

# One-Shot Strategy Validation

## A Statistical Testing Framework for Event-Driven Trade Hypotheses

Technical Reference Document

### 1. Overview

A "one-shot" strategy is a trade thesis triggered by a single event or causal chain, held for a defined window, and not repeated. Because there is no sample of repeated trades from which to compute a Sharpe ratio or win rate, validation requires a different statistical apparatus: you must test the strength of each link in the causal chain individually, then combine the uncertainty.

**Example:** Wheat prices in Canada correlate with McDonald's stock → weather models predict 60% drought probability → drought raises wheat prices 10% → market prices only a 50% drought probability → short MCD. Each of these four assertions is separately testable.

### 2. Anatomy of a One-Shot Strategy

Every one-shot strategy can be decomposed into a directed causal chain. Each node in the chain has an associated claim that can be tested independently:

Node	Claim Type	Example	Test Method
1	Statistical Relationship	Wheat prices → MCD stock returns	Correlation / Granger causality
2	Probabilistic Forecast	60% drought probability from weather model	Brier score, calibration curves
3	Magnitude Estimate	Drought raises wheat prices by 10%	Regression, historical analogues
4	Market Mispricing	Market implies 50% vs. 60% true probability	Implied prob. extraction, binomial test

### 3. Hypothesis Tests by Node Type

#### 3.1 Statistical Relationship (Node 1)

The core question is whether Asset B responds to movements in Variable A with statistical significance.

### Test: Pearson/Spearman Correlation


$H_0$ :  $\rho = 0$  (no linear relationship between log-returns of wheat futures and MCD stock)

$H_1$ :  $\rho \neq 0$

Use Spearman rank correlation if normality is doubtful. Report the two-tailed p-value and 95% confidence interval for  $\rho$ .

### Test: Granger Causality

If the theory is that wheat prices lead MCD returns (not merely correlate contemporaneously), fit a VAR(p) model and apply the Granger F-test. This checks whether lagged wheat price changes add predictive power over MCD returns beyond MCD's own lags.

 **Spurious Correlation Warning:** Always check whether the relationship survives out-of-sample. Split your data: fit on the first 70%, validate on the remaining 30%. If  $p < 0.05$  in-sample but the coefficient reverses sign out-of-sample, the relationship is likely spurious.

### Minimum Data Requirements

Test	Min. Observations	Notes
Pearson/Spearman correlation	$n \geq 30$	Preferably same-frequency data
Granger causality (VAR)	$n \geq 60$ per variable	Lag order $p$ selected by AIC/BIC
Rolling correlation stability	$n \geq 100$	Use 24-month rolling window

## 3.2 Probabilistic Forecast Validity (Node 2)

The weather model asserts a 60% probability of drought. Before using this probability in a trade, assess whether the forecast model is well-calibrated.

### Brier Score

The Brier Score measures the mean squared error of probabilistic forecasts. For a set of  $N$  historical forecasts  $f_i \in [0,1]$  with outcomes  $o_i \in \{0,1\}$ :

$$BS = (1/N) \cdot \sum (f_i - o_i)^2$$

A perfect model has  $BS = 0$ . A climatological baseline (always predict the base rate) gives a reference BS. Report the Brier Skill Score ( $BSS = 1 - BS/BS_{clim}$ ); a positive BSS indicates the model beats climatology.

### Calibration Curve (Reliability Diagram)

Group historical forecasts into bins (e.g., 0–10%, 10–20%, ..., 90–100%). For each bin, plot the mean forecast probability against the observed frequency. A well-calibrated model lies on the diagonal. If forecasts of ~60% have historically resulted in drought only 45% of the time, you must adjust the 60% figure downward before using it in the strategy.

**Practical Step:** Request calibration statistics from the weather provider, or back-test the specific model over at least 10 years of drought events. A minimum of 20–30 events in the relevant probability bin is required for reliable calibration estimates.

### 3.3 Magnitude Estimate (Node 3)

The strategy assumes drought raises wheat prices by approximately 10%. This must be supported empirically.

#### Regression-Based Estimate

**Model:**  $\Delta \text{Wheat\_Price} = \alpha + \beta \cdot \text{Drought\_Severity} + \varepsilon$

Estimate  $\beta$  with OLS on historical drought episodes. Report the 95% confidence interval for  $\beta$ . If the lower bound of the CI includes 0, the 10% assumption is not well-supported.

#### Historical Analogue Method

Identify past events of comparable drought severity and geography. Compute the interquartile range of resulting wheat price changes. The 10% estimate should fall within this range; if it sits above the 75th percentile, it is optimistic and the expected value calculation must be stress-tested.

### 3.4 Market Mispricing (Node 4)

This is often the most testable node. You assert the market prices a 50% drought probability while the true probability is 60%.

#### Extracting Implied Probabilities

Implied probabilities can be extracted from options markets (via put/call pricing on MCD or agricultural futures), prediction markets (if available), or analyst consensus surveys. The extraction method must be documented and its assumptions stated explicitly.

#### Binomial Test for Edge

$H_0$ :  $p_{\text{market}} = p_{\text{true}} = 0.50$  (no edge)

$H_1$ :  $p_{\text{true}} > p_{\text{market}}$  (strategy has positive expected value)

If you have a track record of similar forecast comparisons, you can run a one-sample binomial test or proportion z-test. For a single event, the edge is asserted, not tested; the test applies to your model's calibration history (Node 2).

