

Validraft

FINAL VALIDATION REPORT

Order-Flow Absorption Execution Review

Synthetic scenario - public report format, not a real client result.

UNIVERSE CME equity-index futures · intraday	IN-SAMPLE 2018-01-02 to 2023-12-29
OUT-OF-SAMPLE 2024-01-02 to 2025-12-31	REPORT TYPE Final Validation Report

BLOCKED

Blocking validation issues remain.

MODERATE RISK

A blocking validation gate failed (Cost realism gate, Promotion gate). The run does not pass minimum validation.

1. Executive summary

Verdict: **BLOCKED**

A blocking validation gate failed (Cost realism gate, Promotion gate). The run does not pass minimum validation.

SHARPE (IS)	CAGR (IS)	TOTAL RETURN (IS)	MAX DRAWDOWN (IS)
1.374	12.40%	101.36%	11.07%
SORTINO (IS)	CALMAR (IS)	CVAR 95% (IS)	POSITION CHANGES (IS)
2.463	1.12	-1.01%	2132
SHARPE (OOS)	TOTAL RETURN (OOS)	MAX DRAWDOWN (OOS)	SHARPE RETENTION
0.487	8.45%	12.50%	35.47%

Over the in-sample window 2018-01-02 to 2023-12-29, Order-Flow Absorption Execution Review on CME equity-index futures · intraday produced a Sharpe ratio of 1.374 and a total return of 101.36%, with a maximum drawdown of 11.07% across 2132 recorded position changes. On unseen out-of-sample data (2024-01-02 to 2025-12-31) it recorded a Sharpe ratio of 0.487 and a total return of 8.45%.

What worked

- The available evidence is summarised below; no standalone positive deployment signal was identified.

Main risks / open items

- Benchmark requires review or documented rationale.
- Oos metrics gate requires review or documented rationale.
- OOS performance / Sharpe retention gate requires review or documented rationale.
- Final untouched holdout is not available in this report.
- Production controls is not available in this report.
- Liquidity / capacity review is not available in this report.
- Tail-risk review is not available in this report.

Performance notes

- Out-of-sample Sharpe retained less than half of the in-sample level – a sign of possible overfitting or regime sensitivity.

Decision implication

Do not treat this as validated. The strategy should be revised or re-tested only after the blocking issues are resolved.

Context notes

- Blocking gates failed: Cost realism gate, Promotion gate
- The synthetic gross signal retained limited OOS evidence.
- Spread, queue and slippage stress pushed net Sharpe below zero.

This report is descriptive analysis of historical data: it documents what the strategy produced under the declared run context and gates in place at generation time. It is not investment advice and not a recommendation to trade.

1.1 Validation status summary

This table summarises the validation areas reviewed, the client-facing status, the importance of each control, and the practical implication for interpretation.

Core validation checks

Control	Status	Importance	Interpretation
Data completeness and integrity	VERIFIED	Required	Control completed; no material issue identified.
Point-in-time universe	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Look-ahead / leakage check	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Costs and slippage realism	BLOCKED	Required	Resolve before any promotion (blocking).
Closed-trade ledger	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Trade contribution concentration	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Corporate actions over holding intervals	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Liquidity and capacity	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Statistical performance gates	VERIFIED	Required	Control completed; no material issue identified.
OOS / walk-forward validation	WARNING	Advisory	Review and document rationale.
Final holdout evidence	NOT ASSESSABLE	Advisory	No final untouched holdout is available.

Robustness & stress checks

Control	Status	Importance	Interpretation
Signal permutation	NOT PERFORMED	Advisory	Signal-level predictive validity has not been tested.
PnL permutation	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Parameter sensitivity	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Monte Carlo path-risk	NOT PERFORMED	Advisory	Monte Carlo path-risk has not been run.
Drawdown-breach / stop-out analysis	NOT PERFORMED	Advisory	Drawdown-breach / stop-out probability has not been simulated.
Tail-risk classification	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Historical scenarios / regime stress	NOT ASSESSABLE	Advisory	No diagnostic is available for this area.
Volatility targeting	NOT APPLICABLE	Advisory	Not used by this strategy.

