A/G Bank Scoreboard — Technical Brief

Sanjit Singh Mehat · April 2026 · DOI: 10.5281/zenodo.18489503 · bomeh2017@gmail.com

Updated edition — incorporates walk-forward external validation, expanded benchmarks, and claim-type labels

What the System Measures

The A/G Scoreboard is a two-channel structural monitoring system for every FDIC-insured bank. Ground (G) measures structural viability — low G predicts external failure. The Mirror-Horizon Index (MHI = Appearance / Ground) measures the divergence between surface performance and structural condition — high MHI identifies banks whose observable health exceeds their structural floor. The two channels share zero variables: no feature appears in both scores. G is validated against external outcomes (actual FDIC closure). MHI is validated against the system's own tier-transition dynamics. These are distinct constructs with distinct validation profiles.

Channel Definitions

Channel	Variable	Transform	Agg.
Ground (G)	Tier 1 Leverage Ratio (RBC1AAJ)	Logistic (center=5%, k=1.5)	
	Total Risk-Based Capital (RBCRWAJ)	Logistic (center=10%, k=1.0)	
	Noncurrent Loans % (NCLNLSR)	Inverse (scale=0.5)	
	Net Charge-offs % (NTLNLSR)	Inverse (scale=1.0)	Min
Appearance (A)	Return on Assets (ROA)	Logistic (center=0%, k=2.0)	
	Efficiency Ratio (EEFFR)	Inv. logistic (center=70%, k=0.08)	
	Net Interest Margin (NIMY)	Logistic (center=3%, k=1.5)	Mean

MHI = A / G (Mirror-Horizon Index). Higher MHI = larger gap between appearance and structure. All inputs are from quarterly FDIC Call Reports (banks.data.fdic.gov). Current dataset: 136 quarterly files, 1992Q1–2025Q4, ~1,131,900 bank-quarter observations, 17,362 unique institutions.

Transform rationale: Logistic centers are set at regulatory or industry reference points (5% Tier 1 leverage = well-capitalized threshold; 10% total risk-based capital = well-capitalized; 70% efficiency ratio = standard industry benchmark; 3% NIM = median community bank level). These are not empirically tuned to the outcome variable. Inverse transforms for charge-off and noncurrent-loan rates reflect that deterioration is unbounded upward while structural soundness is bounded.

G aggregation (min): Ground is computed as the minimum of its transformed components, not the mean. This encodes a bottleneck assumption: a bank's structural viability is bounded by its weakest structural dimension. A bank with strong capital but severely impaired asset quality is not structurally sound on average — it is structurally constrained by the impairment. This is a design choice, not an empirical finding. Empirical consequences of relaxing this choice are documented in the Aggregation Operator Note.

Tier Classification

Tier	Condition	Interpretation
Critical	$G < 0.1$	Structural viability near zero
Watch	$0.1 \leq G \leq 0.5$ and MHI > 1.2	Surface outrunning structure (key zone)
Diverging	$G > 0.5$ and MHI > 1.3	Early structural drift

Tier	Condition	Interpretation
Stable	Everything else	—

Each bank also reports the binding G component — which structural variable is the floor. Temporal features include quarter-over-quarter deltas, consecutive Watch/Critical run lengths, trajectory classification (persistent, new entry, escalating, recovered), and historical peak stress markers.

Validation Stack

Tests A, B, and C use the system's own tier transitions as the outcome variable. They confirm internal signal coherence. External-outcome validation (failed-bank anchoring, walk-forward testing) is documented separately in the sections and notes below.

Test	Design	Result
A: Same-band separation **[internal]**	Within G band 0.1–0.5: MHI>1.2 vs MHI≤1.2, 4Q future deterioration	85.2% vs 47.5% Lift: 1.79x
B: Monotonic gradient **[internal]**	MHI deciles (within-quarter), pooled across 34 years	Decile 1: 6.7% Decile 10: 88.1% Zero inversions
C: Persistence **[internal]**	Consecutive Watch quarters vs future deterioration within 4Q	Run 1: 74.2% Run 5: 89.0% Run 10: 92.2%

Outcome definition: "4-quarter future deterioration" means the bank enters the Watch or Critical tier in at least one of the next four consecutive quarters following the observation quarter. This is a binary indicator computed per bank-quarter from the scored historical ledger. The outcome is fixed across all three tests and the baseline comparison. Because the outcome is defined from future tier entry, these tests should be read as temporal-signal and ranking-consistency checks, not as external-event prediction claims.

