ROBUST MULTI-ASSET TREND FOLLOWING IN GLOBAL FUTURES

A systematic study of time-series momentum, volatility normalization, trading costs, diversification, and robustness across global futures

Pratik Bang | August 2026

Primary data provider: Norgate Data
Research stack: Norgate Data + RealTest + Python

First completed study in a continuing independent research program. Results are reported as gross, net of the stated cost model, or constrained portfolio evidence. The historical sample ends on 14 August 2026; the strategy specification is frozen for prospective monitoring after that date.

Executive Summary

This report asks whether systematic futures strategies remain credible after the assumptions that usually make a backtest look attractive are made explicit. The investigation begins with a literature-based 12-month time-series momentum rule across a frozen universe of 75 global futures markets. It then introduces contract economics, currency conversion, lagged volatility forecasts, whole-contract sizing, realistic commissions and slippage, roll activity, concentration limits, leverage constraints, and changing performance through time.

Component	Current research specification
Primary data	Norgate Data daily futures: individual contracts plus pre-built continuous-futures series.
Primary signal	12-month / 252-trading-bar time-series momentum sign.
Direction	Long when current price exceeds its 252-bar lag; short when below.
Baseline volatility	Lagged EWMA of daily point changes, 60-trading-day center of mass, annualized with 252 days.
Risk normalization	Instrument returns scaled to a common ex-ante volatility convention before portfolio aggregation.
Primary continuous series	Norgate back-adjusted continuous-futures for signal research; unadjusted series and individual contracts retained for implementation/roll checks.
Execution convention	Daily implementation with no look-ahead; the exact RealTest execution convention was frozen after Norgate import validation and independently reconciled in Python.
Costs	USD 2.50 commission plus one-tick of slippage per contract leg, with contract-specific tick values, currency conversion, and explicit roll activity.
Research governance	Universe and methodological choices are frozen before accepting strategy performance; changes must be data- or implementation-motivated, not performance-motivated.

A fixed equal-risk portfolio of four strategies preserved the best standalone Sharpe ratio while reducing maximum drawdown by 25.3%, from 20.01% to 14.94%. Applying the final risk target, market and asset-class caps, leverage ceiling, gross notional limit, and whole-contract rounding produced a more implementable account: 6.44% net compound annual growth, 7.19% volatility, a 0.877 Sharpe ratio, and a −11.95% maximum drawdown. These controls were economically consequential rather than cosmetic: the leverage ceiling bound 56 monthly decisions, the market cap bound on 552 dates, the asset-class cap on 4,769 dates, and the gross notional cap on 2,617 dates.

The final conclusion is deliberately narrower than the most attractive backtest. The evidence supports a constrained portfolio of complementary systematic futures strategies under fixed weights; it does not establish future profitability. Every one of the 31 complete rolling five-year windows had positive net growth and a positive Sharpe ratio, but those windows overlap and are not independent trials. The portfolio rules were fixed before the combined results were examined, and no sleeve, weight, or constraint was changed afterward. The decisive next evidence will come from the untouched live ledger that begins after 14 August 2026.

Contents

Pratik Bang | Global Futures Systematic Research | August 2026

1. Introduction

1.1 Motivation

Systematic trend following appears simple: identify the direction of a market, scale the position to its risk, and repeat the process across a diversified futures universe. That simplicity is appealing, but it can also be misleading. A strong historical equity curve may depend on continuous-contract construction, incomplete transaction-cost assumptions, unrealistic contract sizing, excessive leverage, or a small number of unusually profitable markets. A convincing result must survive those questions rather than avoid them. I began this study with a deliberately simple, literature-based benchmark: a 12-month time-series momentum signal applied across 75 global futures markets. The objective was not merely to reproduce an attractive gross backtest, but to determine what remained after progressively introducing the features that make a futures strategy difficult to implement in practice.

1.2 Research Questions

The study asks two related questions. First, how robustly can directional persistence be harvested after accounting for uncertainty in trend horizon, volatility estimation, trading costs, concentration, and implementation constraints? Second, can economically distinct, pre-specified futures strategies improve portfolio risk without relying on retrospective strategy selection or optimized allocation weights? These questions shift the emphasis away from finding the best-looking backtest and toward whether a result remains economically meaningful after its assumptions are made explicit and challenged.

1.3 Research Approach and Contribution

This report does not claim to discover trend following, carry, breakout trading, or cross-sectional momentum. Its contribution is an auditable reconstruction and extension of those ideas using Norgate Data, RealTest, and Python. Contract metadata and price histories are validated before performance is examined; the market universe is frozen without reference to strategy returns; the accepted benchmark is independently reconciled; and complementary strategies are evaluated alone before their correlations and combined behavior are studied. Negative results are retained when they answer a legitimate question. The final architecture is therefore the outcome of a controlled research process, not the combination that happened to maximize historical performance.

1.4 Interpretation of the Evidence

All results are labeled according to whether they are gross, net of the stated transaction-cost model, or produced under the final portfolio constraints. Historical performance is treated as evidence about a fixed research specification; not as a forecast or guarantee. The historical strategy-design phase ends on 14 August 2026. No strategy, weight, target, cap, or parameter may change during the prospective holdout without ending that holdout and beginning a newly documented specification.

2. Literature and Empirical Motivation

2.1 Time-Series Momentum

The baseline follows the time-series momentum literature: each market is traded according to the sign of its own trailing return rather than relative ranking across markets. Candidate explanations include gradual information diffusion, underreaction, persistent macroeconomic trends, institutional positioning, and hedging pressure; these are treated as hypotheses rather than proven causes (Moskowitz, Ooi, and Pedersen, 2012; Hurst, Ooi, and Pedersen, 2017).

2.2 Volatility Scaling

Futures differ materially in price scale, contract multiplier, and volatility. Equal contract counts would therefore create highly uneven risk. The study scales positions using ex-ante volatility forecasts and later tests whether more complex volatility models improve realized risk control.

2.3 Implementation Matters

Continuous-futures construction, contract rolls, settlement conventions, metadata, tick sizes, and transaction costs can materially affect backtests. These are treated as core research variables rather than afterthoughts.

2.4 Complementary Futures Strategies

The later portfolio sequence also tests cross-sectional momentum, which ranks markets relative to one another, and term structure or carry, which seeks compensation embedded in futures curves rather than price direction. These models are motivated by established evidence on momentum and carry, then translated into fixed, implementation-aware futures rules before their results are observed (Asness, Moskowitz, and Pedersen, 2013; Koijen et al., 2018).

The literature provides the economic hypotheses. The empirical test begins by defining exactly which prices, contracts, and markets represent those hypotheses.

3. Norgate Data and Futures Universe

3.1 Primary Data Source

The empirical study uses Norgate Data daily futures. Norgate provides historical individual futures contracts together with pre-built spot-month continuous contracts in unadjusted and arithmetically back-adjusted form. Unless otherwise specified by the provider, the daily Close field is the official settlement price.

3.2 Continuous-Futures Convention

The primary signal research uses Norgate arithmetically back-adjusted continuous-futures to remove artificial roll discontinuities from signal history. Matching unadjusted series and individual deliveries are retained for roll-event validation, term-structure measurement, and implementation robustness.

3.3 Candidate Universe

The initial Norgate inventory contains 224 continuous-futures series: 112 distinct economic futures markets, each represented by both an unadjusted continuous contract and an arithmetically back-adjusted continuous contract. The inventory spans stock indexes, interest rates, agriculture, currencies, energy, metals, and other futures across major global exchanges. The back-adjusted representation is designated as the primary research series; the matching unadjusted series is retained for roll-gap diagnostics and implementation robustness. Economic-market mapping and duplicate handling are completed before any strategy performance is examined.

3.4 Coverage Reset

Development stage universe results were discarded before strategy testing. The universe construction process then regenerated and froze the Norgate symbol mappings, date ranges, and Core, Supplemental, and Excluded labels before any performance evidence was examined.

With the raw inventory defined, the next task is to establish which records can support a credible historical test before any strategy return is examined.

4. Data Validation and Universe Construction

4.1 Validation Framework

Each candidate series was audited for OHLC consistency, missing or non-positive prices, stale observations, quote precision, extreme daily movements, continuity around rolls, and usable history. Contract metadata were checked directly after Norgate import into RealTest.

4.2 Contract Metadata Validation

The Norgate-to-RealTest metadata audit passed across the complete continuous-futures inventory. All 224 imported series had populated futures type, exchange, currency, PointValue, and TickSize fields; no series had a non-positive PointValue or TickSize. The 224 records comprise exactly 112 matched pairs of unadjusted and back-adjusted continuous contracts. This validates the contract-economics layer required for dollar P&L, volatility normalization, tick-aware execution, and transaction-cost modeling before baseline strategy results are accepted.

Table 1. Norgate continuous-futures inventory and metadata validation.

Audit item	Observed result
Imported continuous-series records	224
Distinct economic futures markets	112
Unadjusted continuous series	112
Back-adjusted continuous series	112
Futures type populated	224 / 224
Exchange populated	224 / 224
Currency populated	224 / 224
PointValue positive and populated	224 / 224
TickSize positive and populated	224 / 224

Table 2. Norgate continuous-futures records by asset-class (two continuous constructions per market).

asset-class	Series records	Distinct markets
Stock Index	64	32
Interest Rate	52	26
Agriculture	52	26
Currency	26	13
Energy	16	8
Metal	10	5
Other	4	2

4.3 Back-Adjusted Data-Quality Audit

The standardized data-quality audit was run across all 112 Norgate back-adjusted continuous-futures before any strategy performance was examined. All 112 series passed the structural OHLC consistency test with zero invalid high, low, open, or close relationships. Recent activity was strong for the large majority of markets: 107 of 112 recorded volume on at least 500 of the most recent 504 observations, and only two series produced any recent close-to-close movement greater than eight ATR units.

A key futures-specific result is that 25 back-adjusted series contain historical non-positive adjusted price levels. These observations are not automatically classified as bad data. Norgate uses arithmetic back-adjustment, so repeated historical roll-gap adjustments can legitimately push old adjusted levels through zero even though the underlying tradable contracts were positive. The baseline therefore uses point differences and point-volatility measures, which remain meaningful under additive back-adjustment, and does not compute percentage returns directly from back-adjusted price levels.

The audit isolated a small number of genuinely weak recent representations for further universe screening rather than ad-hoc repair. Eastern Australia Feed Barley (&AFB_CCB) had only 32 active-volume days in the recent 504-bar window, zero 60-day average volume, extensive flat/zero-volume history, and nine recent >8-ATR moves. Eastern Australia Wheat (&AWM_CCB), TecDAX (&FTDX_CCB), and the S&P GSCI futures series (&GD_CCB) also showed materially weaker recent trading activity. These findings will be handled in the pre-performance universe-construction stage alongside duplicate-contract and economic-exposure rules.

Flat bars, zero-volume history, and slow-moving short-rate contracts are treated diagnostically rather than with a single mechanical exclusion threshold. For example, contracts such as short-term interest-rate futures can exhibit many unchanged settlement prices while remaining actively traded. Final inclusion therefore combines structural validity,

recent trading activity, history depth, economic uniqueness, and implementability rather than equating every flat observation with a data error.

4.4 Universe Freeze

The universe construction process froze the baseline before any Norgate strategy performance was examined. Core eligibility required at least 2,500 daily bars, at least 500 active volume observations in the most recent 504 bars, a 60 trading day average volume of at least 5,000 contracts, and no more than two recent close-to-close moves greater than eight ATR units. After these screens, duplicate representations of the same economic exposure were resolved using history, liquidity, and standard contract preference. No strategy return, Sharpe ratio, drawdown, or trend following result was used in the selection.

Table 3. Frozen Norgate universe by asset-class.

asset-class	Core	Supplemental	Primary treatment
Stock Index	17	2	Global indexes; duplicate variants removed.
Interest Rate	19	4	Major sovereign and short-rate curves.
Currency	8	3	8 bilateral FX; USD Index supplemental.
Energy	7	0	7 benchmarks; duplicate WTI removed.
Metal	5	0	5 major metals.
Agriculture	18	0	18 liquid grains, softs and livestock.
Other	1	0	EU carbon allowances.
Total	**75**	**9**	28 additional representations excluded.

4.5 Frozen Research Universe

The frozen baseline contains 75 Core markets and 9 Supplemental markets; 28 of the original 112 back-adjusted representations are excluded. Supplemental markets are retained for specified sensitivity tests when they are economically useful but have shorter histories, lower recent volume, overlapping aggregate exposure, or structurally different futures economics. The complete classification is preserved in the supporting research package.

Once the universe is frozen, correlation reveals whether its apparent breadth represents genuinely different risks or repeated versions of the same exposure.

5. Correlation and Diversification Framework

5.1 Underlying-Market Correlation

The correlation analysis uses the frozen 75 market Core universe and Norgate arithmetically back-adjusted continuous-futures. For each market, daily close-to-close point changes are divided by a lagged EWMA standard deviation of point changes with a 60 trading day center of mass. This avoids raw price level correlation and percentage return distortions that can arise when additive back adjustment pushes old adjusted prices toward or below zero. Correlations are then estimated pairwise using overlapping observations rather than forcing every market to share the latest common start date.

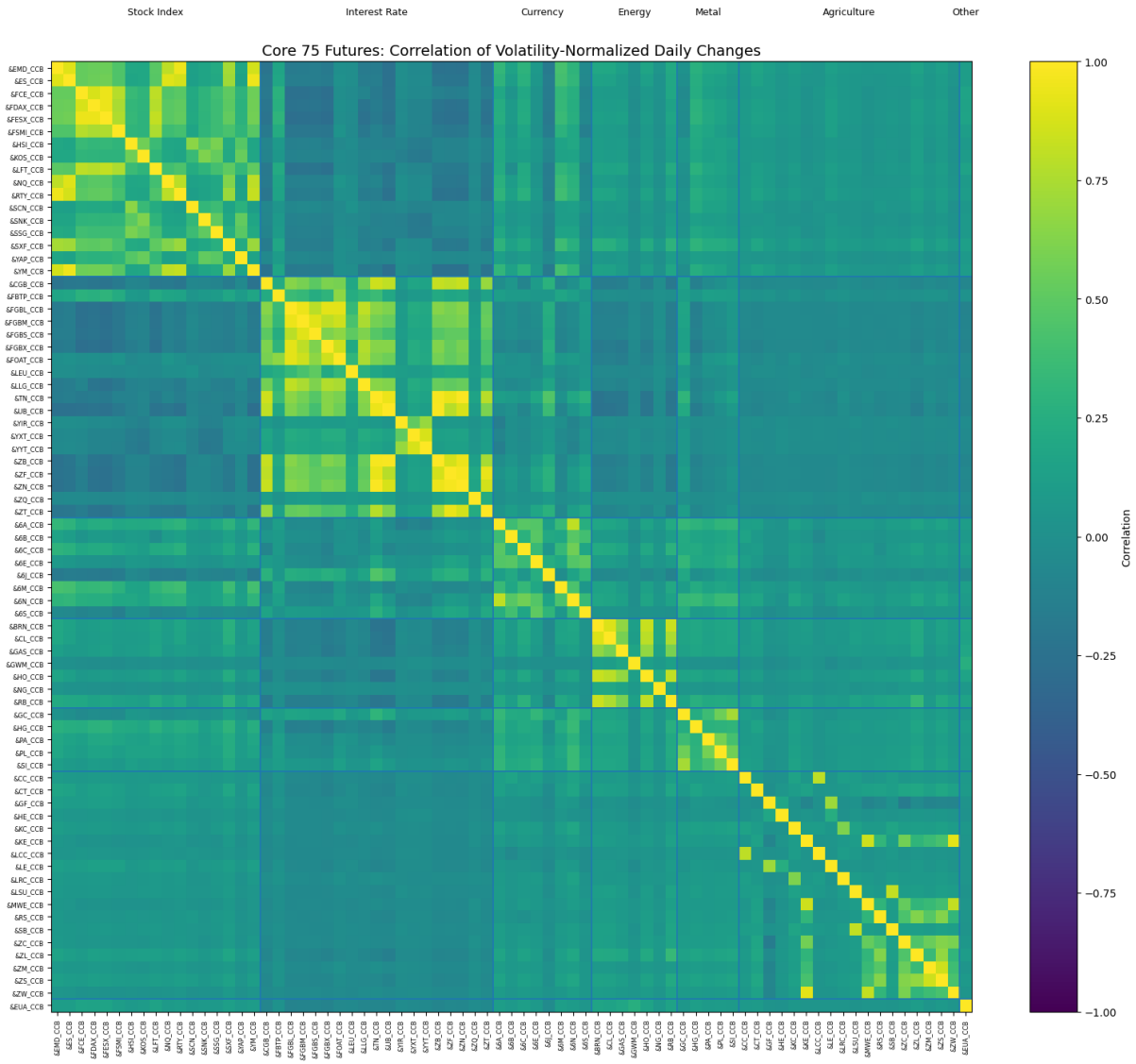


Figure 1. Correlation of volatility normalized daily changes across the frozen 75 market futures universe. Strong clusters within asset classes coexist with weaker dependence across classes.

