AlgoScore v1.2

A Quantitative Framework for Evaluating
Expert Advisor Track Records

Valeriia Mischenko
EasyAlgos — easyalgos.ai
April 2026

Abstract

This paper presents AlgoScore, a deterministic scoring engine that produces a 0–100 quality rating for Foreign Exchange (FX) Expert Advisor (EA) track records. Developed by EasyAlgos—drawing on over a decade of Expert Advisor development, thousands of live client accounts, and tens of billions of dollars in monthly trading volume—AlgoScore evaluates six risk-adjusted metrics across dual time windows, applies Ljung–Box significance testing, and incorporates Bayesian-inspired confidence shrinkage to produce a single, transparent, fully auditable composite score. The methodology penalizes hidden risks including drawdown persistence, negative tail events, serial return smoothing, and floating loss exposure.

Keywords: risk-adjusted performance, Expert Advisor evaluation, Sortino ratio, Ulcer Index, autocorrelation testing, algorithmic trading

1. Executive Summary

AlgoScore is a deterministic scoring system that evaluates Foreign Exchange Expert Advisor track records on a scale of 0 to 100—analogous to a credit score for trading strategy quality. It was developed by EasyAlgos, drawing on over a decade of Expert Advisor development, thousands of live client accounts, and the processing of tens of billions of dollars in monthly trading volume.

The system analyzes six risk-adjusted metrics across two time windows (all-time and recent 365 days), applies statistical significance testing via the Ljung–Box Q statistic, adjusts for track record length using Bayesian-inspired confidence shrinkage, and outputs a single composite score with full transparency into every intermediate calculation. Unlike return-based rankings, AlgoScore penalizes hidden risks: drawdown persistence, crash-like tail events, artificial equity curve smoothing, and floating loss exposure.

Feature	Description
6 Risk-Adjusted Metrics	Sortino, Calmar, Ulcer Index, Neg. Skewness, Autocorrelation, Float Stress
Dual Time Windows	All-time + Recent (365d), adaptively blended by track length
Statistical Rigor	Ljung–Box significance testing; autocorrelation penalized only when $p < 0.05$
Confidence Adjustment	Bayesian shrinkage toward neutral (50) for short track records
Full Transparency	Every intermediate value published; all parameters externalized

Deterministic	Identical inputs always produce identical scores

2. Background and Motivation

The Expert Advisor market in Foreign Exchange trading presents a fundamental transparency problem. Thousands of Expert Advisors are marketed with impressive return figures, yet these numbers are incomplete. A strategy showing 200% annual returns may carry catastrophic tail risk, artificially smooth its equity curve through correlated position layering, or benefit from a short lucky streak with no statistical significance. EasyAlgos, with over 10 years in the space, thousands of active clients, and dozens of billions in monthly trading volume, developed AlgoScore to address the specific challenges of objective Expert Advisor comparison. The system was designed by Valeriia Mischenko, leveraging deep domain expertise in quantitative trading metrics and the specific data characteristics of Myfxbook reporting.

2.1 Limitations of Conventional Metrics

- Total return ignores risk. A 500% return with 80% drawdown differs fundamentally from 100% return with 15% drawdown.
- Profit factor conflates frequency with quality. One large win yields infinite PF with no statistical evidence of edge.
- Maximum drawdown captures only the single worst episode, ignoring frequency and duration of underwater periods.
- Sharpe ratio penalizes upside volatility equally, inappropriate for trend-following or momentum strategies.

3. Design Principles

3.1 Use Myfxbook time-weighted returns (TWR) directly—deposits/withdrawals already excluded.

3.2 Reconstruct calendar-daily series: fill Myfxbook's missing zero-change days with 0% to avoid upward bias in risk metrics.

3.3 Source max drawdown from Myfxbook Stats panel (includes intraday), never from chart data which excludes intraday drawdown.

3.4 Penalize autocorrelation only when Ljung–Box confirms significance at $\alpha = 0.05$; short tracks may show spurious autocorrelation.

3.5 Fully deterministic: same inputs $\rightarrow$ same outputs. All parameters externalized in versioned configuration files.

4. Data Acquisition and Preprocessing

Performance data is sourced via the Myfxbook API. For each Expert Advisor, the engine retrieves daily return series, account-level statistics (including the authoritative maximum drawdown value), daily floating profit/loss data, and where available, additional trade-level details from Advanced Statistics.

Data is normalized into a canonical model: track identifier, date range, daily percentage changes, Stats-reported max drawdown, and daily floating loss values. Calendar reconstruction then creates a complete date array [start, ..., end], filling any missing days with 0.0% and converting to decimal returns (r = pct/100). This step is critical: without it, risk metrics are systematically biased upward because zero-return periods are excluded from variance and drawdown calculations.

5. The Six Metrics

Metric	Category	Weight	Direction
Sortino Ratio	Performance	30%	Higher = better
Calmar Ratio	Performance	20%	Higher = better
Ulcer Index	Risk	20%	Lower = better
Neg. Skewness	Risk	10%	Lower = better
Autocorrelation	Consistency	10%	Lower = better*
Float Stress	Risk	10%	Lower = better

**Penalty only when Ljung–Box p < 0.05*

5.1 Sortino Ratio (30%)

Risk-adjusted return using only downside deviation (MAR = 0). Unlike Sharpe, does not penalize upside volatility.

```
AR = mean(r_d) × 365.25
DD = sqrt(mean(min(0, r_d)²)) × sqrt(365.25)
Sortino = AR / DD (capped at 10.0 if DD = 0)
```

5.2 Calmar Ratio (20%)

Compound annual growth vs. maximum drawdown. Drawdown always from Myfxbook Stats.

```
CAGR = E_end^(1/Y) - 1 Y = days/365.25
Calmar = CAGR / MaxDD_stats
```

