Data-Driven Risk-Adjusted Performance Analysis

A Structural Complement to Classical Risk-Adjusted Ratios

Dvir Ross
Joint work with supervised students Nehoray Sade and Aviv Asraf

Department of Computer Science
Sami Shamoon College of Engineering

IEEE ADTM Conference, 2026

Talk roadmap

Motivation: why standard risk-adjusted ratios can be regime-sensitive.

Proposed structural score based on exponential price-fit stability.

Experimental design across 100 ETFs, 9 market eras, and 180 configurations.

Main empirical result: mean reversion vs. momentum is the key organizing principle.

Implications, limitations, and next steps.

Motivation

Risk-adjusted ratios such as Sharpe, Sortino, Treynor, Calmar, and Information are widely used to compare investment performance under risk.

However, their rankings may change sharply across crashes, rebounds, choppy markets, and persistent trends.

This raises a data-science question: can structural information extracted from the price path complement classical risk summaries?

Main idea

Instead of measuring only return per unit of risk, measure how consistently price follows a stable exponential compounding path.

Research question and contribution

Research question

In which market regimes does a structural exponential-fit score add value relative to classical risk-adjusted ratios?

The goal is *not* to claim a universally superior metric.

The goal is to identify complementary behavior across regimes.

The central empirical finding is that the proposed score behaves like a **mean-reversion-to-trend** signal, whereas the classical ratios behave like **momentum** signals.

Classical risk-adjusted ratios I

Ratio	Formula	Interpretation
Sharpe	$\text{Sharpe} = \dfrac{R_p - R_f}{\sigma_p}$	Excess return per unit of total volatility
Sortino	$\text{Sortino} = \dfrac{R_p - R_f}{\sigma_d}$	Excess return per unit of downside volatility
Treynor	$\text{Treynor} = \dfrac{R_p - R_f}{\beta_p}$	Excess return per unit of systematic risk

What is systematic risk?

Systematic risk is market-wide, non-diversifiable risk. In Treynor, it is measured by β_p, the portfolio sensitivity to broad market movements.

Classical risk-adjusted ratios II

Ratio	Formula	Interpretation
Information	$\mathrm{IR} = \dfrac{R_p - R_b}{\sigma(R_p - R_b)}$	Consistency of outperformance relative to the benchmark
Calmar	$\mathrm{Calmar} = \dfrac{R_p - R_f}{D_{\max}}$	Excess return relative to maximum drawdown

Notation

R_p = portfolio return, R_f = risk-free rate, R_b = benchmark return

σ_p = total volatility, σ_d = downside deviation, β_p = portfolio beta, $D_{\max}$ = maximum drawdown

Proposed structural score

Exponential fit model

Given training-window prices P_t, fit

$$P_t \approx ae^{bt}$$

by ordinary least squares on $\log P_t$ versus time.

$$\text{diff} \cdot r \cdot b = \text{diff} \cdot r \cdot b$$

where

$$\begin{aligned} \text{diff} &= \text{ExpReg}(\text{last}) - P_{\text{last}}, \\ r &= \text{corr}(\log P_t, t), \\ b &= \text{daily log-growth slope}. \end{aligned}$$

r: trend-fit quality

b: direction and strength of trend

diff: signed deviation from fitted trend

Interpretation

If price is below a well-fit upward trend, then $\text{diff} > 0$, $r > 0$, $b > 0$, hence $\text{diff} \cdot r \cdot b > 0$: the model predicts upward reversion toward trend.

From score to prediction and portfolio

For each asset and each measure, the directional forecast is the sign of the score:

$$\hat{y} = \mathrm{sign}(\text{score}) \in \{\mathrm{UP}, \mathrm{DOWN}, \mathrm{FLAT}\}.$$

Positive score $\rightarrow$ predict price increase over the following test window.

Negative score $\rightarrow$ predict price decrease.

Zero or undefined score $\rightarrow$ non-decisive.

Evaluation protocol

Directional accuracy: fraction of decisive predictions that are correct.