Execution, benchmark & governance checks

Control	Status	Importance	Interpretation
Execution realism	NOT APPLICABLE	Advisory	Not relevant for this strategy or evidence package.
Passive benchmark comparison	WARNING	Advisory	Review and document rationale.
Single-benchmark alpha/beta	NOT ASSESSABLE	Advisory	Single-benchmark alpha/beta decomposition is not available in this report.
Multifactor attribution	NOT APPLICABLE	Advisory	Not required for this strategy class.
Process integrity	LIMITED	Advisory	Process integrity evidence is retained internally by Validraft.
Validation approval	BLOCKED	Required	Resolve before any promotion (blocking).

Notes on open or limited items:

- Point-in-time integrity: No diagnostic is available for this area.
- Feature leakage / look-ahead: No diagnostic is available for this area.
- Cost realism: Failed (blocking): spread queue and slippage remove net edge.
- Closed-trade evidence: No diagnostic is available for this area.
- Trade contribution concentration: No diagnostic is available for this area.
- Corporate actions: No diagnostic is available for this area.
- Liquidity / capacity: No diagnostic is available for this area.
- Benchmark: Advisory warning: synthetic opportunity cost context only.
- Alpha / beta: Single-benchmark alpha/beta decomposition is not available in this report.

- Multifactor attribution: Not required by policy for this strategy.
- OOS / walk-forward: Advisory warning: gross oos edge is weak.
- Holdout: No final untouched holdout is available.
- Showing 12 of 23 notes.

1.2 Tail-risk classification

Primary tail-risk drivers:

- Max DD duration (IS) = 201 days (MODERATE; MOD $\geq$ 120 · HIGH $\geq$ 365 · EXT $\geq$ 730).

Risk metric	IS	OOS	Full / MC	Worst band	Evidence
Max drawdown	11.07%	12.50%	12.50%	LOW	OOS metrics
Max DD duration	201 days	188 days	281 days	MODERATE	IS metrics
Conditional DD 95%	9.99%	11.80%	10.95%	LOW	OOS metrics
Excess kurtosis	-0.05	-0.17	-0.07	LOW	IS metrics
Return skewness	0.13	0.03	0.10	LOW	OOS metrics

Threshold bands are evaluated by the report policy; primary driver bullets above show the exact threshold that caused the final band.

Risk classification uses the worst available window/metric. When OOS is available, OOS drawdown and tail metrics are part of the classification rather than an IS-only view.

2. Methodology & data assumptions

Strategy	Order-Flow Absorption Execution Review
Instrument / universe	CME equity-index futures · intraday
Timeframe	event
Session	Synthetic declared session
Dataset / provider	Not available
In-sample period	2018-01-02 to 2023-12-29
Instrument provenance	Not declared; no selection source inferred
Development attempt	Not available from governance ledger
Out-of-sample period	2024-01-02 to 2025-12-31
Cost model	Synthetic conservative net-cost scenario; see gate evidence.
Edge type	order-flow absorption and short-horizon mean reversion
Hypothesis	A gross absorption signal remains economically positive after spread, queue-position and adverse-selection costs.

No point-in-time convention is stated in the retained manifest. PIT/leakage control evidence: no dedicated control-gate evidence is available. Declared execution convention: session=Synthetic declared session. Execution-control evidence: no dedicated control-gate evidence is available. Return/cost basis: Synthetic conservative net-cost scenario; see gate evidence.. Where an out-of-sample window is shown, it was held out from any parameter selection.

2.1 Validation windows

Stage	State	Definition
In-sample (IS)	VERIFIED	Used for development and any parameter selection.
Traditional OOS return series	VERIFIED	Held out from parameter selection; used for validation.
Final OOS / lockbox	NOT PERFORMED	Reserved from research and parameter selection; strongest overfitting control.
Walk-forward permutation diagnostic	NOT PERFORMED	Permutation diagnostic; not a replacement for OOS returns.
OOS performance / Sharpe retention gate	WARNING	OOS Sharpe, CAGR, max drawdown and retention gate.

No final OOS / lockbox is available in this report. This report therefore documents historical and (where present) out-of-sample behaviour, but it does not establish production readiness.

2.2 Data and modelling assumptions

This table lists the data and modelling assumptions that materially affect interpretation of the result. Rules, parameters and assumptions are frozen before final validation; reproducibility documentation is retained internally by Validraft.