5.3 Ulcer Index (20%)

Measures depth and duration of drawdowns (Peter Martin). Captures sustained underwater pain, not just the worst single episode.

```
Peak_t = max(E_s) for s ≤ t
UI = sqrt(mean((100 × (Peak_t - E_t)/Peak_t)²))
```

5.4 Negative Skewness (10%)

Penalizes left-tail fat: occasional large losses disproportionate to typical returns. Only negative skewness is penalized.

```
NegSkew = max(0, -Skewness_FisherPearson)
```

5.5 Autocorrelation (10%)

Detects serial dependence (position averaging, grid strategies). Uses weekly compounded returns to avoid weekend-zero artifacts. Ljung–Box Q test for joint significance across lags 1–5:

```
r_w = ∏(1 + r_d) - 1 per ISO week
Q = n(n+2) × ∑(ρ_k²/(n-k)) k=1..5
p = 1 - χ²_cdf(Q, df=5)
```

If $p \geq 0.05$: subscore = 100. If $p < 0.05$: AutoPos = sqrt(mean(max(0,ρ_k)²)) mapped to subscore. Minimum 26 weeks required.

5.6 Float Stress (10%)

Measures the root mean square (RMS) of daily floating loss as a percentage of equity, computed from Myfxbook's daily floating P/L data. This metric captures how much unrealized drawdown a strategy holds overnight, directly penalizing strategies that carry large floating losses—a hallmark of grid and martingale systems.

```
FloatStress = sqrt(mean(dailyFloatingLoss²))
```

where each daily floating loss is expressed as a percentage of account equity. Lower values indicate strategies that close positions quickly with minimal unrealized exposure; higher values indicate strategies that routinely hold large underwater positions.

Float Stress replaced the previous Side Balance metric (which measured long/short trade count balance) because it is far more informative: it directly measures hidden risk from unrealized losses rather than merely assessing directional concentration by trade count.

6. Metric-to-Subscore Mapping

Each raw value is mapped to 0–100 via piecewise linear interpolation between configurable anchors (Appendix A). Values outside anchor range are clamped. Anchors calibrated against historical Expert Advisor corpus.

7. Time-Window Architecture

Metrics computed on two windows: All-time (full record) and Recent (last 365 days; equals all-time if shorter). Blended adaptively:

```
w_recent = 0.30 × min(1, days/365) if days ≥ 90; else 0
w_alltime = 1 - w_recent
```

Tracks < 90 days: all-time only. Full year: 70/30 blend. Annualization: 365.25 for all tracks.

8. Confidence Layer and Track Length

8.1 Reliability Shrinkage

Bayesian-inspired shrinkage toward neutral (50). Shorter records pulled more strongly toward 50:

```
Reliability = sqrt(days / (days + 365))
AdjustedBase = 50 + (CombinedBase - 50) × Reliability
```

365d → 71%, 730d → 82%, 2000+d → 92% reliability.

8.2 Length Score (Deprecated)

In v1.2, LengthScore is computed for analytics only and is no longer included in the final score formula.

Independent reward for proven track length with logarithmic diminishing returns:

```
LengthScore = 20 × clamp(1 + log2(days/182.625), 0, 5)
```

Track Length	~6 mo	~1 yr	~2 yr	~4 yr	~8+ yr
Length Score	20	40	60	80	100

9. Final Score Computation

```
FinalScore = AdjustedBase ∈ [0, 100]
```

Integer for UI; float retained internally. The final score is determined entirely by risk-adjusted metric quality, with duration influence limited to the reliability shrinkage in Section 8.1.

10. Score Interpretation

Range	Rating	Interpretation
80–100	Excellent	Strong risk-adjusted returns, low drawdown pain, proven track record.
60–79	Good	Solid performance, acceptable risk. May have 1–2 areas for improvement.
40–59	Fair	Mixed results or insufficient history. Elevated risk or short duration.
0–39	Poor	Significant concerns: excessive drawdown, tail risk, smoothing, or short record.

AlgoScore is a quantitative tool, not investment advice. A high score does not guarantee future performance.

11. Transparency and Auditability

Every output includes: raw metrics, subscores per window, effective weights, blend weights, reliability coefficient, adjusted base, length score (analytics only), final score, and all processing flags. All parameters externalized in versioned config files.

12. Limitations and Future Work

Limitations: Relies on Myfxbook-reported data; broker-level manipulation propagates. Float Stress depends on availability of daily floating P/L data from Myfxbook. Weights/anchors calibrated against current Expert Advisor universe. Scores track records, not strategies—different brokers/leverage may differ.

Future: Trade-level analysis (holding time, single-trade contribution). Regime-aware scoring. Peer normalization. Iterative weight optimization based on predictive accuracy.

Appendix A: Subscore Anchor Tables

Piecewise linear interpolation; clamped outside range.

Sortino (↑ better)

Value	0.0	0.5	1.0	1.5	2.0	3.0	4.0
Score	0	35	55	70	80	92	100

Calmar (↑ better)

Value	0.0	0.5	1.0	1.5	2.0	3.0	5.0
Score	0	35	55	68	78	90	100

Ulcer Index (↓ better)

Value	2	4	6	10	15	25	35
Score	100	90	80	60	40	20	0

Neg. Skew (↓ better)

Value	0.0	0.3	0.6	1.0	1.5	2.0	3.0	4.0
Score	100	90	80	65	45	30	10	0

AutoPos (↓ better)

Value	0.0	0.05	0.1	0.15	0.25	0.35	0.5
Score	100	90	75	60	35	20	0

Float Stress (↓ better)

Value	0.2%	1%	3%	6%	12%	20%	30%
Score	100	85	70	50	30	15	0