Long/short backtest: rank by score, long top 10, short bottom 10, weights proportional to |score|, with a strict short gate for the new score.

Data and experimental design

Data

- Daily adjusted close prices from Yahoo Finance
- 100 equity ETFs
- Benchmark: S&P 500 for Treynor and Information ratio

Configuration sweep

- 9 market eras
- 5 training windows: 1, 2, 3, 6, 12 months
- 4 test windows: 1, 2, 3, 6 months
- Total: $9 \times 5 \times 4 = 180$ configurations

Design principle

The sweep is intended to expose *regime dependence*, not to optimize a single static parameter setting.

Main result: what is being compared

Six measures

For each configuration, we compute directional accuracy for:

$$\text{new score } (\text{diff} \times r \times b),\ \text{Sharpe},\ \text{Sortino},\ \text{Treynor},\ \text{Calmar},\ \text{Information}.$$

How the comparison is formed

For each market era, we compare the new score to the classical family average:

$$\text{Advantage} = \text{Accuracy(new score)} - \frac{1}{5} \sum_{\text{classical ratio}} \text{Accuracy(ratio)}.$$

"Classical avg" means the mean of Sharpe, Sortino, Treynor, Calmar, and Information.

Each era is averaged over 20 train/test configurations.

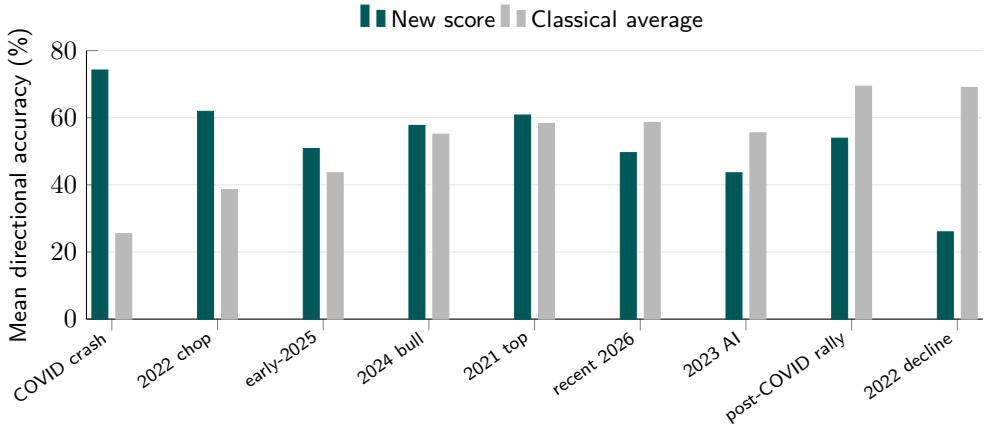
Main result: regime-dependent edge

Market era	New score	Classical avg	Advantage
COVID crash	74.2%	25.5%	+48.7
2022 bear / chop	61.9%	38.6%	+23.3
early-2025	50.8%	43.6%	+7.3
2024 bull	57.7%	55.1%	+2.6
2021 momentum top	60.8%	58.3%	+2.4
recent (2026)	49.6%	58.6%	-9.1
2023 AI recovery	43.6%	55.5%	-12.0
post-COVID rally	53.9%	69.4%	-15.5
2022 bear decline	26.0%	69.0%	-43.0

Positive advantage: the new score outperforms the classical family average.

Negative advantage: the classical family average performs better.

Visual summary of regime behavior



The proposed score is strongest in reversals and choppy regimes.

The classical family is strongest when directional trends persist.

Why Long-Short portfolio return also matters

Directional accuracy vs. economic value

Directional accuracy measures whether a signal predicts the correct sign of the next move.
Long-Short portfolio return measures whether the same signal generates tradable economic value when converted into positions.

For each measure, ETFs are ranked by score on the training window.
The top 10 ETFs form the long basket; the bottom 10 form the short basket.
Within each basket, weights are proportional to |score|, so stronger signals receive more capital.
The strategy is dollar-neutral, with \$1,000 allocated to the long side and \$1,000 to the short side.
Portfolio return is computed as net P&L divided by total deployed capital.

