

Non-Gaussian Likelihoods for Type Ia Supernovae Cosmology: Implications for Dark Energy and H_0

Toby Lovick,^{1*} Suhail Dhawan,² Will Handley^{1,3}

¹ *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

² *Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

³ *Astrophysics Group, Cavendish Laboratory University of Cambridge, JJ Thomson Avenue, Cambridge, CB3 0HE, UK*

16 October 2024

ABSTRACT

The latest improvements in the scale and calibration of Type Ia supernovae catalogues allow us to constrain the specific nature and evolution of dark energy through its effect on the expansion history of the universe. We present the results of Bayesian model comparison on the SNe Ia catalogue Pantheon+, where Flat Λ CDM is preferred by the data over all other models and we find moderate evidence ($\Delta \log \mathcal{Z} \sim 2.5$) to reject a number of alternate dark energy models. The effect of peculiar velocity corrections on model comparison is analysed, where we find that removing peculiar velocity corrections lessens the $\Delta \log \mathcal{Z}$ of each cosmology against Λ CDM and overall reduces the strength of Λ CDM in the model comparison. As well as comparing cosmological models, the Bayesian methodology is extended to testing for non-gaussianity in the distribution of the Pantheon+ Hubble residuals. We find that adding a scale parameter to the Pantheon+ covariances, or alternately using a multivariate Student's t-distribution fits the data better than the fiducial analysis, producing a marginalised evidence increase of $\Delta \log \mathcal{Z} = 1.70$ and 2.70 respectively. This improved treatment of the scatter produces a 5% decrease in the uncertainty of the Hubble constant as compared to SH0ES, finding $H_0 = 73.67 \pm 0.99 \text{ km s}^{-1} \text{ Mpc}^{-1}$ in 5.6σ tension with Planck, guided by the evidence when marginalising over both cosmological and scatter models. We also explore M_B transition models as a potential solution for the Hubble tension, finding no evidence to support these models among the SN data.

Key words: dark energy - cosmological parameters - transients: supernovae - methods: statistical

1 INTRODUCTION

Type Ia supernovae (SNe Ia) are extremely bright thermonuclear explosions of a carbon-oxygen white dwarf in a binary system (see Maguire 2017, for a review of their observational properties). After correcting for the lightcurve shape (Phillips 1993), colour (Tripp 1998) and host galaxy properties (e.g. Kelly et al. 2010; Sullivan et al. 2010), they have a small dispersion in their brightness, making them excellent distance indicators for cosmology (e.g., see Goobar & Leibundgut 2011, for a review of their cosmological utility). They were instrumental in the discovery of universal accelerated expansion (Riess et al. 1998; Perlmutter et al. 1999) and are crucial for precision measurements of dark energy properties, e.g. energy density, equation of state and changes with cosmic time (Brout et al. 2022).

Pantheon+ is the culmination of a series of efforts to standardise the luminosity of SNe Ia, to improve their constraining power as a cosmological probe. The data have been corrected in a number of ways, accounting for statistical and systematic effects on the data: see Brout et al. (2022) Figure 1 for a full list of the companion papers that form the total catalogue. All of these effects lead to the corrected apparent magnitudes in the Pantheon+ catalogue, with their intrinsic scatter reduced to ~ 0.1 mag.

Observations of the SN Ia magnitude-redshift relation, as well as

high-precision early and late universe probes, e.g. the cosmic microwave background power spectrum (Planck Collaboration et al. 2020), baryon acoustic oscillations (du Mas des Bourboux et al. 2020), and the distribution and evolution of galaxies (d'Amico et al. 2020), have led to the inception of a standard cosmological model, termed Λ CDM. It is a model with only 6 parameters, explaining a wide range of datasets and phenomena. However, there exist significant theoretical and observational problems with dark energy that are unsolved by Λ CDM (Perivolaropoulos & Skara 2022), which motivates looking into physics beyond our standard cosmology. The most relevant problem with Λ CDM for this paper is the Hubble tension, the observed 5σ tension between the CMB-measured value of $H_0 \sim 67.4 \pm 0.5 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration et al. 2020), and the distance measure Hubble constant, constrained on the Pantheon+ data as $H_0 = 73.04 \pm 1.04 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Riess et al. 2022). Tensions (Hubble or otherwise) in our current theory motivate searching for physics beyond the standard cosmology; alternate formulations of dark energy.

Despite the mathematical simplicity of a cosmological constant there is no reason a priori to assume that dark energy takes this form, and many alternative theories exist that also function as an accelerating factor in the dynamics of the scale of the universe. The simplest modification is w CDM, where the equation of state parameter $P = w\rho$ is no longer fixed at $w = -1$ for dark energy. Other models allow for a dynamical dark energy, where instead of acting like a cosmological constant, the accelerating power of dark energy

* E-mail: lovicktoby@gmail.com

evolves with cosmological time. A final category of models allows modifications to General Relativity; since our evidence for dark energy comes from Einstein's equations, it's possible to justify the appearance of an accelerating universe via a novel theory of gravity. The list of models considered in this paper are drawn largely from [Dhawan et al. \(2020\)](#) and cover a wide range of physical motivations and theoretical approaches to modelling dark energy.

With the increasing sample size of SN Ia catalogues both current and future ([Rose et al. 2021](#)), as well as an increased focus on the treatment of systematics in the data, it is important to systematically check not only a selection of dark energy models on the data, but also check for the gaussianity of the data. It is possible that the work that has gone in to standardizing SN Ia, as well as combining many different SN Ia surveys into a single catalogue has introduced a non-Gaussian signal into the data - an effect which could change the results of parameter inference and dark energy model comparison. Bayesian evidence allows us to make model comparisons such that the effects of dark energy models and scatter models can be separated and marginalised out.

