Handelsmeisterei
Systematic Trading — Made in Germany

STRATEGY | IMPLEMENTATION

# Strategy Recalibration Breakout

Dr. Sebastian Schmidt

September 2026

ISSUE
04

# Letter from the editor

ISSUE 4

What a summer! The last few months in Germany have been hot, and I've been very busy with work. Once I had published my previous article, Crafting a Trading Strategy, it was time to recalibrate the strategies in my portfolio for the year, following the steps outlined there.

I went down quite a rabbit hole. I expanded the framework to include market-regime definitions and a liquidity analysis, and even changed the algorithm that searches the parameter surface. I'll need to update the previous article to reflect these changes.

Documenting all the strategies has been even more demanding. This led to my first vibe-coding project, using both ChatGPT and Claude. It is a perfect use case because the project is not critical to my trading, and I do not need to understand every line of the generated code.

I ended up building an editorial design system that helps me lay out and render my internal documents and articles for the website. A big shout-out to the two open-source projects Justif and Paged.js, which make it possible to produce professionally typeset PDFs from my notebooks.

I think it was worth the effort. This article presents the validation report for one of my strategies, Breakout, in the same form I use internally. As it turns out, I'll be retiring the strategy from my portfolio. Read on to find out why.

I wish you all the best with your trading this autumn. Stay tuned: I have plenty of interesting topics lined up for the website.

*Dr. Sebastian Schmidt*

EDITOR & AUTHOR

CONTENTS

# Strategy Recalibration Breakout

THIS REPORT *evaluates the Breakout strategy through signal diagnostics, parameter selection, robustness testing, and portfolio integration. It examines whether historical performance holds up under recent conditions and whether the strategy adds value to the portfolio. Despite strong results over the full sample, weaker recent performance and substantial overlap with existing strategies support the decision to retire Breakout.*

## IN BRIEF

**THE QUESTION** Does the strategy survive validation, and does it add value to the portfolio?

**THE PROCESS** A seven-stage framework moves from economic rationale to portfolio integration, testing whether the strategy remains plausible, robust, and implementable throughout development.

Strategy development is treated as a sequence of research questions rather than a search for the strongest backtest. The process begins by defining the tradable universe and economic rationale, then examines whether the signal behaves as intended across instruments, parameters, and market environments.

Only after these diagnostic checks is historical performance used to explore the parameter space. The objective is to identify broad regions in which neighboring specifications produce similar results rather than to select a single optimum. The resulting parameter set is then tested for instrument dependence, liquidity effects, execution sensitivity, calibration stability, and statistical uncertainty. The final stage evaluates the strategy in its intended portfolio context.

Throughout the report, methodology is separated from strategy-specific conclusions: the main text explains the purpose of each stage, while findings specific to this strategy are summarized in the Key Takeaways. The full research framework is described in Crafting a Trading Strategy.

## IN NUMBERS

**29.3%**
return

**−3.2%**
costs

**49.4%**
maximum drawdown

**1.18**
Sharpe ratio

Hypothetical results, net of est. costs; see Appendix A.

# Tradable Universe

*Starting with what can actually be traded.*

The analysis begins with the tradable universe and the historical data available for each instrument. Instruments are included only when they satisfy the applicable data-quality, liquidity, and implementation requirements.

The available universe changes over time as instruments are introduced or discontinued, or as they become liquid enough to trade. Signal calculations may additionally require an initialization or burn-in period before an instrument becomes eligible. The resulting universe therefore reflects both market availability and the data history required by the strategy.

The tradable universe constrains the feasible parameter range. A requirement for long histories reduces instrument availability, while shorter histories permit a broader cross-section.

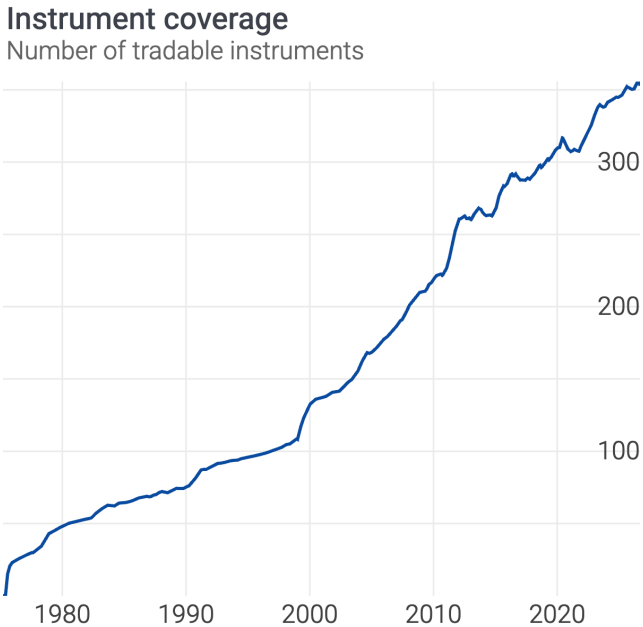


**Figure 1. The tradable universe has grown roughly fivefold since 1980.**
*Note:* Daily counts, 1975–2026, smoothed for readability.

Parameter bounds should therefore balance the amount of information required by the signal against the diversification provided by the available universe.

# Economic Rationale

*Establishing why the signal should exist.*

Strategy development starts with an economic rationale rather than with historical performance. The rationale states why the proposed signal could be expected to contain information about future returns and makes the underlying assumptions explicit.

### ECONOMIC RATIONALE

Financial markets continuously process new information about the economy. While news may arrive instantly, its impact on prices often unfolds gradually. Market participants differ in their information, objectives, risk constraints, and speed of decision-making. As a result, capital is reallocated over time rather than all at once. This gradual adjustment creates persistent price movements that can last for weeks, months, or even years. Trend-following does not attempt to predict these movements before they begin. Instead, it assumes that once a persistent adjustment is underway, it is more likely to continue than to reverse immediately.

# Signal Analysis

*Understanding behavior before measuring returns.*

Before strategy returns are examined, the trading signal itself is analyzed. The objective is to verify that the implementation behaves as intended and that its values are economically and numerically plausible across all instruments.

The first step aggregates signal values across instruments for each lookback. The distribution should change smoothly as the lookback increases. Abrupt jumps, isolated extremes, or discontinuities deserve investigation.

Such irregularities may indicate data problems, insufficient burn-in, numerical instability, or an unintended property of the signal definition (see Figure 2 A).

