Putting Flat ΛCDM In The (Redshift) Bin

E. Ó Colgáin,[1,2] M. M. Sheikh-Jabbari,[3,4] R. Solomon,[5] M. G. Dainotti,[6,7,8] and D. Stojkovic[5]

[1]*CQUeST & Department of Physics, Sogang University, Seoul 121-742, Korea*
[2]*Atlantic Technological University, Ash Lane, Sligo, Ireland*
[3]*School of Physics, Institute for Research in Fundamental Sciences (IPM), P.O.Box 19395-5531, Tehran, Iran*
[4]*The Abdus Salam ICTP, Strada Costiera 11, I-34014, Trieste, Italy*
[5]*HEPCOS, Department of Physics, SUNY at Buffalo, Buffalo, NY 14260-1500, USA*
[6]*National Astronomical Observatory of Japan, 2 Chome-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*
[7]*The Graduate University for Advanced Studies, SOKENDAI, Shonankokusaimura, Hayama, Miura District, Kanagawa 240-0193, Japan*
[8]*Space Science Institute, Boulder, CO, USA*

Flat ΛCDM cosmology is specified by two constant fitting parameters in the late Universe, the Hubble constant H_0 and matter density (today) Ω_m. In the cosmology literature, one typically *assumes* that there is no redshift evolution of cosmological parameters when one fits data sets. Here, in mock observational Hubble data we demonstrate evolution in distributions of best fit parameters with effective redshift. As a result, considerably different (H_0, Ω_m) best fits from Planck-ΛCDM cannot be precluded in high redshift bins. We explore if observational Hubble data, Type Ia supernovae and standardisable quasar samples exhibit redshift evolution of best fit ΛCDM parameters. In all samples, we confirm an increasing Ω_m (decreasing H_0) trend with increasing bin redshift. Through comparison with mocks, we confirm that similar behaviour can arise randomly within the flat ΛCDM model with probabilities as low as $p = 0.0021$ $(3.1\,\sigma)$.

I. INTRODUCTION

Cosmologists are currently debating tensions within the flat ΛCDM cosmology; the two most serious concern the Hubble constant H_0 and the $S_8 \propto \sqrt{\Omega_m}$ parameter [1, 2] [65]. These tensions have been framed as disagreements between the early (high redshift) and late (low redshift) Universe [3]. In particular, local H_0 values [4–8] are universally biased to larger values than Planck-ΛCDM [9]. Observations at different redshifts have shown that H_0 evolves with effective (binned) redshift in the flat ΛCDM model [10–18] (see also [19]). If this trend is not due to observational selection biases, and it is intrinsic, this behaviour is indicative of model breakdown [20, 21].

The flat ΛCDM model Hubble parameter $H(z)$ is specified by two constant fitting parameters (H_0, Ω_m) or (A, B),

$$\begin{aligned} H(z)^2 &= H_0^2\left[1 - \Omega_m + \Omega_m(1+z)^3\right], \\ &= A + B(1+z)^3. \end{aligned} \tag{1}$$

The parameter $A := H_0^2(1-\Omega_m)$ is attributed to dark energy (DE), while the matter sector $B := H_0^2\Omega_m$ scales as $(1+z)^3$ and Ω_m is bounded, $0 \leq \Omega_m \leq 1$. One can relax this constraint by allowing negative energy densities, but interpretation is problematic [66]. Observe that DE becomes irrelevant at higher redshifts, where $A \ll B(1+z)^3$ for reasonable values of Ω_m. On the other hand, note that at higher redshifts $H(z)^2 \sim B(1+z)^3$, so the combination $\Omega_m h^2$, with $h := H_0/100$, is the relevant quantity. Exploiting these facts, it was recently argued that increases in Ω_m (decreases in H_0) with effective redshift may be inherent to the flat ΛCDM model [16]. Here, we study ΛCDM mocks binned by redshift to uncover the mathematical fact that the probability of Planck values $\Omega_m \sim 0.3$ decreases as we increase bin redshift. As a result, some evolution away from $\Omega_m \sim 0.3$ should be expected in best fits of purely high redshift observations.

Armed with this analytic insight, we turn to observed data in order to ascertain whether the same trend exists through comparison to mock simulations. We employ observational Hubble data (OHD), essentially cosmic chronometers [23] and baryon acoustic oscillations (BAO) [24, 25], Type Ia supernovae (SN) [26] and standardisable quasar (QSO) data sets [27]. Throughout we compare values of (H_0, Ω_m) to mock simulations in the *same* redshift range, where the base cosmology for the mock is fixed by the best fit parameters of the *entire* data set. This allows us to confirm evolution between low and high redshifts in the sample.

Ultimately, while the fit of the overall sample to flat ΛCDM is largely dictated by the redshift range with greater density of data points, we will see that in sparser redshift ranges, the data prefers different cosmological parameters. In particular, we find probabilities as low as $p = 0.021$ (OHD), $p = 0.081$ (SN) and $p = 0.019$ (QSOs), respectively, that mock data leads to similar values of (H_0, Ω_m) as observed data. Combining the independent probabilities using Fisher's method, one arrives at the probability $p = 0.0021$ $(3.1\,\sigma)$ that such an evolution indeed exists within flat ΛCDM. An explanation in terms of selection biases is plausible for SN, e. g. [13, 15], but similar effects must impact cosmic chronometers, BAO, etc. Our mock analysis shows that *without selection biases* evolution away from Planck values should be expected.

II. MOCK DATA

Consider a simple data fitting exercise, where one takes Dark Energy Spectroscopic Instrument (DESI) forecasts for $H(z)$ errors $\sigma_{H(z_i)}$ at redshifts z_i in the range $0.05 \leq z_i \leq 3.55$ [22]. Next, adopt Planck values [9], $H_0 = 67.36$, $\Omega_m = 0.315$, for an underlying model and generate $H(z_i)$ values in a nor-

mal distribution about the Planck-ΛCDM model using the errors $\sigma_{H(z_i)}$ as the standard deviation at each z_i. Throughout we fix the parameters for the underlying cosmology and do not pick (H_0, Ω_m) in a distribution. Picking (H_0, Ω_m) in a distribution adds randomness, but this randomness is expected to be subleading to the randomness introduced in the shifts of the data points. For each realisation of mock data, separate the data into four bins, concretely $0 < z < 0.8$, $0.8 \leq z < 1.5$, $1.5 \leq z < 2.3$ and $2.3 \leq z < 3.6$. This ensures similar data quality in each bin. Finally, fit the parameters (H_0, Ω_m) from (1) to the data in each bin with a Gaussian prior on $\Omega_m h^2 = 0.1430 \pm 0.0011$ [9]. Note that the prior only provides guidance for the high redshift behaviour of $H(z)$ and its omission cannot change results (see appendix). Repeat the process a few thousand times and record the distribution of best fit values of (H_0, Ω_m) for each bin.