Assumption	Setting	Why it matters
Data source	Not available	Primary market data source used for the return series.
Universe construction	Synthetic CME equity-index futures events	Defines the instruments eligible for the test.
Calendar / timezone	Not available	Determines trading days, session boundaries and timestamp handling.
Price adjustments / rolls	Not available	Treatment of corporate actions, futures rolls or adjusted prices.
Point-in-time controls	Not available	The retained convention is reported verbatim; missing evidence is not inferred.
Survivorship-bias controls	Not available	Documents whether historical membership and delisted instruments are controlled.
Execution assumption	session=Synthetic declared session	Execution price, timing, fill model and session are reported only when declared.
Costs and slippage	Synthetic conservative net-cost scenario; see gate evidence.	Gross/net status and modeled frictions follow the retained cost evidence.
Benchmark data	Synthetic diversified reference portfolio	Reference series used for passive comparison or alpha/beta analysis.
Audit trail	Retained internally by Validraft	Rules, parameters and assumptions are documented for reproducibility outside the client PDF.

3. Performance measurement

3.1 Returns & risk-adjusted

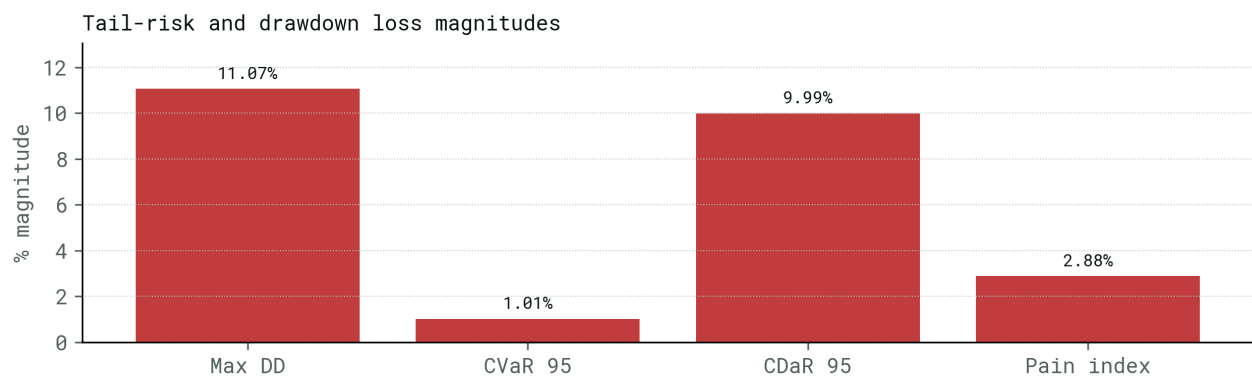
Metric	In-sample	Out-of-sample
Total return	101.36%	8.45%
CAGR	12.40%	4.15%
Sharpe ratio	1.374	0.487
Sortino ratio	2.463	0.83
Calmar ratio	1.12	0.332
Omega ratio ($\theta=0$)	1.242	1.079

3.2 Drawdown & pain

Metric	In-sample	Out-of-sample
Maximum drawdown	11.07%	12.50%
Max DD duration	201 days	188 days
Time to recovery	201 bars	188 bars
Ulcer index	0.0387	0.0611
Pain index	0.0288	0.0454
Conditional drawdown at risk (95%)	9.99%	11.80%

3.3 Return distribution & tail risk

Metric	In-sample	Out-of-sample
Expected Shortfall / CVaR (95%)	-1.01%	-1.10%
Skewness	0.1261	0.0317
Excess kurtosis	-0.0457	-0.1670



3.4 Return observations & position changes

Metric	In-sample	Out-of-sample
Positive return observations (non-zero)	52.37%	50.38%
Negative return observations (non-zero)	47.57%	49.43%
Average positive observation return	0.45%	0.46%
Average negative observation return	-0.40%	-0.44%
Observation profit factor	1.242	1.079

Metric	In-sample	Out-of-sample
Position changes	2132	828
Expected return per position change	0.05%	0.01%
Annualised turnover – unit: gross position units per year; method: sum abs position changes per elapsed year	38	39.52

3.5 Equity & drawdown

Equity and drawdown are shown by analysis window. Drawdown is measured from the running peak inside each window, so the holdout drawdown matches the holdout metrics rather than the earlier in-sample peak.