5.2 Asset-Class Structure

Pairwise history is extensive despite differing contract inceptions: the minimum overlap across the 2,775 unique Core-market pairs is 2,459 normalized daily observations, the median is 6,857, and the maximum is 9,306. Full-sample mean pairwise correlation is 0.084 and the median is 0.052, with a range from −0.276 to 0.985. The highest correlations are economically intuitive clusters rather than accidental duplicate series: for example, adjacent U.S. Treasury maturities, related equity indexes, Brent versus WTI crude, and Kansas City versus Chicago wheat.

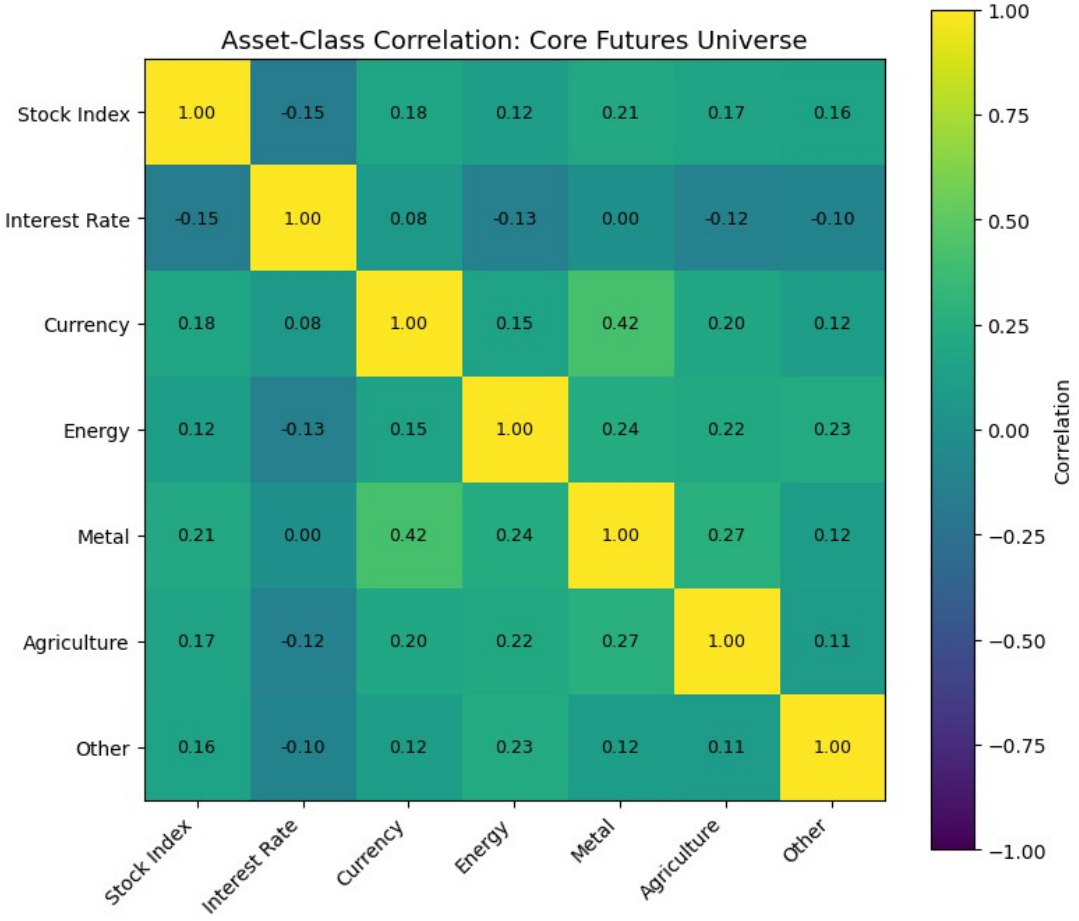


Figure 2. Correlation among equal weighted asset-class composites built from volatility normalized market changes.

5.3 Asset-Class and Time-Varying Correlation

Asset-class aggregation reinforces the diversification result. The highest off-diagonal class correlation is Currency-Metal at 0.419; Stock Index-Interest Rate is −0.153. Rolling 12-month average pairwise market correlation has a full-period mean of 0.078, a minimum of 0.024 on 7 April 2003, and a maximum of 0.154 on 13 March 2023. Thus correlation varies materially through time, but even the highest rolling one-year average remains modest.

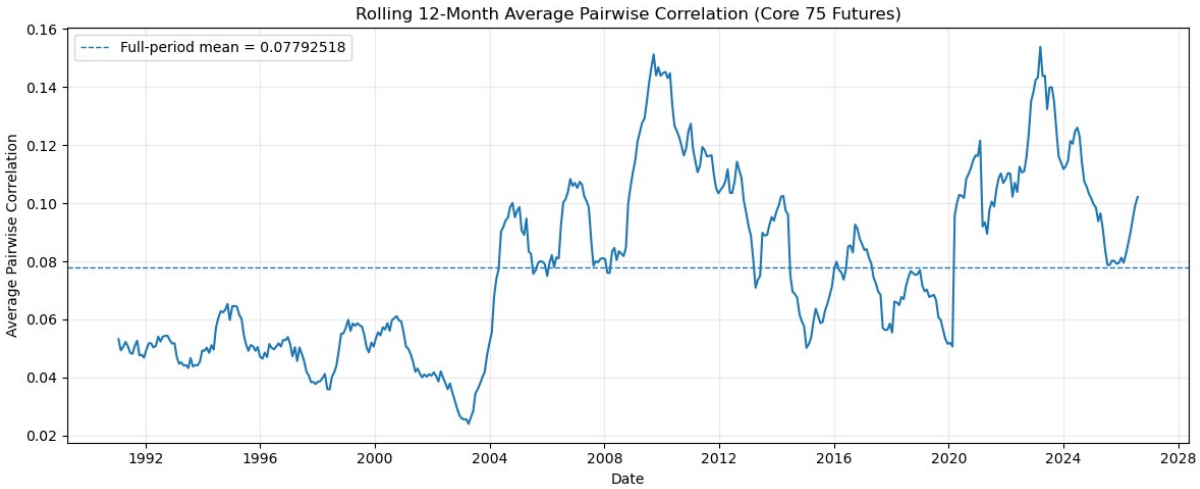


Figure 3. Rolling 12 month average pairwise correlation across the frozen futures universe. The full sample mean is 0.078; the minimum is 0.024 on 7 April 2003 and the maximum is 0.154 on 13 March 2023.

5.4 Strategy-Return Correlation

The dependence analysis separated diversification among underlying markets from diversification among strategy returns. The three directional sleeves remained closely related, with correlations from 0.660 to 0.821. Constrained term structure was materially different, with correlations of only 0.135 to 0.251 and worst decile overlap of 0.103 to 0.117 with the directional strategies. This distinction, rather than a lower standalone drawdown, was the principal evidence for retaining term structure in the combined portfolio.

The opportunity set is now defined. The next issue is procedural: how to prevent reasonable research choices from turning into a retrospective search for the best result.

6. Research Design and Governance

6.1 Chronology

The sequence was fixed before the portfolio tests. It moved from data inventory and market mapping to the data audit, universe freeze, underlying correlation, baseline time-series momentum, volatility and risk scaling, costs, complementary strategies, dependence analysis, fixed portfolio construction, implementation constraints, and finally prospective monitoring.

6.2 Controls Against Overfitting

No performance-based market exclusions. No parameter optimization in the baseline. Robustness tests must be specified as economically meaningful alternatives rather than searches for the best historical result.

6.3 Experiment Log

All material methodological changes and validation outcomes are recorded in Appendix A. Failed experiments are retained when they contain research information, while debugging artifacts remain outside the main empirical narrative.

The first strategy is intentionally simple. It creates a transparent reference against which every later extension can be judged.

7. Baseline Time-Series Momentum Model

7.1 Signal Definition

For each market and decision date, the baseline signal is the sign of the difference between the current continuous-futures price and its level 252 trading days earlier. Positive momentum implies long exposure and negative momentum implies short exposure.

7.2 Timing

The implementation uses only information available at each decision time. RealTest execution timing, signal formation, and next-period implementation were independently reconciled against Python before performance results were accepted.

7.3 Baseline Simplicity

No stops, profit targets, discretionary filters, or optimized overlays are included in the baseline. These may be studied later only as clearly separated extensions.

8. Volatility Estimation and Position Sizing

8.1 Baseline Estimator

The baseline ex-ante volatility estimator is an exponentially weighted standard deviation of daily point changes with a 60-trading-day center of mass, lagged so that no realized movement is used to size itself.

8.2 Risk Normalization

Forecast point volatility is converted to economically meaningful risk using verified Norgate contract metadata. Instrument exposures are scaled to a common ex-ante volatility convention before equal-weight portfolio aggregation.

8.3 Alternative Estimators

Alternative volatility estimators were treated as diagnostics rather than invitations to optimize. The retained lagged EWMA specification was carried through the final architecture because it was causal, interpretable, and compatible with verified futures contract economics; added complexity was not adopted without robust implementation value.

With the signal, timing, volatility estimate, and position convention fixed, the baseline can be evaluated without changing the question after seeing the answer.

9. Baseline Time-Series Momentum Results

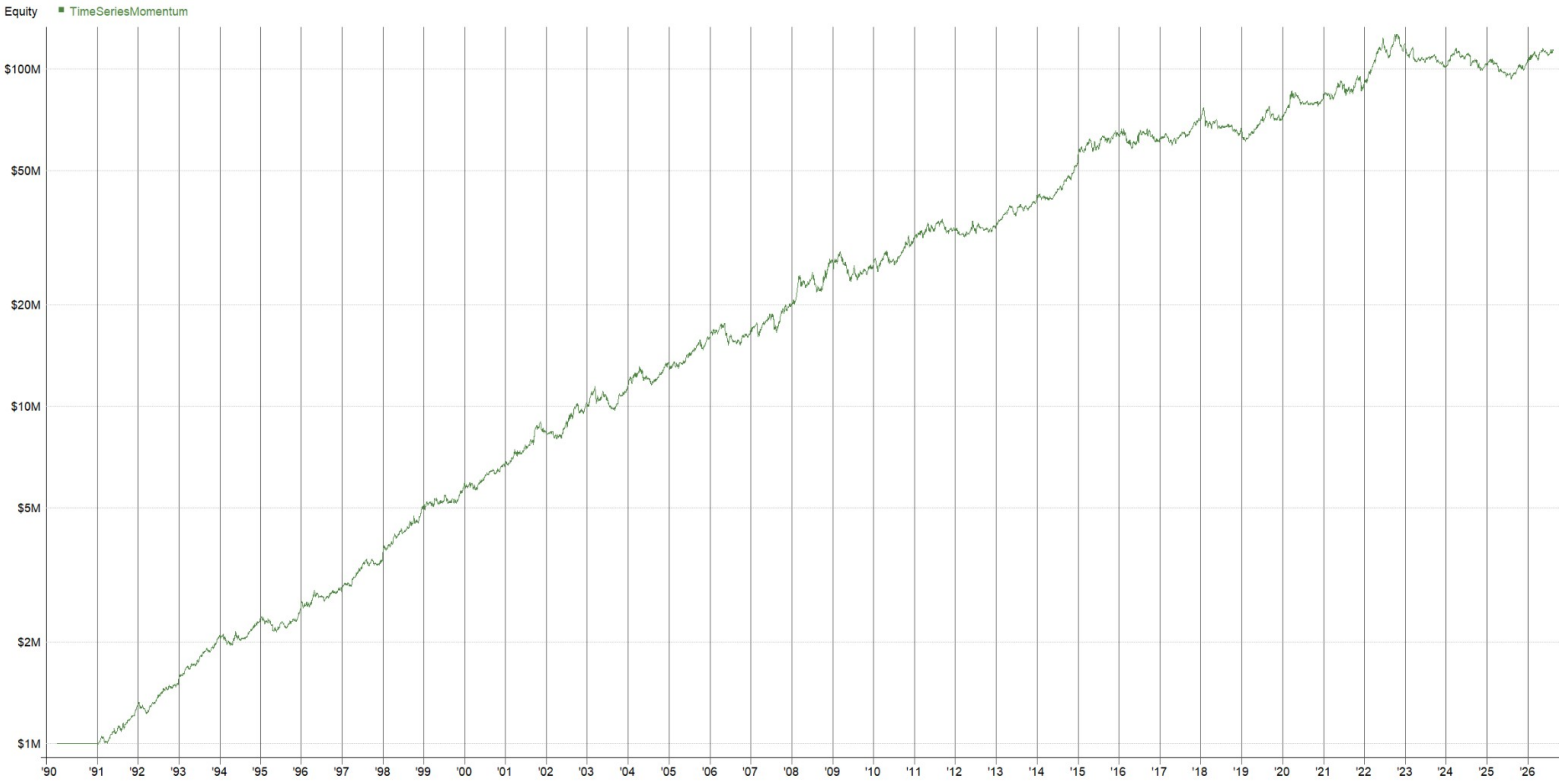
9.1 Headline Result

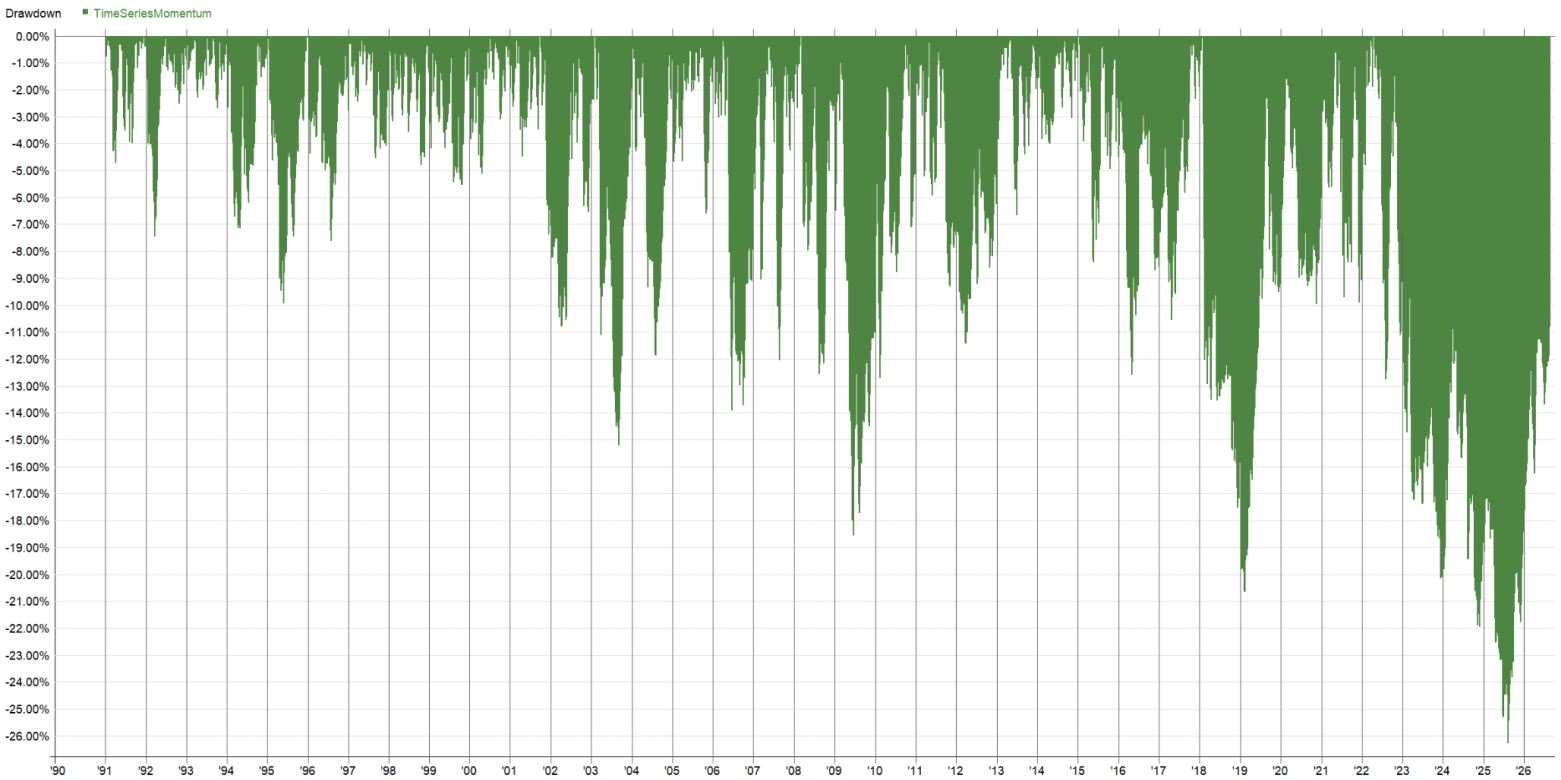
The gross baseline produced a 14.22% annualized return, a 1.25 Sharpe ratio, and a 26.26% worst peak to trough decline. These figures show that the simple 252 trading day time-series momentum rule produced an attractive historical gross profile across the frozen Core 75 universe. They are not an estimate of live net performance. Section 11 shows that the reference cost test materially changes the conclusion for the original daily resizing implementation. RealTest remains the primary execution result because it uses verified futures contract specifications and the required currency conversion data.