The usage of Bayesian techniques in cosmology has greatly increased in recent years ([Hobson et al. 2010](#)). The Bayesian approach is to update our current belief (the prior distribution) using the observed data and their likelihood to form a new belief about a hypothesis (the posterior distribution). This allows us to take an uninformative prior (such as $\Omega_m \sim U[0, 1]$) and update it using the supernova data to infer the value of parameters within a dark energy model. Beyond parameter inference, we can update our level of belief in competing dark energy models, as in [Kurek & Szydłowski \(2008\)](#), [Handley \(2021\)](#). Starting from the belief that all dark energy models are equally likely, we can calculate the posterior probabilities of competing models as they are favoured by the data.

A review of the relative merits and demerits of a Bayesian model selection analysis are explored in detail in [Bernardo & Smith \(2009\)](#); [Lindley \(2000\)](#); [Trotta \(2008\)](#). For comparing various different models, a key advantage is that the Bayes factor, i.e. the logarithm of the ratio of the Bayesian evidence, penalises models with larger degrees of freedom if the additional degrees do not lead to a significant improvement in the quality of the fit ([Hergt et al. 2021](#)). This effectively applies an ‘‘Occam’s razor’’ in the analysis. In this paper, we move beyond just dark energy model comparison and use the Bayes factor to not only compare different cosmological models using SNe Ia data but also different likelihood models and how well they describe the Pantheon+ data. We present our methodology in section 2, our results in section 3 and discuss them in light of the current literature in section 4. Our conclusions are presented in section 5

2 METHODOLOGY

2.1 Bayes’ Theorem

This subsection is an adaptation of [Trotta \(2008\)](#), where all equations and table 1 can be found.

In the context of parameter inference and model comparison, Bayes’ theorem says that the distribution of a set of parameters θ within a model $\mathcal{M}$ gets updated by data d as:

$$p(\theta|d, \mathcal{M}) = \frac{p(d|\theta, \mathcal{M})p(\theta|\mathcal{M})}{p(d|\mathcal{M})}, \quad (1)$$

Table 1. The odds, probabilities, and interpretations of various Bayes factor as given by [Jeffreys \(1961\)](#). The third column is the posterior probability of the superior model in the case of a 2-model comparison, and is offered as an analogy to a p-value.

$ \log B_{01} $	Odds	Probability	Strength of evidence
< 1.0	$\lesssim 3 : 1$	< 0.750	Inconclusive
1.0	$\sim 3 : 1$	0.750	Weak
2.5	$\sim 12 : 1$	0.923	Moderate
5	$\sim 150 : 1$	0.993	Strong

or

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}} \equiv \mathcal{P} = \frac{\mathcal{L}\pi}{\mathcal{Z}}.$$

In this analysis we take the prior distribution of models to be uniform, i.e.

$$p(\mathcal{M}) = \frac{1}{\text{Number of models}}, \quad (2)$$

so that model comparison is quantified by the evidence ratio:

$$\frac{p(\mathcal{M}_0|d)}{p(\mathcal{M}_1|d)} = \frac{p(d|\mathcal{M}_0)p(\mathcal{M}_0)}{p(d|\mathcal{M}_1)p(\mathcal{M}_1)} = \frac{p(d|\mathcal{M}_0)}{p(d|\mathcal{M}_1)}. \quad (3)$$

We compute these evidences by integrating out the posterior distribution over parameter space in equation (1) to find

$$p(d|\mathcal{M}) = \int_{\Omega_\theta} p(d|\theta, \mathcal{M})p(\theta|\mathcal{M})d\theta = \langle \mathcal{L} \rangle_\pi. \quad (4)$$

From here on the evidence of a model is labelled with $\mathcal{Z}$, and all model comparisons are expressed in terms of the natural logarithm $\Delta \log \mathcal{Z} = \log(\mathcal{Z}_0/\mathcal{Z}_1)$. Table 1 shows the Jeffreys’ Scale ([Jeffreys 1961](#)) which we use to judge the significance of our results.

2.2 Occam Penalty

As mentioned previously Bayesian model comparison has a built in Occam’s razor term, that works to rule out over-complicated models. This takes the form of Kullback-Leibler divergence ([Kullback & Leibler 1951](#)), the statistical distance between the posterior and prior:

$$\mathcal{D}_{\text{KL}}(\theta|\mathcal{M}) = \int p(\theta|d, \mathcal{M}) \log \frac{p(\theta|d, \mathcal{M})}{p(\theta|\mathcal{M})} d\theta = \int \mathcal{P} \log \frac{\mathcal{L}}{\mathcal{Z}} d\theta. \quad (5)$$

Splitting up the logarithm demonstrates how this term penalises the evidence:

$$\log \mathcal{Z} = \langle \log \mathcal{L} \rangle_\mathcal{P} - \mathcal{D}_{\text{KL}}, \quad (6)$$

so the evidence of a model is equal to the goodness of fit, as measured by the posterior averaged log likelihood, minus a penalty term, the KL-divergence. $\Delta \log \mathcal{Z}$ can be dissected even further. Analytically integrating over a likelihood that is a multivariate Gaussian in θ gives:

$$\langle \log \mathcal{L} \rangle_\mathcal{P} = \log(\mathcal{L}_{\text{max}}) - \frac{n}{2}, \quad (7)$$

where n is the model dimension. For a likelihood that is not Gaussian in θ , this can be approximated by replacing n with $\hat{n}$, the Bayesian Model Dimensionality ([Handley & Lemos 2019](#)), which is a measure of the effective number of constrained parameters in a theory.

