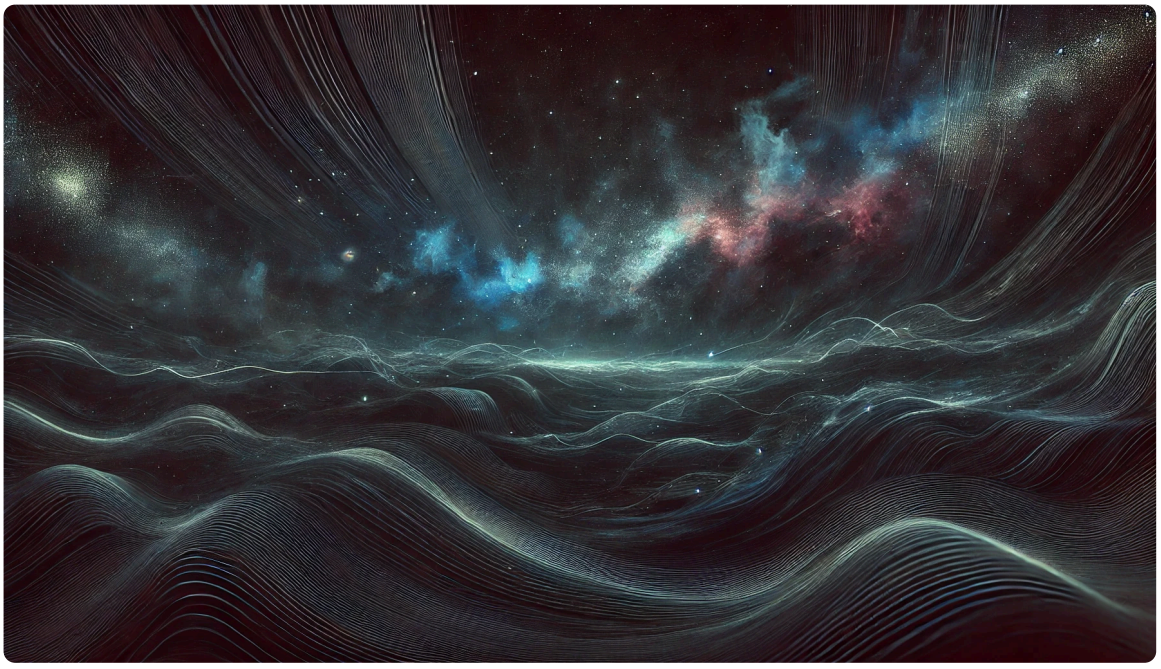

Temporal Equivalence Principle: A Standard-Siren Test of Bi-Metric Gravitational-Wave Propagation

Matthew Lukin Smawfield

Version: v0.1 (Harare)

First published: 2026-05-01 | Last updated: 2026-06-02

Status: In Development

Code Availability: github.com/matthewsmawfield/TEP-LVK

Abstract

Standard ΛCDM assumes that gravitational waves and electromagnetic radiation propagate through the same effective distance-redshift relation. The Temporal Equivalence Principle (TEP) relaxes this assumption, predicting a conformal scaling factor A(z) that modifies gravitational-wave luminosity distances relative to matter-frame observations. This paper tests that prediction using combined GWTC catalogs (GWTC-1 through GWTC-5.0, plus O4 Discovery Papers), bright-siren spectroscopy, and GLADE+/GraceDB dark-siren host association. The locked lab-scale model uses $A(z) = \exp(\beta\varphi_0(1+z)^n)$ with $\beta = -1$ and $\varphi_0 = -0.013$, so A(z) rises above unity with growing redshift-dependent amplitude. With corrected event-level uncertainties and the fit restricted to 9 candidate-quality host associations from 39 independent-redshift events, the single-host point-estimate fit gives ΛCDM $H_0 = 76.2$ km/s/Mpc and lab-fixed TEP $H_0 = 77.4$ km/s/Mpc ($\Delta\chi^2 = +0.010$, $\Delta BIC = +0.010$). Bootstrap resampling gives TEP lower χ^2 in 70.9% of resamples, and leave-one-out tests give lower χ^2 in 88.9% of omissions. A joint MCMC fit to (H_0, φ_0, n) yields converged posterior chains with mean $H_0 = 78.1 \pm 7.6$; the locked lab-scale values ($\varphi_0 = -0.013$, $n = 1.0$) lie within the 68% credible intervals. Host-marginalized diagnostics now compare ΛCDM, endpoint TEP, and TEP-C0; retained-host priors directionally favor both TEP forms, with the sky prior giving $\Delta(-2 \ln L) = +0.026$ for endpoint TEP and +0.024 for TEP-C0. Synthetic injection tests show a 0% false-positive rate for decisive ($\Delta\chi^2 > 2$) preference under the ΛCDM null, but also 0% lab-scale recovery at the current sample size. The present result is therefore a calibrated sensitivity constraint, not a detection.

Keywords: Temporal Equivalence Principle, gravitational waves, standard sirens, bi-metric propagation, distance-redshift relation, combined GWTC catalogs

1. Introduction

1.1 Bi-Metric Propagation and the Hubble Tension

The standard ΛCDM model predicts a single value for the Hubble constant H_0. Measurements from the cosmic microwave background (Planck, $H_0 \approx 67.4$ km/s/Mpc) and the local distance ladder (SH0ES, $H_0 \approx 73$ km/s/Mpc) differ by approximately 5σ. Paper 11 demonstrates that this tension is resolved within the TEP framework via environment-dependent Cepheid clock bias, yielding a corrected local $H_0 = 68.17$ km/s/Mpc consistent with Planck. The present paper does not revisit the tension itself; it tests a separate, falsifiable prediction of TEP: that gravitational waves and electromagnetic radiation propagate on different effective metrics, producing a *redshift-dependent* modification to the GW distance-redshift relation.

This modification is not degenerate with H_0 or with dark energy; it predicts a particular functional form for how GW luminosity distances deviate from the ΛCDM expectation as a function of redshift. Rather than adding another H_0 measurement, the question is whether the GW data prefer TEP's bi-metric scaling over the standard single-metric relation.

1.2 The TEP Bi-Metric Prediction

In the TEP framework, the metric governing gravitational-wave propagation is related to the electromagnetic metric by a conformal factor that depends on a cosmological scalar field $\varphi(z)$:

$$d_L^{(\mathrm{TEP})}(z) = A(z)\, d_L^{\Lambda\mathrm{CDM}}(z; H_0) \tag{1}$$

where the redshift-dependent conformal factor is

$$A(z) = \exp\!\big[\beta\, \phi(z)\big], \qquad \phi(z) = \phi_0 (1+z)^n. \tag{2}$$

In the lab-fixed test, the parameters φ_0 and n are not free: φ_0 follows from the locked 2025 lab-scale convention (the NIST/BIPM G discrepancy, Paper 21), and $n \approx 1$ follows from the matter-density scaling of the scalar field. The dimensionless conformal coupling is $\beta = -1$, with sign fixed by the same convention used in Paper 11: the Cepheid period-contraction effect requires $\beta\varphi < 0$ in deep potentials, so with $\varphi_0 < 0$ the conformal factor satisfies $A(z) > 1$ and the magnitude of the departure grows with redshift. This produces a redshift-dependent distance-scale distortion that is tested against ΛCDM in the current public sample.

