Register-Guard Capital

FORECAST SCORING METHODOLOGY

Adopted by the Chief Executive Officer, September 16, 2026
Amended by the Chief Executive Officer, September 16, 2026
Version II

I. PURPOSE

This document sets out how Register-Guard Capital scores its published forecasts against consensus and against the actual data. The goal is one number that honestly answers a single question: are RGC's calls more accurate than consensus, and by how much? The rules are fixed before the outcomes are known and apply the same way to every call.

II. WHAT IS SCORED

The scored call for each release or meeting is the headline projection in the Guard, or in the pre-meeting outlook of the Federal Reserve Register Guard, published before that event. Only the metrics listed in Appendix A are scored. Supporting figures in the body of a paper are analysis, not scored calls.

A call counts only if it was published before the data release or the FOMC statement. The publication time is recorded.

III. WHAT EACH GUARD MUST STATE

For every scored metric, the headline projection states:

1. A point forecast, at the precision RGC chooses to publish.
2. An 80% range: the range the author is 80% confident will contain the actual. It is RGC's own judgment. It may draw on a quantitative model or depart from one, and it should be as narrow as the author honestly believes. For FOMC decisions, the Federal Reserve Register Guard instead states a probability for each of the five outcomes listed in Section VIII.
3. The consensus figure and its source, chosen under Section IV.

IV. CONSENSUS

Data releases. Consensus is taken from the first available source in this order:

1. Bloomberg survey median
2. FactSet
3. Dow Jones
4. Reuters poll
5. Any other published survey, named in the paper

FOMC meetings. Consensus is the most likely outcome priced by the CME FedWatch Tool when the paper is published. If that is unavailable, it is the majority view of the Reuters poll.

Rules for both:

- The source is chosen separately for each metric. If a source does not publish a figure for a metric, that metric falls to the next source in the order.
- The source used and its value are recorded when the paper is published. The Register scores against that same figure, even if a higher-ranked source turns up later.

V. THE ACTUAL

The actual is the first print, as published on release day by the issuing agency and at the precision the agency publishes. Later revisions do not change the score. A Register may discuss them. For GDP, the scored release is the advance estimate.

VI. SCORING

1. Accuracy vs consensus (the headline)

- For each call, the miss is the distance between the forecast and the actual, for both RGC and consensus.
- Each miss is divided by the scoring unit for its metric (Appendix A). Misses on different metrics can then be added together.
- The headline is 1 minus (the sum of RGC's scaled misses divided by the sum of consensus's scaled misses), across all scored calls.
- +15% means RGC's misses are 15% smaller than consensus's. 0% means equal. A negative value means consensus has been more accurate.
- A tie adds the same miss to both sides, so it counts as exactly as good as consensus.

2. At or better than consensus (tally)

For each metric, this is the number of calls where RGC's miss was less than or equal to consensus's, out of all calls. Ties count toward "at or better."

3. Range record and range score

- **Range record.** For each metric and overall, the number of times the actual fell inside RGC's 80% range, out of the calls that stated one. Well below 80% means the ranges are too narrow. Near 100% means they could be tighter.
- **Range score.** Each range is scored as its width, plus 10 times the distance by which the actual fell outside it, if it did, divided by the metric's scoring unit so that metrics can be combined. Lower is better: a narrow range that contains the actual scores best, and a narrow range that misses scores worst.

- **Benchmark range.** Each ranged call is compared with a benchmark range of the consensus figure plus or minus 1.6 times the metric's unrounded mean absolute consensus miss, as recorded in the data behind Appendix A, which is approximately an 80% range around consensus given its typical miss. The range skill is 1 minus (RGC's average range score divided by the benchmark's average range score). Above zero means RGC's ranges are more informative than a range built from consensus and its historical accuracy.
- The range record, average range score, and range skill are reported for each metric and overall. They do not enter the headline.

4. Accuracy vs a naive forecast

- For every call, a naive forecast is recorded as defined in Appendix B: the last value of the metric published before the release, or no change for the FOMC target range.
- The naive forecast's miss is scaled by the same scoring unit as RGC's and consensus's.
- Two figures are reported: RGC vs naive, 1 minus (the sum of RGC's scaled misses divided by the sum of the naive forecast's scaled misses), and consensus vs naive, computed the same way with consensus's misses.
- Above zero means the forecaster's misses were smaller than simply assuming the last published value repeats. Reading the two figures together shows how much of consensus's accuracy RGC has matched.
- Accuracy vs a naive forecast is reported alongside the headline in Section VI.1 and does not replace it.