This holds for parameters with approximately Gaussian or uniform posteriors, i.e. over “well-behaved” parameter spaces.

$$\langle \log \mathcal{L} \rangle_{\mathcal{P}} = \log \mathcal{L}_{\max} - \frac{\hat{d}}{2}, \quad (8)$$

Substituting this into equation (6) for two different $\log \mathcal{Z}$ values:

$$\log \frac{\mathcal{Z}_0}{\mathcal{Z}_1} = \Delta \log \mathcal{Z} \approx \Delta(\log \mathcal{L}_{\max}) - \frac{\Delta \hat{d}}{2} - \Delta(\mathcal{D}_{\text{KL}}). \quad (9)$$

The evidence difference is (approximately) the best possible fit a model can find, plus a $\mathcal{D}_{\text{KL}}$ penalty, plus a dimensionality penalty.

2.3 Cosmological Model Fitting

Given some parameters θ in a given cosmology, we calculate luminosity distances with

$$d_L(z) = \frac{c(1+z_{\text{HEL}})}{H_0 \sqrt{|\Omega_k|}} \text{sinn} \left(\sqrt{|\Omega_k|} \int_0^{z_{\text{HD}}} \frac{dz'}{E(z'; \theta)} \right), \quad (10)$$

where $E^2(z) = H^2(z)/H_0^2$ is the dimensionless Hubble parameter, which relates the expansion history to cosmological parameters. Note that the heliocentric redshift z_{HEL} and Hubble diagram redshift z_{HD} appear differently in this equation. To account for the geometry of distances in a curved universe, $\text{sinn}(x)$ is $\sin(x)$ for positive curvature $k > 0$, $\sinh(x)$ for negative curvature $k < 0$, and equal to x if $k = 0$. This is related to the distance modulus μ by:

$$\mu(z) = 5 \log_{10} \left(\frac{d_L(z)}{10 \text{ pc}} \right). \quad (11)$$

We now calculate Hubble residuals, the deviation of the observed distances to the model:

$$\Delta_i = \mu_i - \mu_{\text{model}}(z_i) = (m_B)_i - M_B - \mu_{\text{model}}(z_i). \quad (12)$$

The corrected apparent magnitudes are given in Pantheon+ as:

$$m_B^{\text{corr}} = m_B + \alpha x_1 - \beta c - \delta_{\text{bias}} + \delta_{\text{host}}. \quad (13)$$

This equation standardises across light curves via the Tripp parameters α, β (Tripp 1998), as well as correcting for host galaxy/survey specific selection biases (Brout et al. 2022).

For Calibrator observations (SN that have an independent distance measure from Cepheid Variable stars), residuals are calculated as:

$$\Delta = (m_B^{\text{corr}} - M_B) - \mu_{\text{ceph}}. \quad (14)$$

This cosmology-independent term only constrains M_B , which propagates through the data so H_0 can be estimated on the remaining “Hubble Flow” SN. These data are shown in figure 1, along with the redshift cut-off applied to the Hubble Flow SN data at $z = 0.023$ (as implemented in the SH_0ES calculation of H_0 (Riess et al. 2022)). These redshift values have been corrected for the CMB dipole, as well as peculiar velocities (labelled as “zHD” in Pantheon+). We then calculate the likelihood of the data by combining the residuals with $\mathbf{C}$, the covariance matrix of the Pantheon+ data. The process of defining this matrix is detailed in Section 2.2 of Brout et al. (2022), following Brout et al. (2020). When modelling the likelihood as a Gaussian this takes the form of:

$$\log \mathcal{L} = -\frac{1}{2} \left(\Delta^T \mathbf{C}^{-1} \Delta + \log |2\pi \mathbf{C}| \right). \quad (15)$$

This likelihood function or one of the alternate distributions described in section 2.6, as well as the prior distributions detailed in table 2, are fed into the nested sampling software PolyChord (Handley et al. 2015b; Handley et al. 2015a). While nested sampling is a numerical method, all errors in the evidence values are < 0.05 , so are quoted/plotted without error bars.

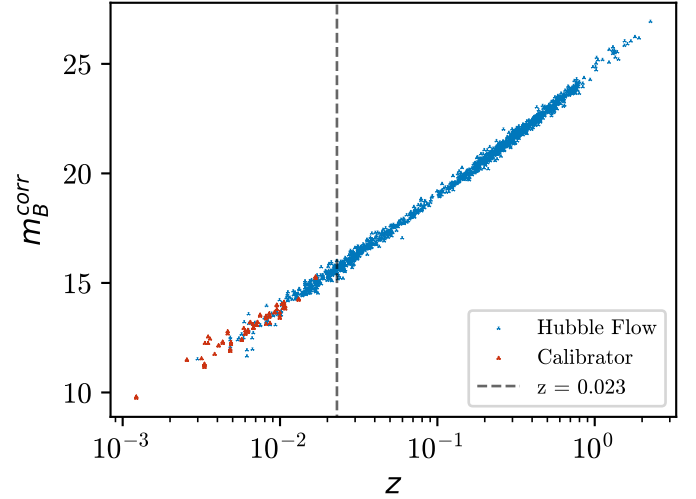


Figure 1. The redshift-apparent magnitude data in the Pantheon+ catalogue. The dashed grey line shows the redshift cutoff applied to the data, at $z = 0.023$. These data are drawn from the “zHD” and “m_b_corr” columns of Pantheon+.

2.4 Dark energy Models

This section describes the set of dark energy models considered in this paper, with brief explanations of the motivation and mechanism of each model. The priors used for each model parameter are given in table 2. From here on, the fiducial 6-parameter cosmology is referred to as Flat/ Λ CDM, to distinguish it from Λ CDM with non-zero curvature.