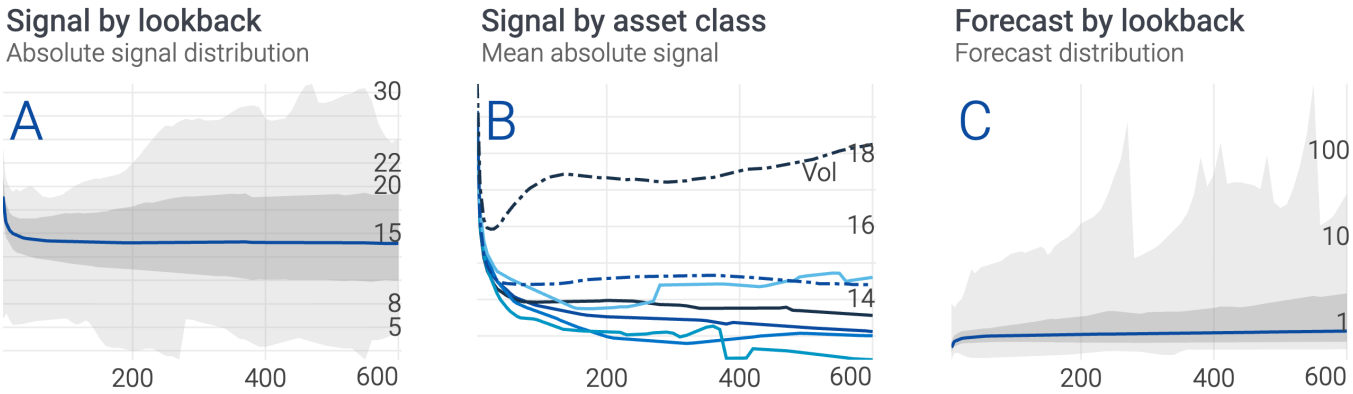


**Figure 2. Signal behavior across lookbacks and asset classes. The lines show the mean (A and B) or median (C), the dark band the 5th–95th percentile, and the light band the minimum-to-maximum range.**
*Note:* Measured across the full universe, 1975–2026; lookbacks 5–600.

## KEY TAKEAWAYS

1. Signal values are smooth across all instruments and lookbacks.
2. Signal magnitudes are similar across asset classes, with one exception: volatility futures exhibit stronger trends because the contracts traded have more distant expiry dates and are subject to a “roll-down” effect.
3. Some forecasts show extreme values, driven by the short data histories of individual instruments such as UMBS-55.

> *Smooth behavior builds confidence—anomalies demand explanation.*

Signal behavior is also compared across different asset classes in Figure 2 B. Large systematic differences should be understood before proceeding. They may be economically justified, but they can also point to a structural problem in the signal or the data.

Finally, the range of the forecast values is examined across the complete parameter space. This provides an additional check on scaling and identifies parameter regions that generate unusually large, unstable, or otherwise implausible exposures.

At this stage, the focus is on signal behavior rather than profitability.

Individual time series provide another diagnostic of the signal. They help identify discontinuities, excessive turnover, unexpected reactions to market data, initialization problems, or other behavior that may indicate an error in the signal construction (Figure 3).

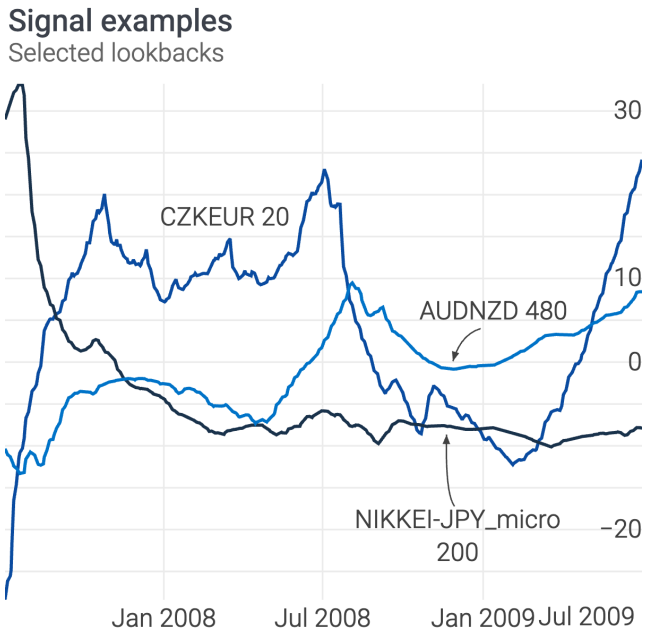


**Figure 3. Three instruments, three lookbacks, a two-year window drawn at random.**
*Note:* Shorter lookbacks are more volatile because they rely on less data.

# Parameter Selection

## *Seeking broad robustness, not a single best fit.*

Once the signal has passed the basic diagnostic checks, historical returns are used to explore the parameter space.

The objective is not to single out the specification with the highest historical performance, but to identify broad regions in which economically reasonable neighboring specifications produce similar results.

To reduce dependence on the particular market environment represented by the full sample, performance is evaluated separately across several market regimes (see Table 1). For each parameter specification, the Sharpe ratio is calculated within each regime and then averaged across regimes. This produces a regime-averaged parameter surface in which each regime receives equal weight, irrespective of its length.

Regime-averaged and full-sample Sharpe ratios are compared.

Return correlations between parameter specifications are analyzed alongside performance. Closely related parameters may produce nearly identical return streams and therefore add little diversification.

| PERIOD | MARKET REGIME |
|---|---|
| 1975–2007 | Pre-GFC Era |
| 2008–2021 | Post-GFC / QE Era |
| 2022–2026 | Post-COVID Inflation Era |

**Table 1. Market-regime classification used for the parameter search.**
*Note:* The periods were defined manually based on major financial-market events and recurring differences in performance across several strategies.

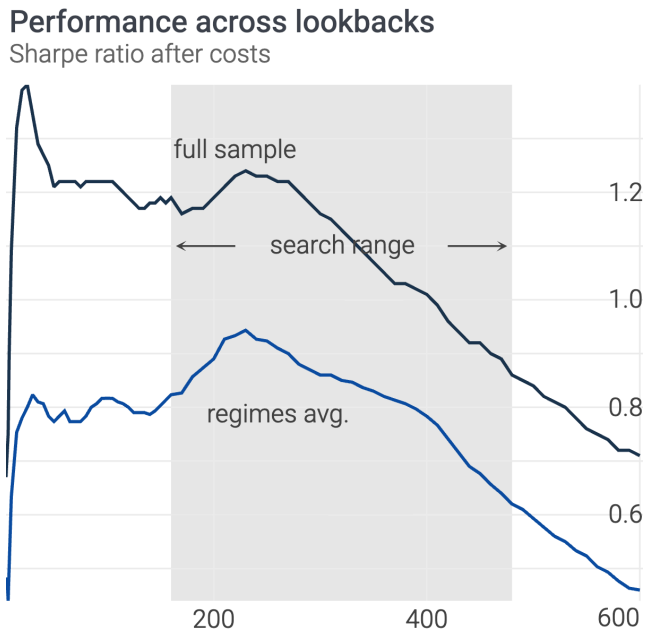


**Figure 4. One robust parameter region across market regimes: 160–480 days.**
*Note:* Full universe, 1975–2026; the regime-averaged line is the mean over the three market regimes in Table 1.

