VALIDATION REPORT · REF CRU-0492

ES intraday mean-reversion system

Order-level reconstruction & survival assessment

PREPARED FOR	INSTRUMENT	SAMPLE WINDOW	ISSUED
Client #0492	ES — MBO	2019 – 2024	2025-06-12
ANALYST	**REVIEW**	**ENGINE**	**PAGES**
M. Aubert	2nd-pass peer	v4.2.1	1 of 6

VERDICT

GO

CLEARED FOR PAPER

The system clears validation. Edge survives order-level reconstruction with realistic fills; losses do not cluster in shock regimes; the live forward window is consistent with the modelled distribution. Cleared to proceed to paper trading at the stated size, subject to the conditions in §8.

01 — EXECUTIVE SUMMARY

The edge is small, stable, and survives the drawdown.

The strategy is a session-anchored mean-reversion system on the E-mini S&P 500 (ES) front month, trading the regular cash session. Reconstructed against market-by-order data over 2019–2024 (1,842 round-trip trades), it produces a net expectancy of `+2.8 pt` per trade after modelled slippage and commission. The headline figure that matters is not the return — it is that the strategy returns **`1.94×`** its worst peak-to-trough drawdown over the window.

A naïve mid-price backtest of the same rules reports a profit factor near 1.9 and a Sharpe above 2. Most of that disappears under realistic queue position and fill assumptions. What remains is a thin but durable edge whose losses are well-behaved — which is the property that keeps a live account funded.

RETURN / UNIT DD	NET EXPECTANCY	CALIBRATION (ECE)	PARITY
1.94×	+2.8 pt	0.041	Conf.

02 — STRATEGY SPECIFICATION (AS SUPPLIED)

Rules under test

Reproduced verbatim from the client submission. No parameters were re-fitted by Crucible; the system was tested exactly as it would run.

Direction	Long & short, fade session extremes
Entry signal	2.0σ deflection from 30-min VWAP
Exit / target	VWAP reversion or +6 pt
Stop	-10 pt hard, flatten 15:55 ET
Session	09:35 – 15:55 ET, RTH only
Position size	1 contract, max 1 concurrent

03 — DATA & RECONSTRUCTION ASSUMPTIONS

How fills were modelled

Every fill is resolved against the reconstructed limit-order book at the decision timestamp. Marketable orders cross the spread and pay the modelled impact; resting orders advance through the queue only as real volume ahead of them trades.

The order-level ES feed is licensed and supplied by Crucible — the client provides no data and bears no data cost. Any ML component is trained on at most the most recent four years; older history reflects a different microstructure and regime and is excluded as it dilutes more than it informs.

Data feed	CME MBO (every order event)
Clock	Event-time, ns resolution
Queue model	FIFO, position-aware
Latency assumed	9 ms order round-trip
Commission	$2.40 / RT (all-in)
Avg modelled slippage	0.41 pt / RT
Data supplied by	Crucible (client cost: none)
ML training horizon	≤ 4 yr (most recent)

04 — HEADLINE PERFORMANCE

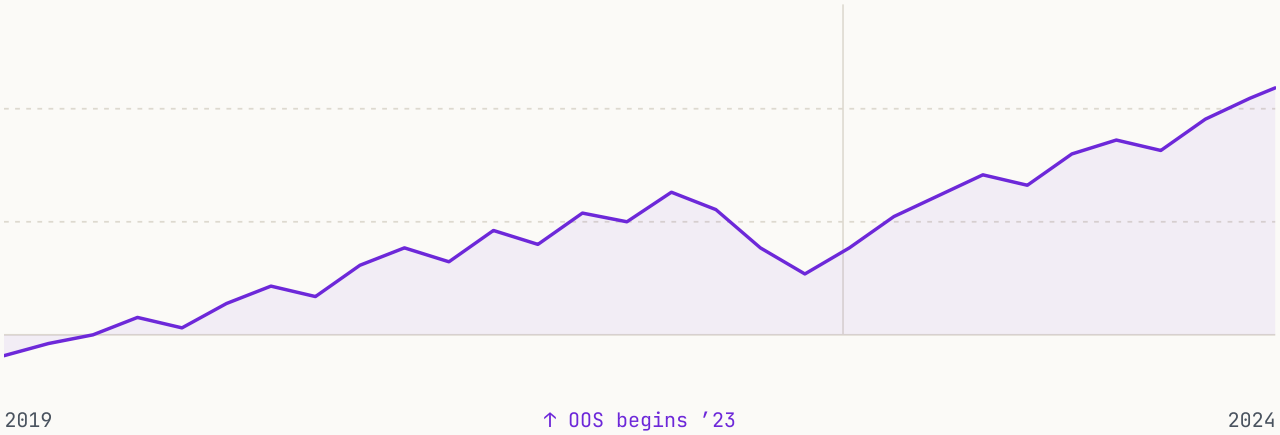
After realistic costs

METRIC	IN-SAMPLE ’19–’22	OOS ’23–’24
Round-trip trades	1,224	618
Win rate	54.9%	53.1%
Net expectancy / trade	+2.9 pt	+2.6 pt
Profit factor	1.39	1.34
Max drawdown	-171 pt	-184 pt
Sharpe (ann.)	1.31	1.21
Return / unit drawdown	2.05×	1.94×

Sharpe shown for reference only. Crucible does not gate on Sharpe — it is sensitive to a handful of fortunate days and silent on tail behaviour.