Λ CDM and w CDM

All of the models considered in this paper act as modifications to Λ CDM, and due to the many successful predictions of Λ CDM they must behave significantly like it to be worth considering. Despite Λ CDM being a 6-parameter model, many of the parameters do not directly affect the expansion history, so as a dark energy model it has a single parameter $\Omega_m = 1 - \Omega_\Lambda$.

The first 4 models of this paper: Λ CDM, Λ CDM, w CDM, and Flat w CDM (FwCDM) treat dark energy as having a static equation of state throughout cosmic evolution. Λ CDM allows for an over or under-dense universe $\Omega_m + \Omega_\Lambda \neq 1$, which we parameterise as curvature density $\Omega_m + \Omega_\Lambda = 1 - \Omega_k$. As well as affecting the evolution of the Hubble parameter, the presence of curvature also affects luminosity distance geometrically, as on a large scale the universe is no longer Euclidean (Ellis & van Elst 2008).

w CDM is a phenomenological model where the equation of state is not fixed to -1 (the value corresponding to a cosmological constant) but is instead a free parameter. We test this model both as a flat cosmology, and with the curvature density as a free parameter.

These four models are nested in one another; starting from curved w CDM setting $w = -1, \Omega_k = 0$, or both leads to Λ CDM, FwCDM and Λ CDM respectively. The dimensionless Hubble Parameter for w CDM is

$$\frac{H^2}{H_0^2} = \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + (1 - \Omega_m - \Omega_k)(1+z)^{3(1+w)}. \quad (16)$$

One-Parameter Slow Roll Dark Energy

Scalar field dark energy theories model current day cosmic acceleration and early time cosmic inflation with the same mechanism: a scalar field rolling down a potential. [Slepian et al. \(2014\)](#) showed that, for a potential satisfying the “Slow-Roll Conditions” that allow it to behave like dark energy, the generic behaviour of this form of dark energy is independent of the initial value and shape of the potential. This generality makes it a highly desirable model, as it has only a single additional parameter δw_0 in its Hubble parameter expression:

$$\frac{H^2}{H_0^2} = \Omega_m(1+z)^3 + (1 - \Omega_m) \left(\frac{(1+z)^3}{\Omega_m(1+z)^3 + 1 - \Omega_m} \right)^{\frac{\delta w}{1 - \Omega_m}}. \quad (17)$$

Growing Neutrino Mass

This model aims to address the “cosmological coincidence” problem; a model that proposes an interaction with the scalar field that halts its evolution can justify why $\Omega_m \sim \Omega_\Lambda$, if the interaction only happens at a typical matter/energy density. In the case of this growing neutrino mass model ([Fardon et al. 2004](#); [Wetterich 2007](#)), the scalar field stops evolving due to neutrinos becoming non-relativistic (at a value of ρ_Λ typical of an epoch transition such as is observed today). This model introduces two additional parameters, the early dark energy density Ω_e , and the current day neutrino density Ω_ν .

In this cosmology the dark energy density has different late/early time behavior, parameterised here by scale factor $a = \frac{1}{1+z}$ for algebraic convenience:

$$\Omega_{de}(a) = \begin{cases} \frac{\Omega_{de} a^3 + 2\Omega_\nu(a^{3/2} - a^3)}{1 - \Omega_{de}(1 - a^3) + 2\Omega_\nu(a^{3/2} - a^3)} & ; a > a_t \\ \Omega_e & ; a < a_t. \end{cases} \quad (18)$$

The transition scale factor where neutrinos become relativistic, $a_t = \frac{1}{1+z_t}$, is found by ensuring the early and late time behaviors are continuous, i.e. equating the early and late time behavior at $a = a_t$. The Hubble expression is then:

$$\frac{H^2}{H_0^2} = \frac{\Omega_m(1+z)^3}{1 - \Omega_{de}(a)}. \quad (19)$$

Algebraic Thawing

Thawing models contain a dark energy with $w = -1$ at high redshift, which departs from this value over cosmic time to less negative values. One such way of realising this evolution is Algebraic Thawing ([Linder 2007](#)), a dark energy that evolves in a similar way to a slow-roll dark energy during matter domination, and as dark energy becomes more dominant the evolution diverges from the scalar field dynamics. Thawing models act like a cosmological constant for most of cosmic history, only diverging at low-redshift, so they match the early universe predictions of Λ CDM. This model contains two additional parameters beyond Ω_m , the current day equation of state w_0 , and p , a shape parameter that controls how fast the equation of state thaws. It has Hubble parameter expression

$$\frac{H^2}{H_0^2} = \Omega_m(1+z)^3 + (1 - \Omega_m) \times \exp\left(\frac{3(1+w_0)}{\gamma p} \left[1 - \left(1 - \gamma + \frac{\gamma}{(1+z)^3} \right)^{p/3} \right]\right), \quad (20)$$

where $\gamma = \frac{1}{1+b}$, $b = 0.3$.