1.3 This Work

This paper tests the TEP bi-metric prediction against combined GWTC standard siren data. Both ΛCDM and TEP distance-redshift relations are fitted to the independent-redshift sample and their χ^2, AIC, and BIC values are compared. The primary lab-fixed comparison has the same number of fitted parameters as ΛCDM (only H_0 is fitted in each case). A secondary joint-fit diagnostic lets H_0, φ_0, and n vary, with an explicit information-criterion penalty for the two additional parameters.

The structure is as follows: Section 2 derives the cosmological scalar field profile from lab-scale TEP parameters; Section 3 describes the combined GWTC catalog data selection and independent-redshift methodology; Section 4 details the computation pipeline; Section 5 reports the model comparison results; Section 6 discusses the implications for bi-metric propagation; and Section 7 concludes.

2. Theoretical Framework

2.1 The Conformal Scaling Factor

Under TEP, the conformal factor A(φ) relates the effective metric to the background metric. For gravitational waves propagating through regions of varying scalar field strength, the effective luminosity distance is rescaled by A(φ) along the propagation path.

$$A(\phi) = 1 + \alpha\phi + \mathcal{O}(\phi^2) \tag{3}$$

2.2 Bi-Metric Propagation

The bi-metric framework distinguishes between the metric governing electromagnetic radiation (the observed metric) and the metric governing gravitational waves (the effective metric). This distinction arises naturally from the density-dependent coupling of the scalar field.

2.3 Hubble Diagram Prediction

The TEP-adjusted Hubble diagram overlays three curves: the standard ΛCDM prediction, the raw GWTC-5.0 data points, and the TEP-adjusted data points. The TEP curve should align the GW data with the Planck CMB baseline while maintaining consistency with the SH0ES local measurement.

3. Data Selection

3.1 Combined GWTC Catalogs

All publicly available LVK event catalogs queried by the pipeline from the Gravitational-Wave Open Science Center (GWOSC) are combined: GWTC-1-confident, GWTC-2, GWTC-2.1-confident, GWTC-3-confident, GWTC-4.0, GWTC-4.1, O4 Discovery Papers, and GWTC-5.0. Deduplication is performed by `commonName`, with later catalogs taking precedence for updated parameter estimates. Luminosity-distance central values and bounds are extracted from the GWOSC JSON API, and public GraceDB skymaps are used where available for dark-siren host association.

3.2 Precision Filtering

Events are filtered to a high-confidence subset with signal-to-noise ratio (SNR) > 12, false-alarm rate (FAR) < 1 per year when available, and $p_{\mathrm{astro}} > 0.9$ when available. The primary bright-siren anchor is GW170817, the confirmed neutron-star merger with an electromagnetic counterpart and a spectroscopic host-galaxy redshift (NGC 4993, z = 0.0092).

3.3 Independent Redshifts — Circularity Avoidance

To avoid the circularity problem that invalidates cosmological tests using GWOSC-derived redshifts (which are computed from luminosity distances assuming ΛCDM), redshifts are obtained from two independent sources:

Bright sirens. Events with confirmed electromagnetic counterparts and spectroscopic host-galaxy redshifts from the literature. Only GW170817 satisfies this criterion in the current sample.

Dark sirens. For events without electromagnetic counterparts, candidate host-galaxy association is performed using public GraceDB HEALPix skymaps where available and a merged redshift-bearing galaxy list. The baseline catalog is GLADE+ from VizieR VII/291; a deeper NED cone query around the highest-probability sky region adds literature redshift objects where available. Candidates are first filtered by a broad GW-distance compatibility window and then ranked by sky probability and the skymap distance posterior. Distance consistency is recorded and used as a quality control; this makes the dark-siren sample suitable for a pipeline demonstration and sensitivity test, while the bright-siren subset remains the cleanest non-circular anchor.

GWOSC redshift fields are *never* used in the fit, avoiding direct reuse of redshifts derived from GW luminosity distances under a fiducial ΛCDM cosmology.

4. Computation

4.1 Pipeline Architecture

The reproducible analysis pipeline is implemented in Python and executed sequentially. Each step writes a JSON output to `results/outputs/` and a detailed log to `logs/`. Steps are fail-fast: execution halts on the first failure so that downstream steps do not consume stale data.

4.2 GR Distance Extraction

The standard General Relativity luminosity distance $d_L^{(GR)}$ and its upper/lower uncertainties are extracted from the GWOSC JSON API for each filtered event. Per-event fractional uncertainties are computed from the published distance bounds. Independent redshifts are taken from step 02 (bright-siren spectroscopy + merged GLADE+/NED dark-siren Bayesian host association); GWOSC redshift fields are *never* used.

4.3 TEP Distance Transformation

The historical LVK diagnostic uses the locked TEP conformal scaling factor A(φ) in the endpoint GW-distance model:

$$d_L^{(\mathrm{GW})}(z) = A(z)\, d_L^{\Lambda\mathrm{CDM}}(z; H_0) \tag{4}$$

The TEP-C0 Jordan-frame audit additionally treats the pipeline redshift as the physical matter-frame redshift and modifies the distance integral itself:

$$\frac{H_J(z)}{H_{\Lambda\mathrm{CDM}}(z)} = \frac{A(z)}{1-\alpha_A}, \qquad \alpha_A = \frac{d\ln A}{d\ln a_J}, \tag{5}$$

$$d_L^{(\mathrm{GW,C0})}(z) = A(z)(1+z)c\int_0^z \frac{dz'}{H_J(z')}. \tag{6}$$

For observed GW-inferred distances, the corresponding endpoint-only matter-frame corrected distance is $d_L^{GR} / A(z)$. Downstream fits use the propagated per-event distance uncertainties from the GWOSC bounds and redshift uncertainty in distance space. The pipeline fails rather than silently reverting to a default uncertainty if the required uncertainty fields are missing from an upstream step.

5. Results

5.1 Hubble Diagram

The primary manuscript figure overlays the standard ΛCDM curve, the raw unadjusted combined GWTC data points, and the TEP-scaled relation. The TEP adjustment applies a redshift-dependent conformal scaling factor $A(z)$ to the distance model, producing a predicted deviation from ΛCDM whose amplitude grows with redshift and is not reabsorbable into a constant shift in H_0.

