

# Backtesting Validation Template

**From Lesson 24: Backtesting Reality - That Perfect Expectancy Is Lying to You**

Use this template to validate backtest results and avoid curve-fitting disasters.



## The Brutal Truth

**If your backtest looks too good to be true, it is.**

You didn't find the holy grail. You curve-fit to noise, not signal.



## PRE-BACKTEST: Strategy Definition

**Strategy Name:** \_\_\_\_\_

**Core Thesis:**

**What edge am I trying to capture?**

**Why should this work?** (Behavioral edge, structural inefficiency, etc.)

**Entry Rules (Must Be Objective):**

1.

2.

3.

**Any subjective discretion?** Yes / No If Yes, this strategy can't be backtested reliably

---

### Exit Rules:

**Stop Loss:** - \_\_\_\_\_

**Take Profit:** - Target 1: \_\_ (% of position) - Target 2: \_\_ (% of position) - Trailing: \_\_ ( \_\_ % of position)

---

### Position Sizing:

- Risk per trade: \_\_%
  - Max portfolio heat: \_\_%
  - Correlation adjustments: Yes / No
- 

### Filters (Market Conditions):

- ☐ Regime filter (trending/ranging)
  - ☐ Volatility filter (ATR threshold)
  - ☐ Time-of-day filter
  - ☐ Volume filter
  - ☐ Other: \_\_\_\_\_
- 



## BACKTEST EXECUTION

---

### Data Quality Check:

- ☐ **Tick data or 1-minute bars minimum** (not daily bars)
- ☐ **Includes bid-ask spread** (execution costs matter)
- ☐ **Commission + slippage modeled** (\$0.50-\$2.00 per side realistic)

- ☐ **Dividends/splits adjusted**
- ☐ **Survival bias avoided** (includes delisted stocks)

**Data Source:** \_\_\_\_\_

**Date Range:** \_\_ to \_\_\_\_

**Total Trading Days:** \_\_

---

### Results (Raw Numbers):

**Total Trades:** \_\_

**Winning Trades:** \_ **Losing Trades:** \_

**Average Winner:** \$ \_ **Average Loser:** \$ \_ **Average R:** \_\_

**Total P&L:** \$ \_ **Total R:** \_

**Largest Winner:** \$ \_ ( \_ R) **Largest Loser:** \$ \_ ( \_ R)

**Max Drawdown:** \_% (\$ \_ ) **Longest Losing Streak:** \_\_ trades

---

### Performance Metrics:

**Sharpe Ratio:** \_ **Sortino Ratio:** \_ **Profit Factor:** \_ (Gross Profit ÷ Gross Loss) **Expectancy:** \$ \_ (Average profit per trade)



## RED FLAG DETECTOR

### Critical Questions (Be Brutally Honest):

#### 1. Sample Size: Is it statistically significant?

- ☐ **<30 trades:** ❌ Not enough data, pure luck
- ☐ **30-100 trades:** ⚠️ Marginal, needs more data







- [ ] **100-300 trades:**  Decent sample
- [ ] **300+ trades:**  Statistically significant

**My Trade Count:** \_\_

**Verdict:** Pass / Fail / Needs More Data

---

## 2. Sharpe Ratio: Is it realistic?

- [ ] **>4.0:**  Almost certainly overfit
- [ ] **3.0-4.0:**  Suspiciously high, verify
- [ ] **2.0-3.0:**  Excellent but rare, double-check
- [ ] **1.5-2.5:**  Realistic for skilled trader
- [ ] **1.0-1.5:**  Solid, achievable
- [ ] **<1.0:**  Not worth trading

**My Sharpe:** \_\_


**Verdict:** Pass / Fail / Suspicious

---

## 3. Parameter Optimization: Did I curve-fit?

**Did I test multiple parameter values?** Yes / No

**If Yes:** - Total parameters tested: \_ - Parameters selected: \_ - Selection criteria: Best P&L / Best Sharpe / Best DD / Other

 **RED FLAG:** If you tested 50+ parameter combinations and picked the best one, you curve-fit.

**How I selected parameters:**

---

---

**Verdict:** Clean / Suspicious / Overfit

#### 4. Look-Ahead Bias: Am I cheating?

##### Do any rules use future information?

Examples of look-ahead bias: - [ ] Using "swing high/low" (only known AFTER the swing completes) - [ ] Using "end of day" signals (can't trade at unknown close) - [ ] Using repainting indicators - [ ] Using future volatility/ATR that wasn't available at trade time

**My strategy has look-ahead bias:** Yes / No


**If Yes, where:**

**Verdict:** Clean / Biased

#### 5. Out-of-Sample Testing: Did I validate on unseen data?

**In-Sample Period:** \_\_ to \_ **Out-of-Sample Period:** \_\_ to \_\_\_\_

METRIC	IN-SAMPLE	OUT-OF-SAMPLE	DIFFERENCE
Avg R	—	—	—
Profit Factor	—	—	—
Sharpe	—	—	—
Max DD	__%	__%	__%