Dark Energy Transitions

Dark energy transition models attempt to resolve the Hubble tension by inducing a large step-like response in $H(z)$ at a low redshift via large fluctuations in the equation of state parameter ([Mortonson et al. 2009](#)). This model adds a single parameter δ parameterising the amplitude of the transition, where the local observed Hubble constant is $H_0 \approx (1 + \delta)\tilde{H}_0$, and $\tilde{H}_0$ is the Hubble constant of a reference Λ CDM model. In this paper the authors suggest a transition at $z_t \approx 0.02$, however our redshift cutoff is above this value, and so we use $z_t = 0.1$ to make use of the low redshift SNe up to $z = 0.1$, as in [Dhawan et al. \(2020\)](#). This model has Hubble parameter

$$\frac{H^2}{\tilde{H}_0^2} = \tilde{\Omega}_m(1+z)^3 + \left[1 + \frac{2\delta \times S(z)}{(1 - \tilde{\Omega}_m)S(0)} \right] (1 - \tilde{\Omega}_m), \quad (21)$$

where

$$S(z) = \frac{1}{2} \left[1 - \tanh\left(\frac{z - z_t}{\Delta z}\right) \right],$$

and Δz is the redshift-width that the transition occurs in, $\Delta z = 0.01$. The local observed Hubble constant and matter density are then given by

$$H_0 = \tilde{H}_0 \sqrt{1 + 2\delta},$$

$$\Omega_m = \tilde{\Omega}_m \sqrt{1 + 2\delta}.$$

Bimetric Gravity

Any deviation from a cosmological constant can also be realised by modifications to the equations of GR. This model of Bimetric gravity adds a massive particle to mediate the gravitational force, by having two interacting metrics for space-time instead of just one ([de Rham et al. 2011](#); [Mörtsell & Dhawan 2018](#)). In this paper we consider the simplest case of Bimetric gravity, including only the linear term of this interaction, which leads to a single additional parameter B_1 (where $B_1 = 0$ collapses the theory back down to Λ CDM). The Hubble parameter expression is given by:

$$\frac{H^2}{H_0^2} = \frac{\Omega_m(1+z)^3}{2} + \frac{B_0}{6} + \sqrt{\left(\frac{\Omega_m(1+z)^3}{2} + \frac{B_0}{6} \right)^2 + \frac{B_1^2}{3}}, \quad (22)$$

where

$$B_0 = 3(1 - \Omega_m) - B_1^2.$$

2.5 Absolute Magnitude Evolution Models

The dark energy transition model attempts to resolve the Hubble tension by inducing a step-like transition in H_0 at low redshift via a spike in $w(z)$. This model is highly phenomenological, and the required change in $w(z)$ is theoretically permitted, but not theoretically motivated. Since M_B and H_0 have a degenerate effect on our Hubble residuals, a step-like change in H_0 can also be realised by a changing absolute magnitude. We test two additional models, both working in a Λ CDM cosmology:

$$M_B = \begin{cases} M_{B1} & \text{if } z \leq 0.1 \\ M_{B2} & \text{if } z > 0.1 \end{cases}, \quad (23)$$

and

$$M_B = M_{B1} + \frac{z}{z_{\max}} \times (M_{B2} - M_{B1}). \quad (24)$$

Equation (23) mimics the effect of the dark energy transition model on H_0 , a sharp step in M_B at $z = 0.1$. This adds an additional step to our distance ladder; the Cepheid distances constrain M_{B1} independent of H_0 , so that the “near” supernovae can constrain H_0 , which is used to constrain M_{B2} . Equation (24) is a first order test for an evolving M_B , by including a steady $M_B - z$ gradient. Kistler et al. (2013), Meng et al. (2011) both suggest that SNe Ia evolve with metallicity (which can change the amount of radioactive material in the explosion, and hence the luminosity of the SN Ia). Since SNe Ia measure dark energy via relative distance measurements, any intrinsic evolution with cosmic time must be accounted for. Both models are considered in an Λ CDM cosmology.

2.6 Likelihood Models

Figure 2 shows the normalised residuals $C^{-1/2}\Delta$ for best fit Λ CDM parameters. The likelihood function in equation (15) assumes that these residuals will form a unit Gaussian, shown by the continuous black line, but the underlying distribution appears to be slightly tighter. Keeley et al. (2022) quantified this discrepancy, showing that the χ^2 of the Pantheon+ data using the given covariance matrix is small when compared to mock Pantheon+-like data. A likelihood function that assigns more accurate weighting to the residuals, so that observations with more or less uncertainty contribute more or less to the likelihood function, will have higher likelihood values, and therefore will be favoured by the evidence.

The discrepancy in figure 2 motivates a series of tests that adjust the form of the likelihood function in equation (15). Here we apply the Bayesian model comparison methodology not just to the cosmological model, but also to the model of the apparent magnitude scatter. A pair of well motivated models are considered: a generalised Gaussian, and the multivariate Student’s t-distribution. Dainotti et al. (2023) demonstrated a preference for a univariate Student’s t-distribution on the normalised residuals, finding a Bayesian Information Criterion of $\Delta_{\text{BIC}} = 134$, which is an approximate method equivalent to $\Delta \log \mathcal{Z} = 67$. Their analysis only normalised the residuals using the diagonal elements of the Pantheon+ covariance matrix, which neglects the correlated sources of error in the data. Additionally the Bayesian Information Criterion is only an approximation to the Bayesian model comparison method, so we repeat their analysis here with the full covariance matrix and a multivariate t-distribution, as well as using evidence instead of the approximate BIC.

Generalised Gaussian

Modelling on 21 cm cosmology data and artificially added scatter, Scheutwinkel et al. (2023) found that on data with an unknown “ground-truth” likelihood function, a generalised Gaussian distribution is a good first-order approximation, and will generally outperform a Gaussian in its Bayesian evidence (unless the underlying scatter is truly Gaussian). In the multivariate case it extends the Gaussian with two additional parameters, scale A and shape B ¹, with likelihood function (Pascal et al. 2013):

$$\mathcal{L} = N(A, B, C) \exp \left(-\frac{1}{2} \left(\frac{\Delta^T C^{-1} \Delta}{A} \right)^B \right). \quad (25)$$

¹ These parameters are often denoted (α, β) in the literature, but these characters are reserved for the Tripp light-curve fitting parameters here, so (A, B) are used instead.