The correlation profile helps determine how densely the parameter space needs to be represented and whether apparently different lookbacks provide diversification.

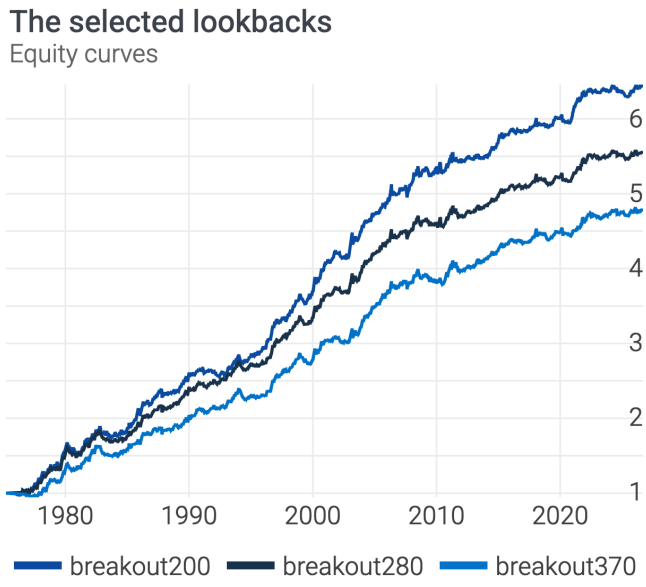


**Figure 5. The equity curves of the selected lookbacks, 1975–2026.**
*Note:* Each lookback is simulated independently.

A robust parameter surface should vary relatively smoothly, with broad regions of similar performance that persist across market regimes. Based on the surface and correlation structure, an ensemble of parameters is selected from the robust region. The ensemble reduces dependence on any single parameter and can improve stability.

Only after the signal behavior and the parameter space have been assessed is the combined ensemble evaluated with a full backtest. These results describe the historical characteristics of the selected ensemble but are not evidence that performance will persist out of sample.

The tear sheet summarizes the historical performance of the selected ensemble (see Figures 7 and 8 and Tables 2 and 3).

The selected lookbacks are examined individually in Figure 5, while Figure 9 shows the correlations between their return streams. Together, these checks indicate whether the ensemble avoids unnecessary redundancy.

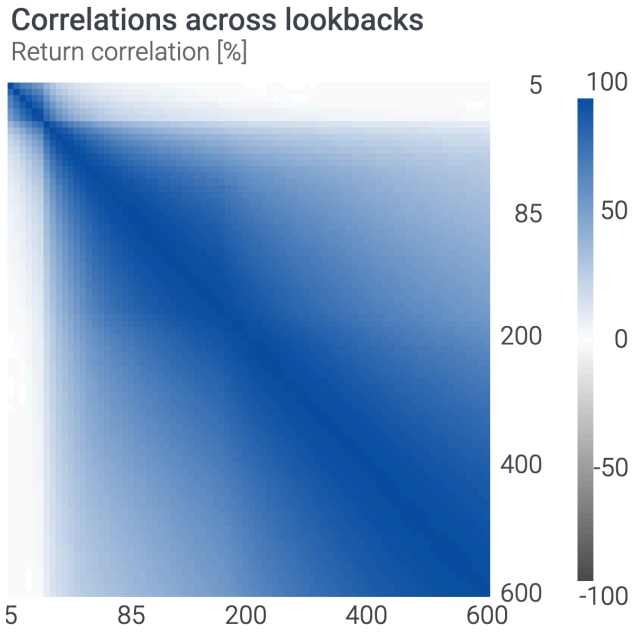


**Figure 6. Correlations decline as lookbacks diverge, indicating greater diversification across distant lookbacks.**
*Note:* Each cell compares two lookbacks over the full sample.

# Tear Sheet Breakout

## Strategy and asset-class performance

Equity curves and drawdowns [%]

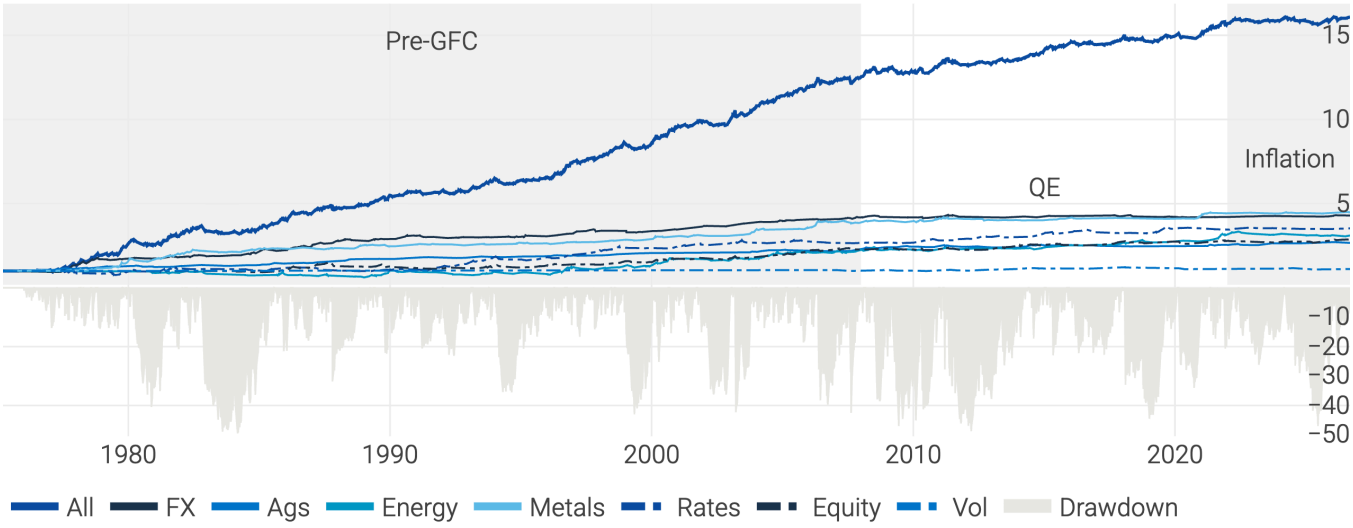


**Figure 7. Selected ensemble (200, 280, 370) and its asset-class components, 1975–2026.**
*Note:* The gray band below the axis is the drawdown.