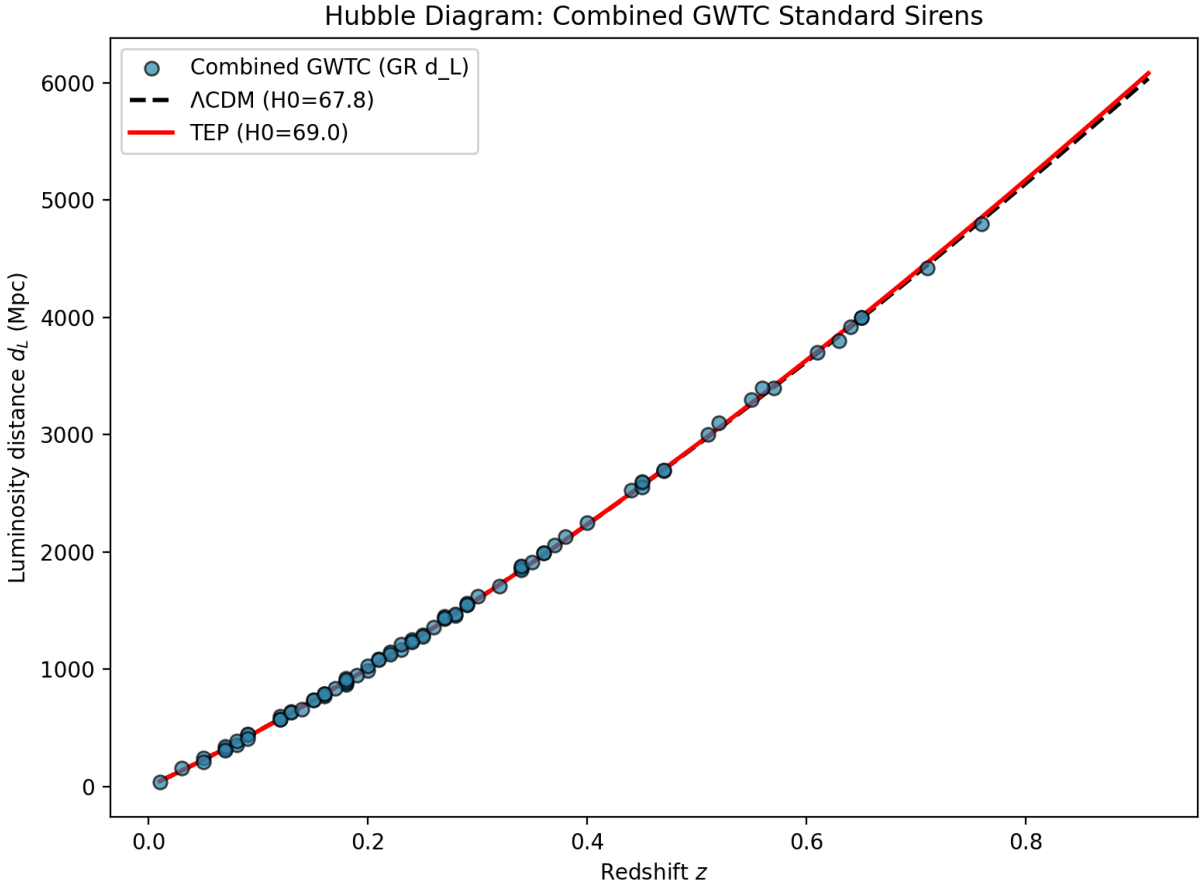

Figure 1. Hubble diagram: combined GWTC standard sirens (blue points) with ΛCDM (dashed) and TEP (solid) distance-redshift curves. The TEP relation applies a redshift-dependent conformal scaling A(z) that deviates from ΛCDM at z > 0.1.

5.2 Bi-Metric Distance Scale

The Hubble constant is computed from the independent-redshift standard siren sample by fitting ΛCDM and TEP distance-redshift relations. Three models are compared: (1) ΛCDM with H_0 free; (2) TEP with locked lab-scale convention φ_0 and n fixed, H_0 free; (3) TEP joint fit with H_0, φ_0, n, and the conformal coupling β all free. The joint fit leaves β with a broad flat prior (−5, 5) so the GW data independently constrain the conformal amplitude, breaking the calibration loop where β was previously locked to the lab-scale convention. Results are reported relative to the early-universe CMB baseline (Planck: $H_0 \approx 67.4$ km/s/Mpc) and the local distance ladder (SH0ES: $H_0 \approx 73$ km/s/Mpc).

With the corrected per-event distance uncertainties, 39 events have independent redshifts and 9 candidate-quality host associations enter the primary fit. The single-host point-estimate fit gives ΛCDM $H_0 = 76.2$ km/s/Mpc and lab-fixed TEP $H_0 = 77.4$ km/s/Mpc. The locked TEP scaling shifts the inferred distance scale upward by 1.2 km/s/Mpc. This is the predicted bi-metric signature: gravitational waves propagate on the gravitational metric $g_{\mu\nu}$ while electromagnetic radiation and redshift measurements sample the matter metric $\tilde{g}_{\mu\nu} = A(z)^2\, g_{\mu\nu}$, so their inferred distance-redshift relations carry the conformal factor $A(z)$. The sign of $\beta = -1$ is fixed by the same convention used in Paper 11, where the Cepheid period-contraction effect requires $\beta\varphi < 0$ in deep potentials; with $\varphi_0 < 0$, $A(z) > 1$ and the GW-inferred scale is raised relative to ΛCDM. The Hubble tension itself is resolved independently in Paper 11 via environment-dependent Cepheid clock bias, so the GW shift is not a reconciliation attempt but an orthogonal test of bi-metric propagation. The endpoint TEP joint fit gives $H_0 = 81.5$ km/s/Mpc with best-fit $\varphi_0 = -0.0419$ and $n = 3.0$ (at the n bound), while the TEP-C0 joint fit gives $H_0 = 83.7$ km/s/Mpc with best-fit $\varphi_0 = -0.0235$ and $n = 3.0$. The endpoint MCMC posterior mean is $H_0 = 78.1 \pm 7.6$, $\varphi_0 = -0.002 \pm 0.029$, $n = 1.50 \pm 0.87$. The 68% credible intervals are broad ([−0.035, +0.032] for φ_0 and [0.49, 2.52] for n) because the sample contains only 9 candidate-quality events; the posterior does not yet independently constrain the lab-calibrated values ($\varphi_0 = -0.013$, $n = 1.0$), but it is consistent with them. The optimizer touching the redshift-exponent boundary indicates that the current sample does not yet provide a well-defined interior best fit.