Gross baseline measure	Result
Annualized compound return	14.22%
Sharpe ratio (return per unit of volatility)	1.25
Worst peak-to-trough decline	−26.26%
Return-to-drawdown ratio (MAR)	0.54
Gross gains divided by gross losses	1.21
Daily observations	9,288
Validation status	PROVISIONAL - gross result; final post-holiday, currency, quantity, and exposure checks remain open

9.2 How to Read the Equity and Drawdown Charts

Figure 4A, 4B, 4C show the baseline equity path, drawdown, and calendar year returns respectively. The logarithmic equity scale makes percentage growth comparable across the sample, while the lower panels show the path of losses from prior peaks and the dispersion of annual outcomes. The account grows from USD 1 million to about USD 114 million over roughly 35 years, but the endpoint mainly illustrates long run compounding. The more decision relevant result is the 26.26% maximum drawdown. The figure is gross and remains subject to costs and implementation constraints.


Equity
TimeSeriesMomentum
$100M
$50M
$20M
$10M
$5M
$2M
$1M
'90 '91 '92 '93 '94 '95 '96 '97 '98 '99 '00 '01 '02 '03 '04 '05 '06 '07 '08 '09 '10 '11 '12 '13 '14 '15 '16 '17 '18 '19 '20 '21 '22 '23 '24 '25 '26


Drawdown
TimeSeriesMomentum
0.00%
-1.00%
-2.00%
-3.00%
-4.00%
-5.00%
-6.00%
-7.00%
-8.00%
-9.00%
-10.00%
-11.00%
-12.00%
-13.00%
-14.00%
-15.00%
-16.00%
-17.00%
-18.00%
-19.00%
-20.00%
-21.00%
-22.00%
-23.00%
-24.00%
-25.00%
-26.00%
'90 '91 '92 '93 '94 '95 '96 '97 '98 '99 '00 '01 '02 '03 '04 '05 '06 '07 '08 '09 '10 '11 '12 '13 '14 '15 '16 '17 '18 '19 '20 '21 '22 '23 '24 '25 '26

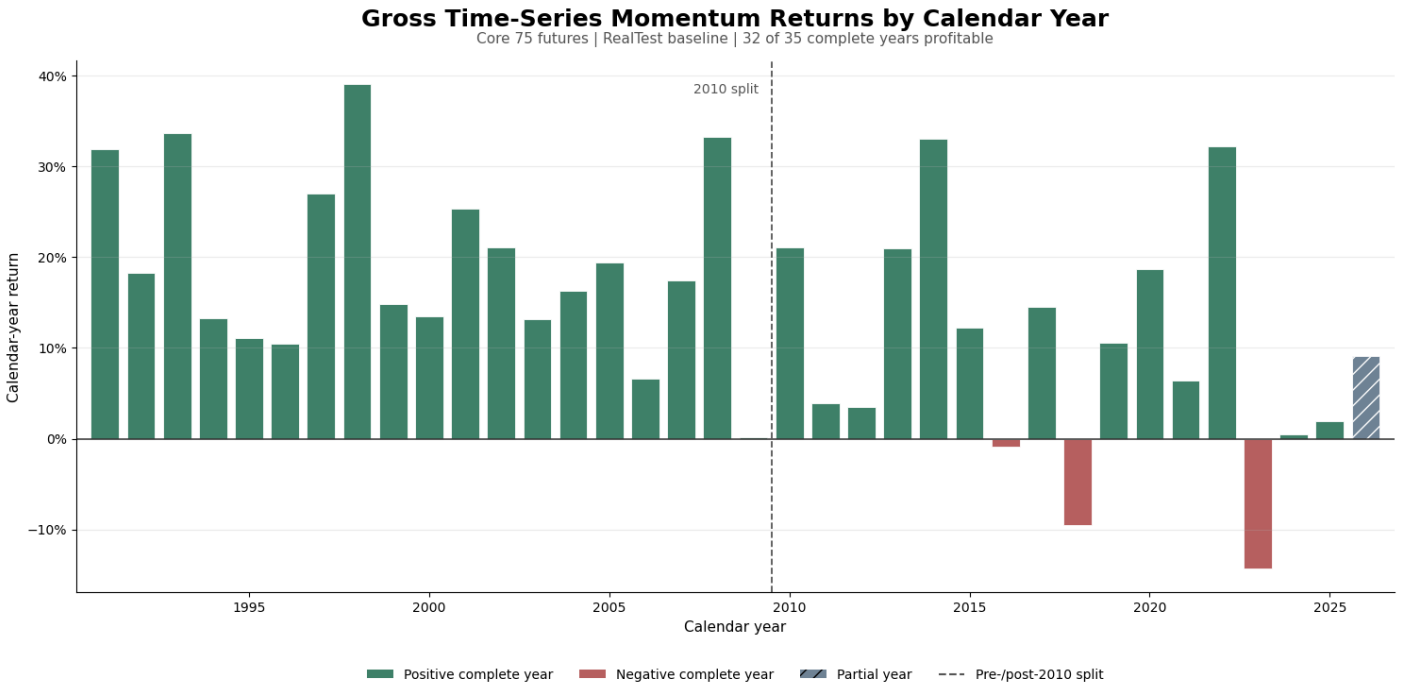


Figure 4A, 4B, 4C. Gross baseline equity, drawdown, and calendar year returns from 1991 to 2026. The logarithmic equity scale shows proportional growth; costs and final portfolio controls are not included.

9.3 Breadth and Remaining Limits

Calendar year and market-contribution evidence establish both temporal and cross sectional breadth. Thirty two of 35 complete calendar years and 57 of 75 Core markets were profitable, and every major asset-class contributed positively. The remaining breadth question is counterfactual: removing a leading market or a full asset-class changes the subsequent equity path and all dynamically sized positions, so attribution alone cannot answer it.

9.4 Independent Reconciliation

The accepted daily benchmark was subsequently reconciled exactly when it became the reference for the locked monthly strategy sequence. Earlier development differences around holidays, currency conversion, and individual position timing are retained in the audit archive. RealTest remains the authoritative execution engine, while Python provides independent signal, accounting, and portfolio checks.

The baseline is historically attractive before costs, but one lookback may conceal horizon dependence. The next tests ask whether continuation persists across nearby horizons and whether a broader signal is actually better.

10. Signal Horizon Research

10.1 Continuation and Reversal Across Horizons

Reader takeaway. Futures returns show a broad tendency to continue in the same direction over shorter horizons, but the pooled evidence is not strong enough to validate one exact lag after correcting for the many horizons tested. The clearest robust result is concentrated in commodity futures around the 10- to 11-month horizon. Extending the search to five years does not reveal a stable or statistically reliable switch from continuation to reversal.

How the Test Was Built

Each market's daily price changes were first placed on a comparable risk scale and then compounded into separate calendar-month returns. The months do not overlap: a daily price move appears in only one monthly observation. This prevents the same information from being counted repeatedly and making the statistical confidence look stronger than it really is. The saved monthly data were independently rebuilt from the daily panel and matched exactly.

What the Regressions Ask

Two complementary tests were used. The magnitude test asks whether a larger past return tends to be followed by a larger current return. The direction based sign test asks a simpler question: does the market's past direction, up or down, help predict its current return? This second specification is grounded in the established time-series momentum literature and was independently reproduced before being extended to additional horizons and asset classes.

H_0: The past direction has no predictive relationship with the current volatility-scaled return.

H_A: A relationship exists: a positive estimate indicates continuation, while a negative estimate indicates reversal. All significance tests are two-sided.

How the Literature Benchmark Is Used

The original published study covers January 1985 through December 2009 and uses 58 futures and forward markets. This report uses a different vendor construction, a 75-market futures universe, and data extending through 2026. Its reported t-statistics therefore serve as historical context, not as numerical targets that this study should be expected to match. Differences in sample period, instruments, return construction, volatility normalization, and statistical uncertainty prevent a clean like-for-like comparison.

The more informative stability test is internal: rerun the same direction-based regressions on the pre-2010 and post-2009 portions of this report's own dataset. Holding the data construction and estimation method constant will show whether continuation weakened, shifted across asset classes, or remained stable in the more recent period. This subperiod comparison is retained as a robustness test; it does not delay the higher-priority contract-level roll-cost audit.

Why the Holm Correction Matters

When many horizons are tested, at least one can look significant purely by chance. The Holm-Bonferroni procedure evaluates the strongest result first and then becomes gradually less strict only if earlier results pass. A result is described as robust in this report only when its Holm-adjusted p-value is 5% or lower. This correction does not change the estimated effect; it changes how much evidence is required before calling that effect a reliable discovery.

What Evidence Shows

Analysis	Plain-English conclusion
All markets, 1-12 months	Continuation is generally positive, but no single pooled lag remains significant after correcting for the 12 horizons tested.
Four asset classes	Commodities provide the clearest evidence. The 10- and 11-month commodity signals survive the stricter correction across all 48 asset-class-by-lag tests. Equity indexes, government bonds, and currencies do not.
All markets, 1-60 months	A few isolated continuation and reversal readings appear, but none survives correction across 60 horizons. The data do not show a stable transition to long-horizon reversal.

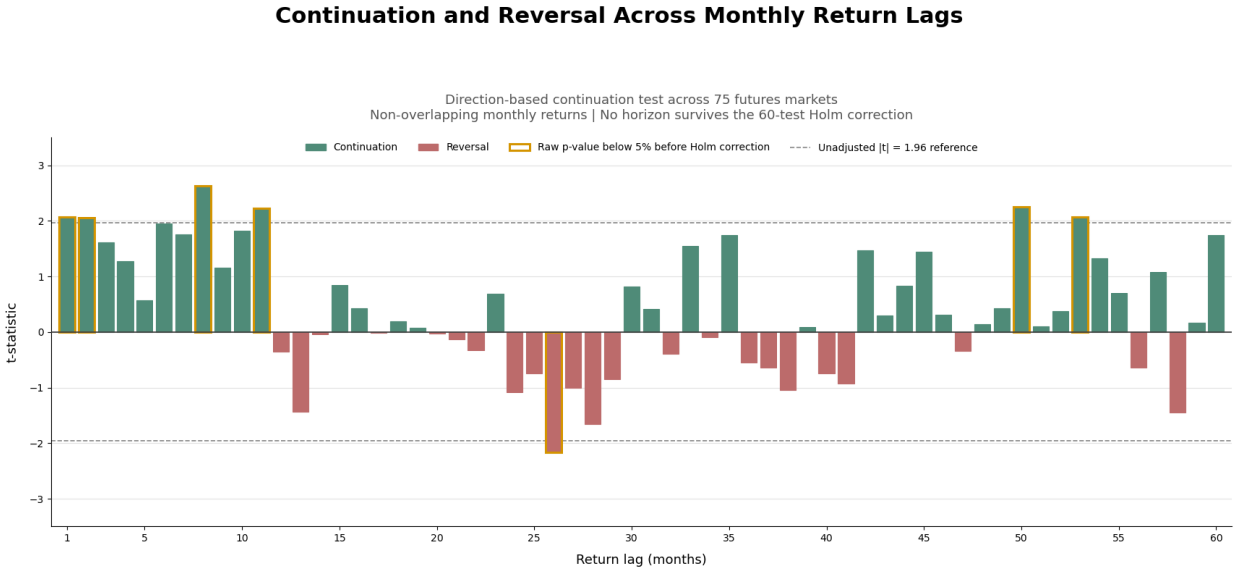


Figure 5. Continuation and reversal evidence across monthly return lags. Gold outlines identify raw p values below 5% before the Holm correction; none remains significant after correcting for all 60 horizons.

Interpretation. The pooled results support continuation as a broad tendency, not as proof that one exact lookback is universally best. The asset-class analysis also shows that the effect is not uniform: the strongest defensible evidence is in commodities near 10 to 11 months. Exact coefficients, t-statistics, p-values, and experiment identifiers remain in the supporting research files rather than interrupting the main narrative.

10.2 Multi-Horizon Strategy Robustness

The earlier multi-horizon test combined four specified directional lookbacks of 21, 63, 126, and 252 trading days while keeping the frozen Core 75 universe, EWMA volatility estimator, FX conversion, capital, and gross sizing convention unchanged. Against the 252 day baseline, the ensemble reduced maximum drawdown from 26.26% to 22.75%, trade count from 167,069 to 128,012, and reported average exposure from 1,806.81% to 1,184.91%. However, annualized return fell from 14.22% to 10.44%, Sharpe ratio fell from 1.25 to 1.20, and MAR fell from 0.54 to 0.46. The ensemble is therefore not a superior replacement for the baseline. Because back-adjusted synthetic prices can distort nominal exposure, RealTest exposure fields are interpreted only as within platform diagnostics, not literal economic notional.

Measure	252-day baseline	Four-horizon ensemble
Annualized return	14.22%	10.44%
Sharpe ratio	1.25	1.20
Maximum drawdown	−26.26%	−22.75%
MAR ratio	0.54	0.46
Profit factor	1.21	1.18
Trades	167,069	128,012
Average exposure diagnostic	1,806.81%	1,184.91%
Maximum exposure diagnostic	11,032.93%	6,013.68%

10.3 Decision Rule

Professional standard. Advance a horizon or horizon combination only if it is supported by neighboring specifications, remains reasonably broad across markets and time, and survives realistic implementation costs. Do not select an isolated lag merely because it has the highest in-sample t-statistic or Sharpe ratio.

Signal robustness is not enough. A futures strategy must also survive contract trading, delivery replacement, and the turnover created by changing target positions.

11. Transaction Costs and Turnover

11.1 Cost Model

The reference cost specification charges USD 2.50 per contract per transaction for commissions and fees, one contract-specific tick of slippage per transaction, and the same commission plus one-tick assumption on both legs of every mandatory roll. Tick values use each market's verified TickSize, PointValue, and FX conversion. This is a transparent common reference case, not a claim that every market had the same historical fee or liquidity.

11.2 Roll and Turnover Accounting

The otherwise identical zero cost control reproduced the accepted gross baseline exactly: 14.22% annualized return, 26.26% maximum drawdown, a 1.25 Sharpe ratio, 167,069 trades, and zero costs. The daily resizing reference run accumulated USD 10.856 million of modeled costs. The 10% no-trade band reduced reported trades to 26,044 and modeled costs to USD 7.321 million. The specified 20% band produced 26,795 trades and USD 7.948 million of modeled costs. Those totals were 2.9% and 8.6% above the 10% band despite the wider threshold. This does not show that the 20% rule mechanically creates more turnover: because sizing references compounded strategy equity, its stronger equity path maintained larger target quantities and contract counts. Cumulative trade counts and dollar costs are therefore path-dependent across divergent equity paths. Roll costs remained dominant under the 20% band at USD 5.311 million, or 66.8% of total modeled costs; ordinary slippage was USD 2.148 million and commissions were USD 0.489 million.