Out-of-sample read-through: Out-of-sample Sharpe degraded materially (0.487 vs 1.374, retention 35.5%). Out-of-sample CAGR was below in-sample (4.15% vs 12.40%). Maximum drawdown was higher in the out-of-sample window (12.50% vs 11.07%).

4. Validation audit

The historical test was evaluated against 7 validation controls; 2 passed, 3 warned, 2 failed. This section documents the controls applied to the reported backtest. Passing controls support the reliability of the historical result; warnings identify areas that require caution when interpreting it. Required controls are core checks for this validation; advisory controls add diagnostic context.

4.1 Validation control summary

Data integrity

Review area	Status	Importance	Client interpretation
Data quality	VERIFIED	Required	Required control verified.

Statistical validation

Review area	Status	Importance	Client interpretation
Out-of-sample performance	WARNING	Advisory	Review caution: gross oos edge is weak

Benchmark readiness

Review area	Status	Importance	Client interpretation
Benchmark	WARNING	Advisory	Review caution: synthetic opportunity cost context only

Performance, execution & capacity controls

Review area	Status	Importance	Client interpretation
In-sample performance	VERIFIED	Required	Required control verified.
Cost realism	BLOCKED	Required	Required control blocked: spread queue and slippage remove net edge

Other validation controls

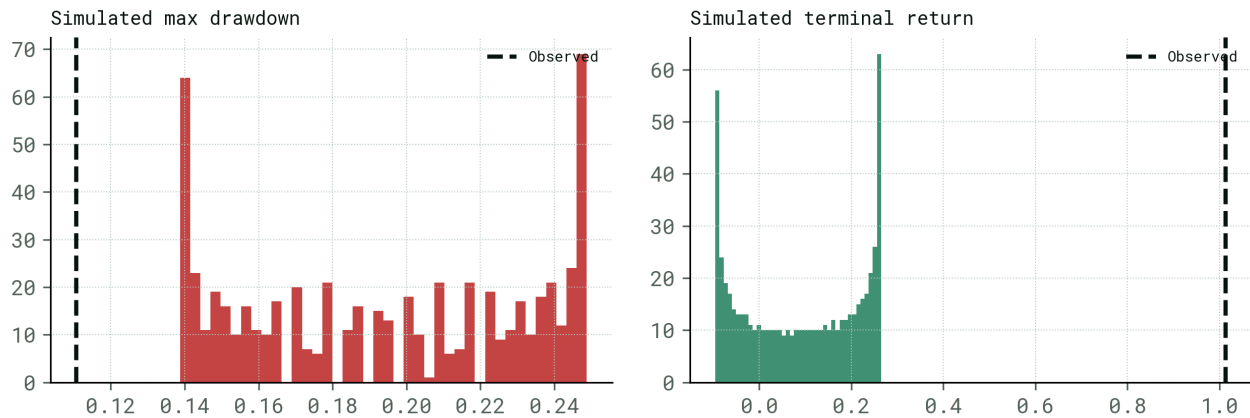
Review area	Status	Importance	Client interpretation
Oos metrics gate	WARNING	Advisory	Review caution: gross oos edge is weak
Promotion gate	BLOCKED	Required	Required control blocked: blocking synthetic validation gates failed

4.2 Signal vs PnL permutation

Two different permutation questions are tested and must not be conflated: PnL permutation asks whether the observed PnL path is rare under a null that scrambles the realized-return path; signal permutation asks whether the signal carries predictive information about future returns by scrambling the signal-to-target link.

Signal-level predictive validity has not been tested. Do not infer a signal edge from PnL behaviour alone.

4.3 Monte Carlo distribution



Distribution of simulated terminal returns and maximum drawdowns. The dashed line marks the observed run result. The empirical result is more credible when it sits inside the bulk of this distribution rather than at an implausible extreme.

4.4 Parameter Sensitivity Review

Field	Value
Status	NOT PERFORMED
Blocking?	no
Required by policy?	yes for configurable or optimized runs; otherwise advisory
Detailed grid	not available
Reason	not run

Parameter sensitivity gate result or detailed heatmaps / parameter stress evidence is not available in this report. No parameter values are invented.

4.5 Path Risk, Monte Carlo and Drawdown-Breach Analysis

This section reports Monte Carlo path risk and the probability of breaching an explicit drawdown/stop-out threshold over the declared horizon. Capital depletion, margin calls and forced liquidation are outside the modeled event.

