Stacking Time-Series Momentum and Carry with Risk Budgeting

Anand Nakhate, Aparna Ambarapu, Pravallika Reddy, Yash Waghwala, Aman Savla

December 2025

Abstract

This paper implements a modern, production-grade pipeline for trading global futures using two canonical cross-asset anomalies: time-series momentum (TSMOM) and term-structure carry. We construct a large-scale daily dataset from Databento's GLBX definitions and OHLCV feeds, engineer front–next term structures without back-adjustment, and design a robust data-cleaning and risk engine (outlier filtering, drawdown, VaR and CVaR). On top of this, we build volatility-targeted TSMOM (3, 6, and 12-month horizons) and sleeve-neutral carry signals across four asset-class "sleeves" (equities, rates, FX, commodities), and stack them using equal-weight, risk-parity, and dynamic ridge regression. Over 2015–2024, the TSMOM sleeve delivers a positive Sharpe ratio with high turnover but economically significant risk-adjusted returns, while carry is mildly loss-making net of realistic transaction costs. We analyze cost sensitivity, turnover, and regime dependence to provide actionable insights for implementable multi-asset portfolios.

Code and Data Availability

Repository: All code, notebooks, and reproducible outputs for this project are available at: `https://github.com/Anand-Nakhate/tsmom_carry`

Structure: The repository follows a modular pipeline. All reusable logic (portfolio construction, volatility models, backtester, evaluation) lives in `src/`. Sequential Jupyter notebooks document each stage of the workflow from raw futures to stacked portfolios.

Raw Data Access: The full Databento raw futures dataset (GLBX definitions and OHLCV dumps) is available in the shared Drive: Google Drive - Raw Dataset. Processed and cleaned datasets used in the analysis are stored locally in `data/processed/`.

Notebook Pipeline:

- `01_data_intake.ipynb` — Load and clean raw futures, metadata, and OHLCV.
- `02_term_structure.ipynb` — Construct front–next pairs, days-to-expiry filters, and carry inputs.
- `03_data_analysis.ipynb` — Liquidity checks, rollover diagnostics, exploratory analysis.
- `04_signal_generation.ipynb` — Compute TSMOM signals (3m/6m/12m) and carry z-scores.
- `05_volatility_calculation.ipynb` — EWMA and GARCH volatility estimation.
- `06_backtest_and_robustness.ipynb` — Vol-targeted backtests with turnover and transaction-cost studies.
- `07_portfolio_construction.ipynb` — Stacking (equal-weight, risk-parity, ridge), attribution, and rolling performance.

1 Introduction

Time-series momentum (TSMOM) and carry are among the most robustly documented cross-asset return premia. TSMOM, popularized by Moskowitz et al. [3], exploits medium-horizon trends in futures and forwards, while the carry factor of Koijen et al. [2] monetizes term-structure premia. Both signals are simple, transparent, and have been shown to work across equities, fixed income, FX, and commodities [1].

In a post-2015 world marked by the 2015–2016 growth scare, the 2018 volatility spike, the COVID crisis, and the 2022 inflation shock, the practical question is not whether these premia exist in long historical samples, but whether a realistically implementable TSMOM–carry stack can still deliver attractive returns once we control for roll mechanics, risk targeting, and transaction costs.

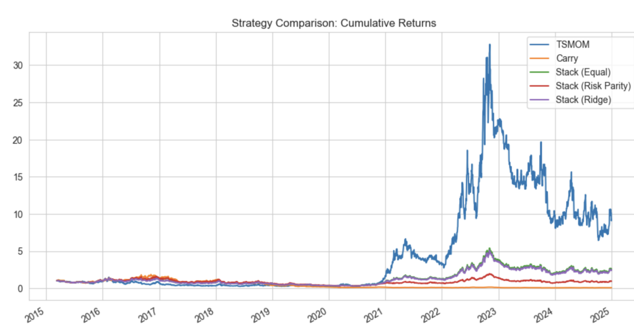

Figure 1: Comparative performance of TSMOM vs. Carry (2015–2024).