11.3 Net Results

Validated results. Under daily resizing and 1x reference costs, annualized return fell to −1.17%, Sharpe to −0.09, and maximum drawdown deepened to −86.23%. The 10% band improved annualized return to 0.41%, Sharpe to 0.09, and maximum drawdown to −74.57%. The 20% band improved those measures further to 1.17%, 0.17, and −65.20%, with a 1.02 profit factor and $510,003 net profit. The direction of improvement is useful evidence that turnover control matters, but the resulting risk-adjusted performance remains economically weak.

Measure	Gross / 0x	Daily resizing / 1x	10% band / 1x	20% band / 1x
Annualized return	14.22%	−1.17%	0.41%	1.17%
Sharpe ratio	1.25	−0.09	0.09	0.17
Maximum drawdown	−26.26%	−86.23%	−74.57%	−65.20%
MAR ratio	0.54	−0.01	0.01	0.02
Profit factor	1.21	0.99	1.01	1.02
Trades	167,069	41,917	26,044	26,795
Net profit	$112,777,281	-$343,318	$153,450	$510,003
Total modeled costs	$0	$10,855,890	$7,321,395	$7,947,831

Interpretation and next action. The pre-specified 10% and 20% band sensitivity is complete. It shows that suppressing small target changes improves net performance, but not enough to make the current highly exposed implementation viable. The 20% result will not be selected merely because it performed better in this sample, and wider bands will not be searched for an in-sample optimum. The next stage is a representative individual-contract roll audit across the four asset classes. It will compare continuous-series signals and P&L with tradable contract prices, actual roll transitions, liquidity, and two-leg roll costs before portfolio-level volatility, leverage, concentration, and margin constraints are finalized.

After measuring the cost burden, the analysis returns to the source of the profits: whether they recur across markets, asset classes, and historical periods.

12. Breadth Across Markets and Time

12.1 Market Breadth

The market-contribution analysis covers 167,069 trades across all 75 Core markets. Exported trade profits sum to USD 112,777,281.31, reconciling to the rounded RealTest net profit within USD 0.31. Fifty seven markets, or 76.0%, contributed positive gross profit and loss, while 18 contributed negatively. The largest positive market generated only 5.5% of total positive market profits, and the five largest together generated 18.9%. The gross result is therefore broad rather than dependent on one or two contracts.

Every broad asset-class contributed positively. Commodity futures generated $65.7 million (58.3% of net profit), with 29 of 31 markets profitable. Government-bond futures generated $27.4 million (24.3%; 15 of 19 profitable), equity-index futures $14.0 million (12.4%; 9 of 17 profitable), and currencies $5.7 million (5.1%; 4 of 8 profitable). Within commodities, agriculture supplied $38.0 million and 17 of 18 markets were profitable; all seven energy markets were profitable.

Attribution caveat. These figures allocate realized dollar P&L within the accepted compounded portfolio; they are not standalone market Sharpe ratios or a prediction of what would happen if a market were removed. Because capital and target quantities are updated dynamically, excluding a market changes the subsequent equity path and sizing of every remaining position. True leave-one-market-out and leave-one-class-out effects therefore require fresh RealTest runs rather than simple subtraction.

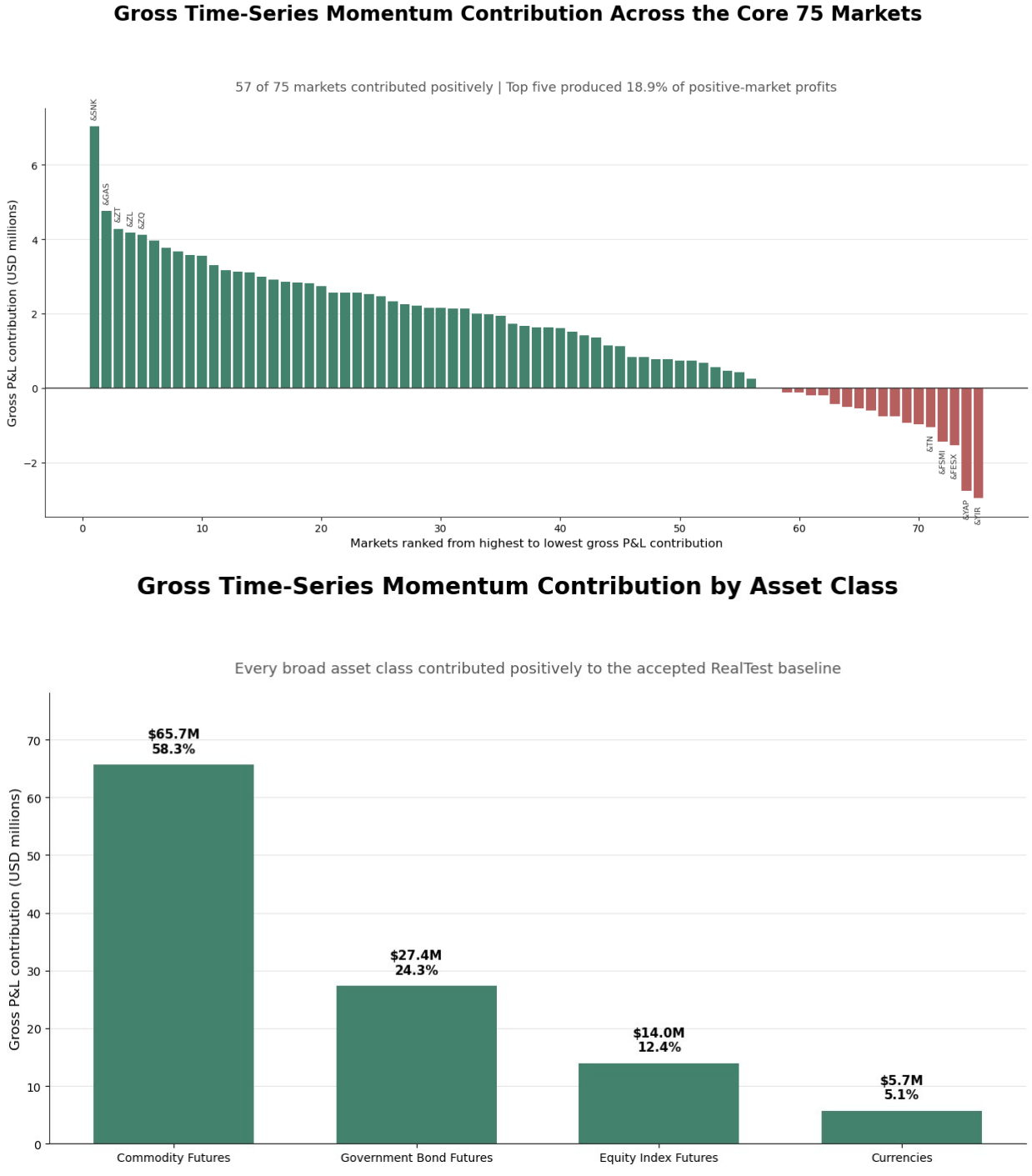


Figure 6. Gross contribution across individual markets and broad asset classes. The result is economically broad, although contributions are not uniform.

12.2 Time Robustness

The calendar year analysis shows that the gross RealTest baseline was profitable in 32 of 35 complete years, or 91.4%. Strong years occurred in different environments, including 2008, 2014, and 2022, so the result is not attributable to one isolated crisis. Temporal breadth is strong, although the magnitude of the edge has weakened materially.

Before 2010, the active sample produced an 18.86% annualized return, a 1.63 Sharpe ratio, and a −18.55% maximum drawdown. From 2010 through August 2026, these measures weakened to 9.13%, 0.83, and −26.26%. Annualized volatility remained similar, 10.60% versus 10.95%, indicating that the deterioration came mainly from lower return rather than a quieter risk environment. The evidence supports a persistent gross effect with economically meaningful decay from 2010 onward; recent performance and implementation costs must therefore receive greater weight in the final strategy decision.

12.3 Counterfactual Scope Boundary

A full dynamic leave-one-class-out rerun was not completed. Static contribution is reported only as attribution, not as a counterfactual removal test, because removing a class changes the compounded equity path and every subsequent target quantity. Concentration is instead addressed directly through the final market and asset-class caps. The missing dynamic rerun remains a documented limitation rather than an inferred result.

Breadth does not guarantee implementability. Contract rolls, curve structure, currency conversion, and whole-contract sizing are therefore tested directly.

13. Implementation and Term Structure Evidence

13.1 Continuous-Series Convention

The continuous series audit established that Norgate supplies the three layers required for an implementation audit: back-adjusted continuous series, unadjusted continuous series, and historical individual deliveries. The subsequent contract linkage reconciliation matched 806 usable contract changes across E mini S&P 500, 10 Year U.S. Treasury Note, Euro FX, and WTI Crude Oil from 1990 through June 2026. On every event, the unadjusted continuous close matched the incoming individual contract close exactly, and the recorded expiry change event agreed with the new continuous contract expiry field. This validates the contract linkage used for the implementation audit.

The audit also shows that a single universal roll rule would be inappropriate. The median switch occurred on the outgoing contract's expiry date for WTI Crude Oil and Euro FX, one day before expiry for E-mini S&P 500, and 21 days before expiry for the 10-Year U.S. Treasury Note. Incoming volume exceeded outgoing volume immediately before the switch in nearly every equity-index, currency, and crude-oil event, but in only 34.5% of Treasury-note events. Roll timing must therefore be taken from the vendor's point-in-time contract mapping rather than reconstructed from one volume-crossover rule.

Market	Events	Median switch before outgoing expiry	Incoming volume higher on prior day
WTI Crude Oil	438	0 days	97.7%
Euro FX	109	0 days	100.0%
E-mini S&P 500	114	1 day	100.0%
10-Year U.S. Treasury Note	145	21 days	34.5%

The outgoing-to-incoming price difference is a term-structure observation, not a transaction-cost. The most extreme event was the April 2020 WTI dislocation: the incoming contract stood 58.06 price points above the outgoing contract, equal to $58,060 per contract. This is precisely why back-adjusted series are retained for signal formation, while individual deliveries are required for carry, roll, and two-leg execution modeling. The execution-cost assumption remains limited to commissions and slippage on the exit and entry legs.

13.2 Term-Structure Signal Pilot

The signal pilot converted the validated roll inventory into 798 non-overlapping, completed roll cycles. The signal was observed immediately before each vendor roll: backwardation, where the nearby contract exceeded the farther contract, implied a long position; contango implied a short position. The forward outcome began at the incoming-contract close on the roll date and ended at the same contract's close immediately before the next roll. This construction excludes the next roll gap from measured profit and avoids mistaking a term-structure discontinuity for an investment return.

The curve implied direction was correct in 435 of 798 cycles, a 54.5% hit rate versus the 50% no predictability benchmark (exact binomial p = 0.0119). In a pooled regression of market standardized forward moves on the sign of the curve, with market fixed effects and standard errors clustered by calendar month, the coefficient was 0.097 risk units (t = 2.46; p = 0.0143). The evidence therefore supports a modest but statistically detectable association between the futures curve and the direction of the following roll cycle move.

Market or test	Cycles	Hit rate	Statistical reading
Euro FX	108	50.0%	t = 0.64; p = 0.522; no standalone evidence
WTI Crude Oil	435	51.7%	t = 1.77; p = 0.078; suggestive but inconclusive
E-mini S&P 500	111	59.5%	t = 1.22; p = 0.226; high hit rate but noisy magnitude
10-Year U.S. Treasury Note	144	62.5%	t = 2.71; p = 0.0077; survives four-test Holm correction
Pooled four-market test	798	54.5%	Fixed-effect coefficient = 0.097; clustered t = 2.46; p = 0.0143

The pooled result is promising but deliberately provisional. Treasury-note futures provide the clearest standalone evidence; only that market survives Holm correction across the four-market-level tests (adjusted p approximately 0.031). Euro FX is indistinguishable from chance, and the equity-index and crude-oil estimates remain imprecise. Moreover, the pilot uses event-level market normalization rather than lagged daily volatility, and it does not yet form a continuously managed portfolio. The signal pilot therefore justifies advancing the curve signal to a daily, volatility-scaled, cost-aware backtest; it does not establish an investable carry strategy by itself.

13.3 Four-Market Term-Structure Prototype

The daily implementation prototype translated the pre-roll curve direction into a daily, executable four-market prototype over the common 15 March 1999 to 27 February 2026 sample. Each market received a fixed 10% annualized standalone-risk budget, for a 40% aggregate budget, using the prior day's EWMA point volatility and whole-contract rounding. Positions changed only at validated contract rolls. Daily futures profit was contracts multiplied by back-adjusted point change and PointValue; back-adjusted price levels were never treated as percentage returns. The cost model charged USD 2.50 commission and one-tick of slippage for every transaction leg, including both legs of mandatory rolls.

The strategy compounded USD 1.0 million to USD 17.31 million before costs and USD 15.00 million after costs. Net CAGR was 10.57%, annualized volatility was 21.36%, Sharpe ratio was 0.58, and maximum drawdown was −43.86%. Total modeled trading costs were USD 1.09 million across 77,612 contract transaction legs, reducing CAGR by 0.59 percentage points. Unlike the daily-resized 252-day time-series momentum cost test, roll-event-only resizing preserved most of the gross return after costs.

Performance measure	Gross	Net after costs
Final equity from USD 1.0M	USD 17.31M	USD 15.00M
Compound annual return	11.16%	10.57%
Annualized volatility	21.29%	21.36%
Sharpe ratio	0.60	0.58
Maximum drawdown	−43.29%	−43.86%
Modeled trading costs	USD 0	USD 1.09M

Contribution was not uniform. After costs, WTI crude oil contributed USD 8.46 million, E mini S&P 500 USD 4.68 million, and Euro FX USD 3.93 million, while the 10 Year Treasury note implementation lost USD 3.07 million. That divergence matters because the event based signal pilot identified Treasury notes as the strongest result when each standardized roll cycle was weighted comparably. The daily prototype instead compounded capital and applied time-varying, volatility scaled whole-contract positions, so outcomes at different dates carried different dollar influence. A cycle level reconciliation was therefore required before the carry specification could be accepted.

Decision. Retain the daily four-market strategy as a successful implementation prototype, not as a final portfolio. It demonstrates that the curve signal can be expressed with low frequency rebalancing and manageable modeled cost drag, but its 0.58 net Sharpe ratio, −43.86% drawdown, four-market scope, and contradictory Treasury contribution prevent standalone acceptance. The next step expanded the validated roll-event measurement to the full Core 75 universe and applied the same frozen signal, risk, currency conversion, and cost conventions before any portfolio combination.

13.4 Full-Universe Roll-Data Validation

The full-universe roll data audit expanded the individually reconciled roll data from the four-market pilot to all 75 Core futures. The inventory contains 13,661 vendor roll events from January 1990 through July 2026, with no duplicate market date observations. As a direct control, the E mini S&P 500, 10 Year U.S. Treasury Note, Euro FX, and WTI crude oil subset reproduces all 806 previously validated contract linkage events and their numeric fields exactly.

A pre-specified quality rule removes five exceptional observations where the continuous series does not match the identified incoming delivery, days to the new expiry are non-positive, or the prior-day roll spread is unavailable. This leaves 13,656 valid events. Of these, 201 have an exactly flat curve and are retained in the audit but produce no directional trade; 13,455 events contain an informative non-zero term-structure signal. Market histories are naturally unbalanced because newer contracts begin later, so portfolio tests use each market only after its own data become available.

The audit also identifies a necessary portfolio control: contract point values and spread values are denominated in each contract's local currency, not uniformly in U.S. dollars. The full-universe strategy therefore converts risk, daily P&L, and execution costs with the contemporaneous FX ratio before aggregation. Large curve dislocations are treated as genuine observations subject to diagnostic review rather than deleted mechanically, while daily P&L continues to use back-adjusted price changes so roll gaps are never counted as investment returns.

13.5 Full-Universe Term-Structure Strategy

The full-universe term-structure strategy converted the validated roll inventory into an event-driven portfolio across the Core 75 universe. Backwardation produced a long signal and contango a short signal. Signals and positions changed only at vendor contract roll events; daily profit used back-adjusted point changes, contract point values, and the contemporaneous currency conversion ratio. The price and currency panels matched exactly across all 602,616 market date rows. Five invalid roll observations identified by the advance quality rule retained the previously known direction, while 201 valid flat curves set the position to zero.

The portfolio compounded USD 1.0 million to USD 9.30 million before costs and USD 4.94 million after costs from 11 January 1990 through 14 August 2026. Gross CAGR was 6.28% with 8.79% annualized volatility, a 0.72 Sharpe ratio, and a −47.29% maximum drawdown. After USD 2.93 million of modeled commissions and one-tick slippage across 228,814 contract transaction legs, net CAGR was 4.46%, the net Sharpe ratio was 0.52, and maximum drawdown was −47.73%. Market and asset-class contributions reconciled exactly to portfolio profit.