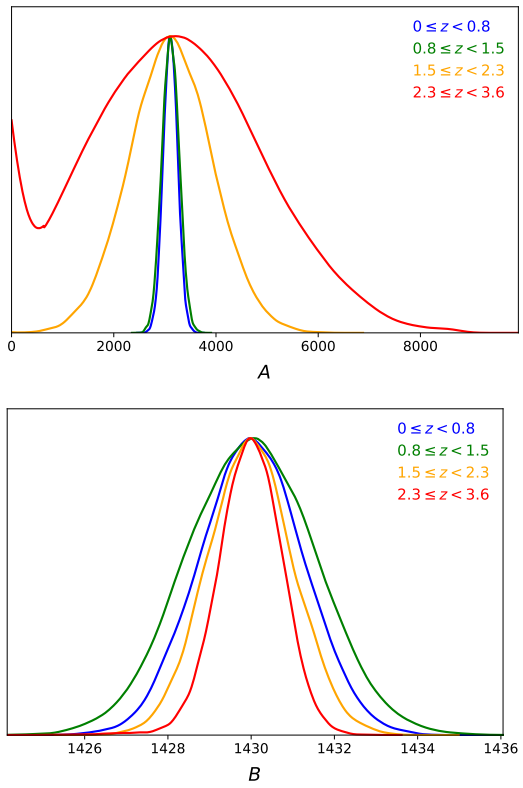

FIG. 1. Distributions of $A = H_0^2(1 - \Omega_m)$ and $B = H_0^2\Omega_m$ parameters reconstructed from mock simulations of the Planck-flat-ΛCDM model in different redshift bins.

Before turning our attention to H_0, Ω_m best fit distributions, let us report on the (unnormalised) distributions for A, B. Fig. 1, produced with *GetDist* [28], demonstrates that both A and B are Gaussian by inspection, except where A is impacted by the boundary at $A = 0$. Note, we have imposed a Gaussian prior on B, so B being Gaussian is expected. Observe that the distributions in A and B spread and narrow, respectively, with increasing bin redshift. Interestingly, the distribution in B spreads from bin 1 to bin 2 before narrowing in bins 3 and 4. This apparently contradicts our claim that A spreads and B narrows, but it can be traced to data quality differences with redshift in the DESI forecast [22]. If one ensures the *same* data quality in all bins, then A spreads and B narrows with redshift. We demonstrate this in the appendix. This outcome is expected as the ΛCDM model (1) transitions from a two-parameter to an effective one-parameter model at high redshift. We have checked that A and B are uncorrelated (see appendix). We also see that A grows a non-Gaussian tail around the $A = 0$ ($\Omega_m = 1$) region at higher redshift bins. This comes about as a Gaussian with a wide spread probes the $A < 0$ region with a growing probability in higher z bins, which we have dubbed a 'pile up' feature. Moreover, the width of the Gaussian distribution for $B = H_0^2\Omega_m$ reduces as we go to higher redshift bins and hence we know B with a better precision in the higher redshift bins. Higher redshift spread in $A = H_0^2 - B$ then yields spread in both H_0 and Ω_m values.

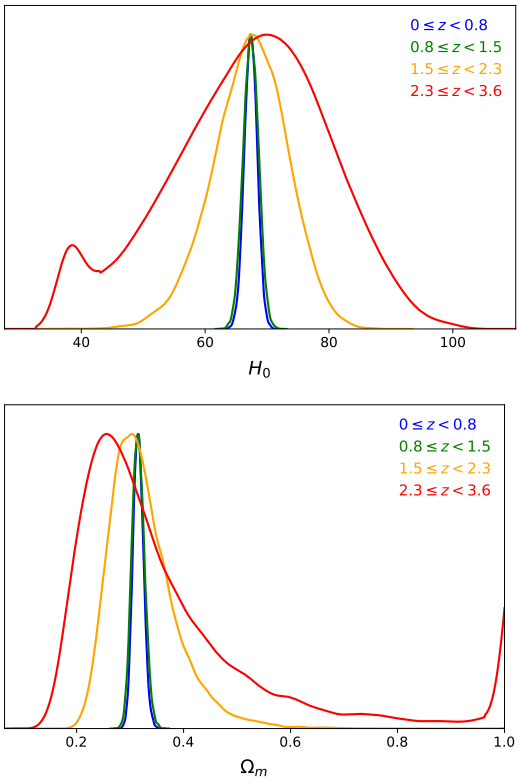

FIG. 2. Distributions of the cosmological parameters in different redshift bins. The 'pile up' at $\Omega_m \sim 1$ and $H_0 \sim 37.8$ km/s/Mpc is due to $\Omega_m > 1$ best fits being restricted to the bound $\Omega_m = 1$.

In Fig. 2 we show the same distribution in (H_0, Ω_m) parameters. It is evident that both H_0 and Ω_m develop long non-Gaussian tails in the direction of smaller H_0 and larger Ω_m despite input Planck values in the mocking procedure, confirming our analytic expectations discussed above. This is easily explained. Since $\Omega_m h^2$ is well constrained, best fit (H_0, Ω_m) values inhabit a $\Omega_m H_0^2 \sim$ constant curve or banana. Nevertheless, as the banana stretches, configurations move from the peak to the extremities, leading to shifts in the peak when projected onto the H_0 and Ω_m axes. Thus, the Ω_m peak shifts to lower values, whereas the H_0 peak shifts to higher values.

This comes from a "projection effect" in the mock data. The pile up at $\Omega_m = 1$ is an artefact of our priors, but this can be relaxed without changing the conclusions. See [63] for a more complete analysis. Our analysis here only concerns $H(z)$, but angular diameter distance $D_A(z)$ constraints, and the combination $H(z) + D_A(z)$, are studied in [63].

Since our mocking procedure is the same in each bin, while data quality does not change greatly (see [22]), one concludes that the behaviour is generic to the flat ΛCDM model. Note also that selection biases do not impact mocks. Moreover, the same argument can be run for *any* mock input parameters (H_0, Ω_m). The main message is that even in a Universe statistically consistent with Planck-ΛCDM by construction, unfamiliar best fit values can easily be returned in data fitting. Furthermore, best fits in the $\Omega_m > 1$ regime of parameter space are possible. See related analysis with Pantheon+ SN [64].

III. OBSERVED DATA

Having uncovered a general feature for $H(z)$ constraints confronted to the flat ΛCDM model, we now explore this prediction with observed OHD. In [63] we show that $D_L(z) \propto D_A(z)$ constraints confronted to flat ΛCDM exhibit similar features, which justifies studying Type Ia SN and QSOs.

A. Comments on Methodology

As we have seen, one encounters non-Gaussian distributions in exclusively high redshift bins (see also [63]). As a result, while best fits, i. e. the extrema of χ^2, are expected to be robust within machine precision [67], estimating errors as is usually done in cosmology is difficult. More explicitly, Fisher matrix leads to unrepresentative Gaussian errors, while Markov Chain Monte Carlo (MCMC) inferences are prone to degeneracies/projection effects that distort inferences. Moreover, with broad distributions it is possible that MCMC inferences are simply tracking the priors and the peaks of distributions are not guaranteed to coincide with the minimum of the χ^2. To overcome this difficulty, it is standard practice in cosmology to combine data sets and/or introduce priors that restore Gaussianity, but it is more agnostic to analyse data sets independently without imposing additional constraints. We highlight the explicit difficulty with MCMC analysis in the appendix.

Given the difficulties with conventional techniques, here we resort to mocks that allow us to generate a large number of best fits that are *statistically consistent (by construction) with no evolution of cosmological parameters*. We make direct comparison between best fits from mocks and observed data *in the same redshift range with the same data quality*. This allows us to rank mock best fits of H_0 and Ω_m in descending and ascending order, respectively, and identify the percentile where observed data best fits appear. This gives us a probability for finding similar best fits *assuming no evolution in the sample*. Note, just as the shape of a PDF of heights of children in a class is irrelevant in such an exercise, the same logic also applies here.

