



Walk-Forward Backtest Harness

The test that refits before every week and scores against the closing line.

A harness that refits on a schedule, prices the week ahead, scores Brier, log loss, calibration and skill against the market, and simulates a staking rule with closing-line value attached. The point of it is to make lying to yourself difficult.

SPORT

Any

FAMILY

Markets and staking

LICENCE

£449

one-off, perpetual

RUNNING IN THE PLATFORM

<https://edgewortharena.com/arena/api/backtests>

Inputs

NAME	TYPE	UNITS	WHERE IT COMES FROM
fixtures	list with results and prices	-	your database
refit_days (optional)	int	days	how often to refit; weekly is the platform's
min_train (optional)	int	matches	before which no price is emitted
staking rule (optional)	callable	-	optional, to simulate returns as well as scores

Outputs

NAME	TYPE	UNITS	NOTES
metrics	dict	Brier, log loss, accuracy, ECE	
market metrics	dict	the same, for the closing line	
brier_skill_vs_market	float	fraction	
betting	dict	bets, staked, profit, ROI, CLV, max drawdown	
predictions	list	-	every priced match, kept so a later rule can be replayed without refitting



Method

- 1 Walk the calendar. At each refit date, fit on everything strictly before it and price the following window.

THIS IS THE SAMPLE

The method runs to 4 steps; one is shown. The assumptions, the limits, the module layout and the reference implementation come with the licence. Everything on the page below this -- what the model was measured at -- is the same in both versions, because that is the part you should not have to take on trust.

What it was measured at

The harness's own output, over every league the platform covers.

MEASURE	RESULT	READING
Matches walked forward	419	
Of those, with a closing price to score against	419	
Refit cadence	weekly	fit on everything strictly before the week
Reported per league	Brier, log loss, accuracy, calibration, skill, ROI, CLV	including when they are negative
Predictions kept	yes	so a staking rule can be replayed later without refitting

Assumptions

What has to be true for the numbers to mean what they say.

- The stored prices are what you could actually have taken. If your feed's opening price lagged the market, the harness will flatter you and the closing-line value column is how you find out.

Limits

Where it is known to be wrong. Published because a model without stated limits is one nobody has pushed on.

- It cannot know about a limit you would have hit, a price you would have been refused, or a market that would have moved as you bet.

LICENCE

Issued to Evaluation copy. One perpetual licence to use and modify this model inside your own organisation. Redistribution of the specification or the reference implementation is not included. No warranty is given: the validation above is the whole of the claim being made.