Measure	Gross	Net after costs
CAGR	6.28%	4.46%
Annualized volatility	8.79%	8.83%
Sharpe ratio	0.72	0.52
Maximum drawdown	−47.29%	−47.73%
Trading costs	USD 0	USD 2.93 million

The important result is not that term structure is a standalone replacement for time-series momentum. It is that the curve signal survives expansion from four markets to the entire frozen universe, retains positive net performance after currency conversion and explicit roll execution, and rebalances far less often than the daily-resized momentum cost test. Diversification remains imperfect: net profits were positive in commodities, equity indices, and currencies, while government bonds lost approximately USD 1.09 million after absorbing USD 1.99 million of trading costs. The next decision is therefore to diagnose bond turnover and then test portfolio combination with the independently validated time-series momentum baseline, rather than tune the carry signal to the full sample.

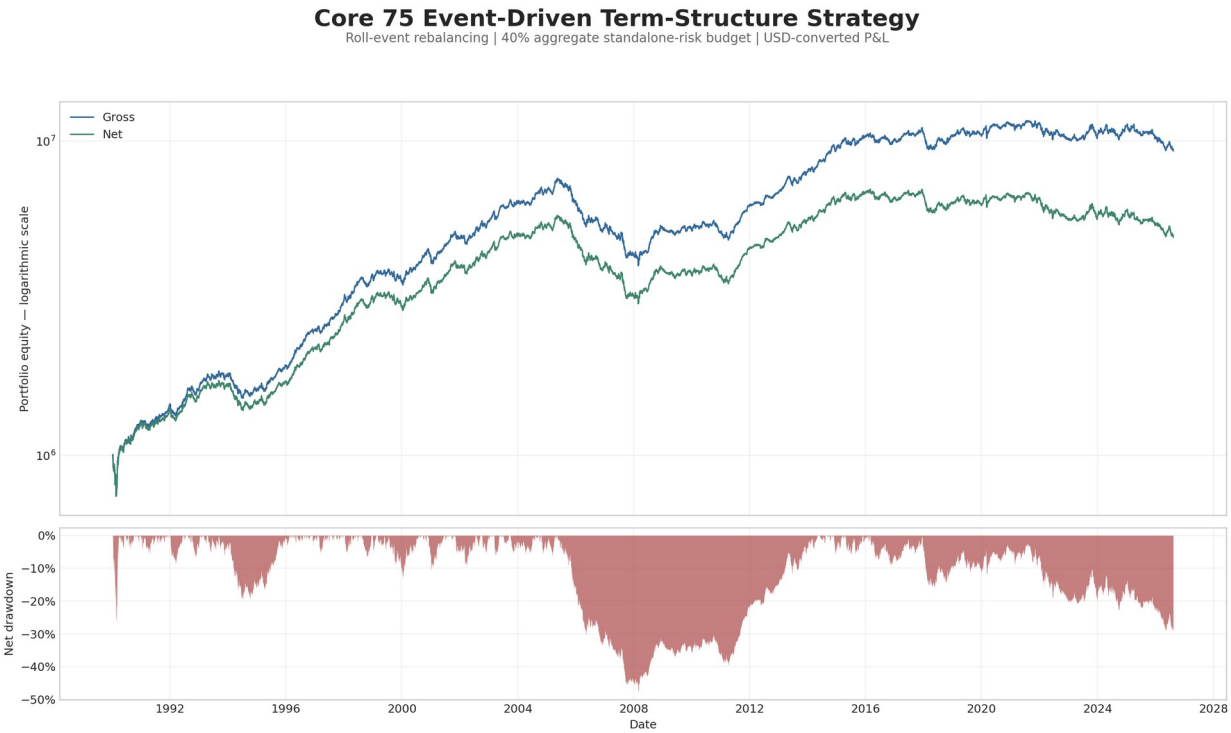


Figure 7. Gross and net equity, net drawdown, and asset-class contribution for the full-universe term-structure strategy. Positions change only at validated roll events.

13.6 Government-Bond Cost and Turnover Diagnosis

The government bond cost audit decomposed every full-universe term structure transaction into commission, one-tick slippage, contract quantity, and event type, then matched each event date to the daily volatility inputs in the price panel. All USD 2.934 million of modeled portfolio costs reconciled exactly between the trade-event and market-contribution files. Event level cost formulas also matched exactly; the largest displayed tick value rounding difference was USD 0.05. This is an accounting decomposition of the accepted term structure specification, not a new backtest.

The cost burden is highly concentrated. Government-bond futures generated 62.2% of all contract legs and 67.8% of total portfolio costs, despite contributing only USD 0.43 million of gross profit. After USD 1.99 million of modeled bond costs, the asset-class lost USD 1.09 million net. Slippage represented 82.1% of bond costs, and mandatory replacement of expiring deliveries represented 83.6% of bond contract legs. The five most expensive bond markets produced 88.9% of bond costs; ZQ alone accounted for USD 1.16 million, or 39.6% of all portfolio costs.

The cross-market pattern identifies the mechanism. Bond cost has a 0.93 Spearman correlation with contract legs and a 0.76 correlation with median absolute target contracts, but essentially no relation to tick value (−0.04). Its correlation with annualized dollar volatility per contract is −0.83: low-dollar-risk contracts require large whole-contract positions under equal-risk sizing, so every scheduled delivery replacement becomes expensive. ZQ is the extreme case, with a median target of 49 contracts, a 95th percentile of 341 contracts, and approximately 12 roll events per year. This is principally a position-scale and replacement-turnover problem, not a general claim that bond tick sizes are unusually costly.

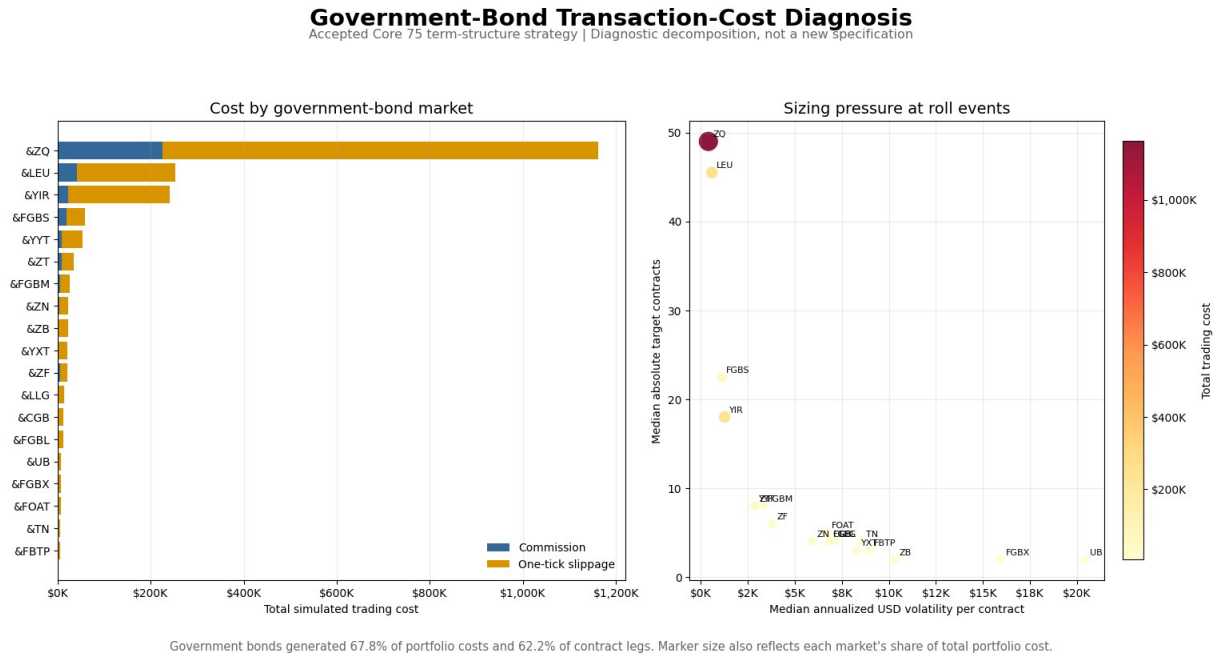


Figure 8. Government bond cost and turnover diagnosis. Low dollar volatility per contract created large target positions and expensive mandatory delivery replacement.

Decision. Retain the full-universe result as evidence that the term-structure signal survives implementation, but do not treat the bond sleeve as institutionally efficient in its original form. The cost diagnosis does not justify deleting ZQ, excluding bonds, or selecting parameters from the same sample. It motivates one implementation test, registered in advance, using a transparent control on contract size applied uniformly before the strategies are combined. Performance is reported both with and without that control, including sacrificed gross return as well as cost reduction.

13.7 Execution Delay

Execution remained causal throughout the completed tests. Monthly signals were formed at calendar month end and implemented on each selected market's first available observation in the following month. Term structure positions changed only at validated vendor roll events. A separate sweep across additional arbitrary delays was not used to select the reported specification.

13.8 Rebalance and Volatility Alternatives

The directional sleeves use monthly implementation, while term structure changes only at validated roll events. The retained lagged EWMA estimator and the uniform dollar volatility floor were carried into the final portfolio. Alternative estimators were diagnostics, not a search for the historical optimum.

The standalone engines are now reconciled under common data and cost conventions. Only at this point does the report ask whether their combination improves the portfolio.

14. Alternative Strategies and Portfolio Construction

14.1 Uniform Sizing Constraint

The uniform sizing constraint was not a search for a more profitable signal. It was the response specified in advance to the bond cost finding: contracts with low dollar volatility required very large positions and made mandatory delivery replacement disproportionately expensive. The same rule was applied to every eligible market. The lagged annualized dollar volatility denominator could not fall below that day's eligible market tenth percentile. This limited contract size without deleting a market, changing a signal, or favoring an asset-class.

The constraint reduced modeled trading costs by 45.0% and contract legs by 42.4%. Net CAGR rose from 4.46% to 5.63%, Sharpe ratio from 0.523 to 0.685, and maximum drawdown improved from −47.73% to −40.43%. Government bonds' share of total costs fell from 67.8% to 35.4%, while the largest single market cost share fell from 39.6% to 7.7%. The rule bound on 1,404 roll events, 10.3% of the validated inventory, so the improvement came from a narrow implementation control rather than a wholesale redesign.

> Key finding. The term structure sleeve did not require the exclusion of government bonds. A uniform control aimed at the economic source of the cost problem preserved the broad universe while materially improving the net result.

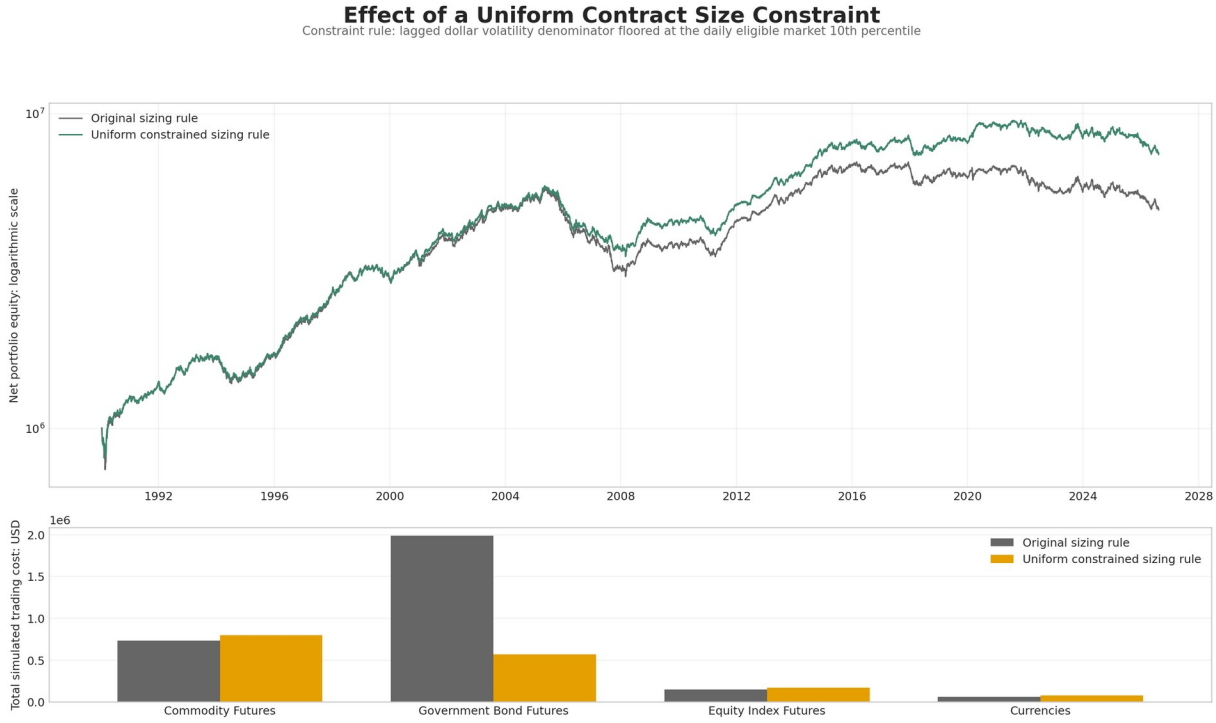


Figure 9. Effect of a uniform contract size constraint. The rule reduced concentration and government bond trading cost without deleting markets or changing signals.

14.2 Trend Horizon Comparison

The trend horizon comparison compared the retained 12-month time-series momentum rule with an equal-risk ensemble of 3-, 6-, and 12-month signs. Holding the universe, monthly implementation, lagged-volatility sizing floor, whole-contract rounding, and costs fixed made this a controlled test of signal horizon rather than a new portfolio design.

The broader ensemble was smoother but weaker. Net volatility declined from 10.44% to 9.20% and total costs fell from USD 5.15 million to USD 2.89 million, yet net CAGR declined from 10.32% to 8.48%, Sharpe ratio from 0.964 to 0.903, and maximum drawdown worsened from −20.01% to −21.82%. The retained 12-month rule also dominated in the earliest and most recent subperiods. The evidence therefore favors the simpler 12-month specification; adding horizons did not create an independent return source.

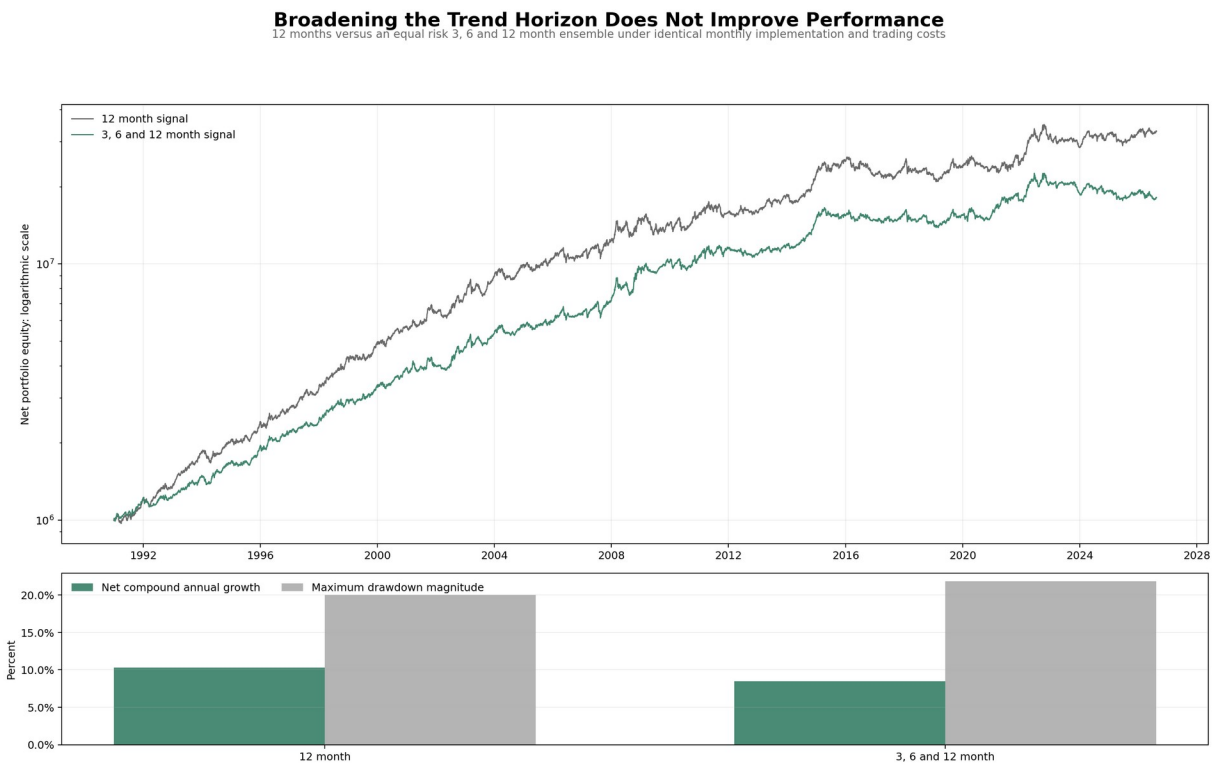


Figure 10. Net performance of the retained 12 month time-series momentum rule and the equal risk 3, 6, and 12 month ensemble under identical implementation and costs.