In all samples, we note that the probabilities (see Tables I, II and III) of finding observed data best fits as extreme in mock data decrease as the effective redshift of subsamples becomes less representative of the full sample. This is expected if the trend is physical. However, the probabilities do not decrease indefinitely, and our results show that the probabilities increase again in the smallest subsamples, which we attribute to noise. It is intuitive that any signal in the data eventually disappears due to statistical fluctuations in small samples. Furthermore, given the probabilities decrease with increasing difference in effective redshift between subsample and the full sample, this means that this probability is bounded below. Thus, we do not pick redshift ranges by hand, but they emerge from the data as the redshift ranges where best fits in a subsample are least representative of the full sample. In other words, one can give a lower bound on probabilities and this lower bound is expected to be well defined.

We impose a strong Planck $\Omega_m h^2 = 0.1430 \pm 0.0011$ prior, which constrains best fits to a curve in the (H_0, Ω_m)-plane. In tandem we start χ^2-minimisation for each realisation of the data, either observed or mock, from the best fits of the full sample. As a result, if there is little or no evolution, one expects the best fits to not move far from the initial guess. In other words, we bias the initial guess towards no evolution. However, our strong Planck $\Omega_m h^2$ prior effectively reduces the fitting procedure to an effective 1-dimensional fit in the (H_0, Ω_m)-plane. What this means in practice is that we may find false minima, but these minima are the closest to the input parameters. Nevertheless, we think false minima are unlikely, given the effective 1-dimensional nature of the fitting. More concretely, note that even in OHD, we are performing an effective 1-dimensional fit with no less than 6 data points and it is hard to imagine that the outcome is not unique modulo machine precision.

B. OHD

Here, we make use of cosmic chronometer [29–35] and BAO data [36–45]. More precisely, we work with the $H(z)$ BAO determinations compiled in Table 2 of [46], where observations have been homogenised to be consistent with a uniform Planck inference of the sound horizon [47]. We added the newer constraint from eBOSS Quasar [48, 49], which we appropriately adjusted for the sound horizon, $H(z = 1.48) = 153.59 \pm 8.27$. Our total sample has 54 OHD sources. Moreover, we have checked that replacing earlier Lyman-α BAO [43–45] with the latest constraints [50] does not change the results, so we work with the earlier determinations collated in [46].

First, we identify the best fit values of the cosmological parameters for the full sample, $(H_0, \Omega_m) = (69.11, 0.299)$, where it is worth noting that $\Omega_m h^2 = 0.1428$, consistent with the prior. Next, we repeat the mocking and binning procedure outlined earlier with the new input parameters $(H_0, \Omega_m) = (69.11, 0.299)$. Following [16] we impose a low redshift cut-

off to remove sources below a given z and isolate high redshift bins. In each bin we compare the best fit values from the real data and flat ΛCDM mocks *in the same bin with the same data quality* in order to establish the probability of recovering *the same or larger* Ω_m and *the same or smaller* H_0 values. In the event of saturation of the bound $\Omega_m = 1$, this means that our probabilities are over-estimated, i. e. too large, since allowing $\Omega_m > 1$ permits further ordering of the values piled up at $\Omega_m = 1$. The results are shown in Table I, where it is clear that (H_0, Ω_m) best fits are evolving in the real data. For easy comparison throughout, we include the best fits for the full sample in tables, but do not assign any probabilities. Understandably, the probability of recovering similar values from mocks decreases with redshift up to a point where statistical fluctuations dominate and the probability increases again. Fig. 3 provides visual confirmation that despite the long tails, a bin exists where the real values are unexpected at $\gtrsim 2\sigma$. This points to redshift evolution in the sample, which may be expected given our discussions in the previous section.

z	H_0 (km/s/Mpc)	Ω_m	Probability
$0 \leq z \leq 2.36$ (54)	69.11	0.299	–
$0.5 \leq z \leq 2.36$ (28)	69.68	0.294	0.646
$0.7 \leq z \leq 2.36$ (18)	65.67	0.331	0.326
$1 \leq z \leq 2.36$ (11)	61.27	0.380	0.258
$1.2 \leq z \leq 2.36$ (10)	53.91	0.491	0.120
$1.4 \leq z \leq 2.36$ (8)	41.55	0.828	0.037
$1.45 \leq z \leq 2.36$ (7)	37.80	1	0.021
$1.5 \leq z \leq 2.36$ (6)	37.80	1	0.069

TABLE I. Best fit cosmological parameters for different redshift ranges of OHD. Throughout, we impose the Planck prior, $\Omega_m h^2 = 0.1430 \pm 0.0011$. Flat ΛCDM simulations based on best fit parameters over the entire redshift range, $0 < z \leq 2.33$, allow us to establish the probability of higher Ω_m and lower H_0 values in real data. Where a discrepancy exists, we quote the largest probability. The OHD count in each bin is denoted in brackets.

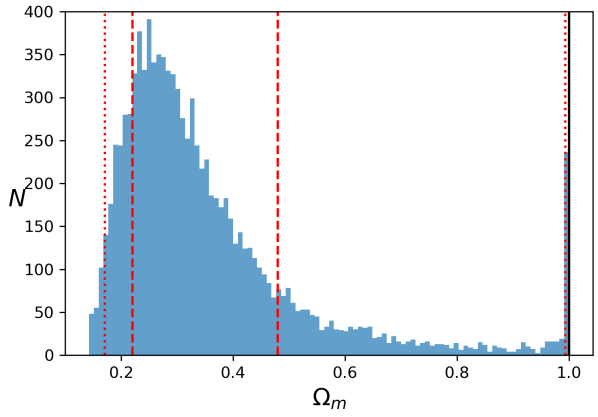

FIG. 3. Comparing 10,000 mock simulations with the best fit value of Ω_m from OHD data (black line) for the bin $1.45 \leq z \leq 2.36$. Dashed and dotted lines denote the $(2.3, 15.9, 84.1, 97.7)$ percentiles.

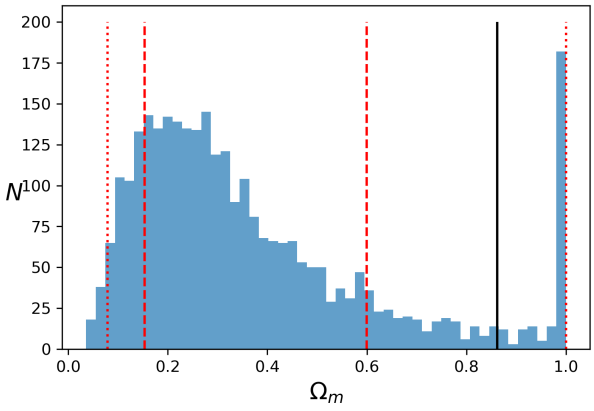

FIG. 4. Comparing 3,000 mock simulations with the best fit value of Ω_m from SN data (black line) for the bin $0.95 < z \leq 2.26$. Dashed and dotted lines denote the $(2.3, 15.9, 84.1, 97.7)$ percentiles corresponding to 1σ and 2σ confidence intervals for a Gaussian distribution.

C. Type Ia SN

We revisit the analysis of the Pantheon data set [26] with 1048 SN conducted in [16] (see also [13, 14]) in order to introduce a high redshift Planck prior on $\Omega_m h^2$ [9]. Note, to do so, we treat the absolute magnitude of Type Ia SN M_B as a nuisance parameter. This gives SN data the freedom to adjust H_0 so that the high redshift behaviour is always the same as Planck, otherwise the analysis is the same as before. Note, we have an additional nuisance parameter, but its role is simply to adopt the value that best accommodates fits in the (H_0, Ω_m)-plane, where we are still confronted with an effective 1-dimensional fit. We identify the best fit parameters $(H_0, \Omega_m, M_B) = (69.26, 0.298, -19.37)$, construct mock realisations in bins, which one compares to the real values. Throughout we allow for statistical and systematic uncertainties by cropping the Pantheon covariance matrix accordingly to fit the redshift bin. The results are shown in Table II and Fig. 4, where the same trend as the OHD data is evident.

