

THE ASH COLE TRADING SERIES

BOOK 5 — COMPANION 2

Backtest Report Template

with Edge Validation Checklist

Ash Cole

Copyright © 2026 Ash Cole. All rights reserved.

This document is the printable companion to Chapters 9 and 10 of Book 5. Fill it out every time you finish a walk-forward run. Do not proceed to Chapter 11 until every row of the Edge Validation Checklist at the end of this document has a yes beside it. The checklist is the gate, and this document is what you hand yourself — or your accountability partner — at the gate.

How to Use This Template

The template is organized in the same order as the Chapter 9 walk-forward run: metadata first, then the per-fold table, then the Monte Carlo summary, then the degrees-of-freedom accounting from Chapter 10, then the feature stability scores, then the final stopping-rule verdict, and finally the Edge Validation Checklist. Work through it top to bottom. Do not skip sections because a number looks good; the point of the template is that every section has to be filled before the last one is meaningful.

The numbers shown in the example rows below are placeholders from the book's own test runs on EURUSD M5. Your results will differ. A number inside square brackets is a field for you to fill in. A number outside brackets is an illustrative example.

1. Run Metadata

Field	Value
Symbol	[e.g. EURUSD]
Timeframe	[e.g. M5]
Training window total	[e.g. 24 months]
Test window per fold	[e.g. 3 months]
Train/test gap	[e.g. 50 bars]
Number of folds	[e.g. 8]
Model	RandomForestClassifier (300 trees, depth 8, min_leaf 20, class_weight balanced)
Feature count	[e.g. 18]
Run date (UTC)	[YYYY-MM-DD]
Code version / git SHA	[short SHA]
Data vendor	[e.g. MT5 broker XYZ]

2. Fold-by-Fold Walk-Forward Results

Record each fold's out-of-sample metrics in the table below. The "Illustrative" row at the top shows the format; delete it and populate with your own numbers.

Fold	Test period	Trades	Win rate	Sharpe	Max DD
Illustrative	2024-01 to 2024-03	47	54%	1.35	4.2%
1	[]	[]	[]	[]	[]
2	[]	[]	[]	[]	[]
3	[]	[]	[]	[]	[]
4	[]	[]	[]	[]	[]
5	[]	[]	[]	[]	[]
6	[]	[]	[]	[]	[]
7	[]	[]	[]	[]	[]
8	[]	[]	[]	[]	[]
Combined	—	[sum]	[avg]	[combined]	[max]

Combined Sharpe is computed over the concatenated out-of-sample trade series, not as a simple average of fold Sharpes. The fold_report.py module in Companion

1 produces this number directly; do not compute it by hand.

3. Monte Carlo Summary

Run the trade-order permutation simulator from Companion 1 with ten thousand iterations on the concatenated out-of-sample trade returns.

Compounding, not additive. Record the percentile distribution below.

Metric	5th %ile	25th %ile	Median	75th %ile	95th %ile
Final equity (× starting)	[]	[]	[]	[]	[]
Max drawdown %	[]	[]	[]	[]	[]
Longest losing streak (trades)	[]	[]	[]	[]	[]

Interpretation rule of thumb: if the 5th-percentile final equity is below your psychological pain threshold, the bot is too risky to live-trade at the current sizing — even if the median looks attractive. Lower the base risk percentage from Chapter 7 and re-simulate; the shape of the distribution is determined by the edge, but the location on the equity axis is determined by sizing.

4. Degrees-of-Freedom Accounting

Adapted from Chapter 10's DF budget into a fill-in sheet — the costs below weight each intervention by how aggressively it mines the data. Fill in each row with the count used during this run. The running total across the column is your DF spend. Stop when the total exceeds the budget for your data size, and do not refill it with another column.

Intervention	DF cost	This run	Running total
Feature added / removed	1 per feature	[]	[]
Feature definition changed	1 per change	[]	[]
Hyperparameter tuned (model)	1 per value tested	[]	[]
Threshold tuned (confidence, sizing)	1 per value tested	[]	[]
Killzone window redefined	2 per change	[]	[]
Risk layer parameter moved	2 per change	[]	[]
Labeling rule altered	3 per change	[]	[]
Test window or gap redefined	5 per change	[]	[]

A budget of ten — the same ceiling Chapter 10 sets — is reasonable for two years of M5 data across a single symbol. If your running total is above the budget, every additional intervention is almost certainly fitting noise, and the right action is to commit to the current configuration, accept whatever it produces, and stop.