14.3 Protocol Lock Before Portfolio Tests

Before the alternative strategy and portfolio tests were run, the protocol lock fixed the Core 75 universe, historical cutoff of 14 August 2026, monthly decision timing, lagged 121 day EWMA point volatility, daily eligible market tenth percentile dollar volatility floor, whole-contract rounding, mandatory roll treatment, and USD 2.50 commission plus one-tick of slippage per contract leg. It also fixed the breakout channel, cross sectional ranking rule, crisis windows, portfolio weights, risk target, caps, leverage constraints, and prospective holdout protocol.

This governance step matters because a portfolio study is unusually vulnerable to selection bias. Once several plausible strategies exist, disappointing results can always be answered by changing a horizon, removing a sleeve, or altering a weight. The portfolio rules were therefore fixed before the joint results were examined. After those results became visible, no sleeve was removed, no weight was changed, and no constraint was altered. From that point onward, success meant faithful execution and reconciliation rather than an attractive backtest.

14.4 price-channel breakout

The pre-specified 252-day entry and 126-day exit price-channel breakout compounded to USD 13.33 million net, with 7.54% CAGR, 10.61% volatility, a 0.717 Sharpe ratio, and a −27.76% maximum drawdown. It traded 134,112 contract legs and incurred USD 1.63 million of modeled costs. The retained 12-month time-series momentum sleeve remained stronger at 10.32% CAGR, 0.964 Sharpe, and −20.01% maximum drawdown. Breakout was retained as a related trend expression for the locked portfolio test, not promoted as a replacement.

14.5 Cross-Sectional Momentum

The cross-sectional momentum test ranked eligible markets each month by the price change over the prior 12 months, excluding the most recent month, divided by lagged volatility. It held the highest quartile long, the lowest quartile short, and the middle half flat. Signals were formed at one month end and mapped causally to each selected market's first observation in the following month. Across 427 execution months, every selected execution was found and the long and short sides remained exactly balanced.

The strategy produced 10.03% net CAGR, 14.28% volatility, a 0.720 Sharpe ratio, and a −45.67% maximum drawdown after USD 6.15 million of modeled costs across 490,664 contract legs. Its return was economically meaningful, but the deep drawdown and high turnover made it less compelling as a standalone strategy than its terminal equity alone would suggest.

14.6 Standalone Reconciliation and Neighboring Rules

The standalone reconciliation reconciled daily profit, costs, contract legs, market and asset-class contributions, rolls, FX conversion, and dates for all four primary sleeves. On the common 1991-2026 window, the retained 12-month time-series momentum rule had the strongest standalone Sharpe ratio at 0.963 and the shallowest drawdown at −20.01%. Breakout, cross-sectional momentum, and constrained term structure remained positive after costs, but none displaced it as the best standalone sleeve.

The robustness grid was deliberately diagnostic. All three breakout pairs remained profitable, with Sharpe ratios from 0.624 to 0.812. cross-sectional momentum was more horizon sensitive: the 6 month signal excluding the most recent month fell to a 0.334 Sharpe ratio, while the comparable 12 month signal remained strongest at 0.720 and the 18 month signal produced 0.560. Because these neighboring rules were not permitted to replace the primary rules, the exercise measures fragility without converting the research into a parameter search.

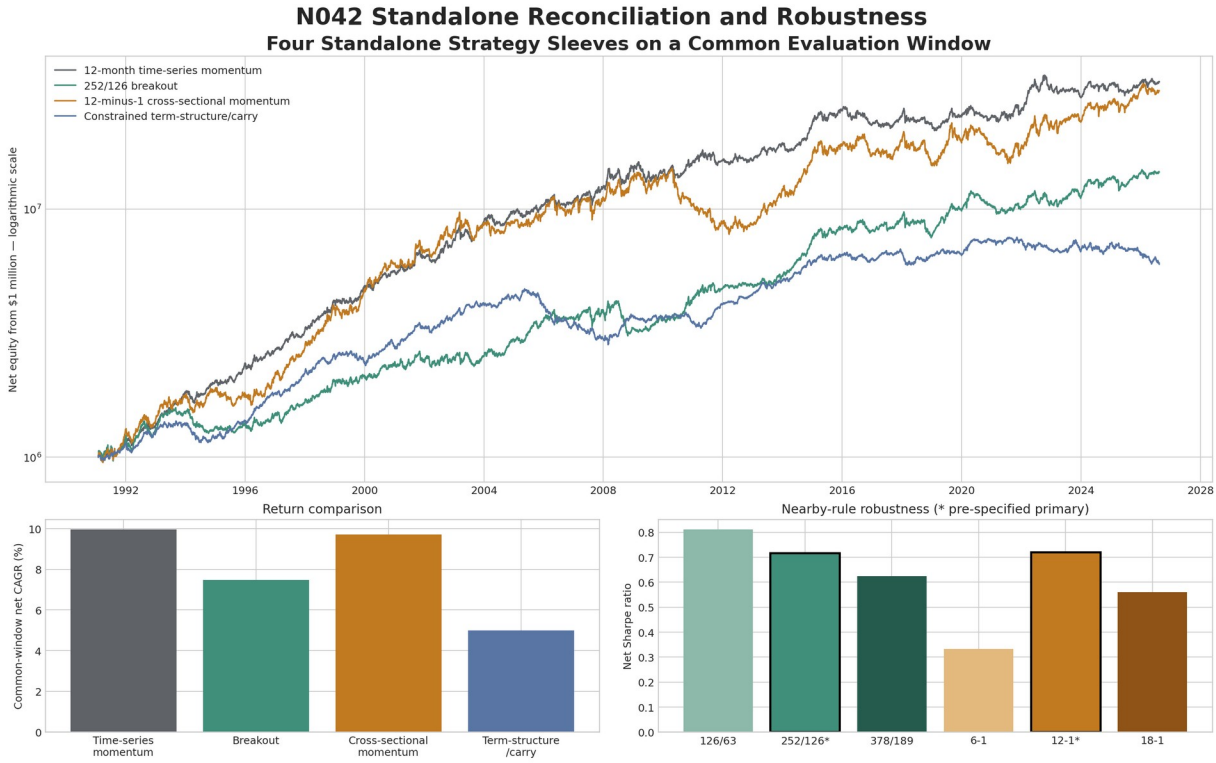


Figure 11. Standalone strategy reconciliation and neighboring rule robustness. The alternative rules are diagnostics and were not eligible to replace the fixed primary specifications.

14.7 Dependence, Crisis Timing, and Drawdown Overlap

The three directional sleeves were closely related: full sample monthly correlations ranged from 0.660 to 0.821. Constrained term structure was the genuine outlier, with correlations of only 0.135 to 0.251 against the directional strategies. Its worst decile month overlap with each directional sleeve was also low, with Jaccard statistics from 0.103 to 0.117. During the 153 losing months for retained time-series momentum, term structure averaged +0.12% and was positive in 53.6% of observations; breakout and cross-sectional momentum were negative on average in the same months.

> Key finding. Strategy diversification is persuasive only when its economic source can be named. The two additional directional models mostly reshape an existing trend exposure, while the futures curve sleeve contributes a genuinely different return source. It was not a universal hedge; term structure lost 9.57% during the 2022 inflation and rate shock while time series and cross-sectional momentum each gained roughly 28%. That regime dependence is why the fixed combination is economically credible rather than cosmetically optimized.

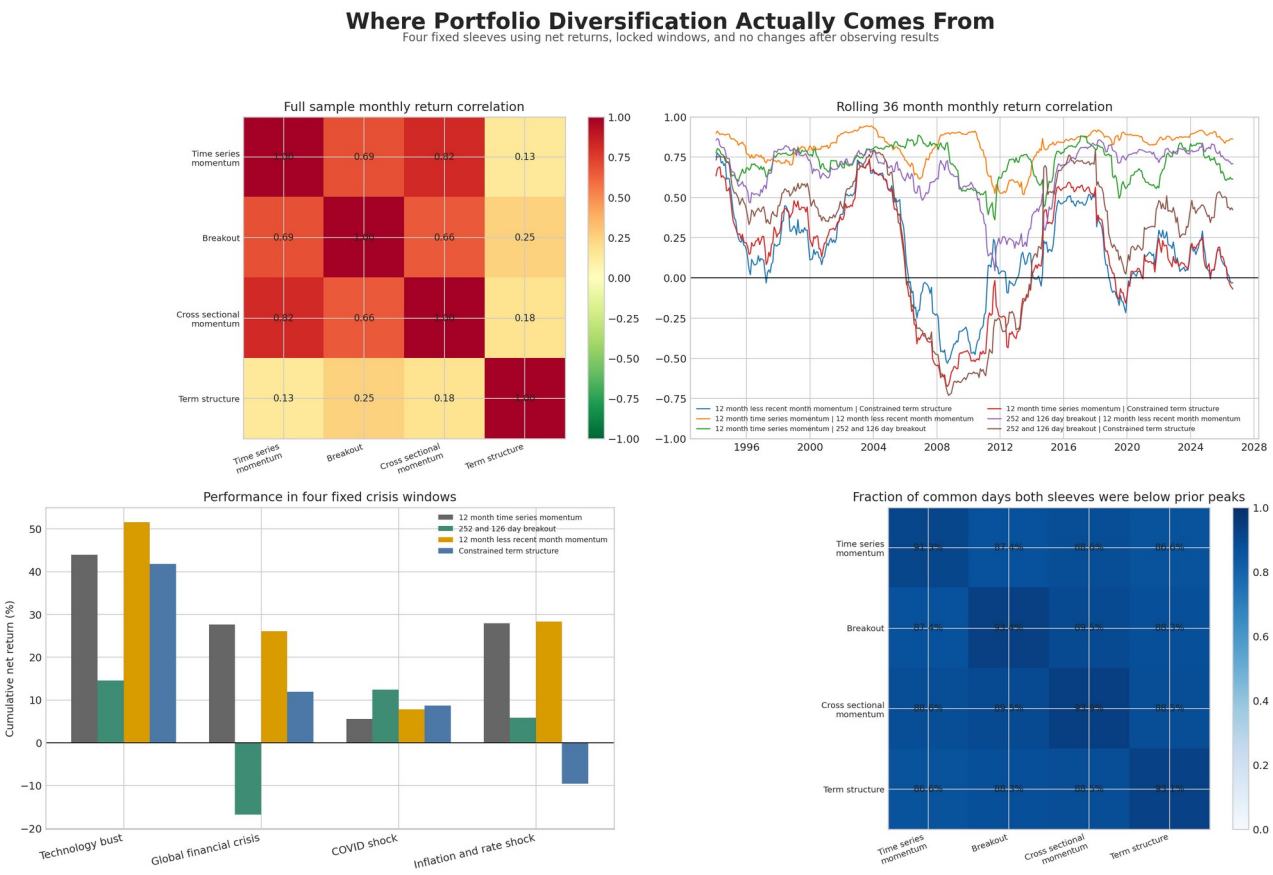


Figure 12. Strategy return correlation, rolling dependence, fixed crisis outcomes, and overlap among the worst monthly losses. Term structure is the principal source of portfolio diversification.

14.8 Fixed Equal-Risk Portfolio

The fixed portfolio used the weights registered before the joint results were examined: one sixth retained time-series momentum, one sixth breakout, one third cross-sectional momentum, and one third constrained term structure. No covariance optimization, historical performance weighting, sleeve removal, or tuning after results were observed was allowed. The fixed return stream portfolio produced 8.08% net CAGR, 8.42% volatility, a 0.964 Sharpe ratio, and a −14.94% maximum drawdown over the common 1991 to 2026 window.

> Key finding. The combination did not manufacture a spectacular Sharpe ratio. It preserved the best standalone risk-adjusted performance, 0.964 versus 0.963 for time-series momentum, while reducing maximum drawdown magnitude by 25.3%, from 20.01% to 14.94%, and lowering volatility from 10.44% to 8.42%. The believable benefit was a shallower loss path, not a backtest miracle.

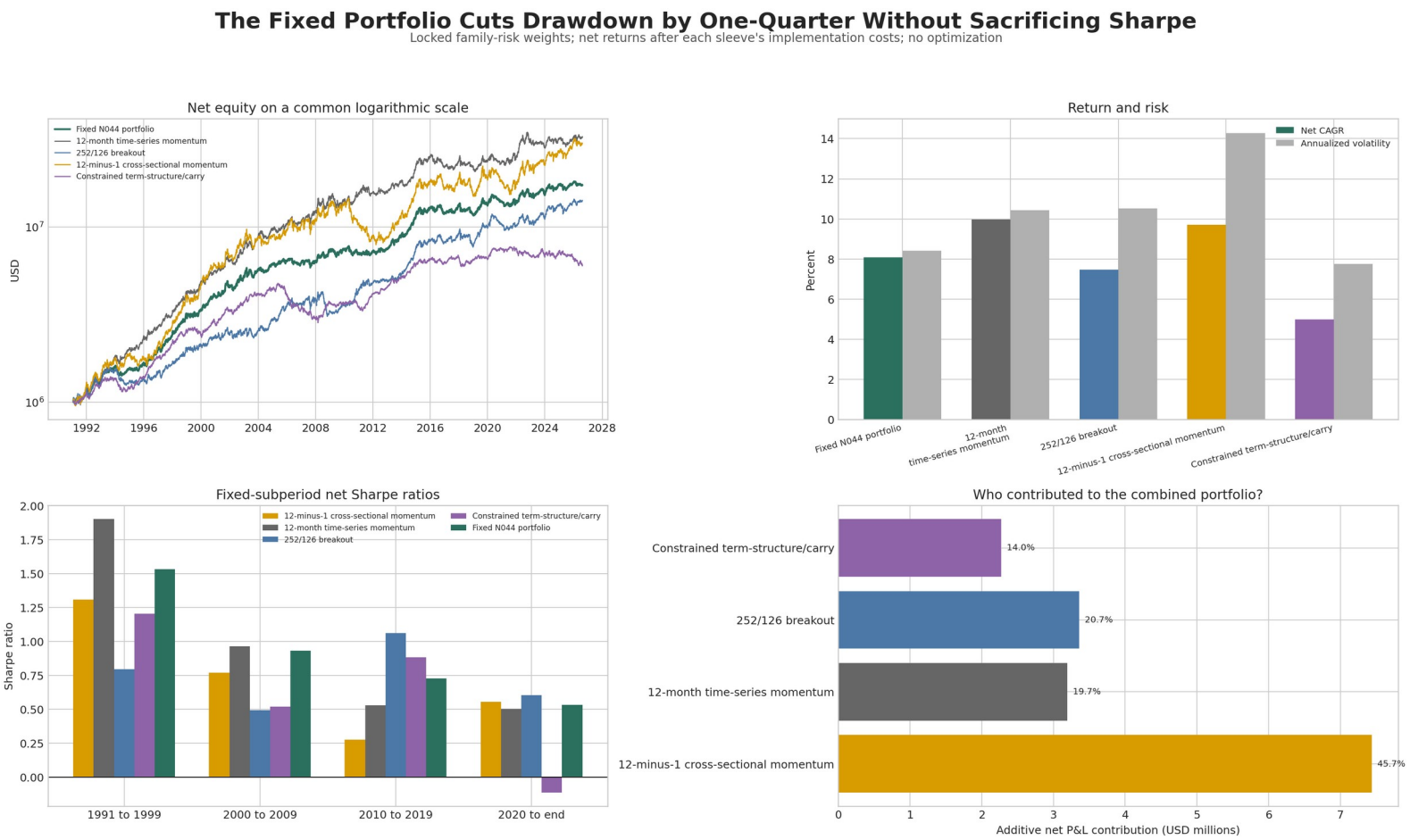


Figure 13. Fixed equal-risk portfolio performance. The combination preserved the best standalone Sharpe ratio while materially reducing drawdown depth.

