3 relative value when correlations temporarily break down.

**Implementation:**

- Compute rolling correlation between S1 and S2 over 20-day windows
- When correlation drops below 0.30 (below historical average), assume mean reversion and trade the spread-of-spreads
- Long the underperforming spread, short the overperforming spread
- Exit when correlation reverts to historical mean

**Challenges:**

- Requires simultaneous trading in four contracts (F1, F2, F3, F4), increasing execution complexity
- Correlations may remain low for extended periods during market stress
- Transaction costs (4 legs) may eliminate edge

## 7.3 Risk Management for Portfolio Managers

Portfolio managers with exposure to front-month futures contracts must anticipate the 28–30 day widening cycle to avoid adverse roll timing.

### 7.3.1 Avoiding Crowded Roll Windows

**Problem:** Rolling during the peak event period (days 1–7 post-expiry, 63.8% clustering) exposes portfolios to elevated spread volatility and potential adverse execution.

**Solution:** Implement staggered rolling schedules:

- Begin rolling 40–45 days before expiry

- Execute 10–20% of position per day over 5–10 days
- Complete rolling by 30 days before expiry, avoiding the peak widening window

**Trade-offs:**

- **Benefit:** Avoid concentrated execution during high-volatility periods
- **Cost:** Hold deferred contracts (F2) for longer periods, increasing exposure to contango costs
- **Complexity:** Requires tracking multiple expiry dates and coordinating execution across contracts

#### 7.3.2 Dynamic Hedging Based on Days-to-Expiry

**Strategy:** Adjust portfolio hedge ratios dynamically based on days-to-expiry, increasing hedges during the 28–30 day window when spread volatility is elevated.

**Implementation:**

- Baseline hedge: 100% delta hedge using F1
- 30–40 days to expiry: Increase hedge to 105–110% to account for anticipated spread widening
- 15–30 days to expiry: Maintain elevated hedge
- $< 15$ days to expiry: Reduce hedge back to 100% as event probability declines

This dynamic approach reduces portfolio variance during predictable high-volatility periods without incurring permanent over-hedging costs.

### 7.4 Market Making and Liquidity Provision

Market makers providing liquidity in calendar spreads can use the systematic patterns to adjust bid-ask spreads and inventory management.

#### 7.4.1 Dynamic Spread Widening

**Strategy:** Widen bid-ask spreads during the 28–30 day window to compensate for elevated volatility and adverse selection risk.

**Implementation:**

- Baseline spread: 1–2 ticks (0.0005–0.0010 USD/lb) during calm periods ($> 40$ days or $< 15$ days to expiry)
- Elevated spread: 2–4 ticks (0.0010–0.0020 USD/lb) during the 20–35 day window
- Maximum spread: 4–6 ticks (0.0020–0.0030 USD/lb) during days 1–7 post-expiry (peak event period)

#### 7.4.2 Inventory Management

**Problem:** Market makers accumulating calendar spread inventory (long S1 or short S1) face directional risk from systematic widening.

**Solution:** Reduce inventory limits during peak event periods:

- Normal inventory limit: 500 spreads
- Reduced limit (20–35 days to F1 expiry): 250 spreads
- Minimal limit (days 1–7 post-expiry): 100 spreads

This reduces exposure during predictable high-risk periods without permanently reducing market-making capacity.

# 8 Limitations and Future Work

## 8.1 Single Commodity Analysis

This analysis focuses exclusively on CME copper futures (HG). While copper is a highly liquid benchmark for industrial metals, conclusions may not generalize to other commodities with different microstructure characteristics:

- **Energy Commodities:** Crude oil (CL), natural gas (NG) have different expiry cycles (monthly for crude, monthly for natural gas), storage economics, and index fund concentration. The 28–30 day pattern may differ due to Cushing storage dynamics (crude) or seasonal demand (natural gas).
- **Agricultural Commodities:** Corn (C), soybeans (S), wheat (W) have seasonal expiry cycles (March, May, July, September, December for corn/soybeans), making continuous rolling less relevant. Harvest timing and weather-driven volatility may dominate over expiry mechanics.
- **Precious Metals:** Gold (GC), silver (SI) have minimal storage costs and serve as financial assets rather than industrial inputs, potentially exhibiting different term structure dynamics and roll patterns.
- **Financial Futures:** Equity index futures (ES, NQ), interest rate futures (ZN, ZB) have different user bases (hedgers vs. speculators) and expiry mechanics (cash settlement vs. physical delivery), likely requiring entirely different analysis frameworks.

**Future Work:** Extend the framework to a cross-sectional panel of 10–20 commodities, testing whether the 28–30 day pattern and 63.8% early-month clustering generalize or are copper-specific. This would enable identification of commodity-level moderating factors (storage costs, expiry frequency, contract liquidity).