A. Available path-risk evidence

Metric	Status	Value	Source
Monte Carlo path-risk diagnostic	VERIFIED	canonical summary available	canonical summary
Evidence schema	VERIFIED	validraft_monte_carlo_evidence_v1	canonical evidence
Number of simulations	VERIFIED	600	canonical evidence
Sampling method	VERIFIED	block bootstrap	canonical evidence
Requested sampling method	VERIFIED	block bootstrap	canonical evidence
Sampling level	VERIFIED	bar return stream	canonical evidence
Simulation input	VERIFIED	synthetic net daily return stream	canonical evidence
Sampling parameters	VERIFIED	block size=20.0000	canonical evidence
Median max drawdown	VERIFIED	19.49%	canonical evidence
Expected max drawdown	VERIFIED	19.48%	canonical evidence
90th percentile max drawdown	VERIFIED	24.63%	canonical evidence
95th percentile max drawdown	VERIFIED	24.81%	canonical evidence
99th percentile max drawdown	VERIFIED	24.87%	canonical evidence
Drawdown-breach threshold	VERIFIED	23.25%	canonical evidence
Drawdown-breach / stop-out probability	VERIFIED	25.67%	canonical evidence

Metric	Status	Value	Source
Maximum accepted breach probability	VERIFIED	35.00%	canonical evidence
Simulation leverage	VERIFIED	1.00	canonical evidence
Leverage model	VERIFIED	unlevered synthetic scenario	canonical evidence
Simulation horizon	VERIFIED	252 bars	canonical evidence

5. Benchmarking & alpha/beta

5.1 Benchmark / attribution readiness

Component	Status	Interpretation
Passive benchmark	WARNING	Explicit opportunity-cost benchmark: Synthetic diversified reference portfolio
Single-benchmark alpha/beta	NOT ASSESSABLE	Single-benchmark regression is not available
Multifactor attribution	NOT APPLICABLE	Equity-style multifactor attribution not applicable to this strategy class.
Overall attribution review	LIMITED	Benchmark / attribution evidence is incomplete.

Attribution notes:

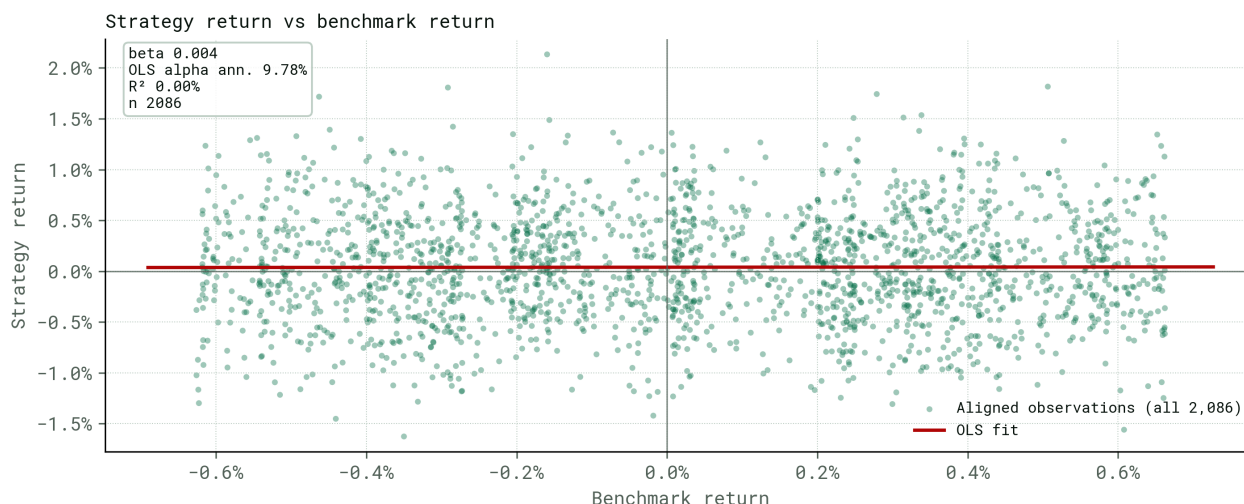
- Passive benchmark: Explicit opportunity-cost benchmark: Synthetic diversified reference portfolio Action: Provide benchmark return series.
- Single-benchmark alpha/beta: Single-benchmark regression is not available Action: Do not treat as multifactor proof.
- Multifactor attribution: Equity-style multifactor attribution not applicable to this strategy class.
- Overall attribution review: Benchmark / attribution evidence is incomplete. Action: Provide missing benchmark and attribution evidence.