VII. CONFIDENCE AND DISPLAY

- The homepage shows two figures: accuracy vs the naive forecast (Section VI.4) and accuracy vs consensus (Section VI.1). Each shows a 90% confidence range and the number of calls. The ranges come from resampling whole releases and meetings, so metrics from the same report move together.
- Colors: each figure is green only when its whole confidence range is above zero, red only when it is entirely below zero. Otherwise the figure shows in plain ink.
- The range record and range skill (Section VI.3) and the FOMC probability score (Section VIII) are shown with the per-metric detail on the methodology page.
- Example: Accuracy vs naive forecast: +29% (90% range +12% to +43%) · Accuracy vs consensus: −29% (90% range −66% to −5%), 47 calls.
- Supporting statistics (significance tests, the per-metric detail) are published on the methodology page.

VIII. FEDERAL RESERVE CALLS

- Decisions are scored in basis points: the distance between the predicted change and the actual change in the target range. They enter the headline like every other metric, one scoring unit per 25 basis points of miss.
- When RGC and consensus call the same decision, the call is a tie and does not move the headline.

- When they differ, the side that is right gains a full unit on the other. A correct call against consensus moves the headline sharply in RGC's favor, and a wrong one moves it sharply against.
- **Probability score.** The probabilities stated under Section III are scored with the Brier score across five outcomes: a cut of more than 25 basis points, a 25 basis point cut, a hold, a 25 basis point increase, and an increase of more than 25 basis points.
    - The score for a meeting is the sum, over the five outcomes, of the squared difference between the stated probability and the result (1 for the outcome that occurred, 0 for the others).
    - Scores run from 0, a perfect forecast, to 2. Lower is better.
    - Any probability the paper does not assign is split equally among the outcomes it does not name.
    - The same score is computed for the CME FedWatch probabilities recorded at publication.
    - The Brier skill score is 1 minus (RGC's average Brier score divided by FedWatch's average Brier score). Above zero means RGC's probabilities were more accurate than the market's.
    - The probability score is reported separately and does not enter the headline.

IX. QUANTITATIVE MODELS

- Where an internal model exists, its point and range at publication are logged next to RGC's call and scored the same way.
- Model results are reported separately, including how often RGC's published call beat the model. They never enter the headline.
- The published call is RGC's decision, whether it follows the model or not.

X. REVISED AND WITHDRAWN CALLS

- Only the last call published before an event enters the headline.
- A longer-horizon call is logged only when a paper explicitly labels a forecast as its call for a later, named release or meeting. Discussion of the expected policy path or later data is analysis, not a scored call. A logged longer-horizon call stays on record and is scored against the event even if later withdrawn, and longer-horizon calls are reported separately.

XI. RECORD BEFORE THIS METHODOLOGY

- Calls from before adoption, including the February and March 2026 Memos, stay in the record and count toward the headline and tally. They are labeled as coming before this methodology.
- They are scored against the consensus figure cited in each paper, and only on the metrics tracked at the time: payrolls, unemployment rate, wages, headline CPI, core CPI and the rate decision. In addition, a seasonally adjusted month-over-month CPI change, headline or core, is scored where the paper stated RGC's figure and a numeric consensus figure for it. Unpublished drafts are not part of the record.

- They stated no ranges, so they are excluded from the range record.

XII. AMENDMENTS

- This document is versioned.
- Changes apply only to calls published after the amendment. Past calls are never rescored under new rules.
- Where a change would materially alter the record, both versions of the result are shown.
- A measure added by amendment that reports on calls without changing how any existing call is scored, such as the naive benchmark in Section VI.4, is computed over all calls, including those published before the amendment.

XIII. STATISTICAL TESTING

1. Edge. For each call, the edge is the consensus scaled miss minus the RGC scaled miss; a positive edge favors RGC. For pooled tests, edges are summed within each release or meeting, and each release or meeting is one observation, because metrics from the same report are not independent.

2. Primary test. The primary test is the Diebold-Mariano test, with the Harvey-Leybourne-Newbold small-sample correction, on the per-release edges, at the 5% level, two-sided. The claim that RGC is more accurate than consensus is made only when this test rejects equal accuracy in RGC's favor.

3. Tests.

Question	Test	Notes
Are RGC's total misses smaller than consensus's? (primary)	Diebold-Mariano test, Harvey-Leybourne-Newbold correction, on per-release edges	The standard test for comparing two forecasters and the formal counterpart of the headline. A t-test on the average edge, corrected for serial correlation and small samples.
Are RGC's misses smaller than a naive forecast's?	Diebold-Mariano test, Harvey-Leybourne-Newbold correction, and Wilcoxon signed-rank test, on per-release differences between the naive forecast's and RGC's scaled misses	Tests the figure in Section VI.4. The same Diebold-Mariano test is reported for consensus vs the naive forecast, for comparison.
What is the plausible range of the headline?	Bootstrap of whole releases and meetings, 10,000 resamples	The 5th and 95th percentiles of the recomputed headline give the published 90% confidence range.
Does RGC win more often than it loses?	Sign test (binomial test at 50%)	Ties are excluded from the test. Ties count favorably in the published tally, but including them in this test would bias it in RGC's favor.