## 8.2 Z-Score Threshold Sensitivity

The event detection methodology uses a fixed threshold of 1.5 standard deviations. While this balances sensitivity and specificity, results may be sensitive to this choice:

- **Higher Thresholds (2.0–2.5 SD):** Would detect fewer, more extreme events, potentially concentrating detection in specific phases
- **Lower Thresholds (1.0–1.25 SD):** Would detect more events, potentially diluting the temporal clustering signal with noise

- **Adaptive Thresholds:** Time-varying thresholds based on recent volatility regimes could improve detection accuracy

**Future Work:** Conduct sensitivity analysis across threshold values (1.0, 1.25, 1.5, 1.75, 2.0 SD), assessing robustness of the 63.8% clustering result. Optimal threshold could be selected via ROC curve analysis using a labeled validation set (e.g., major index roll dates as ground truth).

### 8.3 Open Interest Data Would Strengthen Analysis

This analysis uses price and volume data exclusively. Open interest (OI) data—the total number of outstanding contracts—would provide direct evidence of position migration:

- **OI Migration Timing:** Track when F2 open interest surpasses F1 open interest, providing an unambiguous measure of position concentration shifts
- **OI vs. Volume:** Distinguish between position transfers (roll activity, increases OI in F2) versus round-trip trading (speculative activity, increases volume without OI change)
- **Net Positioning:** Decompose OI into hedger vs. speculator positions (from CFTC Commitment of Traders reports), testing whether spread events coincide with hedger roll timing

**Data Availability Challenge:** High-frequency open interest data (minute-level or even daily) are not widely available in public datasets. CME DataMine provides daily settlement OI, but intraday OI would require proprietary data subscriptions.

**Future Work:** Acquire daily OI data and repeat volume migration analysis using F2/F1 OI ratios. Hypothesized result: OI migration (unlike volume migration) may correlate with spread events, revealing the hidden rolling activity that volume analysis could not detect.

### 8.4 Cross-Market Comparison: Gold and Crude Oil

To validate the generalizability of the expiry mechanics hypothesis, comparative analysis with gold (GC) and crude oil (CL) futures would be highly informative:

- **Gold:** Typically in contango due to financing costs but minimal storage costs. Term structure may be flatter than copper. Gold is more financial asset than industrial commodity, potentially attracting different institutional behavior (hedge funds vs. industrial hedgers).
- **Crude Oil:** Highly liquid with significant index fund presence. Cushing storage dynamics create strong supply-driven term structure effects. WTI often exhibits backwardation during supply disruptions, contrasting with copper's persistent contango.

**Predicted Results:**

- **Gold:** May show weaker clustering (flatter term structure reduces mechanical pressures) but similar multi-spread repetition (systematic expiry mechanics still apply)
- **Crude Oil:** May show stronger clustering (higher index fund concentration during roll windows) but similar timing (28–30 day pattern driven by expiry mechanics, not commodity-specific factors)

**Future Work:** Apply identical methodology to GC and CL futures (2008–2024), comparing detection rates, timing distributions, term structure evolution, and volume correlations. This would enable meta-analysis identifying universal expiry mechanics versus commodity-specific factors.

### 8.5 Incorporation of Macroeconomic Variables

The current analysis treats spread dynamics as intrinsic to contract mechanics. However, macroeconomic conditions may moderate the strength and timing of observed patterns:

- **Interest Rates:** Contango magnitude depends on financing costs. Low interest rates (2008–2021) may flatten term structures, reducing mechanical pressures. Rising rates (2022–2024) may steepen curves.
- **Inventory Levels:** Physical copper inventories (LME warehouses, COMEX warehouses) influence storage costs and spot premia. High inventories may reduce backwardation frequency.
- **Volatility Regimes:** Market-wide volatility (VIX, realized volatility in copper) may amplify or dampen event detection. High-volatility regimes may generate more threshold-crossing events mechanically, unrelated to roll timing.

**Future Work:** Incorporate macroeconomic covariates into regression models predicting event occurrence:

$$P(\text{Event}_{i,t}) = \text{logit}^{-1}(\beta_0 + \beta_1 \text{DaysToExpiry}_{i,t} + \beta_2 \text{InterestRate}_t + \beta_3 \text{Inventory}_t + \beta_4 \text{Volatility}_t) \tag{4}$$

This would enable decomposition of systematic expiry effects (captured by $\beta_1$) from macroeconomic moderators.

## 9 Conclusions

This report presents comprehensive analysis of calendar spread dynamics in CME copper futures over 16 years (2008–2024), processing 202 contracts and detecting 2,737 significant widening events. The analysis adjudicates between two competing hypotheses: whether observed patterns reflect discretionary institutional roll timing or systematic contract expiry mechanics.