5.2 Passive benchmark and alpha/beta

This section isolates benchmark-relative performance, factor attribution, market-neutral residual alpha, and tail dependence against the market. These checks explain whether the strategy is adding idiosyncratic value or merely repackaging common risk exposure.

The passive benchmark is also an economic sanity check: if a strategy does not offer enough compensation for costs, drawdown, operational complexity, and execution risk, a simple passive allocation may be the better opportunity cost.

Benchmark item	Status
Passive opportunity-cost benchmark	Synthetic diversified reference portfolio
Passive benchmark expected	not declared
Benchmark series available	yes – content-addressed identity verified
Formal benchmark gate	diagnostic / not blocking
Factor attribution	not required
Market-neutral claim	no



Each dot is one aligned strategy and benchmark return observation. The fitted line is a single-benchmark OLS view: its slope is beta, and its intercept is the benchmark-relative alpha reported by the gate. This is exposure analysis, not proof of causality.

Review area	Status	Importance	Client interpretation
Benchmark	WARNING	Advisory	Review caution: synthetic opportunity cost context only

5.3 Multifactor attribution

Multifactor attribution asks whether residual alpha survives after correcting for common risk factors. It is distinct from the buy-and-hold benchmark and the alpha/beta decomposition above.

Multifactor attribution status: NOT APPLICABLE

Multifactor attribution is not required by policy for this strategy and was not run.

6. Risk & limitations

- Past simulated performance does not predict future returns; live trading introduces frictions not fully captured by a backtest.
- Results reflect the data-lake snapshot, cost model, and slippage assumptions configured at run time; changing any of these changes the numbers.
- The strategy experienced a maximum drawdown of 11.07% in sample and 12.50% out-of-sample. Position sizing should account for the worse observed drawdown (12.50%) and Monte Carlo path-risk estimates.
- Out-of-sample degradation is material: OOS total return 8.45%, OOS Sharpe 0.487, Sharpe retention 35.47%.
- Tail risk: the conditional VaR at the 95% level is -1.01% per return observation in the worst 5% of observed outcomes.
- Validation thresholds are intentionally institutional; a different gating policy may classify the same run differently.

7. Appendix

Metric notes

These notes summarize how the main performance metrics should be read. Ratios that cannot be computed, such as no losses or no drawdown, are shown as N/A with an explicit reason.

Metric	How to read it	Caveat
Sharpe / Sortino	Return per unit of volatility / downside volatility.	Fragile when the sample is short or tails are extreme.
Calmar	CAGR divided by maximum drawdown.	N/A when the equity path has no drawdown.
CVaR / CDaR	Average loss in the worst return or drawdown tail.	Tail-risk severity, not a normal-period return estimate.
Profit factor	Positive return mass divided by negative return mass.	Uses return observations unless closed-trade PnL is available.

Measurement level

Return-observation metrics are computed from the return series. Trade-level metrics require separate trade or fill evidence with per-trade PnL.

Item	Level	Source	Value / note
Positive return observations	observation-level	returns series	52.37%
Observation profit factor	observation-level	returns series	1.242
Position changes	position-change-level	position series	2132
Expected value per position change	position-change-level	position series	0.05%
Closed-trade PnL metrics	trade-level	not available	Trade-level metrics are not available; observation-level metrics are shown instead.

Metric glossary

Metric	Definition
total_return	Cumulative return over the run (final equity / initial - 1).
cagr	Compound annual growth rate of the equity curve.
sharpe	Annualised mean / stddev of simple net strategy returns (zero risk-free).
sortino	Annualised mean divided by downside deviation (negative-return RMS).
calmar	Full-run CAGR divided by full-run max drawdown.
omega_0	Omega(theta=0) over simple net returns: positive-return mass divided by absolute negative-return mass.