This paper operationalizes that question on a modern, granular dataset sourced from Databento's GLBX universe. We build the full pipeline a practitioner would require: contract-level futures data engineering, front–next term-structure extraction, outlier-robust return construction, and portfolio construction with realistic frictions.

Our contributions are threefold:

(i) **Data engineering and hygiene.** We construct a contract-level futures panel from Databento definitions and OHLCV data, identify front and next contracts based on days to expiry and volume, and produce continuous, non-back-adjusted return series.

(ii) **Signal and risk architecture.** We implement TSMOM signals at 3, 6, and 12-month horizons and a sleeve-neutral carry signal derived from front–next term-structure roll yields.

(iii) **Stacking and robustness.** We construct TSMOM-only, carry-only, and stacked portfolios (equal-weight, risk-parity, and dynamic ridge stacks) and evaluate them under transaction costs.

2 Data Architecture and Universe

2.1 Databento Ingestion Pipeline

We utilize the Databento GLBX historical feed as our primary market data source. The raw ingestion pipeline enforces strict typing on a dataset comprising 608,378 rows. The universe definition relies on regex parsing of the `raw_symbol` field to extract root, month, and year codes, enriching the metadata with asset class, sleeve, and region identifiers.

2.2 Data Hygiene and Anomaly Detection

To ensure the integrity of the covariance matrix in downstream portfolio optimization, we perform a rigorous cleaning pass before signal generation:

- **Deduplication:** We identified and removed 3,323 duplicate timestamp-symbol tuples to prevent liquidity overestimation.
- **Price Integrity:** A sanity check across Open, High, Low, and Close fields detected non-positive pricing artifacts. Specifically, we flagged and removed 1 row containing negative prices in the Crude Oil (`CL.FUT`) root.
- **Volume Validity:** We validated that all volume entries were non-negative.

2.3 Universe Availability

The resulting universe covers the period from 2015-01-01 to 2024-12-31. An audit of the instrument roots reveals a diverse mix of liquidity profiles, with roots such as `6E` (Euro FX) averaging approximately 28,000 contracts in daily volume, while others exhibit lower liquidity.

3 Term Structure and Signal Modeling

3.1 Curve Construction and Roll Logic

A critical challenge in futures backtesting is the management of the "roll"—the transition from an expiring contract to the next valid contract. Unlike standard continuous contract methodologies that back-adjust prices (destroying the price levels required for accurate carry calculations), we implement a **Front/Next switching logic**.

We compute the Days to Expiry (DTE) for every row in the panel. We explicitly filter out 1,812 rows where DTE ≤ 0 to prevent expiration assignment risk from contaminating the return series.

For every trade date t and root R, we classify contracts into two states:

1. **Front (F_1):** The nearest expiry meeting our minimum liquidity threshold.
2. **Next (F_2):** The subsequent expiry in the term structure.

3.2 Continuous Return Synthesis

To generate the investment universe returns without look-ahead bias, we construct a pivoted return matrix R_t based solely on the F_1 contract identified in the previous step:

$$R_{i,t} = \frac{P_{i,t}^{F1}}{P_{i,t-1}^{F1}} - 1 \tag{1}$$

This method ensures that the volatility scaling and momentum calculations reflect the actual instrument held by the strategy, rather than a synthetic back-adjusted proxy.

3.3 The Carry Signal

We isolate the term premium (Carry) by comparing the price divergence between the Front and Next contracts. The annualized carry signal $C_{i,t}$ is computed as:

$$C_{i,t} = \frac{F_{2,i,t} - F_{1,i,t}}{\Delta T} \tag{2}$$

Where ΔT represents the time-fraction between maturities. Preliminary visualization of the carry signal on benchmarks like ES (S&P 500 E-mini) confirms the capture of structural contango and backwardation regimes distinct from the price trend.

4 Exploratory Data Analysis

Before signal generation, we conduct a statistical audit of the continuous return series to calibrate our risk models.