## Return by decade

Net of costs, annualized [%]

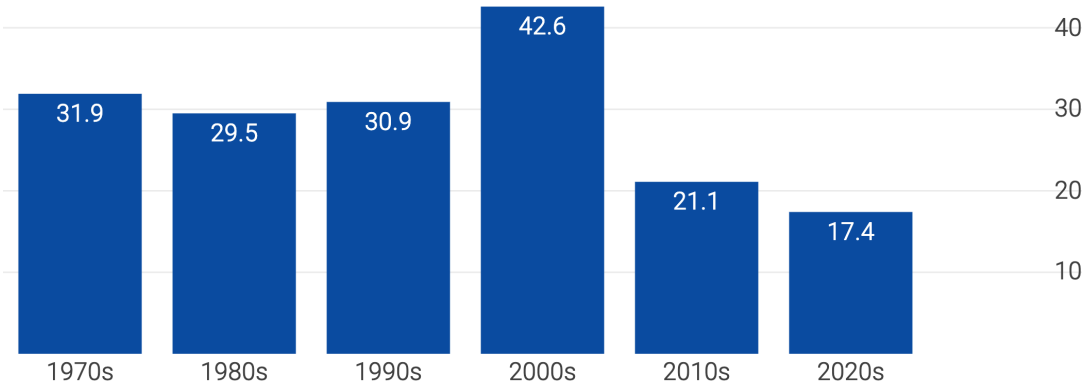


**Figure 8. The first and last decades are only partially covered by the sample.**
*Note:* Each decade evaluated in isolation.

**PERFORMANCE**

| | |
|---|---|
| 29.3% | return |
| −3.2% | costs |
| 24.9% | vol target 25% |
| 1.18 | Sharpe ratio |

Hypothetical results, net of est. costs; see Appendix A.

| RISK | |
|---|---|
| Maximum drawdown [%] | 49.4 |
| Average drawdown [%] | 13.5 |
| Maximum drawdown duration [days] | 760 |
| Average drawdown duration [days] | 26 |

**Table 2. Drawdowns as percentages of capital.**
*Note:* All metrics before dynamic portfolio optimization, sample 1975–2026.

| DISTRIBUTION | |
|---|---|
| Monthly skewness | 0.05 |
| Lower tail | 1.82 |
| Upper tail | 1.44 |
| Winning days [%] | 54.8 |

**Table 3. The shape of daily and monthly returns over the full sample.**
*Note:* All metrics before dynamic portfolio optimization, sample 1975–2026.

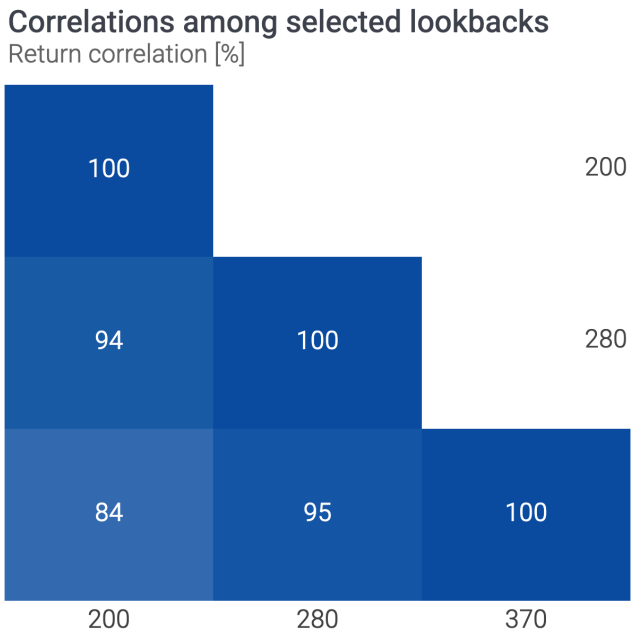


**Figure 9. Correlations decline as two lookbacks move farther apart, from 95% between neighbors to 84% at the extremes.**
*Note:* Darker cells indicate stronger co-movement; the diagonal is self-comparison.

The correlations need not be low; the objective is an ensemble of genuinely distinct lookbacks that adds some diversification.

The resulting parameter set is a practical, robust ensemble of signals rather than an attempt to maximize historical backtest performance. Strategy-specific conclusions are summarized in the Key Takeaways.

### KEY TAKEAWAYS

1. The strategy performs poorly with short lookbacks in more recent regimes. The search is therefore restricted to lookbacks of 160–480 days.

2. Over the full sample, shorter lookbacks produce a Sharpe ratio of up to 1.44, but the improvement is driven by the 1970s and 1980s.

## Validation

*Trying to break the strategy before trusting it.*

Following parameter selection, the strategy is subjected to a series of robustness and statistical validation tests. These tests examine whether the historical results depend excessively on particular instruments, market segments, execution assumptions, calibration choices, or sample periods.

Validation cannot prove that a strategy will remain profitable: markets are non-stationary and historical samples are finite. These tests are therefore diagnostic, seeking weaknesses that would undermine the historical evidence.

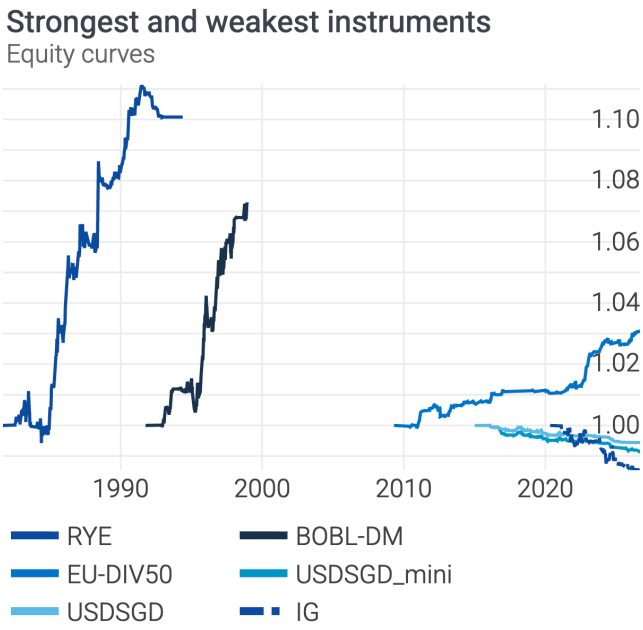


**Figure 10. The strongest and weakest instruments owe their results to short data histories rather than to a signal defect.**
*Note:* Three best and three worst by Sharpe ratio, 1975–2026.