Baseline Comparison

Model	Decile 1	Decile 10	Spread	Profile
MHI	6.7%	88.1%	81.4 pp	Smoothest; strongest mid-range
G-only	5.3%	86.7%	81.4 pp	Matching spread; weaker mid-range
Texas proxy	8.1%	52.6%	44.5 pp	Non-monotonic in middle deciles

Texas proxy = (Nonperforming Assets + OREO) / (Equity + Loan Loss Reserves). All three tested on the same 4-quarter future deterioration task across the full 34-year panel. This outcome is the system's own tier-transition metric (entering Watch or Critical within 4 quarters), not an external event such as actual bank failure. MHI and G-only have similar end-to-end spread on this task, but MHI is the stronger ranking signal by smoothness, monotonicity, and mid-range discrimination. The Texas proxy is materially weaker in both spread and gradient stability.

Sensitivity Checks

Sensitivity checks were run across nearby parameter variations for transform centers and scales, plus selected Ground-aggregation alternatives. Across the local center/scale perturbations, the main signal remained directionally stable: Test A lift stayed between 1.65x and 1.96x, the decile-spread test stayed between 73.9 and 84.2 percentage points, and persistence remained high (Run 1: 71.6%–76.3%; Run 5: 88.5%–90.4%; Run 10: 91.1%–93.0%). In the current quarter, the Critical tail remained narrow while the Watch/Diverging layer moved within an interpretable range: Critical 11–17, Watch 327–638, Diverging 242–422, median G 0.770–0.833, median finite MHI 0.869–0.967. More structural alternatives can move the system more substantially: replacing the bottleneck-style Ground aggregation

with geometric or harmonic mean materially changes the behavior of the instrument — geometric mean collapses the Watch layer by 98%, harmonic by 90% — and should be treated as different modeling assumptions rather than nearby calibration variants.

Metric	Default	Range across local perturbations
Test A lift	1.79x	1.65x–1.96x
Decile-spread test	81.4 pp	73.9–84.2 pp
Persistence, Run 1	74.2%	71.6%–76.3%
Persistence, Run 5	89.0%	88.5%–90.4%
Persistence, Run 10	92.2%	91.1%–93.0%
Current-qtr Critical	13	11–17
Current-qtr Watch	483	327–638
Current-qtr Diverging	334	242–422
Current-qtr median G	0.800	0.770–0.833
Current-qtr median finite MHI	0.924	0.869–0.967

These ranges summarize local center/scale perturbations only. Structural alternatives (geometric mean, harmonic mean) are reported separately in the Aggregation Operator Note because they represent different modeling assumptions, not nearby parameter shifts.

Regime Stability

Regime-split validation indicates that the framework's main empirical signatures persist across distinct historical banking environments. In all nine tested slices, same-band lift remained above 1.5x, and temporal persistence increased materially with consecutive Watch quarters. MHI decile gradients remained strictly monotonic in most regimes, with reduced rank-order precision in the shorter, lower-stress Pandemic and Recent Softening windows. Relative to baseline comparators, MHI and G-only remained the two strongest signals across regimes, while the Texas-style proxy was materially weaker and more crisis-dependent.

Failed-Bank Anchoring [external]

The historical ledger was cross-referenced against the FDIC failed bank list. Most matched failed banks had prior Watch and/or Critical status before failure. Median most-recent Watch lead time: 12 quarters. Median most-recent Critical lead time: 1 quarter. Watch functions as the medium-horizon warning layer; Critical as the near-failure layer.

Example: First City Bank of Florida (CERT 16748) ranked #1 in 2018Q4 with MHI = 66.6 and G = 0.005. The bank was closed by the OCC on October 16, 2020. The scoring, the rank, and the failure date are independently verifiable from the public FDIC ledger and the FDIC failed bank list respectively.

Walk-Forward External Validation [external]

Walk-forward testing evaluated MHI, G-only (1/G), and a Texas-ratio proxy against two targets fully independent of the A/G tier logic: actual FDIC failure within 8 quarters and Texas-ratio distress (ratio > 1.0) within 4 quarters. Three genuinely forward-looking time windows were used (GFC 2006–2012, Post-Crisis 2013–2019, Recent 2020–2025).

