Macro Risk V4 Release Review and Technical Record

2026-02-20

1 Outline and purpose of this report

This report is structured to answer six practical questions:

1. What is this macro engine as a complete system: assumptions, methodology, and design.
2. What are the V4 validation results, and what do they imply for release and promotion.
3. What changed in V4, and why those changes matter.
4. Is V4 materially better than V3 on governance outcomes.
5. What risks remain, and how serious they are.
6. What operating discipline is required after release.

All metrics in this report are sourced from package artifacts under `outputs/macro_engine/validation/`.

2 What this macro engine does as a whole

2.1 System objective

The engine produces quarterly macro paths, interval forecasts, and governance outputs used by downstream risk processes.

2.2 Data and assumption framework

Core data artifacts:

- `data/macro_panel_quarterly_raw.csv`
- `data/macro_panel_quarterly_model.csv`
- `data/macro_panel_metadata.json`

Assumption design combines:

- short-horizon empirical forecasting,
- bridge dynamics,
- long-horizon scenario overlays.

2.3 Methodology and architecture summary

- Forecast horizon: 80 quarters.
- Incumbent regime (`champion_a`): short Q1–Q12, bridge Q13–Q24, long Q25–Q80.
- Challenger regime (`champion_b`): short Q1–Q16, bridge Q17–Q28, long Q29–Q80.
- Candidate short-horizon models: BVAR, AR, RW.
- Champion selection buckets: Q1..Q4 and Q5..Q12.

2.4 Release handoff outputs

Handoff set:

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

Primary targets: `unemployment_rate`, `ust10_rate`, `hpi_yoy` (mapped from `hpi_growth_yoy`).

3 V4 validation results and release implications

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

3.1 Release profile results

Metric	Threshold	Incumbent	Challenger	Result
Release pass	required	True	True	PASS
Minimum required-cell *n_oos*	≥ 40	44	40	PASS
Median rRMSE h1..h2	≤ 1.00	0.9744	1.0000	PASS
Median rRMSE h3..h4	< 0.98	0.9208	0.9333	PASS
Mean CRPS gain h5..h12 vs RW	$\geq 3.0\%$	15.1304%	12.7381%	PASS
Coverage90 pass-rate	≥ 0.75	0.9444	1.0000	PASS
Width ratio mean	≤ 1.35	1.3438	1.2748	PASS
Width ratio per-variable max	≤ 1.60	1.4779	1.3767	PASS
Boundary consistency	pass	True	True	PASS

Scenario checks	pass	True	True	PASS

3.2 Promotion results

Metric	Threshold	Actual
Promotion pass	pass	True
CRPS gain h9..h12 (challenger vs incumbent)	$\geq 5.0\%$	14.1399%
Short-horizon CRPS worsen h1..h4	$\leq 1.0\%$	-8.6351%
Boundary comparator pass	required	True

3.3 Implication for release and promotion

The result is not borderline. Release and promotion both pass with meaningful margin on medium-horizon gain, while short-horizon behavior improves rather than worsens.

4 What changed in V4 and why it mattered

The following changes were material to governance outcomes:

1. Challenger evaluation path was corrected to use challenger champion artifacts end-to-end.
2. Scenario timing checks were aligned with regime start logic.
3. Small-gap tie-break logic was upgraded to use coverage-width balance.
4. Challenger ust10 long-bucket calibration constraints were tightened.
5. Calibration monotonicity checks were added.
6. Boundary consistency checks were added to gate evaluation.

Why this mattered: these changes improved challenger quality and interval discipline, which directly addressed the conditions that previously blocked promotion.

5 Is V4 materially better than V3 on governance

5.1 Direct comparison (V3 vs V4)

Metric	V3	V4	Direction
Challenger release pass	False	True	Improved
Promotion pass	False	True	Improved
CRPS gain h9..h12 (challenger vs incumbent)	1.6426%	14.1399%	Improved

Short-horizon CRPS worsen h1..h4	0.4207%	-8.6351%	Improved
Challenger width ratio mean	1.5285	1.2748	Improved
Coverage-fail count	5	2	Improved

5.2 Conclusion from comparison

On governance metrics that drive release and promotion decisions, V4 is materially stronger than V3.

6 Remaining risks and severity

6.1 Known residual issue

Current coverage-fail cells:

- `hpi_growth_yoy`, h=6, coverage90=0.7727, width-ratio=1.4147
- `hpi_growth_yoy`, h=11, coverage90=0.7955, width-ratio=1.7148

6.2 Severity assessment

- Scope: localized to one variable family and medium bucket.
- Gate impact: no release or promotion failure in current state.
- Severity classification: **medium, controlled**.

7 Next steps

1. Keep V3 available as rollback benchmark.
2. Move long-run structural redesign work to V5 scope.

8 Conclusion and decision

V4 for controlled release.

Reasons:

- release gates pass for incumbent and challenger,
- promotion gate passes,
- no-regression checks pass,
- residual risk is narrow, explicit, and monitorable.

9 Reproducibility commands

```
python3 scripts/fetch_macro_panel_fred.py \
  --raw-output data/macro_panel_quarterly_raw.csv \
  --model-output data/macro_panel_quarterly_model.csv \
  --metadata-output data/macro_panel_metadata.json

python3 scripts/run_macro_forecast_engine.py \
  --config macro_engine_config.json \
  --output-dir outputs/macro_engine

python3 scripts/run_macro_validation.py \
  --config macro_engine_config.json \
  --input data/macro_panel_quarterly_model.csv \
  --output-dir outputs/macro_engine/validation \
  --champion-map-output outputs/macro_engine/champion_map.json \
  --verbose-validation

python3 scripts/export_pd_macro_subset.py \
  --config macro_engine_config.json \
  --input outputs/macro_engine/macro_forecast_paths.csv \
  --model-panel data/macro_panel_quarterly_model.csv \
  --levels-output outputs/macro_engine/pd_regressors_forecast_levels.csv \
  --derived-output outputs/macro_engine/pd_regressors_forecast_derived.csv \
  --metadata-output outputs/macro_engine/pd_regressors_metadata.json
```