D. Standardisable QSOs

Finally we turn our attention to QSOs standardised through the Risaliti-Lusso proposal [52, 53]. We refer readers to the original texts for methodology. Objectively, QSOs constitute emerging cosmological probes [54, 60] and are understandably less well developed than the SN and BAO; nevertheless, even now SN remain a work in progress [55]. In particular, there is considerable intrinsic scatter in the QSO data and there is an ongoing debate about the standardisability of the Risaliti-Lusso QSOs [57–61]. Nonetheless, QSOs support our narrative and this justifies their inclusion. In contrast to OHD and SN, which have lower error-weighted (effective) redshifts of $z_{\rm eff} \sim 0.5$ and $z_{\rm eff} \sim 0.3$, respectively, the QSO sample [27] is larger (2421 sources) and has a higher effective

redshift $z_{\rm eff} \sim 1.4$. Moreover, it is well documented that Ω_m adopts larger values than expected at higher redshifts [53, 56] and that evolution happens within the QSO sample [16, 53]. The key point here is that any evolution of Ω_m with effective redshift may be telling us less about QSOs and more about the flat ΛCDM model.

z	H_0 (km/s/Mpc)	Ω_m	Probability
$0 < z \leq 2.26$ (1048)	69.26	0.298	–
$0.7 < z \leq 2.26$ (124)	64.37	0.345	0.381
$0.8 < z \leq 2.26$ (82)	58.99	0.411	0.258
$0.9 < z \leq 2.26$ (49)	45.88	0.679	0.117
$0.95 < z \leq 2.26$ (34)	40.73	0.862	0.081
$1 < z \leq 2.26$ (23)	43.16	0.768	0.170

TABLE II. Same as Table I but for Pantheon SN. We treat the absolute magnitude M_B as an additional nuisance parameter when we fit mock realisations and real data. We quote the probability of larger values of Ω_m and lower values of H_0. The additional freedom in M_B means that these probabilities always agree. SN count is denoted in brackets.

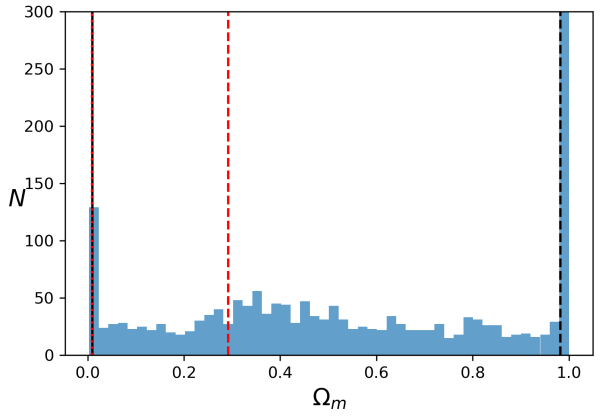

FIG. 5. A comparison between 3,000 mock simulations and the best fit value of Ω_m from QSO data (black line) for the bin $0 < z \leq 0.55$. Dashed and dotted red lines denote the (2.3, 15.9) percentiles corresponding to $1\,\sigma$ and $2\,\sigma$ confidence intervals for a Gaussian distribution. The dashed black line denotes the median, $\Omega_m = 0.982$, which, as expected, is close to the mock input $\Omega_m = 1$. In contrast to Fig. 3 and Fig. 4, the PDF is flat (neglecting the impact of bounds), thus implying that Ω_m errors are large, i. e. Ω_m is poorly constrained, in the redshift range.

Our analysis here follows the earlier sections, but there is a key difference. Risaliti-Lusso QSOs return best fits of $\Omega_m \sim 1$ across the full sample [56–59], whereas at lower redshifts $0 < z \lesssim 0.7$, one recovers Planck values, $\Omega_m \sim 0.3$ [16]; in accord with our earlier discussions and analyses. Thus, we start from the redshift range $0 < z \leq 1.4$ (1326 QSOs), where Ω_m hits the bound $\Omega_m = 1$, and identify the best fit parameters that serve as inputs for mocks, $(H_0, \Omega_m, \beta, \gamma, \delta) = (37.82, 1, 8.64, 0.61, 0.24)$. As before, β is a nuisance parameter degenerate with H_0 (the analogue of M_B in SN), so once again the fit in the (H_0, Ω_m)-plane is effectively 1-dimensional.

z	H_0 (km/s/Mpc)	Ω_m	Probability
$0 < z \leq 0.3$ (56)	406.41	0.009	0.073
$0 < z \leq 0.5$ (177)	353.47	0.011	0.028
$0 < z \leq 0.55$ (233)	433.91	0.008	0.019
$0 < z \leq 0.6$ (279)	381.50	0.010	0.020
$0 < z \leq 0.7$ (398)	73.40	0.265	0.096
$0 < z \leq 0.8$ (543)	58.48	0.418	0.117
$0 < z \leq 1$ (826)	40.69	0.864	0.400
$0 < z \leq 1.4$ (1326)	37.82	1.000	–

TABLE III. Same as Table I but for Risaliti-Lusso QSOs. We treat β, γ and δ (see [52] for definitions) as additional nuisance parameters when we fit mock realisations and real data. We quote the probability of *lower* values of Ω_m and *higher* values of H_0. QSO count is denoted in brackets

To construct the mocks, we generate new UV fluxes F_{UV} by picking values in a normal distribution about the original values with a standard deviation set to the error. Next, we generate corresponding central values for the X-ray fluxes F_X through the relation [52, 53],

$$\log_{10} F_X = \beta + \gamma \log_{10} F_{UV} + (\gamma - 1) \log_{10}(4\pi D_L^2), \qquad (2)$$

where $D_L(z)$ is the luminosity distance, before displacing the values with the standard deviation $\sqrt{\delta^2 + \sigma_i^2}$, where σ_i is the error on $\log_{10} F_{X,i}$ at redshift z_i.

In Table III we show the increasing (decreasing) trend of Ω_m (H_0) with effective redshift. Unexpectedly large values of H_0 and small values of Ω_m are driven partially by poor data quality and the Planck prior on $\Omega_m h^2$. Nevertheless, the trend in central values is the same and one notes that the probability of recovering the best fit values for real data decreases as the effective redshift of the bin decreases, confirming that the best fit values of the entire data set are less representative. In Fig. 5 we provide visual confirmation of this result in a given range, where it is notable that the Ω_m distribution is uniform between the bounds, thus underscoring how poorly QSO data constrains Ω_m in the corresponding redshift range. This is presumably due to the large scatter and fewer QSOs at lower redshifts.

IV. CONCLUSIONS

We explained through analytic arguments and simulations why the Planck value $\Omega_m \sim 0.3$ is less likely when one fits higher redshift binned $H(z)$ observations to the flat ΛCDM model. Our arguments are independent of mock input parameters and simply follow from the irrelevance of the A term in (1) at higher redshifts, and $A \geq 0$, which yield an initial 'pile up' of best fits on $\Omega_m = 1$, before piling up at $\Omega_m \sim 0$ at even higher redshifts. This reduces the probability of recovering the Planck value at high redshift, thus providing an avenue to test the model. Note, it is not enough to find unexpected best fits, but one must prove that those best fits are statistically unlikely assuming no evolution of ΛCDM parameters across data sets. Our warm-up DESI mock analysis here

solely pertains to $H(z)$ constraints, but the same conclusions hold for angular diamater distance $D_A(z) \propto \int_0^z dz'/H(z')$ constraints [63]. The reader will note that $D_L(z) \propto D_A(z)$, so all our observed data is in the $H(z)$ or $D_L(z)/D_A(z)$ class.