5. Feature Stability Scores

For each feature the model uses, compute the rank of its importance in each fold, then the standard deviation of that rank across folds. The stability score is $1 - (\text{std of rank} / \text{number of features})$, clipped to the interval [0, 1]. A stable feature scores above 0.7. A feature below 0.7 is unstable and should be removed before the next run; an unstable feature is a noise detector in disguise. Add rows as needed — the table should contain one row per feature in your run, up to the twenty-feature ceiling from Chapter 3, and no fewer.

Feature	Mean rank	Std rank	Stability score	Keep?
[feature 1]	[]	[]	[]	[]
[feature 2]	[]	[]	[]	[]
[feature 3]	[]	[]	[]	[]
[feature 4]	[]	[]	[]	[]
[feature 5]	[]	[]	[]	[]
[feature 6]	[]	[]	[]	[]

6. Stopping Rule Verdict

From Chapter 10. The rule has three parts, and all three must hold. This is the evidence bar for moving to Chapter 11's demo deployment.

Criterion	Threshold	This run	Pass?
Combined out-of-sample Sharpe	> 1.2	[]	[]
Folds individually profitable	≥ 75% (e.g. 6 of 8)	[]	[]
Average per-fold Sharpe	> 0.5	[]	[]

If all three boxes are yes, the bot has passed the evidence bar. You stop. You stop even if you believe you could do better. You stop especially if you believe you could do better, because that belief is the instinct that spends degrees of freedom without bound. A bot that passed goes to demo. A bot that failed goes back to feature engineering, not to re-tuning.

7. The Edge Validation Checklist

This is the gate. Every row must have a yes. A single no means the bot is not ready, and the correct response is to go back, fix the failing row, and re-run the pipeline — not to waive the row because the others look good. The checklist exists precisely because the moment of maximum temptation to waive a row is the moment at which waiving it is most dangerous.

- ☐ Combined out-of-sample Sharpe > 1.2 , measured across all folds.
- ☐ At least 75% of folds individually profitable (6 of 8, or 3 of 4 with the fourth at least flat).
- ☐ Average per-fold Sharpe > 0.5 , with no single fold carrying the combined number on its own.
- ☐ Monte Carlo 5th-percentile final equity is above starting equity (the bot is profitable in 95% of permutations).
- ☐ Monte Carlo 95th-percentile maximum drawdown is within the trader's stated pain threshold.
- ☐ No single trade contributes more than 10% of total profit.
- ☐ Win rate is consistent across folds — standard deviation of win rate below 10 percentage points.
- ☐ Maximum drawdown was experienced out-of-sample, not only in-sample.
- ☐ Feature count is at most 20 (simplicity principle from Chapter 3).

- [] All features have stability scores above 0.7.
- [] Total degrees-of-freedom spend is within the budget set before the run began.
- [] Strategy survives a one-week walk-forward shift in the start date without the combined Sharpe collapsing.
- [] Costs modeled at 1.5× broker average for spread, commission, and slippage combined.
- [] Bot performs across at least two distinct market regimes (trending and ranging, or ranging and volatile).
- [] The labeling function has not been modified since the most recent complete walk-forward run.
- [] All tests in the Companion 1 test suite pass on the exact code version used for this run.

8. Go / No-Go Decision

The checklist either has all yeses or it does not. If it does, the verdict below is GO and the bot proceeds to demo deployment under the protocol in Chapter 11. If even one row is no, the verdict is NO-GO and the bot returns to feature engineering or data work — not to re-tuning.

Field	Entry
Verdict	[GO / NO-GO]
Primary reason (one sentence)	[]
Reviewer	[name]
Reviewer signature	[]
Date	[YYYY-MM-DD]

A note on the reviewer field: if you trade alone and have no accountability partner, the reviewer is yourself. Write your own name, sign the sheet, and date it. The act of externalizing the decision onto paper is the point, not the presence of a second person — a solo trader who signs their own verdict has still made the verdict real in a way that an unsigned pass never is. The signature line matters more than it looks. Signing the sheet is the moment at which the trader takes personal responsibility for the decision, and the act of signing changes the decision in a measurable way. Do not skip it.

A backtest that cannot be signed off is a backtest that cannot be traded.