Macro Risk Engine V3: Technical Report and Freeze Record

February 17, 2026

1 Executive summary

Within this project line, V3 is the first release with explicit forecast governance. It is designed to produce stable quarterly macro paths for PD input workflows, with explicit acceptance gates and reproducible artifacts.

Current V3 status from the latest validation snapshot:

- Release gate: **pass**
- Promotion gate (challenger replacing incumbent): **fail**
- Diagnostic-only mode: **false** (power requirements satisfied)

This means V3 is suitable to freeze as a production fallback baseline.

2 Scope and output contract

2.1 Scope

The engine forecasts macro variables. It does not estimate PD/LGD models.

2.2 Primary PD handoff contract

Canonical handoff is a quarterly level-path file with:

- `unemployment_rate`
- `ust10_rate`
- `hpi_yoy` (mapped from `hpi_growth_yoy`)

Convenience derived outputs (for quick checks) are included but non-blocking.

3 Data design and assumptions

3.1 Source and frequency

The modeling panel is built from FRED-based series mappings and aggregated to quarterly frequency.

3.2 Modeling panel

The main training table is:

- `data/macro_panel_quarterly_model.csv`

with metadata in:

- `data/macro_panel_metadata.json`

3.3 Transform policy

Rates, spreads, sentiment, and delinquency series are treated in level-style form. Growth and inflation series are mainly YoY transforms.

4 Model architecture from scratch

4.1 Horizon and regime architecture

Forecast horizon is 80 quarters. V3 evaluates explicit regime candidates:

- **Champion A (incumbent)**: Q1–Q12 short model, Q13–Q24 bridge, Q25–Q80 long-run/scenario
- **Champion B (challenger)**: Q1–Q16 short model, Q17–Q28 bridge, Q29–Q80 long-run/scenario
- **Single-stage ablation**: `single_stage_h12_no_bridge`

4.2 Short-horizon model system

V3 does not force one model for every variable. Candidates include BVAR, AR, and RW. Selection is by variable and horizon bucket:

- Bucket 1: Q1–Q4
- Bucket 2: Q5–Q12

Champion tie-break order:

1. Lowest mean CRPS

2. Coverage closeness to 0.90

3. Width-ratio closeness to 1.0

Champion decisions are persisted in:

- `outputs/macro_engine/champion_map.json`

4.3 Bridge and long-run design

After short horizon, the system transitions by bridge dynamics and then long-run scenario behavior. Long-run values combine structural anchors and non-structural regime parameters.

Structural anchors (assumption sets low/base/high): NAIRU, inflation target, neutral real rate, productivity trend, working-age growth, population growth, term premium.

Non-structural long-run parameters: HY spread distribution behavior, mortgage spread, and housing supply drag.

4.4 Scenario layer

Scenarios generated in the forecast output include:

- Baseline
- Mild_Adverse
- Severe_Adverse
- Demographic_LowGrowth

5 Validation framework (V3 governance core)

5.1 Contract and target set

Validation contract version in current artifact: `v3.freeze.1`. Primary gated target set is PD levels only:

- `unemployment_rate`
- `ust10_rate`
- `hpi_yoy`

across horizons h1..h12.

5.2 Power and invariants

- Required-cell power rule: minimum `n_oos` ≥ 40
- Scenario timing and scenario ordering checks
- Boundary smoothness checks at regime seams
- Seeded reproducibility for simulation-based metrics

5.3 Profiles and pass/fail usage

- `release`: production acceptance gate
- `promotion`: challenger replacement gate
- `operational`: reporting-only diagnostics

5.4 Core metrics

- Relative RMSE vs RW (short horizon)
- CRPS gain vs RW (medium horizon)
- Coverage pass-rate (90% interval in configured band)
- Width-ratio guardrails (mean and per-variable)
- Boundary Z metrics (median and max)

6 Why V3 exists: issues found in earlier versions and fixes applied

The V3 build addressed concrete failure modes observed during earlier iterations:

Earlier issue	**Risk**	**V3 response**
Lag-selection mismatch between production and backtest paths	OOS claims not representative of production behavior	Unified lag-parity logic and common validation path
Scenario timing offset at start quarter	Scenario shocks applied one quarter off, distorting stress paths	Deterministic timing/order checks integrated in validation
Small OOS sample and weak evidence quality	Performance claims unstable and easy to over-interpret	Hard power gate with minimum required-cell `n_oos`
Model drift from ad-hoc changes	No stable contract for release decisions	Explicit contract version, threshold profiles, and artifacts
Coverage failures vs overwide intervals tradeoff	Density quality could be gamed by width inflation	Combined coverage + sharpness gates with caps and calibration controls
Boundary discontinuity spikes at regime seams	Artificial jumps contaminating forecast paths	Boundary continuity controls and Z-stat diagnostics

7 Latest V3 evidence snapshot

Source: `outputs/macro_engine/validation/validation_summary.json`

7.1 Release gate (incumbent)

Metric	Value
Minimum required-cell `n_oos`	44
Coverage90 pass-rate	0.8611
Mean width ratio	1.3069
Median rRMSE h1–2	0.9744
Median rRMSE h3–4	0.9208
Mean CRPS gain h5–12 vs RW	15.13%
Boundary median Z / max Z	0.0012 / 0.0350
Release decision	**PASS**

7.2 Promotion gate (challenger vs incumbent)

Metric	Value
Challenger release pass	False
Promotion subset min `n_oos` (h9..12)	40
CRPS gain h9..12 challenger vs incumbent	1.64%
Short-horizon CRPS worsen h1..4	0.42%
Boundary comparator pass	True
Promotion decision	**FAIL**

Interpretation: incumbent is suitable for freeze; challenger is not promoted.