Interpretation

A signal can have high directional accuracy but weak economic impact if it is correct on small moves. Conversely, a signal with lower hit rate can still be profitable if its correct calls occur on larger moves.

Case study A: COVID crash rebound

Test end: 2020-04-15

Training window: 3 months

Test window: 1 month

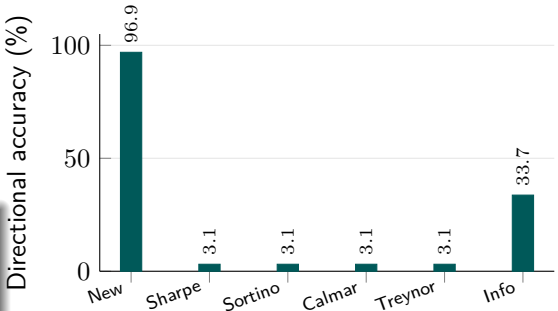
Universe: 98 ETFs

Result

New score accuracy: 96.9%

Classical average: 9.2%

Long/short return: +10.9% vs. -8.8%



Interpretation

After the March 2020 collapse, many ETFs sat far below their exponential trend; the new score correctly captured the rebound, whereas the classical ratios extrapolated the crash.

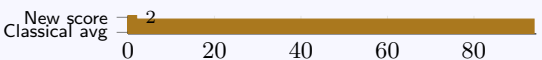
Cases where momentum wins

Case C: post-COVID momentum rally

New score: 2.0%

Classical average: 93.9%

Mechanism: price remained above trend and continued upward.

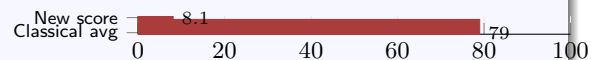


Case D: 2022 sustained decline

New score: 8.1%

Classical average: 79.0%

Mechanism: below-trend names kept falling; reversion caught a falling knife.



Mathematical interpretation

New score family

$$\text{diff} \cdot r \cdot b = \text{diff}\, r\, b$$

- positive when price is below a stable upward trend
- predicts reversion toward the fitted path
- strongest after dislocations and in oscillatory regimes

Classical-ratio family

If a ratio is high after strong recent risk-adjusted returns, then under sign-based prediction it implicitly favors continuation:

high recent quality $\Longrightarrow$ predict continued move.

- this is a momentum-like behavior
- strongest in persistent directional trends

Key organizing principle

The results are not random; they are the signature of a mean-reversion signal competing against momentum.

Why this is relevant to data science

The work turns a financial intuition into a measurable structural feature of time series.

It combines model-based signal construction, large configuration sweeps, and out-of-sample evaluation.

The contribution is not only financial; it is also methodological: identify when different signal families are valid under non-stationary regimes.

Data-science takeaway

A useful predictor may fail globally but still be highly informative conditionally, once the latent regime structure is made explicit.

Limitations

The study uses a static train/test window sweep rather than a sequential walk-forward deployment.

The current evaluation is designed to classify regimes of success and failure, not yet to estimate realistic deployable strategy performance.

The long/short tests do not fully incorporate cumulative trading frictions such as turnover costs or borrow constraints over many cycles.

Future work

Natural extension

Run a walk-forward backtest with sequential re-estimation and rebalancing.

Evaluate whether the regime-dependent edge survives realistic deployment.

Test regime-switching combinations based on realized volatility, trend strength r, or the magnitude of diff.

Measure cumulative P&L, drawdown, turnover, strategy Sharpe ratio, and robustness under transaction costs.

Conclusion

The proposed score $\mathrm{diff} \cdot r \cdot b = \mathrm{diff}\, r\, b$ is best interpreted as a mean-reversion-to-trend signal.

The five classical risk-adjusted ratios behave empirically like momentum signals under the sign-based forecasting framework.

The two signal families are complementary rather than competing.

A regime-aware combination is the most promising path forward.

Thank you

Thank you for your attention.

Questions and discussion