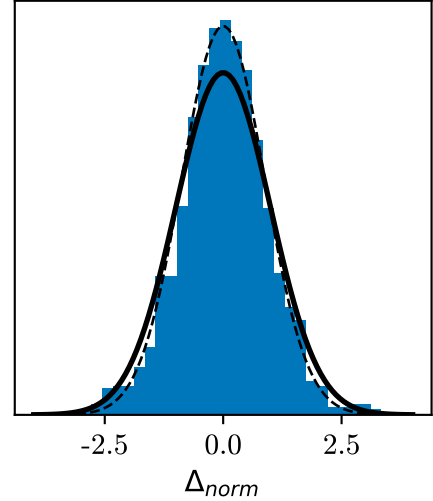


Figure 2. The normalised Hubble residuals of the Pantheon+ data in a best fit Λ CDM universe, over-plotted with a standard Gaussian in the solid black line. The residuals are normalised in a way that takes into account covariances, $\Delta_{\text{norm}} = C^{-1/2}\Delta$. As the figure shows, the data do not normalise perfectly to the Gaussian distribution, with the data showing a higher peak. The dashed line shows a Gaussian with variance $\sigma^2 = 0.88$.

This has covariance matrix M given by:

$$M = A \frac{2^{1/B} \Gamma(\frac{N+2}{2B})}{N \Gamma(\frac{N}{2B})} C. \quad (26)$$

We extend this distribution even further by introducing a “skew” parameter K , to test the residuals for asymmetry. This combines to give the full distribution:

$$\mathcal{L} = \frac{N(A, B, C)}{\prod_i (1 - K[(AC)^{-1/2}\Delta]_i)} \exp \left(-\frac{1}{2} (V^T \cdot V)^B \right), \quad (27)$$

where V are skewed, scaled, normalised residuals:

$$V_i = \frac{1}{K} \log \left(1 - K[(AC)^{-1/2}\Delta]_i \right). \quad (28)$$

If any of the residuals are too large ($K(AC)^{-1/2}\Delta_i > 1$) then the distribution is set to 0, so for higher K , large parts of the parameter space are ruled out.

The full “Doubly-Generalised” distribution tends back to equation (25) as $K \rightarrow 0$, and reverts back to the pure Gaussian in equation (15) for $(A, B, K) = (1, 1, 0)$. For A and B we take priors $A \sim U[0.5, 2]$, $B \sim U(0, 3]$, which are sufficiently wide as these bounds extend out to highly non-Gaussian forms. For the skew parameter we take the thin prior $K \sim U[-0.2, 0.2]$. Since we have > 1000 data points we expect to find values at $> 3\sigma$ even with the best fit parameters, so making $\frac{1}{|K|} < 4$ would have likelihood 0 on almost all of the cosmological parameter space. The univariate case of these distributions are shown for a selection of B, K values in figure 3.

Student’s t

While the generalised Gaussian is a wide-field first attempt at finding non-gaussianity in the data, we can also try a more specific distribution. The 1D t-distribution is constructed from a Gaussian distribution divided by an independent χ^2_ν distribution, so arises as the posterior predictive when working with a Gaussian distribution

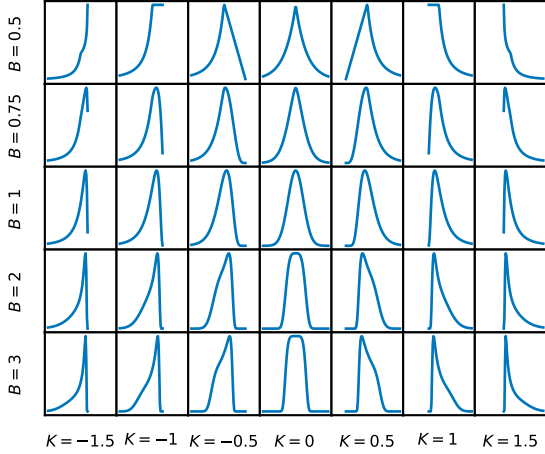


Figure 3. Probability densities of a univariate generalised Gaussian from equation (27) for varying values of B and K . The scale parameter and covariance matrix are set to unity, $A = C = 1$.

of unknown variance. Analogously, the multivariate t-distribution shows up as the posterior predictive in a number of marginalisation problems, in particular when estimating the covariance matrix with ν sums of pairwise deviation products. The process of creating the covariance matrix has uncertainty attached to it, and neglecting this effect by putting in fixed values for the covariance matrix could misrepresent the scatter of the data. This distribution has likelihood function:

$$\mathcal{L} = \mathcal{N}(\nu, C) \left[1 + \frac{1}{\nu} (\Delta^T \Sigma^{-1} \Delta) \right]^{-(\nu+k)/2}, \quad (29)$$

where k is the length of the residuals vector, the degrees of freedom ν is the shape parameter of the distribution, and Σ is related to the covariance matrix by:

$$\Sigma = \frac{\nu - 2}{\nu} C. \quad (30)$$

For this test we use prior $\nu \sim \log U(2, 10^6)$. This bound is chosen because below $\nu = 2$ the t-distribution has no defined variance, and is extended to high ν since we are running this on a large number of data points. In this case a uniform prior would be inappropriate, since it would over-weigh the high ν values, $\nu > 10^5$. At these high values the distribution tends towards a Gaussian, as the aforementioned covariance matrix estimate becomes more accurate. See Feeney et al. (2018) for a detailed explanation of the appropriate prior to use for ν in the univariate case.