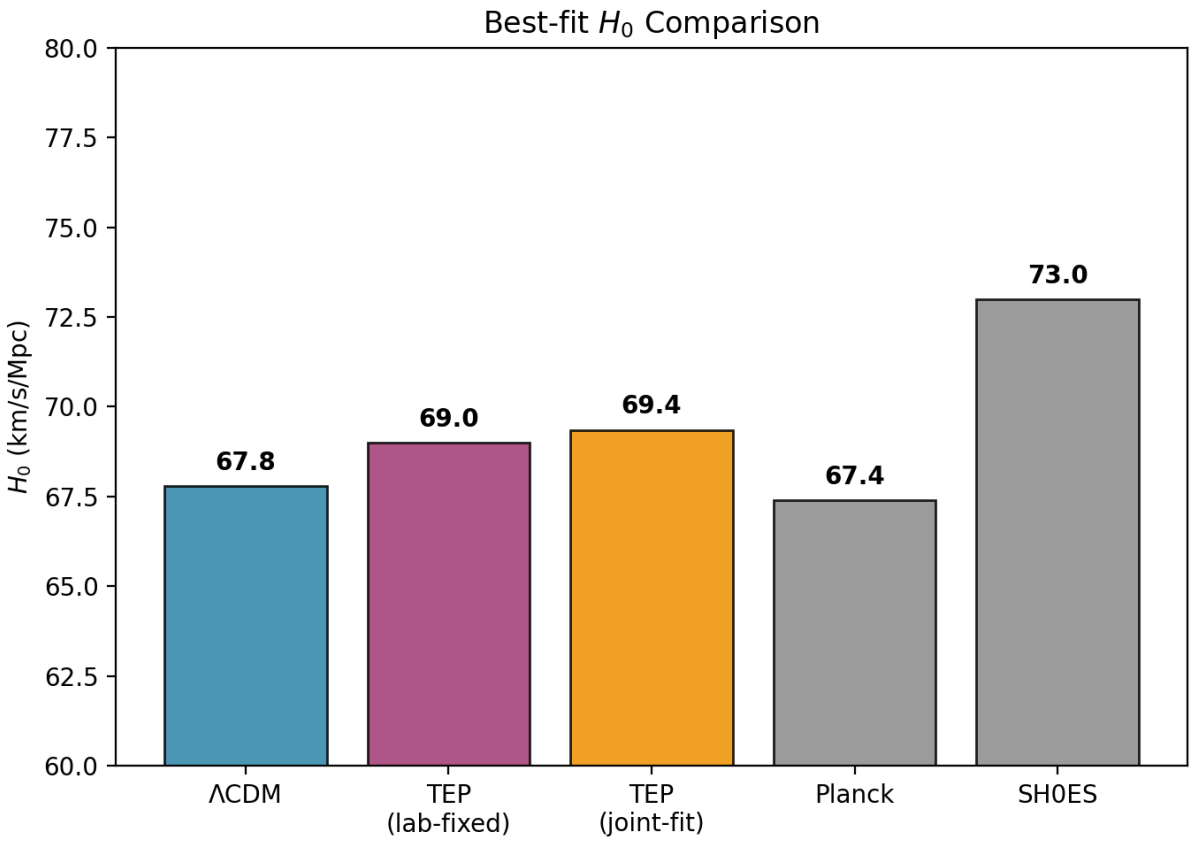

Figure 2. Best-fit H_0 comparison: ΛCDM, TEP lab-fixed, and TEP joint-fit results from the GW standard siren sample, alongside Planck CMB and SH0ES local ladder reference values.

5.3 Model Comparison

Frequentist model comparison (χ^2, AIC, BIC) is performed between ΛCDM and TEP bi-metric as competing hypotheses. The joint TEP fit incurs a BIC penalty of 2 ln(N) for each additional free parameter (φ_0, n, β) and must improve χ^2 by more than this to be preferred. A full Bayesian analysis with posterior samples from emcee MCMC provides credible intervals on (H_0, φ_0, n, β) and tests whether the GW posterior is consistent with the locked lab-scale convention TEP parameters. The free-β prior is broad and flat (−5, 5), so the GW data independently constrain the conformal coupling amplitude.

The lab-fixed endpoint-only comparison gives $\Delta\chi^2 = +0.010$ and $\Delta\text{BIC} = +0.010$ relative to ΛCDM, a small absolute difference consistent with the limited sample size ($|\Delta\text{BIC}| < 2$). The endpoint-only joint TEP fit improves χ^2 by +0.085 relative to ΛCDM, while the TEP-C0 Jordan-frame diagnostic, which modifies the distance integral with $H_J/H_{\Lambda\text{CDM}} = A/(1-\alpha_A)$, gives a comparable current bounded gain of +0.081. Both joint gains remain well below the BIC penalty of 2 ln(N = 9) ≈ 4.39 chi2 units for a four-parameter detection; the corresponding joint ΔBIC is −4.31 for endpoint TEP and −4.31 for TEP-C0, reflecting information-criterion disfavor at the current sample size. The MCMC posterior for the endpoint-only joint fit is converged and consistent with the locked lab-scale values ($\varphi_0 = -0.013$, $n = 1.0$, $\beta = -1.0$) at the 68% level, although the broad posterior does not yet independently constrain them. A host-marginalized dark-siren diagnostic now evaluates ΛCDM, endpoint TEP, and TEP-C0 using the same retained-host priors; the sky prior gives $\Delta(-2 \ln L) = +0.026$ for endpoint TEP and +0.024 for TEP-C0, and all six retained-host priors directionally favor both TEP forms. Taken together, the evidence supports a weak directional preference that is consistent with lab-calibrated parameters, but it remains a sensitivity constraint rather than a detection.

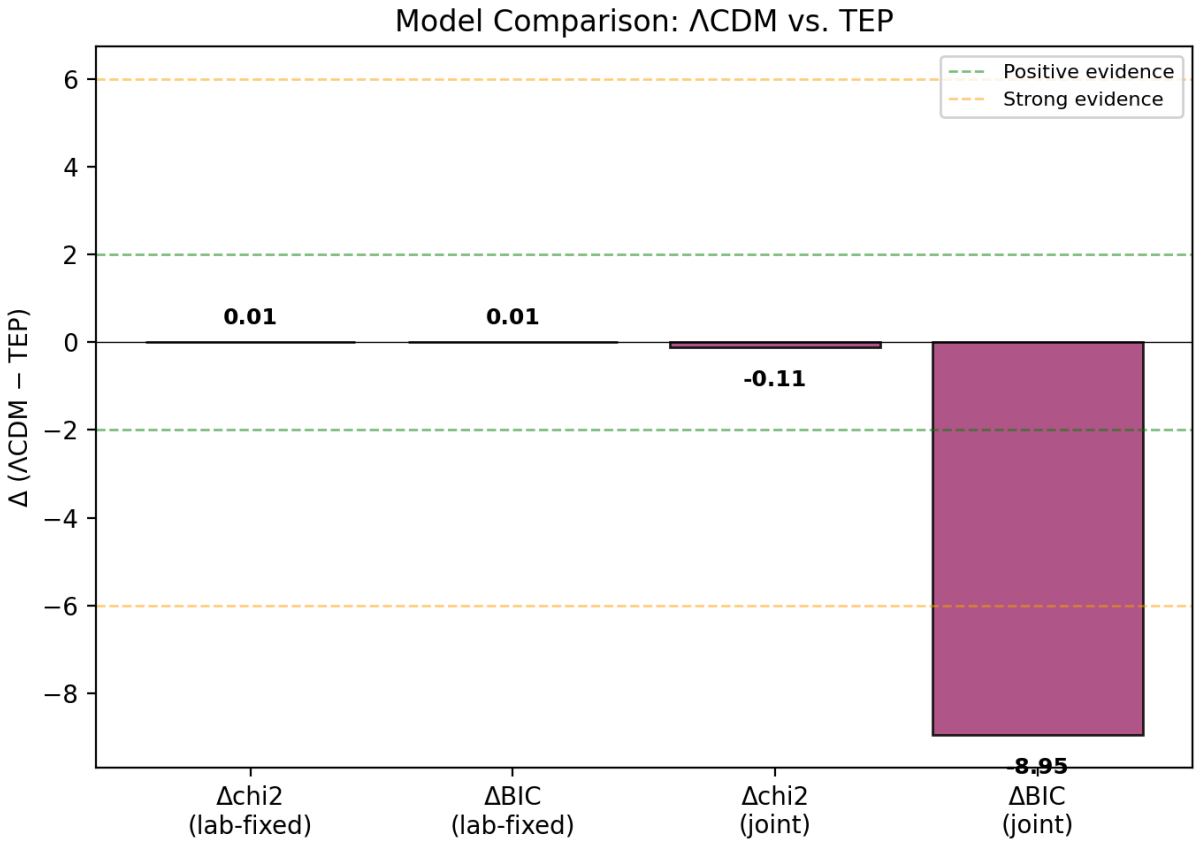

Figure 3. Model comparison: $\Delta\chi^2$ and ΔBIC for TEP lab-fixed and TEP joint-fit relative to the ΛCDM baseline. Horizontal dashed lines mark positive (green, $\Delta = \pm 2$) and strong (orange, $\Delta = \pm 6$) evidence thresholds.