In the second part of our work, we confirmed an increasing Ω_m behaviour in OHD and Type Ia SN with p-values as low as $p = 0.021$ ($2.3\,\sigma$) and $p = 0.081$ ($1.7\,\sigma$), respectively. We resorted to comparison to mock analysis in the same redshift range with the same data quality as a means of circumventing the difficulty estimating errors with non-Gaussian distributions. Using Fisher's method, the combined (lowest) probability for these established cosmological probes is $p = 0.013$ ($2.5\,\sigma$). In QSOs, an intrinsically high redshift emerging observable, we see the opposite trend where discrepant best fit (H_0, Ω_m) values relative to the entire sample appear at lower redshifts with probabilities as low as $p = 0.019$ ($2.3\,\sigma$). Once again combining the probabilities, one finds a (lowest) probability $p = 0.0021$ ($3.1\,\sigma$). One may also benchmark with respect to the bin $0 < z \leq 0.7$, where $\Omega_m \sim 0.3$, in which case the combined (lowest) probability becomes $p = 0.0078$ ($2.7\,\sigma$). As argued in the text, the probabilities we quote are lower bounds. One can of course find redshift ranges with less evolution. Moreover, by working with the full sample and not binning it, one can return to the working *assumption* that there is no evolution in the samples. Our analysis here challenges the working assumption.

Objectively, all observables show signatures of evolution to higher values of Ω_m between low and high redshifts, in line with mock expectations that they can be easily displaced from Planck values. Neglecting selection effects, and more general systematics, across multiple observables (SN, cosmic chronometers, BAO, QSOs), this supports the idea that the flat ΛCDM model is a dynamical model where its fitting parameters, which should be constants, evolve in (cosmic) time. This cautions that cosmological tensions may be an outcome of the flawed assumption that (H_0, Ω_m) are unique within flat ΛCDM. Our analysis can be revisited as data quality improves, e. g. [64].

Acknowledgements We thank Stephen Appleby, Eleonora Di Valentino, Adam Riess, Joe Silk and Jenny Wagner for correspondence. We also thank Bum-Hoon Lee, Wonwoo Lee, Albin Nilsson, Mehrdad Mirbabaei, Somyadip Thakur and Lu Yin for discussion. EÓC was supported by the National Research Foundation of Korea grant funded by the Korea government (MSIT) (NRF-2020R1A2C1102899). DS is partially supported by the US National Science Foundation, under Grant No. PHY-2014021. MMShJ would like to acknowledge SarAmadan grant No. ISEF/M/400121 and ICTP HECAP section where this project was carried out. M.G. acknowledge the support of Division of Science at NAOJ. This article/publication is based upon work from COST Action CA21136 – "Addressing observational tensions in cosmology with systematics and fundamental physics (CosmoVerse)", supported by COST (European Cooperation in Science and Technology).

Appendix A: Removing $\Omega_m h^2$ prior

Removing the Planck $\Omega_m h^2$ prior from Fig. 2 leads to a spreading in all distributions, but qualitatively the features are the same, as expected from the analytic discussions. This can be confirmed in Fig. 6.

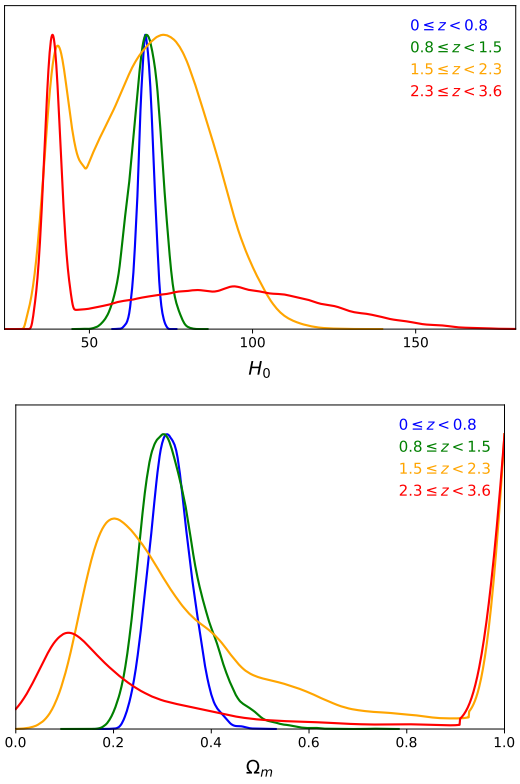

FIG. 6. Same as Fig. 2, but without the $\Omega_m h^2$ prior.

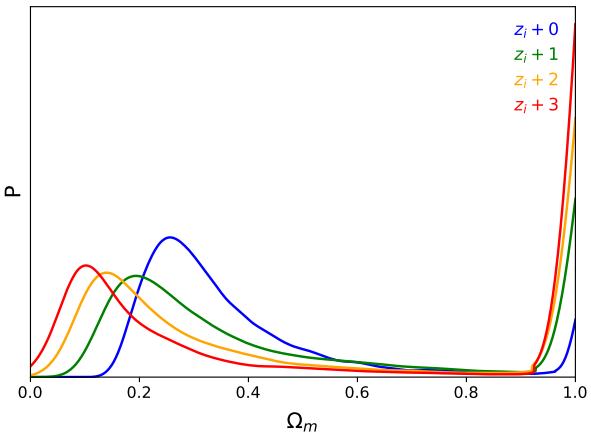

FIG. 7. Probabilities of a given Ω_m best fit value for forecast DESI $H(z)$ data in the range $2.3 \leq z < 3.6$ (blue curve) with Planck input values. Green, yellow and red denote the probabilities if the same forecasted data quality is displaced to higher redshifts.

Appendix B: Confirmation of $P(\Omega_m \sim 0.3) \to 0$

In this section we consider the same mocking procedure but focus exclusively on the fourth DESI bin with redshift range $2.3 \leq z < 3.6$. We now displace the redshifts in intervals of $+1$ without changing the data and document the effect on the distribution of Ω_m best fits over a few thousand mocks. Once again, we assume Planck input parameters, $H_0 = 67.36, \Omega_m = 0.315$ and the Gaussian prior, $\Omega_m h^2 = 0.1430 \pm 0.0011$ [9]. In Fig. 7 we present (normalised) probabilities for Ω_m for Planck-ΛCDM mocks, where the blue curve corresponds to the red curve in Fig. 2. The remaining curves correspond to Ω_m probabilities as we displace the original binned data in redshift. Since we are at high redshift, the probability of $\Omega_m = 1$, $P(\Omega_m = 1) > 0$ and it clearly increases with redshift of the sample. This corresponds to the A distribution spreading to smaller values, where it is impacted by the bound at $A = 0$. However, A must also spread to higher values, and since $B = H_0^2\Omega_m$ is fixed, in the sense that this combination is relevant at higher redshifts, the only way $A = H_0^2 - B$ can spread to higher values is increasing H_0 and decreasing Ω_m. This spreading of the A distribution to larger values explains why the peaks in Fig. 2 get displaced from their input values. In Fig. 7 this trend is more pronounced and it is an obvious implication that at a given high redshift, any knowledge of the input parameters is lost and the probability of recovering the Planck value, $P(\Omega_m \sim 0.3)$ is close to zero.