The tests also assess whether implementation would be difficult. Performance is first examined at the instrument level.

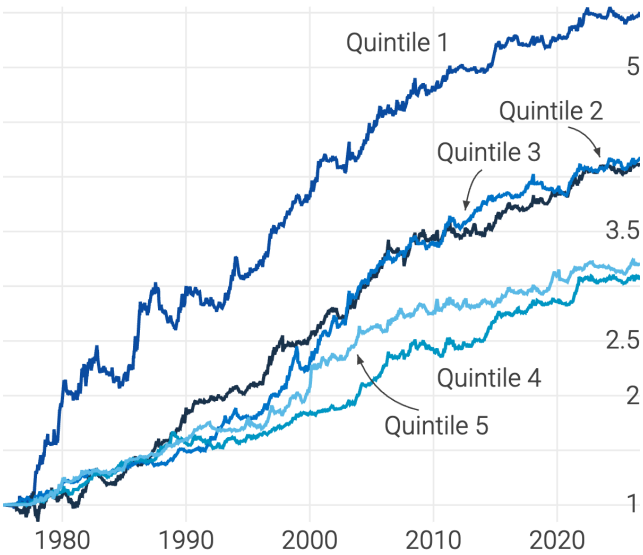


**Figure 11. Full universe split into liquidity quintiles by 20-day average risk volume, 1975–2026.**
*Note:* Quintile 1 holds the most liquid instruments.

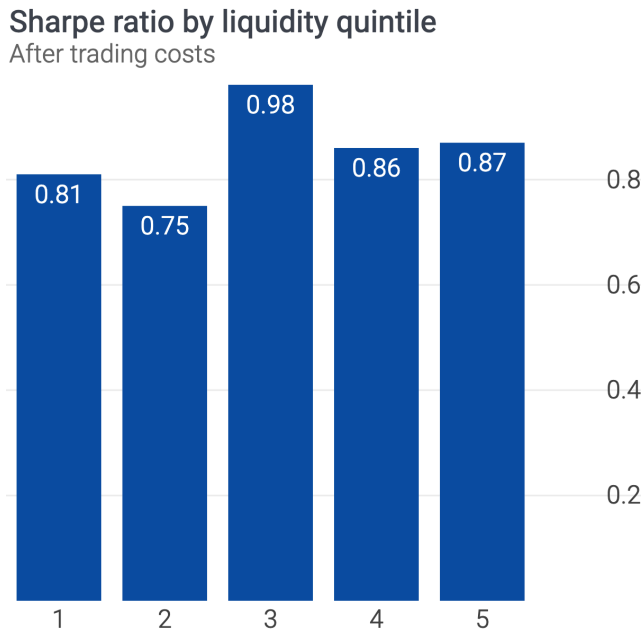


**Figure 12. The Sharpe ratios do not follow the liquidity ordering of the equity curves.**
*Note:* Full universe, 1975–2026, in quintiles by 20-day risk volume; Quintile 1 is the most liquid.

The strongest and weakest contributors are inspected for unusual behavior, data issues, short histories, or structural characteristics that could

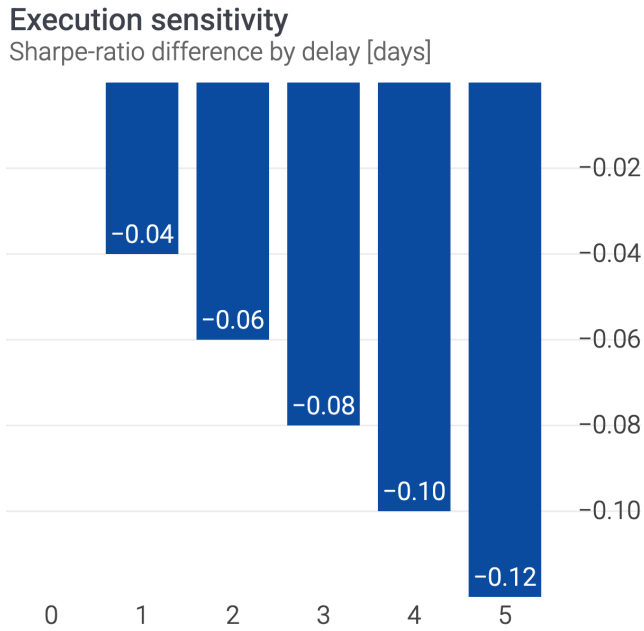


**Figure 13. A one-day execution delay has little impact; performance deteriorates gradually as the lag increases.**
*Note:* Selected ensemble, 1975–2026; orders held back one to five trading days.

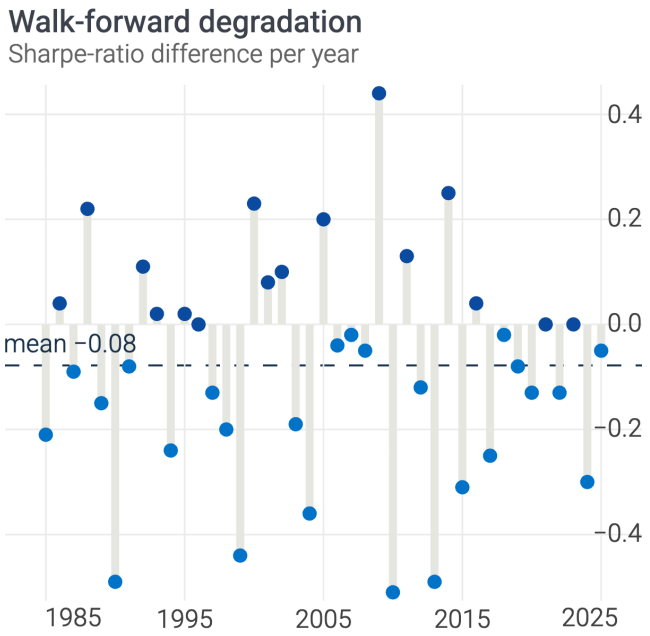


**Figure 14. Rolling refits reveal no calibration problem relative to the full-sample fit.**
*Note:* Yearly refit against the full-sample parameters, 1985–2025.

