

Short-Term Mean Reversion

A mass-experiment framework for buying recent losers, fading winners, and testing whether reversal survives realistic execution, costs, and statistical scrutiny.

SIGNAL FAMILIES

20

UNIQUE SIGNAL SPECS

232

FULL DEEP GRID

13,920

PORTFOLIO MODES

3

RESEARCH PARTNERS

Roberto Cocco & Amin Hydar Ali

Publication standard

The objective is not to find the single highest backtest Sharpe. The objective is to identify whether an economically coherent family remains positive out of sample, across nearby parameters, after turnover, borrow costs, multiple-testing penalties and regime stress.

RESEARCH QUESTION

Can short-term reversal be turned into a robust, investable equity alpha?

The engine tests multiple interpretations of the same broad hypothesis: short-horizon price dislocations may partially reverse over the following sessions.

Short-term mean reversion is not treated as one strategy. The research engine decomposes it into cross-sectional reversal, relative-value signals, residual statistical arbitrage, overnight and intraday anomalies, oscillator-based contrarian signals, price-deviation signals, behavioral effects, liquidity-conditioned reversal, defensive reversal and trend-conditioned pullbacks.

Each signal is evaluated under the same causal execution discipline and portfolio rules, so differences in performance can be attributed more cleanly to the alpha mechanism rather than to inconsistent backtest assumptions.

ENGINE SNAPSHOT

6Y	500	\$20M	\$5
IBKR HISTORY REQUEST	HISTORICAL LIQUIDITY UNIVERSE	MIN TRAILING ADV	MIN STOCK PRICE

Research layer	Default design	Why it matters
Data	IBKR daily OHLCV; frozen source universe preferred	Keeps the research branch reproducible.
Signal timing	Computed at Close T	Prevents future information from entering the rank.
Execution	Portfolio changes at Open T+1	Avoids close-to-close look-ahead assumptions.
Costs	10 bps one-way turnover + 100 bps/year short borrow	Short-horizon alpha is highly cost sensitive.
Validation	2y train / 3m OOS / 10-session embargo	Tests repeated future-like deployment windows.
Governance	252-session diagnostic holdout + later true forward OOS	Separates model selection from independent evaluation.