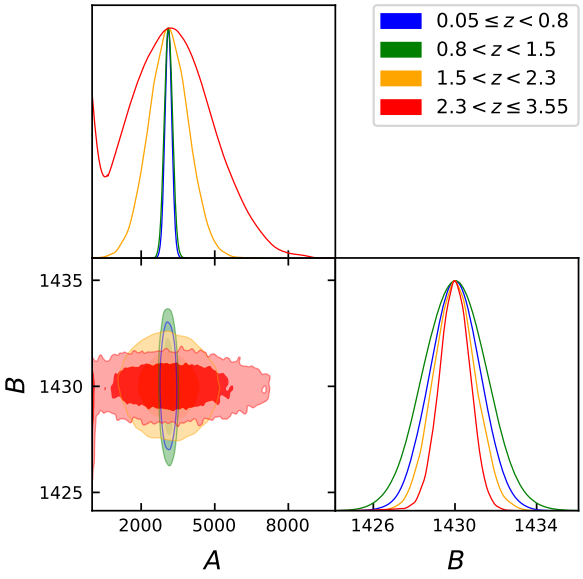

FIG. 8. A corner plot with DESI forecast data demonstrating that (A, B) are uncorrelated across the bins.

Appendix C: Further comments on (A, B)

In this section we show in Fig. 8 that the derived (secondary) parameters (A, B) are uncorrelated. As explained in the main text, we imposed a (strong) Planck Gaussian prior on B, so unsurprisingly B conforms to a Gaussian and A is also Gaussian where it is not impacted by the bound $\Omega_m \leq 1$. Noting that (A, B) are uncorrelated, whereas the transformation from the fitting parameters (H_0, Ω_m) to (A, B) is non-linear, it would be surprising if one encountered Gaussian distributions in all parameters.

There is another loose end to close. The astute reader will notice that the B distribution does not narrow uniformly in Fig. 8. To explain this feature we note that the percentage $H(z)$ errors vary with redshift in the DESI forecast [22] and the errors are smallest at the boundary of bin 1 and bin 2 ($z \sim 0.8$). As a result, bin 1 better constrains B, the relevant high redshift parameter, whereas bin 2 better constrains A, the relevant low redshift parameter. From Fig. 8 one notes that any spread in A is marginal between bins 1 and 2, while the B distribution actually spreads between bins 1 and 2, thus contradicting statements in the text. However, in Fig. 9 we produce four bins with *exactly* the same data quality in each bin by simply displacing the percentage errors in bin 1 in redshift and using them as the basis for mocks in bins 2, 3 and 4. As a result, one has the same percentage errors in each bin, and one notices that A spreads whereas B narrows with effective redshift. In summary, one generically expects a spreading A distribution and narrowing B distribution with effective redshift in the flat ΛCDM model as the model transitions from a two-parameter model to an effective one-parameter model, but this trend may be impacted by data quality.

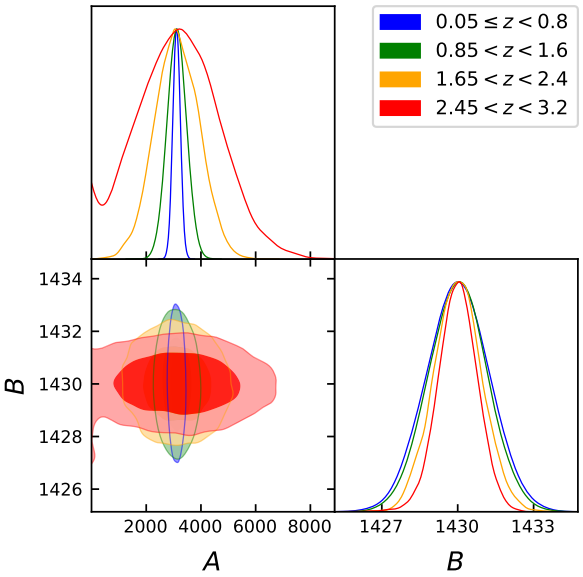

FIG. 9. A corner plot with with uniform data quality in all bins that the A distribution spreads whereas the B distribution narrows once the data quality in all bins is the same.

Appendix D: MCMC analysis

As explained in the text, great care is required with MCMC inferences in non-Gaussian regimes. For this reason we opted to compare our best fits in observed data directly with best fits in mock data *in the same redshift range with the same data quality*. Here we take a look at the inferences one would make with MCMC, where we focus on OHD data, since it is the simplest to analyse because there are no nuisance parameters. Once again, we impose the bounds $0 \leq \Omega_m \leq 1$. We then split the OHD sample at $z = 1.45$, which is of interest since it corresponds to the 7$^{\rm th}$ entry in Table I, where we have recorded the lowest probability of recovering best fits from observed data in mock data. It should be noted that our best fit Ω_m value saturates the bound $\Omega_m = 1$, so we expect that our MCMC distribution has no Ω_m peak because it is precluded by the priors.

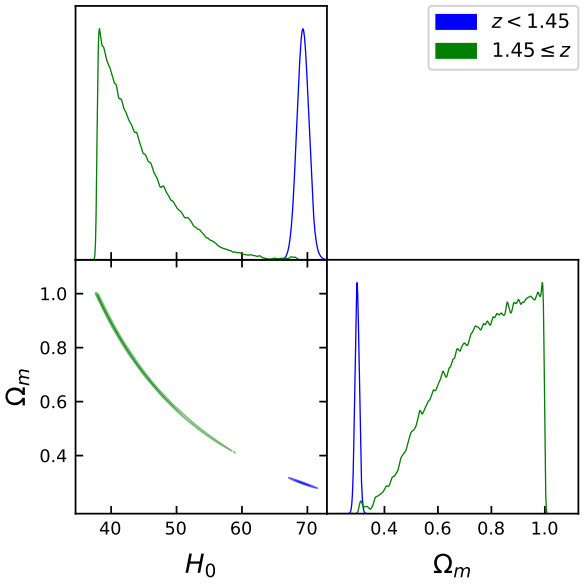

FIG. 10. Inferences of cosmological parameters from MCMC chains in the OHD sample of 54 data points. Evolution of cosmological parameters is evident when comparing low and redshift subsamples.

In Fig. 10 we show the outcome of the MCMC analysis. As expected, the low redshift ($z < 1.45$) sample of 47 OHD data points leads to an Ω_m distribution that is perfectly Gaussian, but the high redshift ($1.45 \geq z$) sample of 7 OHD data points does not. The Ω_m distribution continues to increase towards the $\Omega_m = 1$ bound implying either a peak at the bound or beyond the bound. Priors are clearly impacting the result. In the (H_0, Ω_m)-plane the contours follow a curve of constant $\Omega_m h^2$ due to the Planck prior. This curve is elongated in the high redshift sample and the discrepancy between the low and high redshift subsamples of the full OHD sample is evident in the (H_0, Ω_m)-plane. From the MCMC chains, we infer the constraints on (H_0, Ω_m) from the low redshift sample to be $(H_0, \Omega_m) = (69.32^{+0.90}_{-0.90}, 0.298^{+0.008}_{-0.008})$, whereas the constraints from the high redshift sample are $(H_0, \Omega_m) = (42.82^{+6.56}_{-3.69}, 0.779^{+0.154}_{-0.194})$. Here, we quote 16$^{\rm th}$, 50$^{\rm th}$ and 84$^{\rm th}$ percentiles in line with standard practice, of course assuming a Gaussian distribution. It is clearly wrong to do this as our distributions are non-Gaussian and have been impacted by the $\Omega_m \leq 1$ bound, but the results are indicative. Note that Ω_m inferences are clearly shifted to lower values relative to the best fit $\Omega_m = 1$ (see Table I) due to the presence of the bound. A back of the envelope calculation then places the tension between these results as a 4σ discrepancy in H_0 and a 2.5σ discrepancy in Ω_m, demonstrating evolution in cosmological parameters across the OHD sample.