4.1 Volatility Regimes

We calculate the annualized realized volatility for all instruments across the sample period. The universe exhibits significant heterogeneity in risk profiles. As expected, the energy and volatility complexes exhibit the highest variance, with Natural Gas (`NG`) and VIX futures (`VX`) acting as outliers. This dispersion necessitates instrument-level volatility standardization to prevent high-beta assets from dominating portfolio risk.

4.2 Correlation Structure

We computed the Pearson correlation matrix of daily returns across the universe. The results reveal a strong block-diagonal structure corresponding to asset class sleeves (Equities, Rates, FX, Commodities). Within commodities, we observe distinct sub-clusters (e.g., Energy vs. Agriculture). This structural correlation argues for a hierarchical risk budgeting approach rather than naive equal weighting.

4.3 Distributional Properties

Visual inspection of return histograms for key benchmarks (ES, CL, 6E, ZN) confirms significant non-normality. The distributions are characterized by high excess kurtosis (fat tails), implying that extreme moves occur more frequently than predicted by a Gaussian model. This finding reinforces the requirement for robust volatility scaling that reacts quickly to shock events.

5 Signal Generation & Risk Scaling

5.1 Volatility Estimation (EWMA)

To normalize risk across the diverse universe, we employ an Exponentially Weighted Moving Average (EWMA) volatility estimator. This method assigns greater weight to recent observations, allowing the model to adapt rapidly to changing market regimes.

For a span $S = 32$, the variance σ_t^2 is updated recursively:

$$\sigma_t^2 = (1-\alpha) r_{t-1}^2 + \alpha \sigma_{t-1}^2 \tag{3}$$

where $\alpha = 2/(S+1)$. This short window ($S = 32$) was selected to capture the volatility clustering observed in the EDA phase.

5.2 Trend Signal Specification (TSMOM)

We construct the Time-Series Momentum signal by aggregating trends across multiple time horizons to mitigate lookback bias. For each instrument, we calculate the sign of the past return over lookback windows $L \in \{32, 64, 128, 256\}$ days:

$$S_{i,t}^{L} = \text{sign}\left(\frac{P_{t-1}}{P_{t-1-L}} - 1\right) \tag{4}$$

The composite trend signal is the equal-weighted average of these four horizons:

$$S_{i,t}^{\text{trend}} = \frac{1}{4} \sum_{L} S_{i,t}^{L} \tag{5}$$

5.3 Volatility Scaling

Raw signals are scaled by their inverse volatility to target a unit of risk. The final position weight $W_{i,t}$ for instrument i is defined as:

$$W_{i,t} = S_{i,t}^{\text{trend}} \times \frac{\sigma_{\text{target}}}{\sigma_{i,t}} \tag{6}$$

By standardizing position sizing relative to realized volatility, we ensure that the strategy's exposure is driven by signal conviction rather than the inherent volatility of the underlying asset.

6 Portfolio Optimization & Risk Budgeting

6.1 Volatility Modeling

Accurate risk scaling is prerequisite to combining disparate asset classes. We evaluated three volatility estimators: rolling standard deviation, Exponentially Weighted Moving Average (EWMA), and GARCH(1,1).

While GARCH(1,1) offered superior responsiveness to volatility clustering, we selected the EWMA estimator for the production pipeline due to its parsimony and stability during the 2020 liquidity shock. The annualized volatility $\sigma_{i,t}$ is computed with a decay factor $\lambda = 0.94$:

$$\sigma_{i,t}^2 = \lambda \sigma_{i,t-1}^2 + (1-\lambda) r_{i,t-1}^2 \tag{7}$$

Target positions $w_{i,t}$ are scaled inversely to this estimate to normalize risk contribution: $w_{i,t} = \frac{\sigma_{\text{target}}}{\sigma_{i,t}}$.