G-only is the dominant external predictor: 13.0pp decile spread on GFC failure (the only metric achieving full monotonicity on that target). MHI achieves 5.7pp — narrower than G-only and Texas proxy (10.3pp). Same-band lift inverts on external targets: within G band 0.1–0.5, Watch banks (high MHI) are less likely to fail externally than non-Watch banks (low MHI). High Appearance provides a real economic buffer against failure even when structural Ground is compromised.

MHI measures performed continuation (surface outrunning structure). G measures structural viability. These are related but distinct constructs. The first is internally validated. The second is externally anchored. See Walk-Forward External Validation memo for full results.

Expanded Benchmark Comparison

Seven metrics were tested on both internal and external targets: MHI, G-only, Texas proxy, NCL-only (single noncurrent-loans variable), Mean-G (mean of G components instead of min), ROA-only (simple profitability), and G+A blend (additive combination). On the GFC failure target, the ranking by decile spread is: Mean-G (13.8pp) > G-only (13.0pp) > NCL-only (11.9pp) > ROA-only (10.8pp) > Texas proxy (10.3pp) > MHI (7.2pp) > G+A blend (1.6pp). The G+A blend — which combines both channels additively — is the weakest performer everywhere, validating the design decision to keep G and A as separate channels. See Expanded Benchmark Comparison note for full results.

Explicit Limitations

- Does not predict specific failure events. Measures structural deterioration.
- Does not compare against CAMELS (requires confidential supervisory data).
- Has been benchmarked against G-only (1/G), Mean-G, NCL-only, ROA-only, Texas-ratio proxy, and G+A blend on both internal and external targets. G-only is the strongest external predictor (13pp spread, monotonic on GFC failure). MHI adds internal-tier discrimination but not incremental external failure prediction.
- Walk-forward external validation (against actual FDIC failure and Texas-ratio distress) has been completed. G alone is the dominant external predictor; MHI adds internal-tier discrimination but not incremental external failure prediction beyond G.
- Recovery and de-escalation exist. The system is probabilistic, not deterministic.
- Sensitivity to nearby parameter shifts has been explored. Threshold-freeze validation has confirmed that production parameters generalize under strict temporal restriction (see supporting documentation). Broader robustness testing across alternative aggregation families remains open, and whether materially different threshold choices would produce equivalent results has not been tested.

Current State: 2025Q4

4,408 banks scored. 13 Critical, 483 Watch, 334 Diverging, 3,578 Stable. Median G = 0.800, median finite MHI = 0.924.

The system has softened for 12 consecutive quarters from a 34-year structural high-water mark in 2022Q4 (median G = 0.868, Watch = 275). For comparison, 2019Q4 had median G = 0.771 and Watch = 658. Current 2025Q4 readings (median G = 0.800, Watch = 483) sit between those two benchmarks. The Watch cohort is actively deteriorating (mean delta_MHI = +0.105, mean delta_G = −0.051) and recruiting new entrants from the Stable tier.

This is not a crisis signal. At the 2010Q1 peak, median MHI was 1.282 and Watch count was 3,287. Current readings are far below that. The system indicates post-2022 softening with a real but contained Watch/Diverging layer, not systemic collapse.

The system qualifies traditional structural risk rather than replacing it. It imposes a new ordering among superficially similar banks — not a claim of universal dominance over existing supervisory or proprietary models.

Supporting Documentation

- **Leakage / Bias Audit Memo:** Construction integrity, leakage controls, outcome-definition analysis
- **Failed-Bank Case Note:** Peak-quarter snapshots of extreme-signal failed banks
- **External Anchor Case Set:** Quarter-by-quarter tier-evolution trajectories for failure and distress cases
- **Walk-Forward External Validation:** External-outcome walk-forward results against FDIC failure and Texas-ratio distress

• **Aggregation Operator Note:** Bottleneck axiom, four-operator comparison (min, geometric, harmonic, weighted min), and defense of min for the instrument's stated purpose

• **Expanded Benchmark Comparison:** Seven-metric comparison on internal and external targets across four time windows

---