The main message to be taken away is the difficulty performing an MCMC analysis when distributions become non-Gaussian either because of a fundamentally non-Gaussian distribution or the impact of the priors. Nevertheless, one can follow the standard procedures and still find evidence for evolution in cosmological parameters across the sample. MCMC may lead to misleading results, but evolution is still evident.

[1] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess and J. Silk, Class. Quant. Grav. **38** (2021) no.15, 153001 [arXiv:2103.01183 [astro-ph.CO]].

[2] E. Abdalla, G. Franco Abellán, A. Aboubrahim, A. Agnello, O. Akarsu, Y. Akrami, G. Alestas, D. Aloni, L. Amendola and L. A. Anchordoqui, *et al.* JHEAp **34** (2022), 49-211 [arXiv:2203.06142 [astro-ph.CO]].

[3] L. Verde, T. Treu and A. G. Riess, Nature Astron. **3**, 891 [arXiv:1907.10625 [astro-ph.CO]].

[4] A. G. Riess, W. Yuan, L. M. Macri, D. Scolnic, D. Brout, S. Casertano, D. O. Jones, Y. Murakami, L. Breuval and T. G. Brink, *et al.* [arXiv:2112.04510 [astro-ph.CO]].

[5] W. L. Freedman, "Measurements of the Hubble Constant: Tensions in Perspective," Astrophys. J. **919** (2021) no.1, 16 [arXiv:2106.15656 [astro-ph.CO]].

[6] D. W. Pesce, J. A. Braatz, M. J. Reid, A. G. Riess, D. Scolnic, J. J. Condon, F. Gao, C. Henkel, C. M. V. Impellizzeri and C. Y. Kuo, *et al.* "The Megamaser Cosmology Project. XIII. Combined Hubble constant constraints," Astrophys. J. Lett. **891** (2020) L1 [arXiv:2001.09213 [astro-ph.CO]].

[7] E. Kourkchi, R. B. Tully, G. S. Anand, H. M. Courtois, A. Dupuy, J. D. Neill, L. Rizzi and M. Seibert, Astrophys. J. **896** (2020) no.1, 3 [arXiv:2004.14499 [astro-ph.GA]].

[8] J. P. Blakeslee, J. B. Jensen, C. P. Ma, P. A. Milne and J. E. Greene, Astrophys. J. **911** (2021) no.1, 65 [arXiv:2101.02221 [astro-ph.CO]].

[9] N. Aghanim *et al.* [Planck], Astron. Astrophys. **641** (2020), A6 [arXiv:1807.06209 [astro-ph.CO]].

[10] K. C. Wong, S. H. Suyu, G. C. F. Chen, C. E. Rusu, M. Millon, D. Sluse, V. Bonvin, C. D. Fassnacht, S. Taubenberger and M. W. Auger, *et al.* Mon. Not. Roy. Astron. Soc. **498** (2020) no.1, 1420-1439 [arXiv:1907.04869 [astro-ph.CO]].

[11] M. Millon, A. Galan, F. Courbin, T. Treu, S. H. Suyu, X. Ding, S. Birrer, G. C. F. Chen, A. J. Shajib and D. Sluse, *et al.* Astron. Astrophys. **639** (2020), A101 [arXiv:1912.08027 [astro-ph.CO]].

[12] C. Krishnan, E. Ó Colgáin, Ruchika, A. A. Sen, M. M. Sheikh-Jabbari and T. Yang, Phys. Rev. D **102** (2020) no.10, 103525 [arXiv:2002.06044 [astro-ph.CO]].

[13] M. G. Dainotti, B. De Simone, T. Schiavone, G. Montani, E. Rinaldi and G. Lambiase, Astrophys. J. **912** (2021) no.2, 150 [arXiv:2103.02117 [astro-ph.CO]].

[14] N. Horstmann, Y. Pietschke and D. J. Schwarz, [arXiv:2111.03055 [astro-ph.CO]].

[15] M. G. Dainotti, B. De Simone, T. Schiavone, G. Montani, E. Rinaldi, G. Lambiase, M. Bogdan and S. Ugale, Galaxies **10** (2022), 24 [arXiv:2201.09848 [astro-ph.CO]].

[16] E. Ó Colgáin, M. M. Sheikh-Jabbari, R. Solomon, G. Bargiacchi, S. Capozziello, M. G. Dainotti and D. Stojkovic, [arXiv:2203.10558 [astro-ph.CO]].

[17] J. Wagner, [arXiv:2203.11219 [astro-ph.CO]].

[18] X. D. Jia, J. P. Hu and F. Y. Wang, [arXiv:2212.00238 [astro-ph.CO]].

[19] J. P. Hu and F. Y. Wang, [arXiv:2203.13037 [astro-ph.CO]].

[20] C. Krishnan, E. Ó Colgáin, M. M. Sheikh-Jabbari and T. Yang, Phys. Rev. D **103** (2021) no.10, 103509 [arXiv:2011.02858 [astro-ph.CO]].

[21] C. Krishnan and R. Mondol, [arXiv:2201.13384 [astro-ph.CO]].

[22] A. Aghamousa *et al.* [DESI], [arXiv:1611.00036 [astro-ph.IM]].

[23] R. Jimenez and A. Loeb, Astrophys. J. **573** (2002), 37-42 [arXiv:astro-ph/0106145 [astro-ph]].

[24] H. J. Seo and D. J. Eisenstein, Astrophys. J. **598** (2003), 720-740 [arXiv:astro-ph/0307460 [astro-ph]].

[25] D. J. Eisenstein *et al.* [SDSS], Astrophys. J. **633** (2005), 560-574 [arXiv:astro-ph/0501171 [astro-ph]].

[26] D. M. Scolnic *et al.* [Pan-STARRS1], Astrophys. J. **859** (2018) no.2, 101 [arXiv:1710.00845 [astro-ph.CO]].

[27] E. Lusso, G. Risaliti, E. Nardini, G. Bargiacchi, M. Benetti, S. Bisogni, S. Capozziello, F. Civano, L. Eggleston and M. Elvis, *et al.* Astron. Astrophys. **642** (2020), A150 [arXiv:2008.08586 [astro-ph.GA]].

[28] A. Lewis, [arXiv:1910.13970 [astro-ph.IM]].

[29] D. Stern, R. Jimenez, L. Verde, M. Kamionkowski and S. A. Stanford, JCAP **02** (2010), 008 [arXiv:0907.3149 [astro-ph.CO]].

[30] M. Moresco, A. Cimatti, R. Jimenez, L. Pozzetti, G. Zamorani, M. Bolzonella, J. Dunlop, F. Lamareille, M. Mignoli and H. Pearce, *et al.* JCAP **08** (2012), 006 [arXiv:1201.3609 [astro-ph.CO]].

[31] C. Zhang, H. Zhang, S. Yuan, T. J. Zhang and Y. C. Sun, Res. Astron. Astrophys. **14** (2014) no.10, 1221-1233 [arXiv:1207.4541 [astro-ph.CO]].

[32] M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro and D. Wilkinson, JCAP **05** (2016), 014 [arXiv:1601.01701 [astro-ph.CO]].