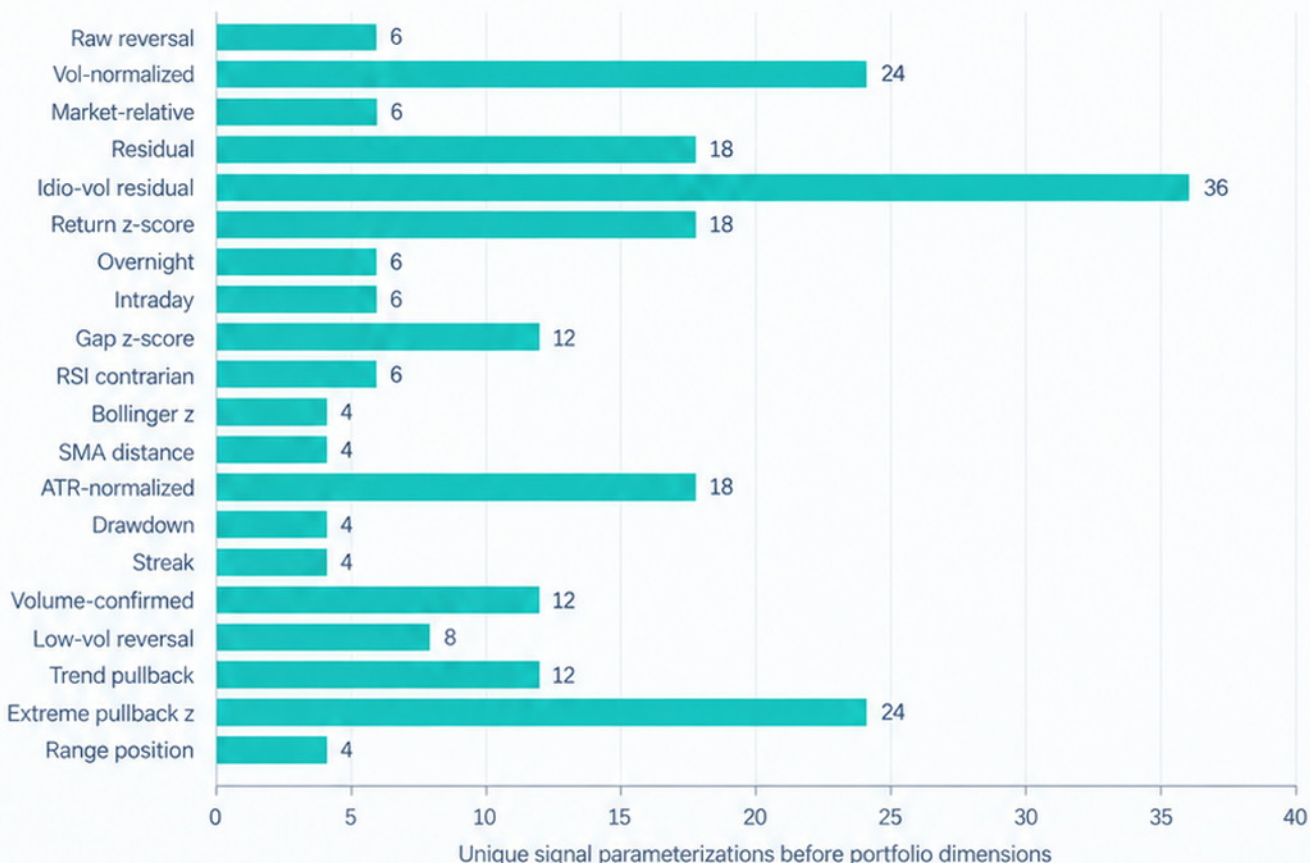
What is intentionally not optimized

The default research does not search for leverage. Gross exposure is fixed at 1.0. This prevents a weak signal from appearing attractive simply because the winning configuration used more risk.

Twenty signal families, twelve alpha groups, one common test harness

The DEEP library contains 232 distinct causal signal definitions before adding portfolio breadth, rebalance frequency and neutralization.

Breadth of the short-term mean-reversion signal library



Those 232 signal definitions are crossed with five Top-N choices, four rebalance cadences and three portfolio modes. The complete DEEP grid therefore contains 13,920 configurations before the optional MAX_EXPERIMENTS cap.



Why classification matters

A robust conclusion should appear at the family level. If only one narrow combination wins while neighboring configurations and related variants fail, the result is more likely to be parameter luck than a durable source of alpha.

Cross-sectional reversal, relative value and statistical arbitrage

Family	Alpha group	Parameter range	Economic question
Raw reversal	CROSS_SECTIONAL_REVERSAL	1, 2, 3, 5, 7, 10d	Ranks stocks by the negative of recent return: recent losers score highest.
Vol-normalized reversal	CROSS_SECTIONAL_REVERSAL	1-10d; vol 10/20/40/60	Scales the recent move by the stock's own realized volatility.
Market-relative reversal	RELATIVE_VALUE	1-10d	Fades stock-specific under/overperformance relative to the cross-sectional median.
Residual reversal	RESIDUAL_STAT_ARB	1-10d; beta 20/60/120	Removes market-beta return before measuring the reversal signal.
Idio-vol residual reversal	RESIDUAL_STAT_ARB	1-10d; residual vol 20/60	Normalizes residual reversal by residual volatility.
Return z-score reversal	TIME_SERIES_EXTREME	1-10d; history 20/60/120	Measures whether the completed recent return is extreme versus its own trailing history.
ATR-normalized reversal	CROSS_SECTIONAL_REVERSAL	1-10d; ATR 14/20/60	Compares the short-term return to current trading range / ATR.

The central statistical-arbitrage question

If residual and idiosyncratic-volatility variants outperform raw reversal, the evidence points toward stock-specific dislocation rather than simply buying whatever fell most with the market.