14.9 Final Portfolio Controls

The controlled portfolio converted the fixed return stream into one causal whole-contract account. It applied the frozen 10% portfolio volatility target, a two times scaling ceiling, market risk and asset-class caps, a six times gross notional cap, and nearest feasible whole-contract rounding without reallocating released risk. The controlled account produced 6.44% net CAGR, 7.19% volatility, a 0.877 Sharpe ratio, and a −11.95% maximum drawdown, with USD 1.20 million of modeled costs.

The two times scale ceiling bound on 56 monthly decisions; the market risk cap bound on 552 event dates, the asset-class cap on 4,769, and the gross notional cap on 2,617. Rounding suppressed 11,023 contracts cumulatively. The lower return and Sharpe ratio are the price of a feasible, constrained account rather than evidence of a failed idea; importantly, maximum drawdown improved again from −14.94% to −11.95%. The negative WTI settlement on 20 April 2020 was preserved in profit and loss and treated by absolute economic exposure for the notional cap.

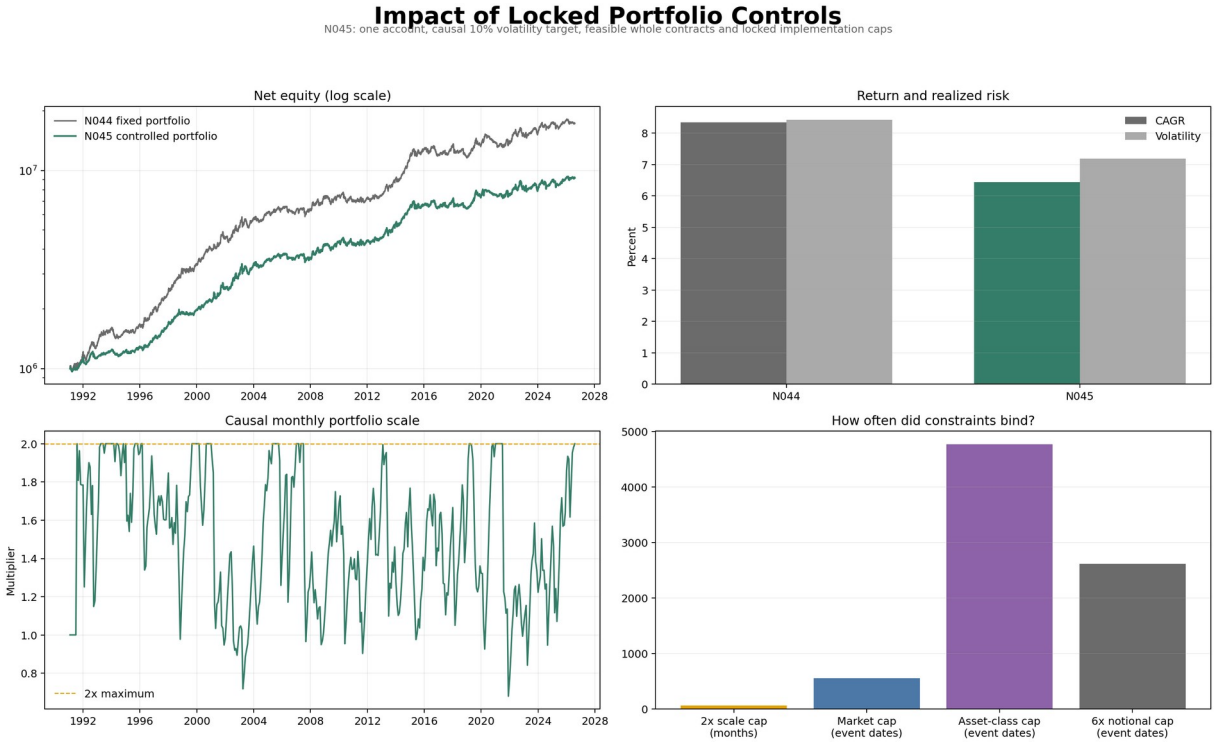


Figure 14. Effect of portfolio risk targeting, market and asset-class caps, gross notional control, leverage limits, and whole-contract feasibility.

A fixed portfolio is credible only if its behavior is not confined to one favorable era. Rolling historical windows provide a demanding final diagnostic before prospective monitoring begins.

15. Historical Robustness and Prospective Holdout

15.1 Rolling Five-Year Robustness

The rolling window analysis evaluated the frozen controlled account in 31 complete five-year windows, advanced one year at a time. Every window had positive net CAGR and a positive Sharpe ratio. The median window produced 5.07% CAGR, 7.35% volatility, a 0.833 Sharpe ratio, and a −9.88% maximum drawdown. The weakest risk-adjusted window began in 2015 and ended in 2019, with 2.35% CAGR, a 0.344 Sharpe ratio, and a −11.74% drawdown; the strongest began in 1996 with a 1.717 Sharpe ratio.

15.2 Frozen live-holdout

Key finding. The historical record is broad rather than dependent on one favorable five-year episode, but the windows overlap and are not 31 independent experiments. The weak window beginning in 2015 is therefore as informative as the perfect positive window count: diversification remained profitable, yet its efficiency fell far below the best historical periods. The live ledger begins strictly after the accepted 14 August 2026 cutoff and currently contains zero observations by design. No success or failure claim is permitted before 252 common trading observations; 756 is the preferred long horizon evaluation. No strategy, weight, target, cap, or parameter may change during that holdout.

15.3 Statistical Interpretation

The rolling windows describe historical stability, not independent out-of-sample trials. The strongest evidence comes from the sequence of advance registration, exact accounting reconciliation, neighboring rule diagnostics, fixed weights, explicit implementation constraints, and a genuinely untouched prospective ledger. Until that ledger contains sufficient data, the report makes no performance claim for the period after August 2026.

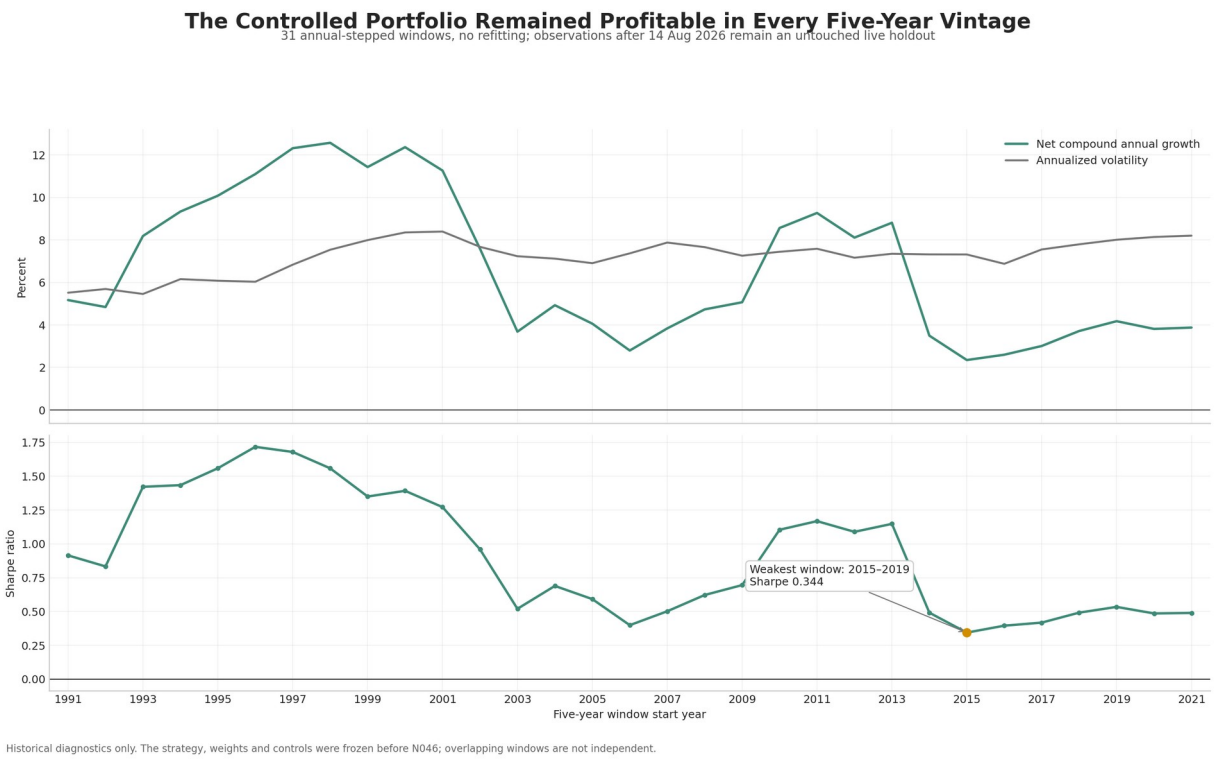


Figure 15. five-year rolling results for the frozen controlled portfolio. Every complete window was profitable, but the windows overlap and are not independent trials.

The evidence can now be interpreted as a whole: what survived, what failed, what portfolio benefit was actually earned, and what remains unknowable from historical data.

16. Discussion, Limitations, and Conclusion

16.1 Discussion

The research began with a simple question: whether global futures trend following survives realistic implementation. It ended with a more useful answer. The 12-month time-series momentum rule remains the strongest standalone sleeve. Broader trend horizons and price-channel breakout do not improve it. Cross-sectional momentum is profitable but costly and drawdown-prone. Term structure is weaker alone, yet it supplies the return source that the directional models do not. The final portfolio works because it combines different economic mechanisms under fixed weights, then accepts the return sacrificed by real contract, concentration, and leverage constraints.

16.2 Limitations

The results remain conditional on daily settlement data, vendor continuous-series construction, lagged-volatility estimates, whole-contract rounding, contemporaneous FX conversion, and the assumed commission and one-tick slippage model. Historical windows overlap, crisis labels cover only a small set of pre-registered episodes, and the term-structure sleeve relies on vendor roll mappings. Most importantly, no historical study, however carefully reconciled, substitutes for a long untouched live record. The report therefore treats the final account as a research specification ready for prospective monitoring, not as proof of future returns.

16.3 Conclusion

The completed evidence supports a constrained global futures portfolio that combines several strategies. The case does not rest on every added model improving performance. It rests on rejecting combinations that merely looked broader and retaining the one sleeve that was demonstrably different. The final controlled account earned 6.44% net CAGR with 7.19% volatility, a 0.877 Sharpe ratio, and a −11.95% maximum drawdown over the historical sample. All 31 complete rolling five-year windows were positive, although the weakest Sharpe ratio was only 0.344. These are credible historical results under a locked and reconciled process. The decisive next evidence will come from the untouched ledger beginning after August 2026.

17. Future Research Program

This report closes one historical research cycle, not the broader research program. Its main contribution is a validated data and implementation foundation that can support new hypotheses without reopening the accepted results. My next studies will therefore remain separate and versioned. Each will begin with an economic rationale, a fixed protocol, explicit reconciliation tests, and an untouched evaluation period.

17.1 Conditional Volatility and Risk Forecasting

The first extension will test whether richer conditional volatility models improve risk estimation and portfolio control. The current implementation uses a lagged exponentially weighted volatility estimate because it is transparent, causal, and operationally stable. Future comparisons will include autoregressive conditional heteroskedasticity, or ARCH, and generalized autoregressive conditional heteroskedasticity, or GARCH, alongside the existing estimator and simpler realized volatility benchmarks. The relevant question is not which model fits historical volatility most closely. It is whether better forecasts survive out-of-sample and produce more stable realized risk, less unintended leverage variation, and better whole-contract implementation after costs.

17.2 Additional Independent Return Sources

The next strategy work will focus on economically distinct systems rather than additional versions of the same directional trend rule. Candidate areas include short horizon reversal, richer curve and carry signals, seasonal effects supported by an economic rationale, and volatility or liquidity aware signals. New systems will first be studied as standalone return streams. Only then will they be assessed by correlation, crisis timing, drawdown overlap, turnover, and incremental contribution to the frozen portfolio. The priority is low structural dependence on the existing sleeves, not a larger strategy count. A model will not be included merely because it raises historical return.

17.3 Academic Research and Replication

I also intend to deepen the academic component of the project. This means continuing to study the primary literature on trend following, momentum, carry, volatility forecasting, portfolio construction, transaction costs, and market microstructure. Selected published results will be reproduced on the frozen futures data, with particular attention to whether the evidence survives the report's contract conventions, causal timing, currency conversion, whole-contract sizing, and cost model. Replication will remain separate from strategy selection so that a literature review does not become a source of retrospective parameter search.

17.4 Prospective Evidence and Research Governance

The post cutoff live ledger remains the most important next test. Historical extensions will be versioned, and no new model will be permitted to rewrite the results of this report. Each future study will preserve its own hypothesis, decision rule, cost model, data cutoff, and acceptance criteria. Over time, the project should become a sequence of comparable research modules connected by a common data, accounting, and governance framework.

The objective is not to produce a larger collection of backtests. It is to build an evolving systematic futures research program in which statistical methods, economic reasoning, implementation realism, and prospective evidence improve together. This report is the first completed foundation for that program.

References

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and Momentum Everywhere. Journal of Finance, 68(3), pages 929 to 985.

Hurst, B., Ooi, Y. H., and Pedersen, L. H. (2017). A Century of Evidence on Trend-Following Investing. Journal of Portfolio Management, 44(1), pages 15 to 29.

Koijen, R. S. J., Moskowitz, T. J., Pedersen, L. H., and Vrugt, E. B. (2018). Carry. Journal of Financial Economics, 127(2), pages 197 to 225.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). time-series momentum. Journal of Financial Economics, 104(2), pages 228 to 250.

Appendix A. Research Audit Trail and Reproducibility

Internal audit IDs appear only in this appendix so that every result can be traced to its saved code and output. They are not part of the reader facing research sequence.

This appendix records the final Norgate based research chronology. Development stage vendor debugging is archived separately and is not part of the final empirical evidence.