### 9.1 Principal Findings

**Expiry Mechanics Dominate.** The supervisor's hypothesis is validated through three independent lines of evidence: (1) 63.8% of events cluster in the first seven days after previous contract expiry when properly defined ($\chi^2 = 1195.2$, $p < 0.001$), contrasting sharply with the 3.2% clustering under the first-appearance proxy; (2) the 28–30 day timing pattern systematically repeats across all 11 spreads (S1–S11), ruling out S1-specific institutional behavior; (3) orderly term structure evolution with 74% contango prevalence contradicts forced liquidation hypotheses.

**Rolling Activity Remains Hidden.** Despite strong temporal clustering in spread dynamics, volume migration analysis reveals zero correlation between F2/F1 volume ratios and event occurrence (Pearson $r = 0.0043$, $p = 0.369$). F1 remains volume-dominant during events (median ratio 0.426 vs. 1.692 outside events), demonstrating that institutional rolling activity is successfully hidden through sophisticated execution strategies (algorithmic trading, dark pools, continuous rolling).

**Methodological Sensitivity is Critical.** The definition of "contract month" fundamentally determines interpretation: days since first becoming F1 (3.2% clustering) versus days since previous expiry (63.8% clustering). This 20-fold difference demonstrates that methodological choices regarding temporal proximity to expiry are not neutral—they encode implicit assumptions about whether patterns reflect lifecycle mechanics versus calendar-based behavior.

## 9.2 Implications for Interpretation

The original hypothesis that "clustering at 19 days indicates institutional roll timing" must be revised. What we detect through statistical thresholds are not discrete roll execution events but rather **systematic volatility associated with contract lifecycle transitions**. These transitions occur approximately 28–30 days before each contract expires (equivalently, 5 days after the previous contract expires), regardless of institutional behavior.

This does not mean that institutional rolling activity is absent or unimportant. On the contrary, institutions *do* roll positions and *do* impact spreads. However, they execute rolls using sophisticated strategies that successfully avoid creating detectable volume signatures while expiry mechanics remain unavoidably visible in spread dynamics.

The framework thus characterizes a **dual nature** in spread dynamics: systematic patterns that cannot be hidden (intrinsic to contract mechanics) coexist with discretionary activity that can be hidden (through sophisticated execution). This duality explains the apparent paradox of strong temporal clustering combined with absent volume correlation.

## 9.3 Practical Contributions

This analysis provides quantitative foundations for multiple practical applications:

- **Continuous Futures Construction:** Validates early rolling (5–10 days before expiry) used by major index providers, contradicts last-trading-day methods
- **Systematic Trading Strategies:** Enables calendar spread trading strategies anticipating systematic widening during the 28–30 day window
- **Risk Management:** Informs dynamic hedging strategies that increase hedge ratios during predictable high-volatility periods
- **Market Making:** Justifies dynamic bid-ask spread widening and reduced inventory limits during peak event periods

## 9.4 Methodological Contributions

The framework demonstrates several methodological advances applicable beyond copper futures:

1. **Vectorized Contract Identification:** Deterministic F1–F12 chain construction using days-to-expiry matrices enables sub-second computation on 16-year datasets, replacing iterative loops with $O(n \times m)$ array operations.
2. **Multi-Spread Comparative Analysis:** Testing hypotheses across S1–S11 simultaneously enables causal inference distinguishing institutional behavior (S1-specific) from expiry mechanics (systematic repetition).
3. **Dual Proxy Analysis:** Testing multiple definitions of temporal proximity (first-appearance vs. post-expiry) reveals methodological sensitivity and prevents premature conclusions based on single metrics.
4. **Term Structure Evolution Tracking:** Full forward curve (F1–F6) analysis through lifecycle phases enables detection of forced liquidations, panic-driven behavior, and phase-dependent structural changes.

## 9.5 Limitations and Extensions

The analysis focuses exclusively on copper futures, uses price and volume data (but not open interest), and employs a fixed z-score threshold (1.5 SD). Future work should extend the framework to:

- Cross-sectional panel of 10–20 commodities (energy, agriculture, precious metals, financials)
- Incorporate open interest data to directly observe position migration
- Sensitivity analysis across detection thresholds (1.0–2.5 SD)
- Incorporate macroeconomic covariates (interest rates, inventory levels, volatility regimes)

## 9.6 Final Remarks

The central finding of this analysis is that **systematic expiry mechanics create patterns in calendar spread dynamics that cannot be eliminated through strategic execution choices**. The 63.8% clustering in the first seven days after previous contract expiry, the systematic 28–30 day repetition across all spreads, and the orderly term structure evolution are fundamental properties of futures market microstructure.

These patterns exist regardless of institutional behavior and must be incorporated into any model of futures price dynamics, continuous series construction, or systematic trading strategy. While rolling activity itself can be hidden through sophisticated execution, the mechanical consequences of approaching expiry—declining liquidity, position concentration, term structure adjustments—remain visible and quantifiable.

This framework provides essential foundations for understanding how futures markets function during roll periods, enabling more accurate pricing models, more efficient execution strategies, and more robust risk management practices across commodity, energy, and financial futures markets.