Overnight, intraday, behavioral and conditioned reversal

Family	Alpha group	Parameter range	What it tests
Overnight reversal	OVERNIGHT_ANOMALY	1-10d	Fades cumulative close-to-open gaps.
Intraday reversal	INTRADAY_ANOMALY	1-10d	Fades cumulative open-to-close moves.
Gap z-score reversal	OVERNIGHT_ANOMALY	1/2/3/5d; 20/60/120 history	Targets abnormal rather than merely large gaps.
RSI contrarian	OSCILLATOR_CONTRARIAN	2/3/5/7/10/14	Tests oversold/overbought behavior through a bounded oscillator.
Bollinger z reversal	PRICE_DEVIATION	5/10/20/40d	Fades standardized distance from the rolling mean.
SMA-distance reversal	PRICE_DEVIATION	5/10/20/40d	Fades percentage displacement from the moving average.
Drawdown reversal	PRICE_DEVIATION	5/10/20/40d	Ranks pullbacks from recent highs.
Streak reversal	BEHAVIORAL_CONTRARIAN	3/5/7/10d	Tests whether runs of up/down days overshoot.
Range-position reversal	PRICE_DEVIATION	5/10/20/40d	Fades position near the extremes of the recent high-low range.
Volume-confirmed reversal	LIQUIDITY_CONDITIONED	1/3/5/10d; vol history 20/60/120	Gives more weight to dislocations accompanied by abnormal volume.
Low-vol reversal	DEFENSIVE_REVERSAL	1/3/5/10d; vol 20/60	Combines reversal with preference for lower-volatility names.
Trend pullback	HYBRID_PULLBACK	1/3/5/10d; trend 63/126/252	Buys pullbacks inside positive trends and fades rallies inside negative trends.
Extreme pullback z	HYBRID_PULLBACK	1/3/5/10d; trend 63/126/252; z 20/60	Combines trend direction with standardized short-term dislocation.

PARAMETER DESIGN

Which dimensions matter - and what each one is actually testing

Parameter variation should map to an economic hypothesis. The engine therefore records every experiment by signal family, horizon, normalization, conditioning, portfolio mode, breadth and rebalance cadence.

Dimension	Values	Research interpretation
Signal lookback	1-10 days for most reversal families; longer for price-deviation families	How quickly does reversal decay?
Auxiliary window	20/60/120 or 63/126/252 depending on family	How much historical context is needed to define an extreme or a trend?
Volatility window	10/14/20/40/60 depending on family	Does risk normalization improve comparability across stocks?
Beta window	20/60/120	Does removing systematic market movement strengthen the signal?
Top-N	5 / 10 / 15 / 20 / 30 per side or long book	Concentration versus diversification.
Rebalance cadence	1 / 2 / 3 / 5 sessions	Signal decay versus transaction-cost burden.
Portfolio mode	Long-only / dollar-neutral L/S / beta-balanced L/S	Directional anomaly versus market-neutral alpha.

THE FACTORS MOST LIKELY TO DETERMINE WHETHER THE ALPHA SURVIVES

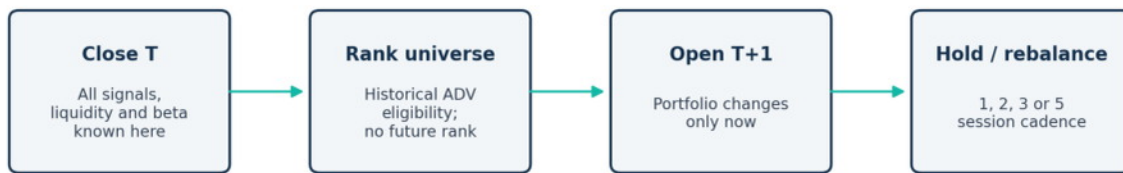
- Horizon: short-term reversal can decay rapidly; a 1-day signal and a 10-day signal are economically different effects.
- Normalization: raw return, realized volatility, ATR, beta residual and residual volatility determine whether the signal is truly extreme after accounting for risk.
- Conditioning: trend, volume and low-volatility tilts test whether reversal is conditional rather than universal.
- Neutralization: dollar or beta balancing determines how much performance is simply equity-market exposure.
- Breadth: Top 5 may capture the most extreme dislocations but can become unstable; Top 30 dilutes alpha but reduces idiosyncratic risk.
- Turnover and borrow: the shorter the horizon, the more implementation assumptions dominate gross alpha.

A good parameter result looks like a plateau

The preferred outcome is not one isolated optimum. A credible result should remain economically useful across adjacent lookbacks, Top-N choices and rebalance frequencies within the same signal family and portfolio mode.

Three implementations of the same signal

Mode	Construction	Interpretation
LONG_ONLY	100% gross long	Buys the highest-scoring recent losers. Useful for testing the anomaly in an implementable cash-equity form.
LONG_SHORT_DOLLAR	~50% long / ~50% short	Long losers and short winners with approximately zero dollar net exposure.
LONG_SHORT_BETA60	100% gross; long/short legs beta-balanced	Uses trailing 60-day betas known at Close T to reduce systematic market beta.



No signal formed with data later than Close T. Execution feasibility at T+1 may filter a name, but never re-rank it.