Audit ID	Research change / test	Observed result	Decision
N001	Import Norgate futures inventory into RealTest and Python	224 continuous-series records imported, representing 112 economic markets with matched adjusted/unadjusted series.	PASS: inventory established
N002	Validate futures metadata: type, exchange, currency, PointValue, TickSize	224/224 populated Type, Exchange, Currency, PointValue, and TickSize; no non-positive contract metadata.	PASS: contract-economics layer validated
N003	Run standardized OHLC / activity / quote-precision audit	112/112 structural OHLC pass; 107/112 >=500 recent active-volume days. Additive back-adjustment explains non-positive historical adjusted levels in 25 series. A small set of weak/illiquid representations flagged for universe screening.	PASS: proceed to objective universe construction
N004	Apply pre-performance history/liquidity screens, resolve duplicate representations, and freeze Core/Supplemental universe	75 Core, 9 Supplemental, 28 Excluded. Core thresholds: >=2,500 bars, >=500/504 recent active days, >=5,000 60-day average volume, <=2 recent >8-ATR moves; duplicate contract representations resolved without performance information.	PASS: universe frozen before strategy performance
N005	Estimate market-level, asset-class, and rolling correlation on frozen Core 75 using lagged-EWMA-normalized daily point changes	2,775 unique pairs; mean 0.084, median 0.052, range −0.276 to 0.985. Asset-class max off-diagonal 0.419. Rolling 12M mean 0.078, min 0.024 (7 Apr 2003), max 0.154 (13 Mar 2023).	PASS: diversification structure documented before strategy performance
N006	Validate 252-bar time-series momentum signal and execution timing across Python and RealTest	Signal counts reconcile after accounting for CSV export precision; 18,900 observations correctly identified as insufficient 252-bar history. RealTest retains greater internal precision than exported rounded prices.	PASS: signal logic validated; RealTest authoritative for execution
N007	Validate lagged EWMA point-volatility normalization and aligned global-calendar portfolio construction	The intended zero-return treatment for eligible closed markets was not preserved in the saved Python portfolio. On partial-market holidays, closed markets were omitted and risk was redistributed across the markets with bars.	REOPENED - global-calendar portfolio construction not validated
N008	Run first gross 252-day time-series momentum baseline and temporal/asset-class decomposition	The preliminary 12.77% CAGR, 10.45% annualized volatility, 1.203 Sharpe, and −22.25% maximum drawdown inherit the calendar-reallocation defect.	WITHDRAWN - do not use as performance evidence

Audit ID	Research change / test	Observed result	Decision
N009	Reconcile Python and RealTest gross baseline returns date by date	9,264 matched dates; daily-return correlation 0.92885; mean absolute difference 0.10484 percentage points; RMSE 0.25890 percentage points; maximum difference 4.79549 percentage points. Largest discrepancies occur on partial-market holidays when Python EligibleMarkets falls to 1-6 while RealTest retains the broader book.	FAIL - repair calendar alignment, rerun both implementations, and reconcile before acceptance
N010	Record provisional RealTest execution of the frozen 252-day time-series momentum gross baseline	RealTest annualized ROR 14.22%, Sharpe 1.25, maximum drawdown −26.26%, MAR 0.54, profit factor 1.21 over 9,288 periods; no transaction costs. MetaStock spot FX is used through FxRatio. Reported AvgExp and MaxExp are not interpreted as economic notional because back-adjusted synthetic prices can distort nominal exposure statistics.	PROVISIONAL - retain as current execution candidate; validate calendar, FX, quantity, and exposure before acceptance
N011	Correct Python global-calendar aggregation and rerun partial-market-holiday eligibility checks	Corrected file contains 9,264 dates. EligibleMarkets: mean 64.98, minimum 27, maximum 75. On 14 Apr 1995, 21 Apr 2000, 25 Dec 2000, 9 Apr 2004, and 25 Dec 2007, counts are 45, 60, 62, 66, and 70 rather than collapsing to one or two. Corresponding daily returns range from −0.116% to +0.107%, eliminating the earlier holiday-driven extremes.	PASS: calendar aggregation defect corrected; rerun daily reconciliation and continue unless a major residual issue remains
N012	Rerun Python-RealTest daily reconciliation after correcting Python global-calendar aggregation	Across 9,264 matched dates, correlation improved from 0.92885 to 0.95875; mean absolute difference declined from 0.10484 to 0.08542 percentage points; RMSE declined from 0.25890 to 0.19318 percentage points; maximum absolute difference declined from 4.79549 to 4.26238 percentage points. Former holiday failures now reconcile closely. Largest residuals cluster mainly on post-holiday sessions.	PARTIAL PASS: calendar defect resolved; keep RealTest authoritative, continue horizon research, and retain post-holiday/FX/position reconciliation as a nonblocking validation item
N013	Estimate the initial non-overlapping monthly continuation-versus-reversal horizon map across the frozen Core 75	Signal-free 40%-volatility-scaled monthly returns; exact lags 1-12 months; market fixed effects and two-way clustering by market and month. Significant positive coefficients at lags 1, 2, and 10 months; 12-month coefficient negative but insignificant. No reliable reversal through 12 months.	IN PROGRESS - apply multiple-testing correction, extend to longer lags and asset classes, then test pre-specified tradable horizons in RealTest

Audit ID	Research change / test	Observed result	Decision
N014	Correct the 1-12 month horizon-map inference for testing multiple lags	No lag remains significant at 5% under Holm or Benjamini-Hochberg false-discovery-rate correction. Lowest Holm-adjusted p-value: 0.1223 at 10 months. Lowest false-discovery-rate-adjusted p-value: 0.0819 at one and 10 months.	PASS: downgrade isolated lag significance to suggestive; continue with longer lags, asset classes, joint tests, and pre-specified RealTest horizons
N023	Run a pre-specified 21/63/126/252-day direction ensemble in RealTest using the same frozen Core 75 data, EWMA volatility scaling, FX conversion, and gross sizing convention as the 252-day baseline	Annualized ROR 10.44%, Sharpe 1.20, maximum drawdown −22.75%, MAR 0.46, profit factor 1.18, and 128,012 trades. Reported average and maximum exposure declined to 1,184.91% and 6,013.68%, versus 1,806.81% and 11,032.93% for the baseline.	PASS: robustness check completed. Retain the 252-day rule as the primary gross baseline; retain the ensemble as a lower-intensity variant for cost- and constraint-aware testing.
N024	Apply and validate a reference transaction-cost model on the 252-day RealTest baseline; charge USD 2.50 per contract per transaction, one-tick of slippage per transaction, and both legs of each mandatory roll	The 0x control exactly reproduced the gross baseline. At 1x, annualized return fell from 14.22% to −1.17%, Sharpe from 1.25 to −0.09, and maximum drawdown worsened from −26.26% to −86.23%. Total modeled costs were $10.856 million: 70.8% roll costs, 23.8% ordinary slippage, and 5.5% commissions.	FAIL CURRENT IMPLEMENTATION - gross edge is not viable under the 1x reference costs. Preserve the signal, then test a cost-aware no-trade band and explicit portfolio risk constraints before judging the strategy family.
N025	Add a 10% no-trade band to the validated 252-day 1x reference-cost implementation while preserving the frozen signal, risk allocation, universe, and cost assumptions	Trades declined from 41,917 to 26,044 (−37.9%) and total modeled costs from $10.856 million to $7.321 million (−32.6%). Annualized return improved from −1.17% to 0.41%, Sharpe from −0.09 to 0.09, and maximum drawdown from −86.23% to −74.57%. Roll costs remained 67.8% of total modeled costs.	PARTIAL PASS: the no-trade band materially reduces turnover and cost drag but does not restore economic viability. Continue a small pre-specified band sensitivity; do not optimize to the best historical result.
N026	Widen the pre-specified no-trade band from 10% to 20% while preserving the signal, risk allocation, universe, compounding, and reference-cost assumptions	Annualized return improved to 1.17%, Sharpe to 0.17, maximum drawdown to −65.20%, profit factor to 1.02, and net profit to $510,003. The run reported 26,795 trades and $7.948 million of modeled costs. Trades and dollar costs rose versus N025 because the stronger compounded equity path maintained larger contract quantities; rolls remained 66.8% of total costs.	PARTIAL PASS: net performance improved but remains too weak. Close the band sensitivity without optimizing wider thresholds; proceed to a representative individual-contract roll and execution audit.

Audit ID	Research change / test	Observed result	Decision
N027	Measure calendar-year and pre-/post-2010 temporal robustness using the authoritative RealTest gross-equity path	32 of 35 complete years profitable. Pre-2010: 18.86% annualized return, 1.63 Sharpe, −18.55% maximum drawdown. Post-2009: 9.13%, 0.83, and −26.26%. Full active sample reconciles to 14.22%, 1.25, and −26.26%.	PASS: temporal breadth is strong, but post-2009 decay is material. Retain the gross effect while requiring cost-aware and recent-sample validation.
N028	Export and reconcile all trades from the accepted gross RealTest baseline; attribute realized P&L by market and asset-class	167,069 trades and 75 markets; exported P&L reconciles within $0.31. 57 markets profitable. Top five generated 18.9% of positive-market profits. All four broad asset classes contributed positively: commodities 58.3%, bonds 24.3%, equities 12.4%, and currencies 5.1%.	PASS: gross contribution is broad across markets and asset classes. Retain the attribution as descriptive evidence and run counterfactual exclusion tests in RealTest because dynamic sizing makes simple subtraction invalid.
N029	Audit Norgate availability of individual-delivery histories for representative equity-index, government-bond, currency, and commodity futures	Confirmed individual deliveries for E-mini S&P 500 from 1997, 10-Year U.S. Treasury Note from 1982, Euro FX from 1999, and Light Sweet Crude Oil from 1984, alongside unadjusted and back-adjusted continuous forms.	PASS: required contract-level data are available. Import the four delivery chains and reconcile roll dates, contract prices, volume/open interest, expiration metadata, and two-leg execution costs.
N030	Reconcile Norgate continuous-series contract changes with outgoing and incoming individual deliveries for ES, ZN, 6E, and CL	806 usable events with valid dates and no duplicate market-date rows. Every continuous close exactly matched the incoming individual contract. Median switch timing was 0 days before outgoing expiry for CL and 6E, 1 day for ES, and 21 days for ZN. Incoming volume was higher in 97.7%, 100.0%, 100.0%, and 34.5% of events, respectively.	PASS: contract linkage and roll timing validated. Use vendor mapping for roll dates; keep term-structure gaps separate from execution costs.
N031	Test whether pre-roll backwardation or contango predicts the direction of the next non-overlapping roll-cycle move in ES, ZN, 6E, and CL	798 completed cycles. Pooled directional accuracy was 54.5% (exact p = 0.0119). The market-fixed-effect coefficient was 0.097 standardized risk units with calendar-month-clustered t = 2.46 and p = 0.0143. ZN was the strongest market and survived Holm correction across four decompositions.	PASS AS SIGNAL PILOT - retain term structure as a candidate diversifier. Next require a daily lagged-volatility, cost-aware implementation before portfolio acceptance or full-universe scaling.

Audit ID	Research change / test	Observed result	Decision
N032B	Implement the four-market curve-direction signal as a daily, whole-contract, lagged-volatility strategy with roll-only resizing and reference costs	1999-2026 common sample. Net CAGR 10.57%, annualized volatility 21.36%, Sharpe 0.58, and maximum drawdown −43.86%. Costs reduced CAGR by 0.59 percentage points. CL, ES, and 6E contributed positively; ZN lost USD 3.07M.	PASS AS IMPLEMENTATION PROTOTYPE - cost drag is manageable, but do not accept or scale until N032C reconciles the negative ZN contribution against the positive N031 event evidence and tests temporal stability.
N033	Expand the validated four-market roll inventory to all 75 Core futures and reconcile the pilot subset	13,661 vendor roll events with no duplicate market-date rows. The four pilot markets reproduced all 806 N030 events exactly. Five observations failed pre-specified quality checks, leaving 13,656 valid signal observations; 201 valid flat curves produced zero signals.	PASS: full-universe roll linkage accepted. Retain vendor mapping, local-currency contract metadata, and explicit FX conversion for portfolio implementation.
N034	Build and reconcile the Core 75 daily back-adjusted price, volatility, contract-metadata, and FX panels	N034C contains 602,616 market-date rows across 75 markets from 1990-2026. N034D supplies the contemporaneous currency-conversion series. The N035 merge produced no missing FX rows and reconciled all market-date inputs.	PASS: daily P&L, lagged volatility, point value, tick size, and FX conversion inputs accepted for N035.
N035	Implement the event-driven term-structure signal across the frozen Core 75 universe with whole contracts and explicit two-leg roll costs	Gross CAGR 6.28%, volatility 8.79%, Sharpe 0.72, and maximum drawdown −47.29%. After USD 2.934M of costs over 228,814 contract legs, net CAGR was 4.46%, Sharpe 0.52, and maximum drawdown −47.73%.	PASS AS FULL-UNIVERSE IMPLEMENTATION - retain as a candidate diversifier; diagnose the disproportionate government-bond cost burden before portfolio combination.
N036	Reconcile and decompose N035 government-bond transaction costs by event type, quantity, commission, slippage, and contract-level risk	All USD 2.934M reconciled exactly. Bonds generated 62.2% of contract legs and 67.8% of costs; 82.1% of bond cost was slippage and 83.6% of bond legs were mandatory replacements. ZQ alone produced 39.6% of total portfolio cost. Bond cost correlated 0.93 with legs and −0.83 with dollar volatility per contract.	PASS AS DIAGNOSTIC - the burden is a concentrated position-scale and replacement-turnover problem. Pre-register one uniform constraint test; do not delete markets or optimize thresholds from this sample.
N037	Uniform contract-size constraint	Costs −45.0%; contract legs −42.4%; net Sharpe 0.685; drawdown −40.43%.	PASS: accepted uniform implementation control
N038	12-month versus 3/6/12-month trend horizon	Broader ensemble reduced volatility and cost but lowered CAGR and Sharpe and worsened drawdown.	RETAIN 12-month time-series momentum
N039	Lock remaining research protocol	Universe, signals, weights, crises, costs, controls, and holdout frozen	PASS: preregistration locked

Audit ID	Research change / test	Observed result	Decision
		before N040.	
N040	252/126 price-channel breakout	Net CAGR 7.54%; Sharpe 0.717; drawdown −27.76%.	PASS: viable related trend sleeve
N041	12-minus-1 cross-sectional momentum	Net CAGR 10.03%; Sharpe 0.720; drawdown −45.67%; causal monthly mapping validated.	PASS: retain for locked combination
N042	Standalone reconciliation and robustness	All primary sleeves reconciled; neighboring breakout rules robust; cross-sectional momentum more horizon-sensitive.	PASS: no primary rule replaced
N043	Dependence, crises, and drawdown overlap	Directional sleeves correlated 0.660-0.821; carry only 0.135-0.251 versus directional sleeves.	PASS: carry identified as distinct diversifier
N044	Fixed equal-risk multi-strategy portfolio	Sharpe 0.964; drawdown −14.94%, 25.3% shallower than strongest standalone sleeve.	PASS: fixed architecture accepted
N045	Risk targeting, caps, and leverage constraints	Net CAGR 6.44%; Sharpe 0.877; drawdown −11.95%; all constraints reconciled.	PASS: controlled one-account portfolio accepted
N046	Rolling robustness and live-holdout	All 31 complete five-year windows positive; median Sharpe 0.833; live ledger empty at cutoff.	PASS: historical research complete; prospective monitoring begins

Appendix B. Contract Metadata and Market Mapping

The universe was frozen at 75 Core and 9 Supplemental Norgate back-adjusted futures before any strategy performance was examined. Representative contract metadata are shown below; the complete row level classification and its reasons are preserved in the supporting research package.

asset-class	Economic market	Norgate symbol	Exchange	Currency	PointValue	TickSize	History	Status
Stock Index	E-mini S&P 500	&ES_CCB	CME	USD	50	0.25	7284 bars	CORE
Interest Rate	10-Year U.S. T-Note	&ZN_CCB	CBOT	USD	1000	0.0156	9209 bars	CORE
Currency	Euro FX	&6E_CCB	CME	USD	125000	5e-05	6942 bars	CORE
Energy	Crude Oil - Light Sweet	&CL_CCB	NYMEX	USD	1000	0.01	9196 bars	CORE
Metal	Gold	&GC_CCB	NYMEX	USD	100	0.1	9197 bars	CORE
Agriculture	Corn	&ZC_CCB	CBOT	USD	50	0.25	9221 bars	CORE

Frozen Core symbols by asset-class:

Stock Index (17): &EMD_CCB, &ES_CCB, &FCE_CCB, &FDAX_CCB, &FESX_CCB, &FSMI_CCB, &HSI_CCB, &KOS_CCB, &LFT_CCB, &NQ_CCB, &RTY_CCB, &SCN_CCB, &SNK_CCB, &SSG_CCB, &SXF_CCB, &YAP_CCB, &YM_CCB

Interest Rate (19): &CGB_CCB, &FBTP_CCB, &FGBL_CCB, &FGBM_CCB, &FGBS_CCB, &FGBX_CCB, &FOAT_CCB, &LEU_CCB, &LLG_CCB, &TN_CCB, &UB_CCB, &YIR_CCB, &YXT_CCB, &YYT_CCB, &ZB_CCB, &ZF_CCB, &ZN_CCB, &ZQ_CCB, &ZT_CCB

Currency (8): &6A_CCB, &6B_CCB, &6C_CCB, &6E_CCB, &6J_CCB, &6M_CCB, &6N_CCB, &6S_CCB

Energy (7): &BRN_CCB, &CL_CCB, &GAS_CCB, &GWM_CCB, &HO_CCB, &NG_CCB, &RB_CCB

Metal (5): &GC_CCB, &HG_CCB, &PA_CCB, &PL_CCB, &SI_CCB

Agriculture (18): &CC_CCB, &CT_CCB, &GF_CCB, &HE_CCB, &KC_CCB, &KE_CCB, &LCC_CCB, &LE_CCB, &LRC_CCB, &LSU_CCB, &MWE_CCB, &RS_CCB, &SB_CCB, &ZC_CCB, &ZL_CCB, &ZM_CCB, &ZS_CCB, &ZW_CCB

Other (1): &EUA_CCB

Supplemental (9): &HTW_CCB, &VX_CCB, &CRA_CCB, &SJB_CCB, &SO3_CCB, &SR3_CCB, &BTC_CCB, &DX_CCB, &ETH_CCB

Appendix C. Supporting Files and Diagnostics

The supporting research package contains the data-quality audit, universe classification, correlation outputs, signal and execution reconciliations, volatility diagnostics, trade-event ledgers, contribution files, strategy return series, portfolio controls, and robustness tables used in this report.