3 RESULTS

3.1 Cosmological Model Comparison

In this section we present the results of the model comparison and parameter estimation on each of the cosmological models.

Table 3 gives the posterior estimates for each cosmological parameter with a Gaussian likelihood, as well as the relative evidences of each model. The parameters are quoted with their median values, and

Table 2. Prior distributions for all model parameters in the dark energy models as well as likelihood models tested in this work. For parameters without strong theoretical bounds wide uniform priors are used, and in the case of w the prior allows for non-accelerating cosmologies.

Parameter	Prior	Model
M_B	U[-20,-18]	All
H_0	U[50,100]	All
Ω_m	U[0,1]	All
w	U[-2,-2]	wCDM (Flat and Curved)
Ω_k	U[-0.5,0.5]	wCDM and Λ CDM
w_0	U[-2,2]	Algebraic thawing
p	U[-4,4]	Algebraic thawing
B_1	U[0,6]	Bimetric gravity
δw_0	U[-2,1]	One parameter slow-roll dark energy
Ω_e	U[0,0.25]	Growing ν mass
Ω_ν	U[0,0.4]	Growing ν mass
δ	U[-0.4,0.6]	Dark energy transition
A	U[0.5,2]	Generalised Gaussian
B	U(0,3]	Generalised Gaussian
K	U[-0.2,0.2]	Generalised Gaussian
ν	logU(2,10 ⁶)	Student's t

the intervals given are Highest Posterior Density Intervals (Hyndman 1996), a form of Bayesian credibility interval.

Figure 4 shows an example of the 1 and 2D posteriors for F Λ CDM and FwCDM, with the median/interval values over plotted. We can see that both models find a consistent H_0 value, showing the independence between w and H_0 , as well as the effect on the posterior uncertainty of Ω_m from adding the additional parameter.

The relative evidence values are plotted in figure 5, where 4 of the models are disfavoured at a moderate evidence level (shown by the dashed blue line). The black lines in figure 5 show the relative $\mathcal{D}_{KL}$ of each model, the size of the ‘‘Occam Penalty’’ of each model compared to F Λ CDM. As well as the evidence values favouring F Λ CDM, the log $\mathcal{L}_{max}$ values show that none of the models are able to find a better fit than F Λ CDM with their additional parameters.

3.2 Peculiar Velocity Corrections

The peculiar velocities (PVs) of the SNe can be up to 300 km s⁻¹, which accounts for $\sim 10\%$ of the total redshift at $z = 0.01$. For higher redshifts, PVs still contribute to the total uncertainty, and the v_{pec} values used to correct the measured redshifts of the SNe in Pantheon+ are evaluated in Peterson et al. (2022) and Carr et al. (2022). If the process of modelling the large-scale velocity field of the universe out to high redshift has assumed a standard cosmological model, this may have an effect on model comparison, by artificially increasing the fit and evidence for F Λ CDM. By analysing the Pantheon+ data with and without the peculiar velocity corrections, we can explore whether any possible source of bias is displayed within the relative Bayesian evidences of the models. This concern is addressed in Carr et al. (2022), where parameter estimation is run by minimising a χ^2 , finding that the tested parameters H_0 and w do not shift appreciably. In this work we are interested in the strength of model comparison instead of parameter estimation, so we recheck this bias with our methodology to see if any trace of an F Λ CDM assumption in velocity estimation shows up in the Bayesian Evidence of non-F Λ CDM models.

For this test, instead of using the fiducial case, wherein the redshifts are corrected for the peculiar velocities, we perform model comparison using the CMB frame redshifts with no correction. Additionally,

Table 3. Parameter estimations and evidences (relative to Λ CDM) for each model. All parameter values are quoted as median $\pm$ Bayesian credibility interval. Included also are the odds of each model, given by $\mathcal{Z}$: $\mathcal{Z}_0$.

Model	H_0	Ω_m	Parameter 2	Parameter 3	$\Delta \log \mathcal{Z}$	Odds
Flat Λ CDM	$73.71^{+1.01}_{-1.03}$	$0.33^{+0.02}_{-0.02}$	–	–	0	–
Λ CDM	$73.67^{+1.03}_{-1.03}$	$0.32^{+0.05}_{-0.05}$	$\Omega_k = 0.03^{+0.12}_{-0.13}$	–	-1.1	<1:2
Flat w CDM	$73.67^{+1.03}_{-1.02}$	$0.32^{+0.07}_{-0.06}$	$w = -0.97^{+0.16}_{-0.15}$	–	-2.2	<1:9
w CDM	$73.69^{+1.02}_{-1.06}$	$0.29^{+0.07}_{-0.06}$	$\Omega_k = 0.09^{+0.32}_{-0.22}$	$w = -1.07^{+0.49}_{-0.28}$	-2.3	<1:9
Bimetric Gravity	$73.77^{+1.0}_{-1.02}$	$0.35^{+0.03}_{-0.04}$	$B_1 = 0.55^{+0.23}_{-0.55}$	–	-1.7	<1:5
Slow Roll Inflation	$73.69^{+1.07}_{-1.02}$	$0.32^{+0.05}_{-0.05}$	$\delta w_0 = 0.02^{+0.16}_{-0.13}$	–	-2.1	<1:8
Algebraic Thawing	$73.69^{+1.00}_{-1.10}$	$0.33^{+0.05}_{-0.05}$	$w_0 = -0.99^{+0.17}_{-0.14}$	$p = 0.55^{+2.71}_{-2.22}$	-2.3	<1:10
Growing Neutrino Mass	$73.56^{+1.01}_{-1.05}$	$0.25^{+0.07}_{-0.05}$	$\Omega_e = 0.13^{+0.07}_{-0.09}$	$\Omega_\nu = 0.11^{+0.05}_{-0.11}$	-0.6	<1:1
Dark Energy Transition	$73.68^{+1.49}_{-1.50}$	$0.33^{+0.04}_{-0.04}$	$\delta = 0.00^{+0.03}_{-0.03}$	–	-2.6	<1:12

