

VEKTOR
BY
INVESTPUPPY

VEKTOR RESEARCH SERIES · 2026

Evaluating Systematic Strategy Performance

For DPMs, family office analysts, and institutional allocators

IP-WP-PERF-260501-1.0

June 2026

ABSTRACT

We wrote this paper because we expect to be asked how to evaluate a systematic platform with no live track record. The honest answer requires a framework, not a talking point. This is the framework. Systematic strategies are frequently evaluated using the wrong metrics, applied incorrectly, against inappropriate benchmarks. This produces two failure modes: good strategies are rejected because they score poorly on a metric they were not designed to optimise, and poor strategies are adopted because they score well on metrics that do not capture their actual risk. This paper gives a DPM the vocabulary and framework to evaluate any systematic listed equity strategy — and specifically to evaluate the strategy outputs that Vektor produces. The metrics covered here are the same metrics Vektor uses in its optimisation process, which means they are directly applicable to interpreting every output the platform generates.

investpuppy.com · contact@investpuppy.com · IP-WP-PERF-260501-1.0 · Copyright 2026 InvestPuppy

1

THE RIGHT QUESTION

Before choosing a metric, choose the question you are trying to answer.

The most common evaluation mistake is applying a metric without first stating what question it is supposed to answer. A DPM evaluating a systematic strategy — including a strategy produced by Vektor — typically has three distinct questions: Does this strategy generate returns that justify its risk? Does it perform consistently across different market conditions? And will it continue to perform when deployed with real capital?

These are three different questions requiring different metrics. Sharpe ratio addresses the first — and is the same metric Vektor uses as its optimisation objective, making it directly applicable to every strategy output the platform produces. Drawdown analysis addresses the second. Signal decay analysis and orthogonal signal validation address the third. Applying any single metric to all three questions will produce a misleading evaluation.

2

SHARPE RATIO — THE PRIMARY METRIC

What the Sharpe ratio measures, what it does not measure, and how to interpret it for a systematic listed equity strategy.

The Sharpe ratio is the ratio of excess return (above the risk-free rate) to the standard deviation of returns. It answers: how much return does this strategy generate per unit of risk taken? A Sharpe ratio of 1.0 means one unit of excess return for every unit of volatility — a reasonable target for a systematic equity strategy. Vektor uses Sharpe ratio as the objective function in both the efficient frontier optimisation (identifying the

max-Sharpe allocation from 10,000 portfolio configurations) and the technical indicator grid search (selecting the highest-Sharpe indicator configuration per instrument).

INTERPRETING SHARPE RATIOS IN BACKTESTS

Backtest Sharpe ratios are consistently higher than live Sharpe ratios because backtests do not include transaction costs, market impact, or execution slippage in full. A live Sharpe ratio of 50–70% of the backtest Sharpe ratio is a reasonable expectation for a well-constructed systematic strategy. A backtest Sharpe above 2.0 should prompt questions about overfitting.

WHAT THE SHARPE RATIO DOES NOT MEASURE

The Sharpe ratio assumes normally distributed returns. Momentum signal strategies often have skewed distributions — relatively small, frequent gains and occasional larger losses. For strategies with non-normal distributions, the Sortino ratio — which uses downside deviation rather than total standard deviation — is a more appropriate risk-adjusted measure.

3

DRAWDOWN ANALYSIS

Maximum drawdown and drawdown duration — the metrics that reveal how a strategy behaves in adverse conditions.

Maximum drawdown measures the largest peak-to-trough decline in portfolio value over a given period. Drawdown duration — time to recover from a drawdown to a new high — addresses the second dimension: how long did recovery take? A strategy with a 20% maximum drawdown recovering in three months is a very different client experience from one with a 20% drawdown taking three years to recover.

DRAWDOWN IN TRENDING VS MEAN-REVERTING MARKETS

Technical indicator strategies — particularly momentum signals like SMA, EMA, and MACD — tend to generate larger drawdowns in mean-reverting market conditions. Drawdown analysis should be conducted across multiple market regimes, not just the dominant regime in the backtest period.

CALMAR RATIO

The Calmar ratio — annualised return divided by maximum drawdown — combines return and drawdown in a single metric. A Calmar ratio above 1.0 is generally acceptable for a systematic equity strategy. Like the Sharpe ratio, backtest Calmar ratios are typically higher than live ratios.

4

BENCHMARK COMPARISON

How to choose the right benchmark for a systematic multi-instrument strategy — and why a single index is usually the wrong choice.