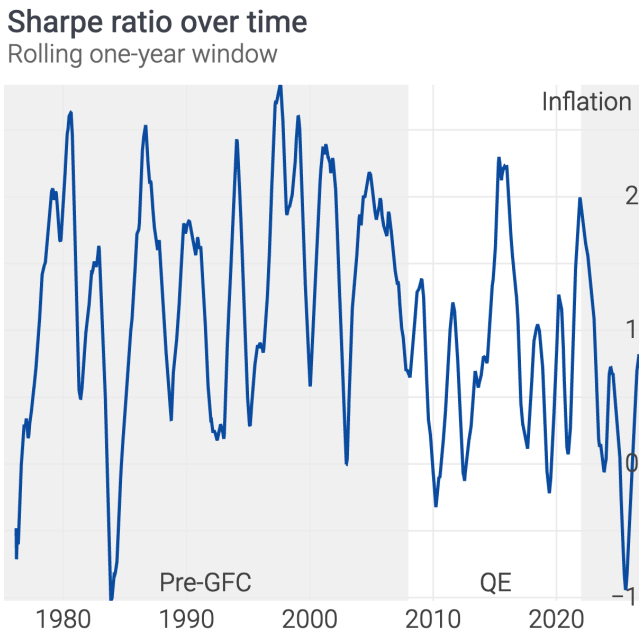


**Figure 15. The rolling Sharpe ratio varies materially over time, with weaker readings in recent years.**
*Note:* Trailing one-year Sharpe ratio, smoothed exponentially.

disproportionately influence aggregate performance. The objective is to determine whether extreme results have a plausible explanation.

Instruments are not removed simply because they performed poorly. The universe is then grouped by liquidity to test whether strategy performance depends systematically on the most or least liquid instruments.

Material differences can indicate implementation constraints, underestimated trading costs, capacity effects, or structural characteristics of particular market segments.

| PERIOD | SHARPE | 95% CI | SHARPE > 0 |
|---|---|---|---|
| 1975–2026 | 1.18 | [0.85, 1.46] | ✓ Yes |
| 2008–2026 | 0.75 | [0.26, 1.26] | ✓ Yes |
| 2022–2026 | 0.45 | [−0.53, 1.49] | × No |

**Table 4. Sharpe ratio robustness across sample periods.**
*Note:* Estimates use daily returns and a stationary block bootstrap (mean block length 20 days; 10,000 simulations).

Execution sensitivity is tested by systematically delaying strategy orders. This approximates implementation failures such as missed fills, temporary infrastructure problems, or delayed order processing. A strategy whose economics deteriorate dramatically after a small delay may require substantially greater execution precision than suggested by a standard end-of-day backtest.

A walk-forward procedure evaluates parameter stability. At each recalibration date, the parameters are chosen using only the information available at the time and are then applied to the following period. This approximates the sequence of decisions that would have been made in real time and provides an indication of the strategy's estimation risk.

Statistical uncertainty is assessed with a stationary block bootstrap that preserves some of the time dependence in the returns. Table 4 compares confidence intervals across the full sample and two more recent periods.

Performance is also examined through time rather than only over the full historical sample.

Rolling statistics help identify changes in the strategy's risk-adjusted return and highlight periods during which its historical behavior differs materially from the long-run average. Such variation should be interpreted in the context of estimation uncertainty and possible changes in market structure.

### KEY TAKEAWAYS

1. All outliers can be explained: they involve instruments with short data histories that performed exceptionally well before 2000 or exceptionally poorly after 2010.
2. Sharpe ratios do not vary systematically with liquidity quintile. This analysis therefore provides no clear basis for selecting instruments by liquidity to improve risk-adjusted performance.
3. Delaying execution reduces performance only slightly, though the effect grows with the lag. The strategy is therefore not very sensitive to execution timing.
4. The walk-forward analysis reveals no calibration problems.
5. For 2022–2026, the estimated Sharpe ratio is positive, but its 95% confidence interval includes zero.

## Portfolio Integration

*Testing whether the strategy adds something different.*

Portfolio integration evaluates the strategy in its intended setting: a portfolio of strategies. Strong stand-alone results do not automatically imply a useful portfolio addition. A strategy with modest returns may still contribute when its gains and losses differ from those of existing strategies. The assessment therefore focuses on its incremental contribution to the combined portfolio.

Return correlations (Figure 16) provide a first indication of overlap with existing strategies. High correlations suggest that the strategy may largely reproduce exposures already present in the portfolio. Lower correlations indicate potential diversification but must be weighed against the strategy's own performance.

Portfolio effects are then evaluated by comparing the portfolio with and without the strategy. The portfolio inclusion tear sheet (Figures 17 and 18, Tables 5 and 6) summarizes changes in returns, volatility, drawdowns, and the shape of the return distribution.

Results should be examined across market regimes as well as over the full sample. A small historical improvement may conceal weaker contributions in recent periods. The final decision therefore weighs the size and consistency of the portfolio benefit against additional trading costs and implementation requirements. Inclusion requires a clear contribution to the portfolio's objectives.