FIG. 1 — NET CUMULATIVE P&L (PT), AFTER COSTS

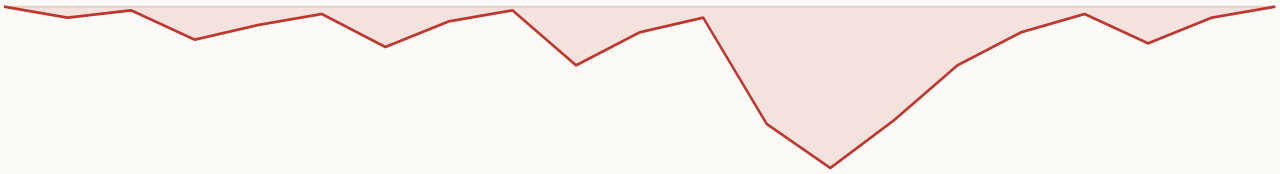
+5,158 pt



05 — DRAWDOWN ANALYSIS

Depth, duration, recovery

FIG. 2 — UNDERWATER CURVE (DRAWDOWN FROM PEAK, PT)



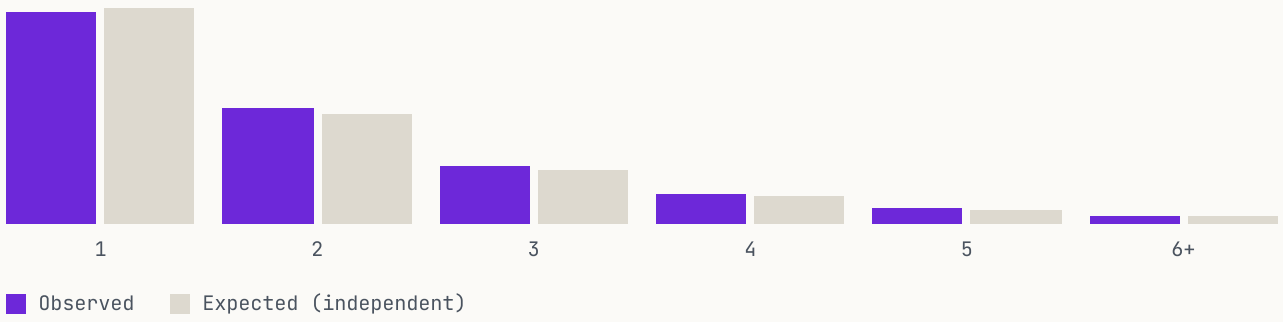
Max peak-to-trough drawdown	-184 pt (≈ -$9,200)
Drawdown duration (deepest)	31 trading days
Time to recover	19 trading days
2nd / 3rd worst drawdown	-129 pt / -97 pt
Return / unit drawdown (window)	**1.94×**

06 — SURVIVAL DIAGNOSTICS

Do the losses behave?

A strategy dies when losses arrive together — a benign average drawdown hides a fatal cluster. We test whether losing trades are independent, and whether the edge holds when volatility spikes.

FIG. 3 — CONSECUTIVE-LOSS RUN LENGTH: OBSERVED VS. INDEPENDENCE

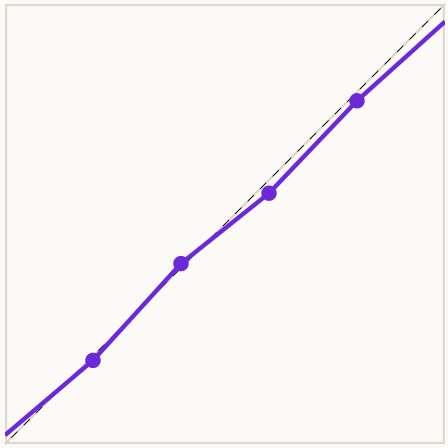


Longest losing streak	7 (expected ≈ 8)
Ljung-Box on trade signs (p)	0.34 — no autocorrelation
Expectancy, VIX > 30 days	+1.9 pt (n = 214)
Shock-regime behaviour	**Losses non-clustered**

07 — CALIBRATION & LIVE PARITY

Does the model tell the truth?

FIG. 4 — RELIABILITY (PREDICTED VS. REALISED WIN PROB.)



ECE = 0.041

LIVE / BACKTEST PARITY

A 62-day forward window was run live on a paper account in parallel with the model. Realised results fall inside the modelled 90% interval on every headline measure.

	MODELLED	LIVE
Expectancy	+2.8	+2.6
Win rate	54%	53%
Slippage / RT	0.41	0.44
Parity		Confirmed

08 — CONDITIONS & LIMITATIONS

What this verdict does not cover

- Verdict is bounded to 1 contract. Edge degrades above ≈ 4 concurrent contracts as queue position and impact rise; re-validate before scaling.
- Tested on RTH only. Overnight / globex behaviour is out of scope.
- Slippage modelled at the observed 9 ms latency. A materially slower execution path will erode the +2.8 pt expectancy.
- Edge is thin. A single missed flatten at 15:55 ET can exceed several days of expectancy — operational discipline is part of the strategy.

VERDICT

GO

Cleared for paper trading at 1 contract, RTH, under the conditions in §8. Recommended review at +3 months or on a 1.5× breach of the −184 pt drawdown, whichever is first.

M. Aubert · Lead analyst

Peer review · 2025-06-12

APPENDIX — GLOSSARY

Return / unit drawdown

Window return divided by worst peak-to-trough loss. How much edge you earn per unit of pain endured.

ECE

Expected calibration error — average gap between the model's stated win probability and the realised rate.

MBO

Market-by-order data — every individual order event, enabling true queue-position modelling.

Parity

Agreement between modelled and live-forward results inside a stated confidence interval.