8 PD-target champion map snapshot

From current `champion_map.json`:

- `unemployment_rate`: BVAR in both buckets; Q5–Q12 calibration scale 1.205
- `ust10_rate`: AR in both buckets; Q5–Q12 calibration scale 1.55
- `hpi_growth_yoy`: BVAR in both buckets; Q5–Q12 calibration scale 0.90

9 Integrated scorecard and independent-review notes

9.1 Historical transition note (V3.1 recovery to V3 freeze)

Earlier recovery-stage summaries recorded a period where coverage improved but width inflation still failed release checks. That transition work introduced tiered profiles, boundary continuity

enforcement, and calibration controls. The frozen snapshot in this report reflects the post-recovery state where release now passes on the primary PD level targets.

9.2 Detailed promotion bottleneck from scorecard diagnostics

Promotion remains blocked for two concrete reasons in the challenger comparison:

- Challenger release profile does not pass.
- Challenger medium-horizon CRPS gain vs incumbent on h9..h12 is 1.64%, below the 5.0% promotion threshold.

Short-horizon challenger degradation is controlled (0.42% worsen on h1..h4, within threshold), and boundary comparator conditions pass, so the blocker is model-quality gain rather than governance mechanics.

9.3 Coverage-fail diagnostics retained

Coverage-fail diagnostics identify localized misspecification zones rather than broad instability. Current retained diagnostic table highlights fail cells in:

- `hpi_growth_yoy`: h5, h6, h11
- `ust10_rate`: h3, h4

9.4 Ablation context retained

PD-target ablation artifacts are preserved to document regime/scope tradeoffs. The recorded winner in retained artifacts is Champion A with full variable scope, while release-like checks on alternative candidates remain non-passing.

9.5 Next-change priority for future branch work

The highest-value single experiment remains challenger `ust10_rate` density improvement in bucket Q5–Q12. Target outcome is to reduce challenger width pressure while lifting medium-horizon CRPS gain, so promotion thresholds can be approached without loosening governance.

10 Shippable outputs in frozen V3

10.1 Canonical PD handoff

- `outputs/macro_engine/pd_regressors_forecast_levels.csv`
- `outputs/macro_engine/pd_regressors_forecast_derived.csv`
- `outputs/macro_engine/pd_regressors_metadata.json`

10.2 Governance and diagnostics

- `outputs/macro_engine/validation/validation_summary.json`
- `outputs/macro_engine/validation/backtest_metrics.csv`
- `outputs/macro_engine/validation/calibration_factors.json`
- `outputs/macro_engine/validation/pd_ablation_results.csv`
- `outputs/macro_engine/champion_map.json`

10.3 Legacy baseline table retained

- `outputs/macro_engine/bvar_oos_backtest_table.csv`

10.4 Validation artifact retention policy in frozen package

The frozen V3 package keeps both final-result validation artifacts and extended diagnostics. Core files (`validation_summary`, `backtest_metrics`, `calibration_factors`, `pd_ablation_results`) support the headline release decision. Extended files (boundary culprits, coverage fail cells, DM table, seed registry, scenario checks, and assumption/regime rerun folders) are retained for deeper audit trails.

11 Limitations and next-work boundary

- This package uses revised-history macro data, not real-time vintage data. Forecast quality under true real-time information sets may differ.
- Promotion remains blocked. Challenger regime does not yet meet replacement criteria, mainly because medium-horizon density improvement is insufficient.
- The primary validation focus is the three PD level targets. This is intentional for handoff quality, but it does not imply all secondary macro variables are equally optimized.
- Calibration improves coverage behavior but introduces model-risk tradeoffs in width. Width controls are enforced, but further refinement may still improve sharpness efficiency.
- Scenario design remains rule-based and assumption-driven in long horizon. It is not a structural macroeconomic simulator and should be interpreted as disciplined scenario path generation.
- Derived PD deltas are convenience outputs only. Final PD feature policy, transformations, and segmentation remain the responsibility of the downstream PD modeling layer.
- This freeze captures one validated snapshot in time. Re-estimation on new data vintages may alter champion choices and gate outcomes.

12 Freeze statement

V3 is designated as the stable fallback baseline for this repository. It is intended for reproducible reruns, controlled handoff, and governance traceability when experimental branches (V4+) are under active change. Version progression should therefore follow a branch-and-compare workflow: retain V3 unchanged for baseline reference, and evaluate all new modeling ideas in V4+ against this frozen benchmark before any future promotion.