A benchmark serves one purpose: to establish what return a passive investor would have earned by holding a comparable portfolio. A single-country index like the STI is the correct passive benchmark only if the strategy holds all instruments in proportion to their index weights. A concentrated ten-instrument portfolio selected from a 500-instrument universe will have very different sector and size exposure from the index.

CONSTRUCTING A BLENDED BENCHMARK

The appropriate benchmark for a Vektor strategy is a blended index weighted to match the portfolio's sector composition. This isolates the contribution of instrument selection from sector allocation. A strategy that outperforms on a sector-blended benchmark is generating genuine stock selection alpha.

INFORMATION RATIO

The information ratio — excess return over benchmark divided by tracking error — measures consistency of outperformance. A high information ratio means the strategy outperforms consistently, not just in certain periods.

5

SIGNAL DECAY

The most underappreciated risk in technical indicator strategies — and how to detect it early.

Signal decay occurs when a technical indicator loses its predictive power over time. A momentum indicator reliable during a trending market regime will underperform — and may reverse — during a mean-reverting regime. A signal optimised on three years of data may already be decaying by the time it is deployed.

EARLY WARNING SIGNS

The earliest observable signal of decay is a reduction in XGBoost confidence scores for an instrument's assigned indicator. A second signal: actual portfolio performance diverging from the projected performance implied by the backtest Sharpe ratio.

VEKTOR'S STRUCTURAL RESPONSE

Vektor's primary structural response to signal decay is re-optimisation at each strategy build. Indicator parameters are not fixed at initial setup — they are re-selected by the grid search each time the strategy is rebuilt, using the most recent three years of data. A second structural response is the XGBoost signal validation layer, which filters low-confidence signals before they reach execution. A third layer — news

sentiment analysis aggregated over a rolling window — is on the near-term roadmap. It provides an orthogonal, regime-adaptive signal drawn from unstructured text rather than price history, addressing signal decay from a fundamentally independent data source. See WP-11: Signal Robustness Architecture.

6

THE LIVE TRACK RECORD

The only truly honest evaluation of a systematic strategy is what it does with real capital in live markets.

All backtest metrics are approximations of what a strategy might do. They are constrained by the backtest period's market regime, assumptions about execution, and the optimisation choices that produced the strategy. A live track record removes all of these constraints.

Vektor does not yet have a live track record with real client capital. The platform is operational in pre-track-record mode using IBKR paper trading and the full workflow has been tested and validated. The Proof Partners programme exists specifically to begin building a live track record under real conditions, with partners who understand they are participating in the early stage of platform development.

7

PRACTICAL EVALUATION CHECKLIST

Six questions to ask when evaluating any systematic listed equity strategy — including strategies produced by Vektor.

1. What is the Sharpe ratio, and over what period?

A three-year backtest Sharpe above 1.0 is a reasonable starting point. Ask what the Sharpe looks like over sub-periods — a strategy scoring 2.0 over three years but 0.3 in the most recent year is showing signs of decay.

2. What is the maximum drawdown, and how long did recovery take?

Assess both dimensions. A deep, short drawdown and a shallow, long drawdown may have the same maximum drawdown figure but very different client experience implications.

3. What is the benchmark, and is it appropriate?

The benchmark should match the strategy's sector and geographic exposure, not just the exchange. A sector-blended benchmark is more meaningful than a single index for a concentrated portfolio.

4. What is the correlation between instruments?

A portfolio holding ten instruments with average pairwise correlation above 0.7 is effectively concentrated in fewer positions. Correlation matrix verification is a required step in Vektor's optimisation workflow.

5. Has the strategy been tested out-of-sample?

A backtest optimised on the full history is always overfit to some degree. Walk-forward validation on held-out historical periods is a conventional response to this concern — but it tests historical consistency, not forward predictability. Future prices are not drawn from the same distribution as any historical window. Vektor's response is a multi-layer validation architecture that combines mathematical price-based indicators with an orthogonal text-based sentiment layer. The two layers cannot share the same failure mode — which is the genuine test of signal robustness. See WP-11: Signal Robustness Architecture.

6. What are the transaction costs?

A strategy requiring frequent rebalancing may look strong on a pre-cost Sharpe ratio but weak after costs. Transaction cost modelling is not yet embedded in the optimisation engine — it is applied at the mandate design stage. This is a known gap. Cost-adjusted optimisation is on the near-term roadmap.

For signal optimisation methodology, see WP-02 and WP-04. For capital allocation precision, see WP-03. For risk and limitation disclosure, see WP-09. All documents at investpuppy.com.

investpuppy.com · contact@investpuppy.com · IP-WP-PERF-260501-1.0 · Copyright 2026 InvestPuppy