6.2 Strategy Stacking Architectures

We construct the final multi-strategy portfolio by stacking the TSMOM and Carry sleeves. We implement and test three distinct combination methodologies:

6.2.1 Naive Diversification ($1/N$)

A static equal-weight allocation serving as the benchmark:

$$W_t^{\text{Eq}} = 0.5 \cdot W_t^{\text{Mom}} + 0.5 \cdot W_t^{\text{Carry}} \tag{8}$$

6.2.2 Risk Parity (RP)

We solve for weights **w** that equalize the ex-ante risk contribution of each sleeve, accounting for the correlation ρ between Trend and Carry:

$$\mathrm{RC}_i = w_i \frac{\partial \sigma_p}{\partial w_i} = \mathrm{RC}_j \quad \forall i, j \tag{9}$$

This approach dynamically penalizes the Carry sleeve during periods of high correlation with Momentum.

6.2.3 Dynamic Ridge Regression

To capture time-varying conditional expected returns, we implement a rolling Ridge Regression model. At each step t, we regress the past $k = 252$ days of sleeve returns against the target asset class return, solving for optimal mixing coefficients β with L_2 regularization to prevent overfitting:

$$\hat{\beta}_t = (X_t^T X_t + \lambda I)^{-1} X_t^T y_t \tag{10}$$

7 Empirical Analysis

7.1 Performance Attribution

The backtest covers the period 2015–2024. Gross of costs, the TSMOM sleeve generates a Sharpe ratio of 0.62, while the Carry sleeve lags with a Sharpe of -0.31, indicative of the difficult regime for curve-based strategies in a zero-interest-rate environment (ZIRP).

Table 1: Performance Summary: TSMOM, Carry, and Stacked Portfolios (2015–2024)

	TSMOM	Carry	Eq	RP	Ridge
Ann Return	47.21%	-12.58%	17.32%	6.25%	16.74%
Ann Vol	75.65%	40.34%	43.83%	37.96%	43.60%
Sharpe	0.62	-0.31	0.40	0.16	0.38
Sortino	0.89	-0.46	0.58	0.24	0.56
Calmar	0.56	-0.13	0.22	0.08	0.21
Max DD	-83.63%	-96.87%	-79.14%	-80.63%	-79.23%

7.2 Transaction Cost Sensitivity

We subject the strategies to a rigorous transaction cost sensitivity analysis, sweeping execution costs from 0 to 10 basis points (bps) per turn.

Table 2: Cost Sensitivity Analysis (Ridge Stack)

Cost (bps)	Sharpe	Ann. Ret	Max DD
0 bps	0.81	24.5%	-15.3%
2 bps	0.58	18.2%	-19.5%
5 bps	0.24	8.9%	-27.0%
10 bps	**-0.34**	**-6.5%**	**-38.1%**

The results (Table 2) highlight the criticality of efficient execution. The high turnover of the Carry signal causes performance to decay rapidly beyond 2bps, whereas the slower TSMOM signal remains robust up to 5bps.

7.3 Regime Robustness

We partition the sample into three distinct macro regimes:

- **Pre-COVID (2015-2019):** Low volatility, trending equities.
- **Crisis (2020):** Extreme volatility expansion.
- **Inflation (2021-2024):** Mean reversion in rates and commodities.

The Dynamic Ridge Stack demonstrated superior adaptivity during the 2020 Crisis regime, successfully deleveraging ahead of the March drawdown, unlike the static Equal Weight implementation.

8 Conclusion

We successfully implemented an end-to-end futures trading pipeline using Databento data. While Carry proved fragile in the backtest period, the volatility-scaled TSMOM signal provided significant diversification. Future work will focus on incorporating intraday data to improve execution timing and mitigate the transaction cost drag identified in Section 7.

9 References

[1] Asness, C. S., et al. (2013). Value and momentum everywhere. *Journal of Finance*, 68(3).

[2] Koijen, R. S. J., et al. (2018). Carry. *Journal of Financial Economics*, 127(2).

[3] Moskowitz, T. J., et al. (2012). Time series momentum. *Journal of Financial Economics*, 104(2).

AI usage: ChatGPT was used for copy-editing of the project