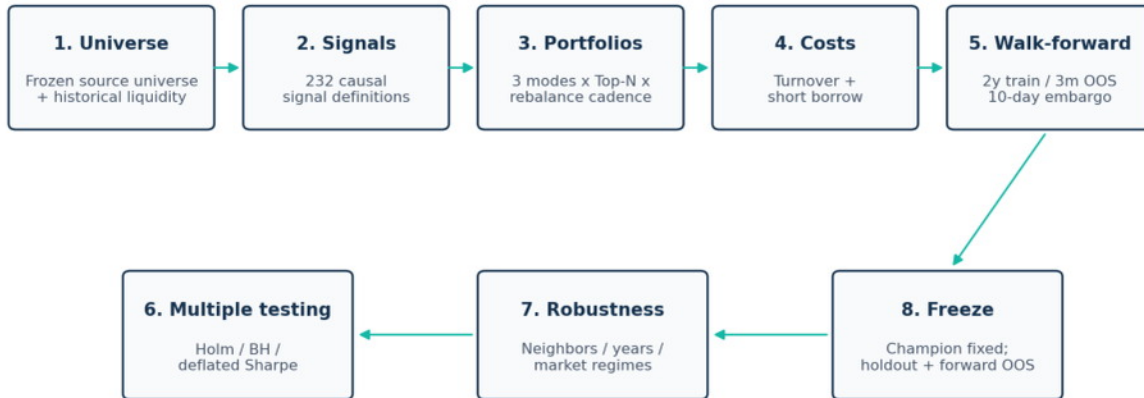
At each signal date the engine first applies historical price, ADV and history requirements, reconstructs the most liquid eligible universe, ranks only on Close-T information, and then executes the selected names at the next available session open.

Execution feasibility is not alpha

The next open can be used to determine whether a selected stock has a valid executable price. It is never used to improve or re-rank the signal. That distinction is essential for avoiding subtle look-ahead bias.

Walk-forward selection is separated from fixed-candidate OOS evidence

Research governance pipeline



FOLD DESIGN

The development sample is divided into repeated walk-forward folds. Each fold uses approximately 504 sessions of training data, a 10-session embargo, and then 63 sessions of out-of-sample testing. The test window advances by 63 sessions.

Inside each training fold, candidates are ranked by a pre-specified composite score. Separately, every candidate is also evaluated on every OOS fold so the final frozen champion is chosen from fixed-candidate OOS evidence rather than only from the adaptive meta-strategy.

Score	Pre-specified weighting
Training selection score	35% Sharpe; 20% CAGR; 15% Sortino; 15% MaxDD; 10% Calmar; 5% lower turnover
Frozen-champion robust score	30% pooled OOS Sharpe; 20% OOS CAGR; 15% OOS Sortino; 15% OOS MaxDD; 10% positive-fold rate; 5% median fold Sharpe; 5% lower turnover

Why two scores?

The training score answers: 'what would have been selected at that historical moment?' The robust score answers: 'which fixed specification behaved most consistently across the OOS evidence we actually observed?'

STATISTICAL DISCIPLINE

Thousands of experiments require stronger evidence, not weaker standards

Searching a large grid increases the probability of finding an impressive backtest by chance. The engine therefore raises the hurdle as the number of tested configurations grows.

Control	Purpose	Interpretation
One-sided t-statistic	Initial diagnostic for positive mean return.	Weak on its own; not treated as sufficient.
Holm adjustment	Family-wise multiple-testing correction.	Controls the chance of at least one false discovery.
Benjamini-Hochberg	False-discovery-rate adjustment.	Useful when many related hypotheses are tested.
Expected max null Sharpe	Raises the benchmark Sharpe as the number of trials increases.	Makes a 10,000-experiment search harder to 'pass'.
Probabilistic / deflated-Sharpe-style probability	Accounts for sample size, skew and kurtosis versus the elevated null benchmark.	Tests whether the observed Sharpe is meaningfully beyond search luck.
Moving-block Reality Check	Bootstraps the maximum t-stat across the entire candidate family.	Preserves some serial dependence through block resampling.

PARAMETER-NEIGHBORHOOD STABILITY

After the champion is selected, nearby configurations are drawn from experiments that were actually tested in the same signal family and portfolio mode. The research asks whether neighboring lookbacks, Top-N choices and rebalance frequencies remain positive and whether their median Sharpe is reasonably close to the champion.

Family evidence > leaderboard rank

For publication, a family with moderate but broad OOS performance is more credible than one spectacular strategy surrounded by failing neighbors.

