NIST Standard Simulated Dataset v1.1.0

Methodology for Assessing Reliability and Validity (Reproducible Version)

Data version: v1.1.0 | Seed: 20251211 | Build time (UTC): 2025-12-09T18:51:00Z

I. Purpose and Scope

This methodology document standardises and reproduces experiments to evaluate the reliability (consistency, stability, dependability) and validity (accuracy, authenticity, objectivity) of the NIST standard simulation dataset (v1.1.0) for 'Grey Zone Hybrid Warfare System Modelling'. The methodology adheres to the principle of the smallest reproducible unit (script + parameters + fingerprint), enabling direct application in paper appendices and peer review verification.

II. Data and Version Information

• Dataset: v1.1.0 (includes version field, seed, provenance fingerprint)

• Key files: market_event_study.csv, pdl_events.csv, drloop_metrics.csv, provenance.json, params.json

• Time window: $t_0 \pm \{-5, -3, -1, 0, +1, +3, +5\}$ (event studies)

• SFR Rounds: 1..6 (SFR > 1 from Round 3 onwards)

III. Reliability Assessment Methodology

(1) File Integrity and Reproducibility

Verify provenance.json contains SHA-256 fingerprints for all CSV files and check field completeness. Cross-reference parameters (seed, t_0, event window, etc.) with params.json to ensure reproducibility.

(2) Internal Consistency (Target Sectors, Cross-Event Windows)

Under S1 (sector valuation shock), calculate the pairwise correlation coefficient r for abnormal returns (AR) of two stocks per target sector at $t_0 \pm$event window. Higher r

values (with sufficient n_dates) indicate consistent direction and magnitude of responses to events within the same sector, reflecting structural shocks rather than random noise.

(3) Key Series Stability (Sovereign Level)

Under S2 (sovereign credit shock), verify that sovereign CDS values exhibit strict monotonic upward movement at t_0, t_0+1, and t_0+3 to satisfy the mechanism constraint 'credit kill speed $\lambda>0$'.

IV. Validity Assessment Methods

(1) Construct Validity H1: Target vs Non-Target Industries (t_0)

On day t_0 of S1, compare the abnormal returns (AR) between the target and non-target industries using Welch's t-test (for unequal variances) and Cohen's d effect size; employ bootstrapping (N=5000, fixed seed) to provide 95% confidence intervals for the mean difference (Target−NonTarget) and Cohen's d.

(2) Construct validity H2: Credit kill speed λ

Linear slope fitting is performed on sovereign CDS sequences at t_0, t_0+1, and t_0+3 to derive λ (unit: bps/day). It is anticipated that $\lambda>0$ in S2 and $|\lambda_S2|>|\lambda_S1|$.

(3) Mechanism/content validity: Transmission visibility × Interdependence and Industry shock

Verify the composite propagation/coupling indicator (Visibility × Interdependence) from PDL against Pearson correlation with the industry average AR on t_0. Anticipate negative correlation (higher weights, greater declines).

(4) Normative Validity (Stylised Facts)

The target industry should exhibit a significant negative AR near t_0; sovereign-level CDS for S2 should sustain upward pressure post-t_0.

(5) Placebo and Robustness Tests

Select pseudo t_0 dates (e.g., t_0-5) and replicate H1 comparisons. Non-significance is expected to control false positive risks from 'significance everywhere,' confirming main effects concentrate at true t_0.

V. Statistical Implementation and Parameters

• Welch t-test, Cohen's d effect size

• Bootstrap: N=5000, 95% confidence level seed fixed (adjustable in script)

• Correlation analysis: Pearson correlation between industry-aggregated AR and (visibility × interdependence)

• λ estimation: Least squares slope of sovereign CDS at t_0, t_0+1, t_0+3

VI. Reproducibility Steps (Code and Output)

Execute script assess_reliability_validity.py in the same directory as data files:

1) Read market_event_study.csv / pdl_events.csv / drloop_metrics.csv / provenance.json;

2) Calculate reliability and validity metrics (including bootstrap confidence intervals);

3) Output rv_summary.json for archiving and print detailed results to console.

VII. Key Code Snippets

The complete code version is provided with the data package (assess_reliability_validity.py). The following are excerpts from the core statistical functions:

```
def cohens_d(x, y):
    x, y = np.array(x), np.array(y)
    nx, ny = len(x), len(y)
    vx, vy = x.var(ddof=1), y.var(ddof=1)
    denom = ((nx-1)*vx + (ny-1)*vy)
    if (nx + ny - 2) <= 0 or denom <= 0:
        return float('nan')
    s = np.sqrt(denom / (nx + ny - 2))
    return (x.mean() - y.mean()) / s if s > 0 else float('nan')

def bootstrap_ci_stat(a, b, stat_fn, n_boot=5000, alpha=0.05, seed=20251211):
    rng = np.random.default_rng(seed)
    a = np.array(a); b = np.array(b)
    nA, nB = len(a), len(b)
    boots = []
    for _ in range(n_boot):
        sa = rng.choice(a, size=nA, replace=True)
        sb = rng.choice(b, size=nB, replace=True)
        boots.append(stat_fn(sa, sb))
    lo = float(np.percentile(boots, 100*alpha/2))
    hi = float(np.percentile(boots, 100*(1-alpha/2)))
```

```
return lo, hi
```

VIII. Restrictions and Ethical Considerations

• The data constitutes a simulated dataset, with authenticity reflected in its alignment with mechanisms and stylised facts; any extrapolation to real markets must be undertaken with caution.

• The monotonic upward trend of λ serves as a constraint to minimise the mechanism, as disclosed; should parameters be altered, the version and fingerprint must be updated accordingly.

• The assessment verifies directionality and relative magnitude only, and does not constitute a judgement on real entities or investment advice.

Appendix: English Brief of Methods

Reliability: provenance hash coverage; intra-sector pairwise correlations across the event window; sovereign CDS monotonic rise at t_0+[0,1,3]. Validity: Welch t-test and Cohen's d (with 5,000 bootstrap resamples); λ from sovereign CDS slope; correlation between (visibility×interdependence) and sector AR; stylized facts; placebo test at pseudo t_0.