Question	Test	Notes
Does RGC win by more than it loses?	Wilcoxon signed-rank test	Accounts for the size of each win and loss without assuming the edges are normally distributed. Checks the primary test in small samples.
Is RGC right more often than consensus on FOMC decisions?	McNemar's test	Uses only the meetings where RGC and consensus disagreed, comparing RGC right and consensus wrong against the reverse. Its large-sample form is a chi-squared test.
Are RGC's 80% ranges honest?	Exact binomial test of coverage against 80%	Preferred over the chi-squared goodness-of-fit test while counts are small. The two agree once there are dozens of ranged calls.
Does RGC's edge differ by release type?	Chi-squared test of independence on win, tie and loss by release type; Fisher's exact test while any expected count is below five	Compares employment, inflation, GDP and FOMC results.
When RGC differs from consensus, does the actual fall on RGC's side?	Binomial test at 50%	Measures the value of RGC's independent view, apart from the size of the miss.
Does RGC's published call beat its internal models?	Diebold-Mariano test and Wilcoxon signed-rank test, with model misses in place of consensus misses	Reported separately under Section IX.
Are RGC's 80% ranges more informative than a consensus-based range?	Wilcoxon signed-rank test on per-call differences between the benchmark range score and RGC's range score	Tests the range skill in Section VI.3. Pooled across metrics, differences are summed within each release as for the primary test.
Are RGC's FOMC probabilities more accurate than FedWatch?	Wilcoxon signed-rank test on per-meeting differences between the FedWatch Brier score and RGC's Brier score	Tests the Brier skill score in Section VIII.

4. Multiple comparisons. Only the primary test supports the claim of greater accuracy. All other tests are supporting evidence. Per-metric tests are adjusted with the Holm method.

5. Sensitivity. The headline and primary test are also reported with each metric's scoring unit halved and doubled in turn, all other units held fixed. (Scaling every unit by the same factor leaves the headline unchanged.)

ADOPTED AS AMENDED:
Dakota Charles Baker, Chief Executive Officer of Guard Capital Company
September 16, 2026

APPENDIX A: SCORED METRICS

Release	Metric	Miss measured in	Scoring unit
Employment Situation	Nonfarm payroll change	thousands	63,000
Employment Situation	Unemployment rate	percentage points	0.1
Employment Situation	Average hourly earnings, year-over-year	percentage points	0.15
Consumer Price Index	Headline CPI, year-over-year	percentage points	0.1
Consumer Price Index	Core CPI, year-over-year	percentage points	0.1
Consumer Price Index	Headline CPI, month-over-month, seasonally adjusted	percentage points	0.1
Consumer Price Index	Core CPI, month-over-month, seasonally adjusted	percentage points	0.1
Gross Domestic Product	Real GDP growth, advance estimate, seasonally adjusted annual rate	percentage points	0.53 (provisional)
FOMC	Target range decision	basis points	25

A scoring unit is the typical size of a consensus miss for its metric: the mean absolute difference between the published consensus figure and the first print, over January 2015 through September 2026, excluding 2020 and 2021, rounded to two significant figures. Historical consensus figures are taken from the highest-ranked survey available in public reporting at the time of each release. Where historical coverage is incomplete, the unit is computed on the releases available. No unit is smaller than the metric's reporting increment. The unrounded mean absolute misses are also recorded, for use in the benchmark range under Section VI.3.

Two exceptions apply:

- The FOMC target range decision uses 25 basis points, the size of a standard policy move, because consensus misses on decisions are too rare to measure.
- The GDP unit is provisional: no historical GDP consensus archive was available, so it is set from the mean absolute miss of the Federal Reserve Bank of Atlanta GDPNow final nowcast against the advance estimate. It will be replaced by amendment under Section XII once consensus-based figures are available.

Units are computed once at adoption and change only by amendment under Section XII. The data and calculations behind each unit are archived with Register-Guard Capital's publication records.

APPENDIX B: NAIVE FORECASTS

For each call, the naive forecast is the most recent value of the metric published before the release, taken from the Federal Reserve Bank of St. Louis ALFRED vintage as of the day before release, computed and rounded the way the issuing agency publishes it.

Metric	Naive forecast	ALFRED series
Nonfarm payroll change	Prior month's change, as last published	PAYEMS
Unemployment rate	Prior month's rate	UNRATE
Average hourly earnings, year-over-year	Prior month's year-over-year change	CES0500000003
Headline CPI, year-over-year	Prior month's year-over-year change	CPIAUCNS
Core CPI, year-over-year	Prior month's year-over-year change	CPILFENS
Headline CPI, month-over-month, seasonally adjusted	Prior month's change	CPIAUCSL
Core CPI, month-over-month, seasonally adjusted	Prior month's change	CPILFESL
Real GDP growth, advance estimate	Prior quarter's growth rate, as last published	A191RL1Q225SBEA
FOMC target range decision	No change (0 basis points)	none