STRESS AND ROBUSTNESS

A strategy is not credible until the assumptions are made worse

Test	Configuration	Question
Transaction costs	5 / 10 / 15 / 25 / 50 bps per one-way turnover	How quickly does net alpha disappear as execution becomes less favorable?
Short borrow	50 / 100 / 200 / 500 bps per year	Does the short book still add value under materially higher financing cost?
Year-by-year analysis	Development period split by calendar year	Is performance concentrated in one exceptional episode?
Market regimes	Bull/bear and high/low-volatility diagnostics	Does the effect depend on a narrow market state?
Moving-block Monte Carlo	5,000 paths; 10-session blocks	How wide is the distribution of plausible CAGR, Sharpe and drawdown paths?

PRE-SPECIFIED GOVERNANCE GATES

Gate	Required level
Fixed OOS Sharpe	≥ 0.60
Positive OOS fold rate	$\geq 60\%$
Maximum drawdown	not worse than -40%
Positive neighborhood rate	$\geq 60\%$
Deflated-Sharpe-style probability	$\geq 90\%$
Reality Check p-value	$\leq 10\%$
CAGR at 25 bps cost	> 0
Monte Carlo positive-CAGR probability	$\geq 80\%$

Passing is not proof

The gates are governance aids. A strategy can pass and still fail live. Their purpose is to prevent weak, unstable or cost-fragile findings from being promoted as robust research.

How findings should be presented on a quant-research site

The site should publish the research process and the failures, not just the winning equity curve.

Publication block	What the reader should see
1. Research hypothesis	State the behavioral or structural mechanism before showing the result.
2. Experiment taxonomy	Show which families and parameter ranges were tested.
3. OOS leaderboard	Clearly label development OOS metrics; do not mix them with the diagnostic holdout.
4. Family summary	Publish median OOS Sharpe/CAGR, positive rate and best robust score by family.
5. Parameter stability	Show whether nearby configurations confirm or contradict the champion.
6. Cost / borrow stress	Display the break-even point rather than only the base-cost result.
7. Regime behavior	Explain when the strategy tends to work and when it fails.
8. Statistical controls	Report adjusted p-values, deflated-Sharpe-style probability and Reality Check.
9. Holdout	Show only after the development ranking is frozen; label it diagnostic.
10. Forward OOS	Once new data accumulate beyond the freeze date, make this the primary independent evidence.

Recommended headline format

Use conclusions such as: 'Residual reversal showed the broadest OOS stability across 18 signal definitions and three portfolio constructions.' Avoid headlines such as: 'Best strategy returned X%' unless the result is accompanied by the complete validation context.

What this framework can and cannot establish

- The IBKR/current frozen universe is not a true point-in-time survivorship-free security master. Historical liquidity reduces, but does not eliminate, survivorship and membership bias.
- Daily OHLCV bars cannot reproduce intraday path, auction queue priority, market impact, exact historical short-locate availability or every corporate-action microstructure detail.
- The historical diagnostic holdout improves governance but is not equivalent to genuinely unseen future data. Only dates after the stored forward cutoff become true forward OOS.
- Statistical significance does not guarantee economic persistence. Short-horizon anomalies can decay as participants adapt and as implementation costs change.

PLANNED RESEARCH SERIES

Series	Theme	Research focus
Research Note 01	Short-Term Mean Reversion	Cross-sectional, residual, overnight, price-deviation and pullback reversal.
Research Note 02	Cross-Sectional Momentum	Long-horizon ranking, gap filtering, trend quality and risk overlays.
Research Note 03	Pairs & Relative Value	Correlation / cointegration screening, spread z-scores and market-neutral construction.
Research Note 04	Volatility Breakouts	Compression, range expansion, breakout confirmation and execution robustness.
Research Note 05	Multi-Alpha Portfolio	Correlation of OOS alpha streams, risk budgets, diversification and ensemble construction.

Editorial objective

The personal site should function as a transparent research archive: hypothesis -> experiment design -> OOS evidence -> robustness -> limitations -> forward update. Consistency of process across strategies is itself part of the research credibility.

Methodology source

SHORT_TERM_MEAN_REVERSION_MASS_RESEARCH_IBKR.py | Institutional-style research engine | September 2026

Disclosure

This publication documents quantitative research methodology and simulated historical results. It is not investment advice, an offer, a recommendation, or a representation that any strategy will achieve similar results in live trading.