Metric	Definition
cvar_95	Signed Expected Shortfall of simple net returns over the empirical 5% left tail; the tail count uses $\text{ceil}(0.05 * n \dots$
cdar_95	Conditional Drawdown at Risk: average of the empirical 5% largest positive drawdowns; the tail count uses $\text{ceil}(0.0 \dots$
ret_skewness	Bias-corrected sample skewness of simple net strategy returns.
ret_kurtosis_excess	Bias-corrected sample excess kurtosis of simple net strategy returns.
max_drawdown	Largest peak-to-trough drawdown observed in the equity curve, using normalized initial equity 1.0 as the first pea...
max_dd_duration_bars	Bars elapsed during the longest underwater drawdown segment.
win_rate	Legacy alias for observation_win_rate.
profit_factor	Legacy alias for observation_profit_factor.
num_trades	Legacy alias for position_change_count; number of position-state changes.
ev_per_trade_empirical	Legacy alias for ev_per_position_change_return: arithmetic sum of simple net return observations per position-stat...
annualized_turnover	Sum of position changes (or trade notional) annualised.
max_dd_duration_days	Calendar-day length of the longest underwater drawdown segment.
time_to_recovery_bars	Bars between the deepest single trough and the bar that first recovers to the prior peak.
ulcer_index	Root-mean-square drawdown severity across the full equity path, using normalized initial equity 1.0 as the first r...
martin_ratio	Full-run CAGR divided by Ulcer Index.
mar_ratio	Full-run CAGR divided by full-run max drawdown.
gain_to_pain_ratio	Arithmetic sum of simple net returns divided by the absolute sum of negative returns.
burke_ratio	CAGR divided by the square root of summed squared positive drawdowns (unmodified Burke ratio).
pain_index	Average positive drawdown over the full equity path.
equity_curve_r_squared	R^2 of log-equity against a linear time trend over the full equity path.
model_r_squared	Single-factor model R^2 when a benchmark return series is available.
model_adj_r_squared	Adjusted single-factor model R^2 when a benchmark return series is available.

Metric	Definition
n_return_bars	Number of return observations in the equity curve used for performance metrics.
backtest_years	Observed backtest length in elapsed calendar years from ts_event when available; observation-count annualisation i...
min_btl_years	Framework 4.15 approximate minimum backtest length needed to support the observed annualised Sharpe given return s...
min_btl_coverage_ratio	Observed backtest_years divided by min_btl_years.
loss_rate	Fraction of non-zero return observations with negative strategy_ret unless the engine explicitly reports trade-lev...
avg_win	Average positive non-zero return observation unless the engine explicitly reports trade-level PnL.
avg_loss	Average negative non-zero return observation including sign unless the engine explicitly reports trade-level PnL.
observation_win_rate	Positive share of non-zero simple net return observations.
observation_loss_rate	Negative share of non-zero simple net return observations.
observation_profit_factor	Positive strategy_ret mass divided by absolute negative strategy_ret mass.
observation_avg_win	Mean positive simple net return observation.
observation_avg_loss	Mean negative simple net return observation, including its sign.
position_change_count	Number of changes in the position-state series.
ev_per_position_change_return	Arithmetic sum of simple net returns divided by position_change_count.
trade_count	Closed trades represented by an engine-owned trade ledger; null without such a ledger.
trade_win_rate	Positive share of non-zero closed-trade outcomes.
trade_loss_rate	Negative share of non-zero closed-trade outcomes.
trade_profit_factor	Gross positive closed-trade outcome divided by absolute gross negative outcome.
trade_expectancy	Mean net outcome per closed trade; interpret only with trade_metric_unit.
trade_avg_win	Mean positive closed-trade outcome.
trade_avg_loss	Mean negative closed-trade outcome, including its sign.
trade_metric_unit	Unit shared by trade_expectancy/trade_avg_win/trade_avg_loss, such as account_currency.
annualization_factor	Declared observation factor used by annualised return-distribution ratios.

Metric	Definition
annualization_method	Provenance of the annualization_factor; must be explicit for client release.
duration_method	Method used for CAGR, elapsed years and turnover annualisation.
return_type	Return convention used by metrics; v2 engine output is simple_net.
risk_free_rate_annual	Annual risk-free assumption used by risk-adjusted ratios.
metrics_contract_version	Version of the metric semantics contract.
turnover_metric_unit	Unit of annualized_turnover, for example gross position units, capital multiples or portfolio rebalances per year.
turnover_method	Exact numerator and normalization used for annualized_turnover.

Disclaimer

Validraft delivers descriptive analysis of historical data, not investment advice and not a recommendation to trade. Past simulated performance does not guarantee future results.