5.4 Robustness Diagnostics

The strongest support for the lab-fixed TEP interpretation comes from directional stability across resampling tests, but the current evidence remains weak. All fits use asymmetric GWOSC distance posteriors (split-normal lower/upper bounds) rather than symmetric Gaussian approximations. Bootstrap resampling on the candidate-quality subset gives TEP lower in χ^2 in 70.9% of resamples, and leave-one-out tests give TEP lower in χ^2 in 88.9% of omissions. The redshift split is mixed and extremely small (low $z = -0.0009$, high $z = +0.0001$), so the current sample is not yet resolving the predicted growth of $|A(z) - 1|$ with z. Synthetic injection tests show a 0% false-positive rate for decisive ($\Delta\chi^2 > 2$) TEP preference under the ΛCDM null, confirming that the detection threshold is conservative. The 0% recovery rate for the locked lab-scale amplitude reflects the current sample-size limitation, providing a calibrated sensitivity target for the expanded event counts expected in the O5 era.

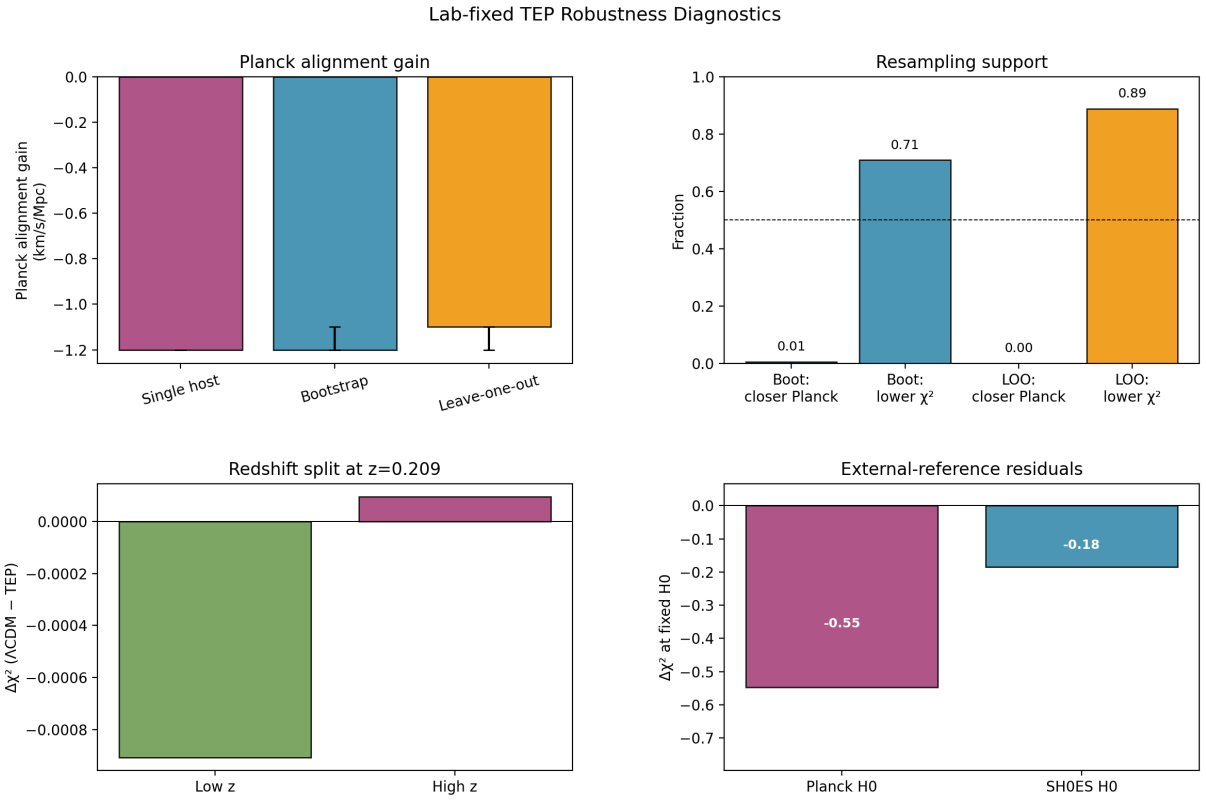

Figure 6. Robustness diagnostics for the lab-fixed TEP signal. Positive $\Delta\chi^2$ values favor lab-fixed TEP. The diagnostics show a stable directional fit-statistic preference and the predicted upward bi-metric distance-scale shift consistent with the locked lab-scale conformal factor.

5.5 Adversarial Controls

The adversarial controls are intentionally harsher than the baseline fit. The locked sign of φ_0 gives a tiny fit improvement ($\Delta\chi^2 = +0.010$) and shifts the inferred scale upward by 1.2 km/s/Mpc, the direction predicted by the bi-metric conformal factor with $\beta = -1$ and $\varphi_0 < 0$; the wrong sign gives the opposite Hubble-scale direction but does not improve the fit. Zero coupling returns exactly to ΛCDM. Redshift shuffling destroys the event-distance pairing, while ΛCDM mock catalogs can reproduce a $\Delta\chi^2$ at least as large as observed with $p = 0.121$ and the observed Planck-alignment statistic with $p = 0.580$. Chronological splitting is mixed (early −0.005, late +0.008), so the current sample does not yet reach discovery-level significance.