[33] A. L. Ratsimbazafy, S. I. Loubser, S. M. Crawford, C. M. Cress, B. A. Bassett, R. C. Nichol and P. Väisänen, Mon. Not. Roy. Astron. Soc. **467** (2017) no.3, 3239-3254 [arXiv:1702.00418 [astro-ph.CO]].

[34] N. Borghi, M. Moresco and A. Cimatti, Astrophys. J. Lett. **928** (2022) no.1, L4 [arXiv:2110.04304 [astro-ph.CO]].

[35] K. Jiao, N. Borghi, M. Moresco and T. J. Zhang, [arXiv:2205.05701 [astro-ph.CO]].

[36] E. Gaztanaga, A. Cabre and L. Hui, Mon. Not. Roy. Astron. Soc. **399** (2009), 1663-1680 [arXiv:0807.3551 [astro-ph]].

[37] A. Oka, S. Saito, T. Nishimichi, A. Taruya and K. Yamamoto, Mon. Not. Roy. Astron. Soc. **439** (2014), 2515-2530 [arXiv:1310.2820 [astro-ph.CO]].

[38] Y. Wang *et al.* [BOSS], Mon. Not. Roy. Astron. Soc. **469** (2017) no.3, 3762-3774 [arXiv:1607.03154 [astro-ph.CO]].

[39] C. H. Chuang and Y. Wang, Mon. Not. Roy. Astron. Soc. **435** (2013), 255-262 [arXiv:1209.0210 [astro-ph.CO]].

[40] S. Alam *et al.* [BOSS], Mon. Not. Roy. Astron. Soc. **470** (2017) no.3, 2617-2652 [arXiv:1607.03155 [astro-ph.CO]].

[41] C. Blake, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, D. Croton, T. Davis, M. J. Drinkwater and K. Forster, *et al.* Mon. Not. Roy. Astron. Soc. **425** (2012), 405-414 [arXiv:1204.3674 [astro-ph.CO]].

[42] L. Anderson, E. Aubourg, S. Bailey, F. Beutler, A. S. Bolton, J. Brinkmann, J. R. Brownstein, C. H. Chuang, A. J. Cuesta and K. S. Dawson, *et al.* Mon. Not. Roy. Astron. Soc. **439** (2014) no.1, 83-101 [arXiv:1303.4666 [astro-ph.CO]].

[43] J. E. Bautista, N. G. Busca, J. Guy, J. Rich, M. Blomqvist, H. d. Bourboux, M. M. Pieri, A. Font-Ribera, S. Bailey and T. Delubac, *et al.* Astron. Astrophys. **603** (2017), A12 [arXiv:1702.00176 [astro-ph.CO]].

[44] T. Delubac *et al.* [BOSS], Astron. Astrophys. **574** (2015), A59 [arXiv:1404.1801 [astro-ph.CO]].

[45] A. Font-Ribera *et al.* [BOSS], JCAP **05** (2014), 027 [arXiv:1311.1767 [astro-ph.CO]].

[46] J. Magana, M. H. Amante, M. A. Garcia-Aspeitia and V. Motta, Mon. Not. Roy. Astron. Soc. **476** (2018) no.1, 1036-1049 [arXiv:1706.09848 [astro-ph.CO]].

[47] P. A. R. Ade *et al.* [Planck], Astron. Astrophys. **594** (2016), A13 [arXiv:1502.01589 [astro-ph.CO]].

[48] J. Hou, A. G. Sánchez, A. J. Ross, A. Smith, R. Neveux, J. Bautista, E. Burtin, C. Zhao, R. Scoccimarro and K. S. Dawson, *et al.* Mon. Not. Roy. Astron. Soc. **500** (2020) no.1, 1201-1221 [arXiv:2007.08998 [astro-ph.CO]].

[49] R. Neveux, E. Burtin, A. de Mattia, A. Smith, A. J. Ross, J. Hou, J. Bautista, J. Brinkmann, C. H. Chuang and K. S. Dawson, *et al.* Mon. Not. Roy. Astron. Soc. **499** (2020) no.1, 210-229 [arXiv:2007.08999 [astro-ph.CO]].

[50] H. du Mas des Bourboux, J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, T. Etourneau, J. M. Le Goff, A. Cuceu, C. Balland and J. E. Bautista, *et al.* Astrophys. J. **901** (2020) no.2, 153 [arXiv:2007.08995 [astro-ph.CO]].

[51] G. E. Addison, Y. Huang, D. J. Watts, C. L. Bennett, M. Halpern, G. Hinshaw and J. L. Weiland, Astrophys. J. **818** (2016) no.2, 132 [arXiv:1511.00055 [astro-ph.CO]].

[52] G. Risaliti and E. Lusso, Astrophys. J. **815** (2015), 33 [arXiv:1505.07118 [astro-ph.CO]].

[53] G. Risaliti and E. Lusso, Nature Astron. **3** (2019) no.3, 272-277 [arXiv:1811.02590 [astro-ph.CO]].

[54] M. Moresco, L. Amati, L. Amendola, S. Birrer, J. P. Blakeslee, M. Cantiello, A. Cimatti, J. Darling, M. Della Valle and M. Fishbach, *et al.* [arXiv:2201.07241 [astro-ph.CO]].

[55] C. Meldorf, A. Palmese, D. Brout, R. Chen, D. Scolnic, L. Kelsey, L. Galbany, W. Hartley, T. Davis and A. Drlica-Wagner, *et al.* [arXiv:2206.06928 [astro-ph.CO]].

[56] T. Yang, A. Banerjee and E. Ó Colgáin, Phys. Rev. D **102** (2020) no.12, 123532 [arXiv:1911.01681 [astro-ph.CO]].

[57] N. Khadka and B. Ratra, Mon. Not. Roy. Astron. Soc. **497** (2020) no.1, 263-278 [arXiv:2004.09979 [astro-ph.CO]].
[58] N. Khadka and B. Ratra, Mon. Not. Roy. Astron. Soc. **502** (2021) no.4, 6140-6156 [arXiv:2012.09291 [astro-ph.CO]].
[59] N. Khadka and B. Ratra, Mon. Not. Roy. Astron. Soc. **510** (2022) no.2, 2753-2772 [arXiv:2107.07600 [astro-ph.CO]].
[60] M. G. Dainotti, V. Nielson, G. Sarracino, E. Rinaldi, S. Nagataki, S. Capozziello, O. Y. Gnedin and G. Bargiacchi, [arXiv:2203.15538 [astro-ph.CO]].
[61] V. Petrosian, J. Singal and S. Mutchnick, [arXiv:2205.07981 [astro-ph.CO]].
[62] R. C. Nunes and S. Vagnozzi, Mon. Not. Roy. Astron. Soc. **505** (2021) no.4, 5427-5437 [arXiv:2106.01208 [astro-ph.CO]].
[63] E. Ó Colgáin, M. M. Sheikh-Jabbari and R. Solomon, Phys. Dark Univ. **40** (2023), 101216 [arXiv:2211.02129 [astro-ph.CO]].
[64] M. Malekjani, R. M. Conville, E. Ó Colgáin, S. Pourojaghi and M. M. Sheikh-Jabbari, [arXiv:2301.12725 [astro-ph.CO]].
[65] S_8 tension is less well established, see [62].
[66] Later we show that mock realisations can easily violate this bound at higher redshifts. If the same trend is observed in observed data, does this immediately falsify flat ΛCDM?
[67] One can test this by initialising the χ^2-minimsation algorithm from different points in parameter space and checking that one recovers best fit parameters that are close in value. See analysis in [64].