 **RED FLAG:** If out-of-sample results are >20% worse, strategy is overfit.

**Performance Degradation:** \_\_%

**Verdict:** Pass / Fail

#### 6. Walk-Forward Analysis: Is it robust over time?

**Did you test rolling periods?** Yes / No

**If Yes, results by year/quarter:**

PERIOD	TRADES	AVG R	PROFIT FACTOR	TOTAL R
Q1	—	—	—	—
Q2	—	—	—	—
Q3	—	—	—	—
Q4	—	—	—	—


**Consistency Check:** - [ ] All periods positive R - [ ] No single period dominates results - [ ] Similar expectancy across periods

**Verdict:** Consistent / Inconsistent

## 7. Market Conditions: Did one regime inflate results?

**Performance by Regime:**

REGIME	TRADES	AVG R	PROFIT FACTOR
Trending Up	—	—	—
Trending Down	—	—	—
Ranging	—	—	—
Volatile	—	—	—

 **RED FLAG:** If 80%+ of profit came from one regime that rarely occurs.



**Verdict:** Balanced / Regime-Dependent

## 8. Cost Reality Check: Did I model slippage?

**Modeled Costs Per Trade:** - Commission: \$ \_ - Slippage: \$ \_ - Total: \$ \_

**Realistic Costs (Manual Entry):** - Should be \$1-3 per side minimum

**Did I use realistic costs?** Yes / No

**If I removed slippage, results would:** - Improve by: \_\_%  (unrealistic) - Stay similar   
(robust)

**Verdict:** Realistic / Optimistic





## **VALIDATION SCORECARD**

**Score each category:**

CATEGORY	SCORE	WEIGHT
Sample Size (100+ trades)	__/5	×2 = __
Sharpe Ratio (1.5-2.5 realistic)	__/5	×1 = __
No Curve-Fitting	__/5	×2 = __
No Look-Ahead Bias	__/5	×2 = __
Out-of-Sample Validation	__/5	×2 = __
Walk-Forward Consistency	__/5	×1 = __
Regime Balance	__/5	×1 = __
Realistic Costs	__/5	×1 = __

**Total Score:** \_\_ / 60

### **Score Interpretation:**

**50-60 points:**  Strategy is likely valid, proceed to paper trading **40-49 points:**  Some concerns, address red flags before live trading **30-39 points:**  Multiple issues, significant revision needed **<30 points:**  Overfit or biased, start over with cleaner approach

**My Score:** \_\_

**Decision:** Proceed / Revise / Scrap



## NEXT STEPS

---

### If Score $\geq 50$ (Proceed to Paper Trading):

- ☐ Paper trade for 30-60 trades minimum
- ☐ Track live vs. backtest performance
- ☐ Document any execution challenges
- ☐ Verify slippage assumptions
- ☐ Check emotional difficulty of following rules

**Paper Trading Start Date:** \_\_\_\_\_

**Expected Completion:** \_\_\_\_\_ (after 30+ trades)

---

### If Score 40-49 (Revise):

**Red flags identified:** 1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_

**Revision plan:** 1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_

**Re-test date:** \_\_\_\_\_

---

### If Score $< 40$ (Scrap or Major Overhaul):

**Major issues:** 1. \_\_\_\_\_ 2. \_\_\_\_\_

**Core problem:** Curve-fit / Look-ahead bias / Unrealistic / Sample too small

**Lesson learned:**

---




Next strategy idea:



## Paper Trading Comparison

After 30-60 paper trades:

METRIC	BACKTEST	PAPER TRADING	DIFFERENCE
Trades	—	—	—
Avg R	—	—	—
Profit Factor	—	—	—
Total R	—	—	—
Max DD	__%	__%	__%

 **RED FLAGS in Paper Trading:** - Expectancy drops >20%: Strategy may not translate to live - Avg R drops >20%: Slippage worse than expected - Max DD increases >30%: Risk model inadequate - Can't follow rules: Strategy too complex or emotionally difficult

**Performance Difference:** \_\_%

**Decision:** Go Live / More Paper Trading / Revise / Scrap



## Key Principles

**1. Be Skeptical of Great Results** - If backtest shows >5R expectancy, assume error - If Sharpe >3.0, verify extensively - Best strategies are "good enough," not perfect

- 2. Out-of-Sample Testing is Non-Negotiable** - Always reserve 30% of data for validation  
- Never optimize on full dataset
- 3. Walk-Forward Analysis Reveals Truth** - Test rolling periods - Consistent performance  
= robust - One great year = curve-fit
- 4. Model Reality, Not Fantasy** - Use real slippage (\$1-3 per side) - Include commission -  
Test during different market conditions
- 5. Paper Trade Before Live** - 30-60 trades minimum - Real-time execution reveals issues -  
Psychology can't be backtested

---

**The goal isn't a perfect backtest. It's a robust strategy that works in real time.**

**Be honest. Be skeptical. Be thorough.**

---

© Signal Pilot Education Hub

**Signal Pilot Education Hub**

[signalpilot.io](https://signalpilot.io) | [education.signalpilot.io](https://education.signalpilot.io)

Educational purposes only. This is not financial advice. Always do your own research.