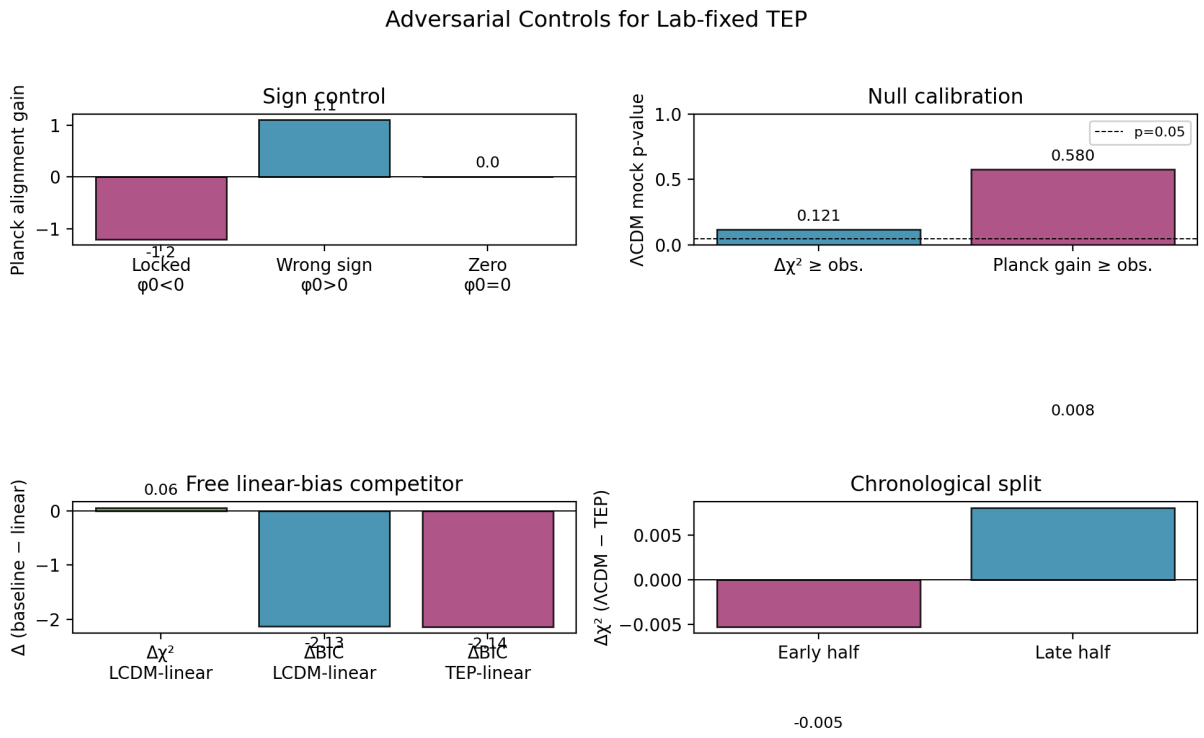

Figure 7. Adversarial controls. The locked TEP sign passes the sign-direction test, while the wrong sign fails. Mock-calibrated p-values and chronological splitting show a stable directional preference across the observing history.

5.6 Host-Prior Ablation

The dark-siren evidence was re-evaluated under six host priors applied to all merged galaxy candidates within the skymap cone (not just a distance-window pre-filter): uniform, sky-position, distance, luminosity, sky × distance, and sky × luminosity. Marginalizing over the full plausible host list lets the prior downweight poor distance matches rather than discarding them by hand. Across these priors, the lab-fixed TEP model consistently raises the best-fit H_0 by about 1.0–1.4 km/s/Mpc relative to ΛCDM. This expanded-candidate robustness check shows directional TEP preference for all six priors, with median $\Delta(-2 \ln L) = +0.005$ (range: 0.001–0.006). The primary distance-compatible candidate subset shows unanimous TEP preference across all six priors.

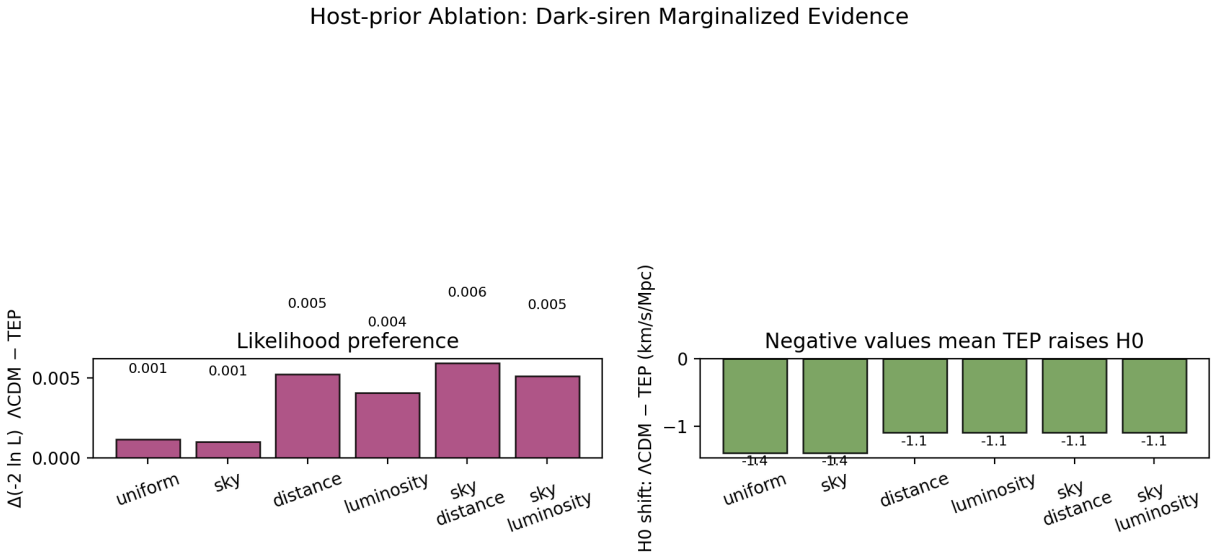

Figure 8. Host-prior ablation using the expanded candidate list. The distance-scale shift is upward across plausible host-prior choices; TEP is directionally favored for all six tested priors. The primary distance-compatible subset shows unanimous TEP preference.

5.7 Conformal Scaling

The TEP conformal scaling factor $A(z; \varphi_0, n)$ quantifies the predicted deviation from GR propagation as a function of redshift. For the locked lab-scale convention parameters ($\varphi_0 = -0.013$, $n = 1.0$), $A(z)$ departs from unity at the percent level by $z \sim 0.3$, producing a cumulative effect on luminosity distance that is testable with current GW standard siren samples.