### Expected Value Calculation

Scenario	Prob. (True)	Prob. (Market)	Trade P&L
Drought occurs, MCD falls	60%	50%	+X
No drought, MCD rises	40%	50%	-X

$EV = 0.60 \cdot (+X) + 0.40 \cdot (-X) = 0.20X \rightarrow$  Positive edge of 0.20X per dollar risked, assuming the magnitude and correlation nodes hold.

## 4. Combining Uncertainty Across Nodes

Each node introduces uncertainty. A valid strategy must have a positive expected value even after accounting for compounded uncertainty across all nodes. The recommended approach is a Monte Carlo simulation over the joint distribution of node parameters.

### Algorithm

1. Fit distributions to each node's estimate and its uncertainty: e.g.,  $p \sim \text{Normal}(0.35, 0.12)$ ,  $\beta_{\text{wheat}} \sim \text{Normal}(0.10, 0.04)$ ,  $p_{\text{true}} \sim \text{Normal}(0.60, 0.05)$ .
2. Draw  $N = 10,000$  samples from each distribution simultaneously.
3. For each sample, compute the trade's expected P&L using the payoff formula.
4. Report: mean EV, 5th–95th percentile range, and the fraction of simulations in which  $EV > 0$  (the "probability of positive edge").

**Decision Rule:** Proceed with the strategy only if (a) mean EV > transaction costs, and (b) the probability of positive edge exceeds your pre-specified threshold (commonly 75–80% for single-event trades).

## 5. User Input Template

To apply this framework, the analyst must supply the following data for each node. Incomplete inputs should trigger explicit uncertainty widening in the Monte Carlo.

Node	Required Input	Format / Example
------	----------------	------------------

Causal relationship	Historical time series of both variables	CSV: date, wheat_price, MCD_return (monthly, ≥5 yrs)
Forecast probability	Model output + historical calibration data	p=0.60; calibration: 15/25 ≈ 0.60 in [55–65%] bin
Magnitude estimate	Historical episodes + estimated change	Mean: +10%, Std: ±4%, based on 8 drought years
Market pricing	Implied probability extraction method + value	Options-derived: 50%; method: put-call parity
Trade structure	Entry, exit, position size, stop-loss	Short MCD at entry, cover at event resolution

## 6. Strategy Validity Criteria Summary

A one-shot strategy is considered statistically valid if all of the following conditions are met:

#	Criterion	Threshold	Status Field
1	Causal relationship is statistically significant	$p < 0.05$ , out-of-sample $R^2 > 0$	PASS / FAIL
2	Relationship is stable across sub-periods	Rolling $p$ does not change sign	PASS / FAIL
3	Forecast model is well-calibrated	BSS > 0, calibration curve near diagonal	PASS / FAIL
4	Magnitude estimate CI excludes zero	Lower 95% CI bound > 0	PASS / FAIL
5	Market-implied probability is measurably lower	$\Delta p \geq 5$ percentage points, documented method	PASS / FAIL
6	Monte Carlo EV positive at 5th percentile	EV(5th pct) > transaction costs	PASS / FAIL
7	Probability of positive edge exceeds threshold	$P(\text{EV} > 0) \geq 75\%$	PASS / FAIL

## 7. Limitations and Known Failure Modes

### Correlation Without Causation

A statistically significant correlation between wheat prices and MCD stock does not guarantee that a drought in Canada will move MCD stock in the expected direction, especially if supply

chains have changed, hedging policies have shifted, or the correlation was partially driven by a common macro factor.

### Model Uncertainty in Probability Estimates

Weather models provide ensemble probabilities, not ground truth. If the model's calibration history is short or covers a different climate regime, the stated 60% figure carries substantial model risk that cannot be fully quantified.

### Simultaneous Resolution Risk

If many market participants have access to the same weather model data, the mispricing (50% vs. 60%) may close before the event resolves, eliminating the edge without requiring the underlying hypothesis to be wrong.

### Single-Event Inference

The fundamental limitation of one-shot strategies is that ex-post validation is impossible without repetition. A loss does not falsify the thesis (it may be the 40% scenario), and a win does not confirm it. The statistical work described in this document validates the priors and causal chain, not the outcome.

## 8. Recommended Statistical Tests — Quick Reference

Test	Python Function	R Function	Key Output
Spearman correlation	<code>scipy.stats.spearmanr()</code>	<code>cor.test(method='spearman')</code>	$\rho$ , p-value, 95% CI
Granger causality	<code>statsmodels.tsa.stattools.grangercausalitytests()</code>	<code>vars::causality()</code>	F-stat, p-value
Brier score	<code>sklearn.metrics.brier_score_loss()</code>	<code>DescTools::BrierScore()</code>	BS, BSS
OLS regression	<code>statsmodels.OLS().fit()</code>	<code>lm()</code>	$\beta$ , SE, 95% CI
Binomial test	<code>scipy.stats.binomtest()</code>	<code>binom.test()</code>	p-value, CI for p
Monte Carlo EV	<code>numpy.random</code> + custom loop	<code>replicate()</code> + custom	Mean EV, percentiles

*End of Document*