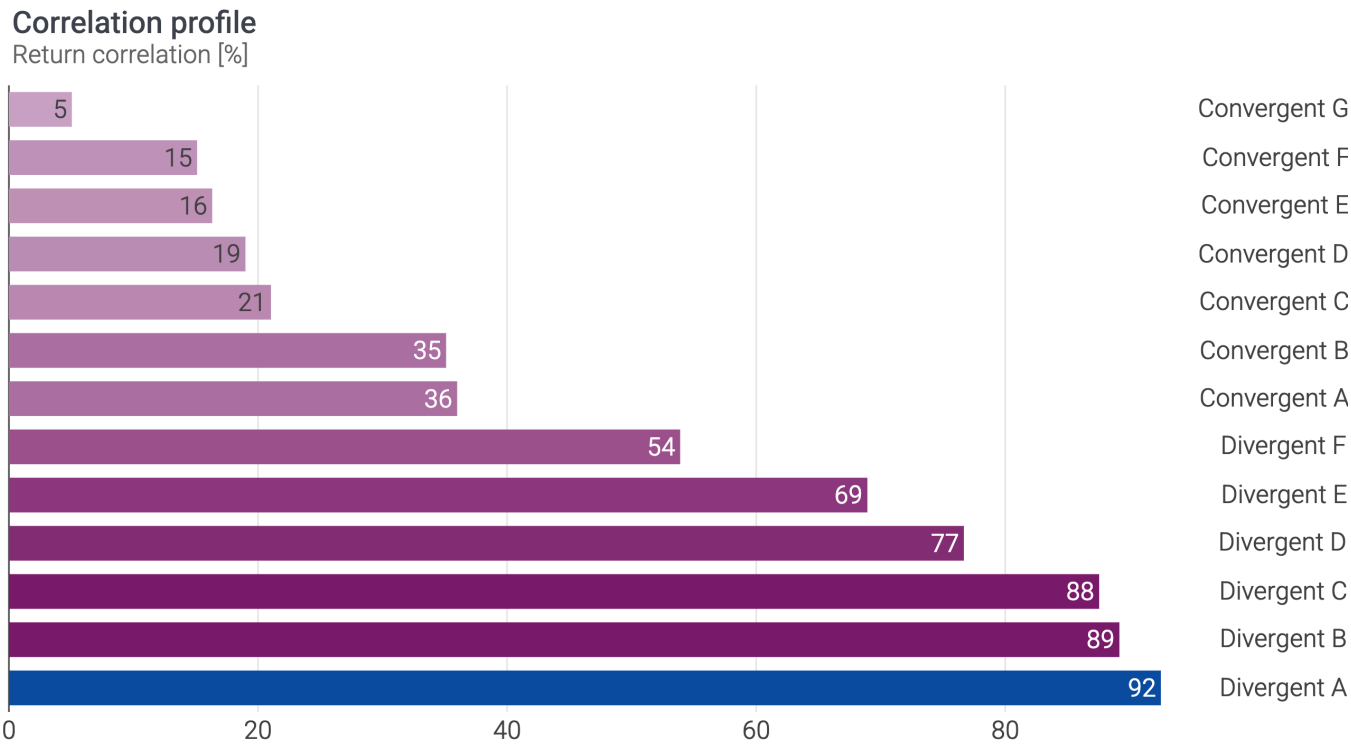


**Figure 16. High overlap with three other trend strategies.**
*Note:* This strategy against each existing strategy, 1975–2026.

## KEY TAKEAWAYS

1 With longer lookbacks, Breakout is highly correlated with three other divergent strategies. Shorter lookbacks provide more diversification but perform poorly in the challenging market regime that began in 2022.

2 The strategy will be retired because it adds no material value to the existing portfolio. Including it reduces performance from 2022 onward and worsens several risk metrics.

3 The strategy may be revived later, but only if the equal weighting of strategies changes: very similar strategies could be grouped to share one weight, and that feature needs additional implementation work first.

# Tear Sheet Portfolio Inclusion

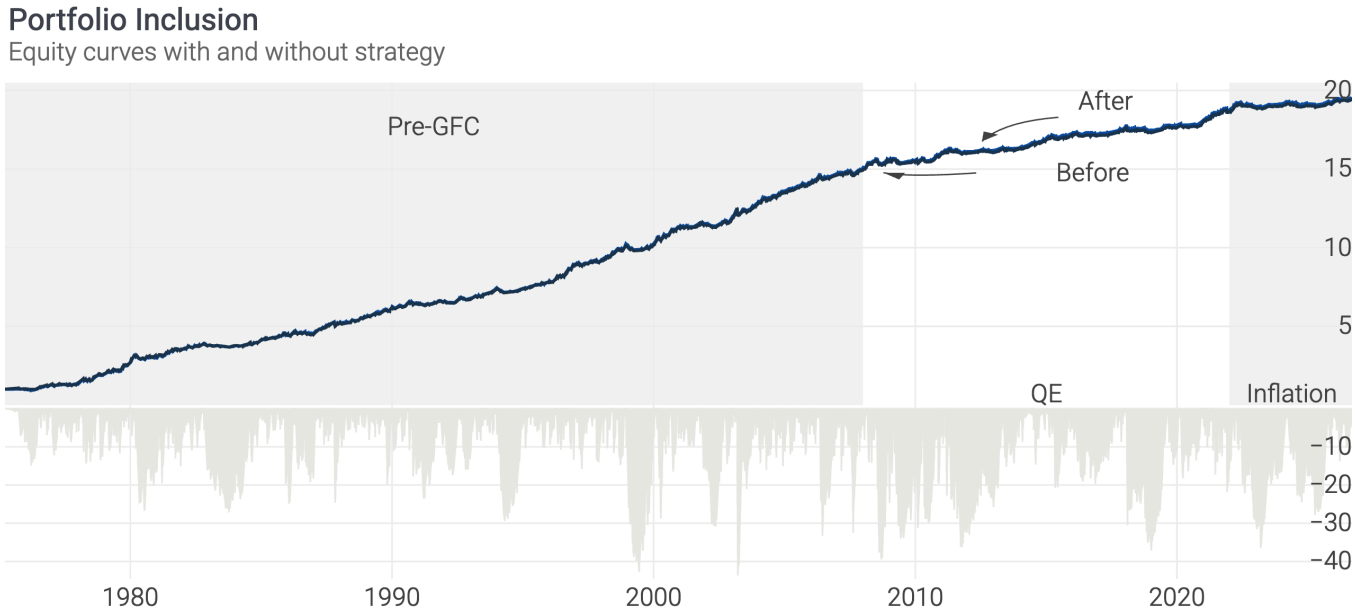


**Figure 17. Equity curves of the portfolio with and without the strategy, 1975–2026.**
*Note:* Calculated using production parameters before dynamic optimization.

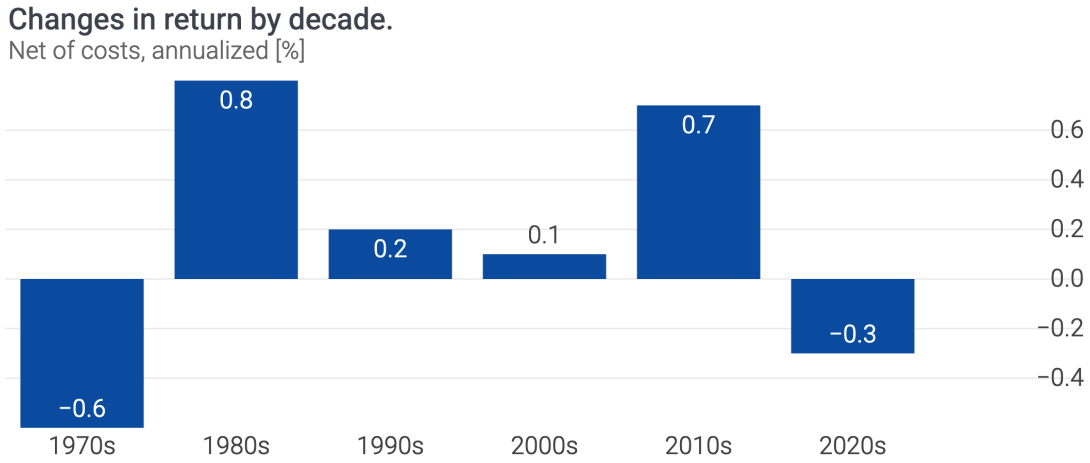


**Figure 18. Change in annualized return by decade after including the strategy.**
*Note:* Each decade evaluated in isolation; the first and last are only partially covered by the sample.