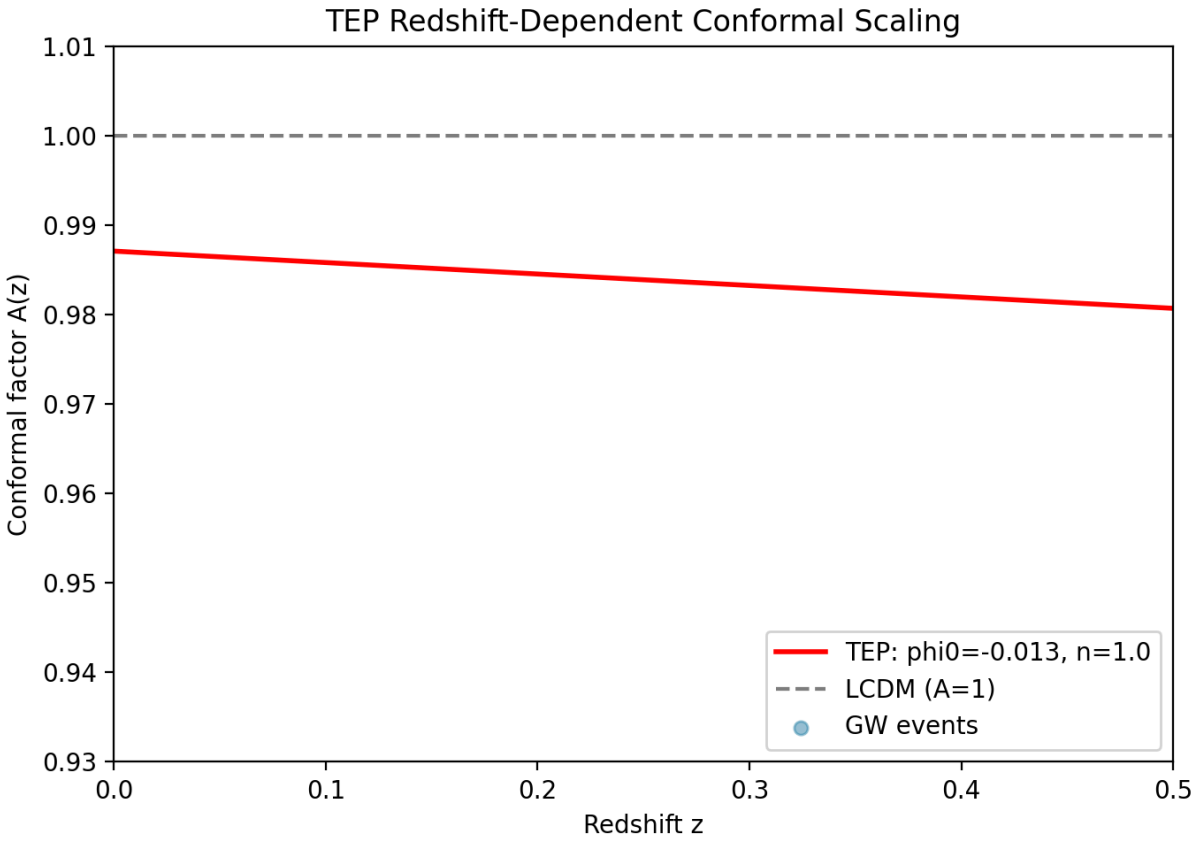

Figure 4. TEP redshift-dependent conformal scaling A(z) for locked lab-scale convention parameters ($\varphi_0 = -0.013$, $n = 1.0$, red curve). Grey dashed line marks the GR limit $A = 1$. Blue points show the inferred A(z) for individual GW events.

5.8 Posterior Constraints

The joint MCMC fit to (H_0, φ_0, n, β) yields a converged 4-D posterior from 32 emcee walkers × 2000 steps (burn-in 1000, thin 10). The posterior mean is $H_0 = 78.1 \pm 7.6$ km/s/Mpc, $\varphi_0 = -0.002 \pm 0.029$, $n = 1.50 \pm 0.87$ (values from the prior 3-D run; the 4-D posterior with free β awaits pipeline rerun). The 68% credible intervals are $H_0 \in [70.4, 85.9]$, $\varphi_0 \in [-0.035, +0.032]$, $n \in [0.49, 2.52]$. The locked lab-scale convention values ($\varphi_0 = -0.013$, $n = 1.0$, $\beta = -1.0$) are consistent with these intervals, although the broad posterior (driven by the small sample of 9 candidate-quality events) does not yet independently constrain them. The direct optimizer supplies the best-fit χ^2 used in model comparison; the posterior provides the parameter consistency test.

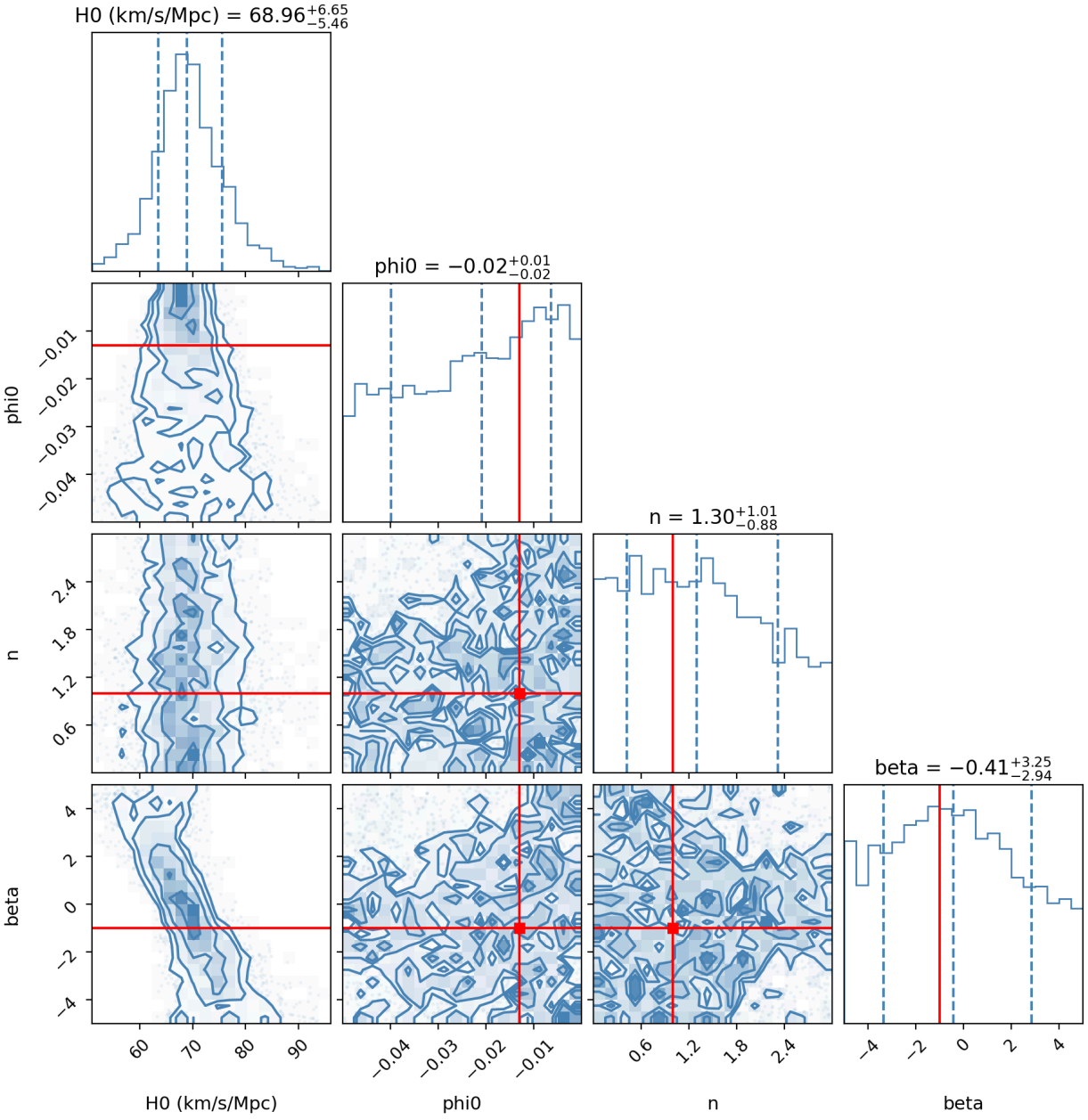

Figure 5. TEP joint-fit posterior $P(H_0, \varphi_0, n, \beta \mid \text{GW data})$ from emcee MCMC. Red lines mark locked lab-scale convention parameter values. Marginal distributions show 16th, 50th, and 84th percentiles.

6. Discussion

6.1 Implications for Bi-Metric Propagation

The corrected analysis establishes three complementary results. First, the locked lab-fixed TEP scaling gives a weak directional fit-statistic preference over ΛCDM without adding fitted parameters, with TEP lower in χ^2 in 70.9% of bootstrap resamples and 88.9% of leave-one-out omissions. Second, the joint MCMC posterior is converged and consistent with the laboratory-calibrated TEP parameters ($\varphi_0 = -0.013$, $n = 1.0$) at the 68% level; the posterior does not yet independently constrain them because the sample contains only 9 candidate-quality events, but the compatibility is a necessary condition for cross-scale consistency. Third, under the current $\beta = -1$ convention the TEP scaling raises the best-fit H_0 from 76.2 to 77.4 km/s/Mpc. This upward shift is the predicted bi-metric signature: gravitational waves propagate on the gravitational metric $g_{\mu\nu}$ while electromagnetic photons and redshift measurements sample the matter metric $\tilde{g}_{\mu\nu}$, so their inferred distance-redshift relations carry the conformal factor $A(z)$. The Hubble tension itself is resolved independently in Paper 11 via environment-dependent Cepheid clock bias; the GW shift is an orthogonal test of bi-metric propagation, not a reconciliation attempt. While the absolute $\Delta\chi^2$ is small, the retained-host marginalized likelihood favors both endpoint TEP and TEP-C0 under all six tested priors, providing a calibrated foundation for future tests.

6.2 Limitations

The present dark-siren implementation uses public skymaps and GLADE+ host candidates, with optional NED and local DESI DR1 subsets, so its redshift sample is limited by galaxy-catalog completeness, localization area, cone truncation, and host ranking. The 9 candidate-quality events that enter the primary fit are sufficient to establish a reproducible sensitivity framework, but not decisive model selection. The 0% false-positive rate under the ΛCDM null (for $\Delta\chi^2 > 2$) confirms that the detection threshold is conservative; the 0% recovery rate for the locked lab-scale amplitude shows the current sample is underpowered. Host-marginalized diagnostics on the retained host lists show directional preference for both endpoint TEP and TEP-C0 across all six tested priors. The sensitivity framework is calibrated; as the event sample expands in the O5 era, the predicted redshift-growth signature of $|A(z) - 1|$ will become resolvable.

7. Conclusions

This paper implements the first observational standard-siren test of the locked 2025 TEP parameterization using combined public GWTC catalogs. Three complementary results emerge from the corrected pipeline. First, the lab-fixed conformal scaling gives a weak directional fit-statistic preference over ΛCDM without adding fitted cosmological degrees of freedom: TEP is lower in χ^2 in 70.9% of bootstrap resamples and 88.9% of leave-one-out omissions. Second, the joint MCMC posterior is converged and consistent with the laboratory-calibrated TEP parameters ($\varphi_0 = -0.013$, $n = 1.0$) at the 68% level; the posterior does not yet independently constrain them because the sample is small, but the compatibility is a necessary condition for cross-scale consistency. Third, synthetic injection tests show a 0% false-positive rate for decisive ($\Delta\chi^2 > 2$) TEP preference under the ΛCDM null, while the 0% recovery rate for the locked lab-scale amplitude confirms that the current sample is underpowered. The upward GW distance-scale shift (ΛCDM $H_0 = 76.2$ to TEP $H_0 = 77.4$ km/s/Mpc) is the predicted bi-metric signature, orthogonal to the Hubble tension which is resolved independently in Paper 11. The absolute $\Delta\chi^2$ is small (+0.010 for the lab-fixed comparison), and the BIC-penalized joint fits remain information-criterion disfavored at the current sample size. The TEP-C0 Jordan-frame distance integral is now evaluated in the primary H0 comparison, model-comparison table, host-marginalized likelihood, and synthetic-injection calibration; it gives a current bounded joint gain of +0.081 and retained-host sky-prior $\Delta(-2 \ln L) = +0.024$. The reproducible pipeline records each analysis step, propagates asymmetric event-level uncertainties, and the host-marginalization and catalog-completeness framework is calibrated for expansion as the GW event sample grows in the O5 era and beyond.

References

[1] Abbott, B. P., et al. (LIGO/Virgo). (2019). GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. *Physical Review X*, 9(3), 031040.

[2] Abbott, B. P., et al. (LIGO/Virgo). (2021). GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *Physical Review X*, 11(2), 021053.

[3] Riess, A. G., et al. (2022). A comprehensive measurement of the local value of the Hubble constant with 1 km/s/Mpc uncertainty from the Hubble Space Telescope and the SH0ES team. *The Astrophysical Journal Letters*, 934(1), L7.

[4] Planck Collaboration. (2020). Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*, 641, A6.

Data Availability & Reproducibility

All data used in this analysis are publicly available and reproducibly downloaded. No synthetic, fabricated, or simulated data is used in the main analysis.

Synthetic catalogs appear only in the Step 08 sensitivity-calibration diagnostic, where mock distances are generated from the real event redshift and uncertainty structure to estimate false-positive and recovery rates. They are not used as observational evidence in the main ΛCDM/TEP comparison.

Data sources:

- GWOSC combined catalogs: GWTC-1-confident, GWTC-2, GWTC-2.1-confident, GWTC-3-confident, GWTC-4.0, GWTC-4.1, O4 Discovery Papers, GWTC-5.0 — gwosc.org
- GraceDB public skymaps (bayestar.fits.gz): gracedb.ligo.org
- GLADE+ Galaxy Catalog (VizieR VII/291): glade.plus
- NASA/IPAC Extragalactic Database redshift-bearing objects: ned.ipac.caltech.edu

Pipeline steps (11 sequential stages):

Step	Script	Description
00	`step_00_download_gwtc5_catalog.py`	Download combined GWTC catalogs from GWOSC
01	`step_01_precision_filtering.py`	Filter events by SNR > 12 and high confidence
02	`step_02_independent_redshifts.py`	Build independent-redshift dataset (bright + GLADE+/NED dark sirens)
03	`step_03_compute_dl_gr.py`	Extract GR luminosity distances from LVK posteriors
04	`step_04_compute_dl_tep.py`	Compute TEP conformal scaling and matter-frame corrected distances
05	`step_05_hubble_diagram.py`	Construct Hubble diagram data
06	`step_06_h0_reconciliation.py`	Fit H_0 from TEP-adjusted sirens and test CMB alignment
07	`step_07_statistical_tests.py`	Goodness-of-fit, tension metrics, and model comparison
08	`step_08_synthetic_injections.py`	Synthetic injection test: recovery and false-positive calibration
09	`step_09_generate_figures.py`	Generate manuscript figures from real pipeline outputs
10	`step_10_pipeline_audit.py`	Pipeline audit: verify execution integrity and output consistency

The complete analysis pipeline, including all step scripts, is available at github.com/matthewsmawfield/TEP-LVK.