**PERFORMANCE**

**0.3%**
Δ return

**0.0%**
Δ costs

**0.4%**
Δ volatility

**−0.01**
Δ Sharpe ratio

Hypothetical results, net of est. costs; see Appendix A.

| Δ RISK | |
|---|---|
| Maximum drawdown [%] | 0.8 |
| Average drawdown [%] | 0.3 |
| Maximum drawdown duration [days] | 0 |
| Average drawdown duration [days] | −1 |

**Table 5. Change in risk metrics when including this strategy in the portfolio.**
*Note:* All metrics before dynamic portfolio optimization, sample 1975–2026.

| Δ DISTRIBUTION | |
|---|---|
| Monthly skewness | −0.02 |
| Lower tail | 0.01 |
| Upper tail | −0.02 |
| Winning days [%] | −0.1 |

**Table 6. Change in the shape of the return distribution after including the strategy.**
*Note:* All metrics before dynamic portfolio optimization, sample 1975–2026.

# Appendix A: Definitions and Methodology

| | |
|---|---|
| **BACKTESTING CONVENTIONS** | |
| Capital and risk target | Results assume no compounding, an initial normalized capital of 1, and a 25% volatility target. The base currency is USD and all prices of futures contracts are converted to their USD value using the respective end-of-day exchange rate. |
| Costs | All performance figures are stated after estimated trading costs unless a figure or table states otherwise. |
| Sharpe ratio | Annualized from daily returns, after estimated trading costs, without geometric compounding, and assuming a risk-free rate of 0%. No excess-return adjustment is applied, so the figure is a return-to-risk ratio rather than a risk-premium ratio and is not directly comparable to published Sharpe ratios that subtract a cash rate. |
| Drawdown | Maximum and average drawdown are reported as percentages of capital. Drawdown duration is reported in trading days and measures the time from a peak until that peak is regained. |
| Correlations | Calculated from strategy returns after estimated trading costs. |
| Monthly skewness | Skewness of monthly returns, where a month's return is the sum of its daily returns without compounding. Computed as the third standardized moment of the monthly distribution (population form, no sample correction). Positive values indicate occasional large gains against many small losses, negative values the reverse. Monthly rather than daily returns are used because daily skewness is dominated by single-day noise. |
| Lower and upper tail | Tail ratios of daily returns after estimated trading costs, with the mean removed. The lower tail is the 1st percentile divided by the 30th percentile, the upper tail the 99th percentile divided by the 70th, each divided by 4.43, the corresponding ratio of a normal distribution. A value of 1 means the tail is as heavy as a normal distribution's; values above 1 mean fatter tails, values below 1 thinner tails. The ratios are scale-free, so they can be compared across strategies with different volatilities. |
| **DATA AND UNIVERSE** | |
| Instrument | A futures contract on a particular underlying asset. The same asset may be traded as two different instruments — for example two contract sizes on the same underlying, such as mini and micro contracts. |
| Data sources | Data providers for futures data are NorgateData Pty. Ltd. and Commodity Systems Inc. |
| Tradable universe | An instrument is tradable when it satisfies the applicable market-data and liquidity requirements. The universe changes over time as instruments are introduced, discontinued, or become sufficiently liquid to trade. |
| Burn-in | Signal calculations require an initialization period before an instrument becomes eligible to trade. Observations during burn-in are neutral and do not create signals that will be traded on. |
| Liquidity quintiles | Instruments are assigned to quintiles using their most recent 20-day average risk volume in USD. Quintile 1 contains the most liquid instruments. |
| **PARAMETER SELECTION AND VALIDATION** | |
| Strategy family | A strategy family groups strategies with similar behavior. There are currently two families: divergent and convergent. Divergent strategies seek to profit from persistent price movements, while convergent strategies seek to profit from prices reverting toward an equilibrium level. |
| Strategy | Several parameters are combined into an ensemble, which reduces dependence on any one parameter estimate. This ensemble of parameters is called a strategy and also provides some diversification. Each version of the strategy using a particular parameter value (or lookback) is called a strategy variant. |
| Parameter | A parameter refers to a specific lookback period used to calculate a trading signal. |

# Appendix B: Research and Publication Conventions

| | |
|---|---|
| **RESEARCH STATUS AND LIMITATIONS** | |
| Result status | Unless explicitly stated otherwise, the strategy results presented are based on hypothetical historical simulations and do not represent the results of actual trading. Historical simulation cannot fully reproduce the effects of live execution, liquidity, funding, market impact, capacity constraints, operational events, or changing market structure. |
| Interpretation | Historical performance is used as diagnostic evidence within the research process, not as proof that a strategy will remain profitable. Markets are non-stationary, samples are finite, and estimated relationships may change over time. |
| Trading risk | Futures trading may involve leverage and substantial losses. Diversification and volatility targeting may alter the distribution of risk but do not eliminate the possibility of loss. |
| Research purpose | This article is provided for informational and research purposes only. It does not constitute investment, legal, or tax advice, a recommendation, or an offer or solicitation to buy or sell any financial instrument or strategy. |
| **FIGURES, TABLES, AND SOURCES** | |
| Default conventions | Unless a figure or table states otherwise, the definitions, performance conventions, cost assumptions, and calculation methods in the preceding appendix apply throughout the article. Captions therefore state only information needed to interpret the specific exhibit or an exception to the default conventions. |
| Sources and calculations | Unless otherwise stated, market data are from NorgateData Pty. Ltd. and Commodity Systems Inc. Data transformations, strategy simulations, calculations, tables, and figures are produced by Dr. Sebastian Schmidt. References to third-party data providers identify the source and do not imply endorsement. |
| Sample periods | The applicable sample period is stated in the surrounding text, figure, or table where necessary. Effective samples may differ across instruments because of inception dates, data availability, liquidity requirements, and burn-in periods. |
| Rounding | Reported values may be rounded for readability. Calculations and comparisons are based on the underlying unrounded values unless stated otherwise. Apparent differences caused solely by rounding are not economically meaningful. |
| Exceptions | Figure- or table-specific notes override these default conventions where they differ. Methodological exceptions that materially affect interpretation are stated with the relevant exhibit. |
| **PUBLICATION CONTROL** | |
| Document identification | The publication month, issue number, and author shown on the cover identify the article. A materially revised publication should be issued with an updated date, issue, or version identifier so that the research state can be distinguished from earlier releases. |
| Terminology | Defined terms are used consistently throughout the article. Where a term has a strategy-specific meaning, the definition in the preceding appendix governs unless the text explicitly states otherwise. |
| Reproduction and attribution | Third-party data remain subject to the terms of the respective data providers. Reproduction of article content should preserve relevant source attribution and identify Dr. Sebastian Schmidt as the author of its calculations and research. |