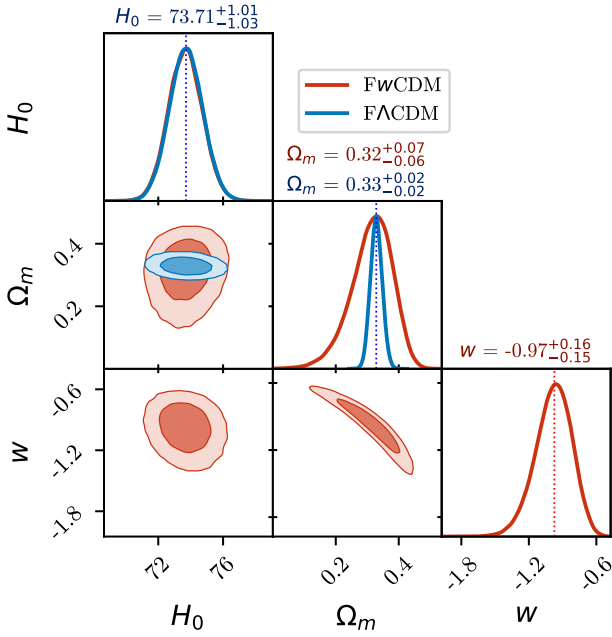


Figure 4. The posterior distribution of H_0 , Ω_m and w for a Gaussian likelihood. Over plotted are the H_0 and Ω_m posteriors from Λ CDM, finding similar values for both values since w is constrained tightly to the null $w = -1$.

we remove the component of the covariance matrix representing the error due to peculiar velocity estimation. This matrix, containing both the statistical and systematic components of this error, can be found in the “groupings” section of the Pantheon+ release. The relative evidences are shown in figure 6, with the relative evidences from figure 5 over-plotted. Note that each set of tests (CMB frame and fiducial) are plotted relative to their own Λ CDM, as the evidences of the CMB frame tests are much lower than the fiducial. While the distribution of models is roughly the same for both, in the uncorrected data all models (except Bimetric gravity) perform slightly better as compared to Λ CDM, with $|\Delta \log \mathcal{Z}| \sim 1.28$ lower on average.

3.3 Alternate Likelihood Models

The initial test is repeated, taking the same Hubble Flow and Calibrator SN set, but now through each of the alternate likelihood models described in section 2.6. For the generalised Gaussian, we run a test with each combination of the parameters free/fixed, giving seven al-

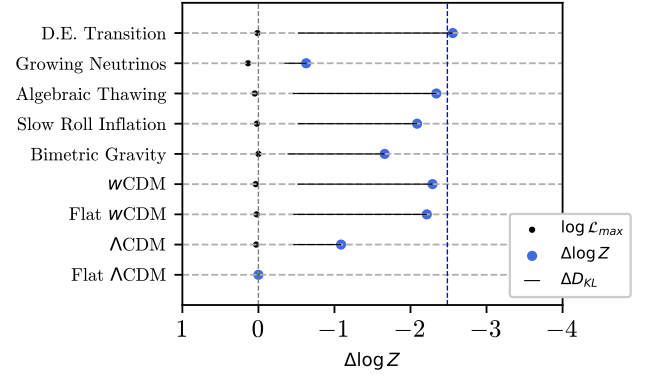


Figure 5. The evidences (relative to Λ CDM) of each model, run on the set of all SNe with $z > 0.023$. The black lines show the $\mathcal{D}_{KL}$ of the models, i.e. the Occam penalty applied to them by the Bayesian evidence. Also shown is the “Moderate Evidence” cut-off at $\Delta \log \mathcal{Z} = -2.5$, which 1 models fall beyond, a dark energy transition. The black dots show the maximum likelihood point (again relative to Λ CDM) achieved by each model.

ternate forms of the likelihood function, and these tests are referred to by their free parameters: A , B , K , AB , AK , BK , ABK . This gives nine total likelihoods, along with the regular Gaussian and the Student’s t . The evidences of each cosmology/likelihood model pair are given in table 4, and all cosmology-marginalised parameter values are given in table 6.

For the set of generalised Gaussian models, the results of the model comparison in figure 10 can be explained by the estimates of each of the parameters A , B , and K , as well as their relationships to each other. As shown in figure 7 A is centered on 0.88 ± 0.03 , finding a distinctly non-null value, and this results in a better fit. On the other hand, both B and K are tightly constrained to their null values: $B = 1.02 \pm 0.005$ and $K = 0.02 \pm 0.02$. This means that the B , K , and BK tests all underperform the Gaussian; they introduce an unnecessary parameter and are punished according to Occam’s razor.

For a multivariate distribution with $N = 1448$, the covariance of the generalised Gaussian, given in equation (26), is extremely sensitive to B , leading in this case to a very tight posterior centred on $B = 1$. This can also be seen when considering that for a normal distribution, the distance term $(\Delta^T C^{-1} \Delta)^B \sim N^B$, so any non-unity B in the exponent has an exponential effect on the size of this term. As well as this, A and B have a degenerate effect on the scale of the covariance, which means that in the AB test this degeneracy breaks the $A = 0.88$ preference that had benefited the model. This also results in the full ABK model finding a disfavoured evidence, in combination with the data finding a null value of the skew.

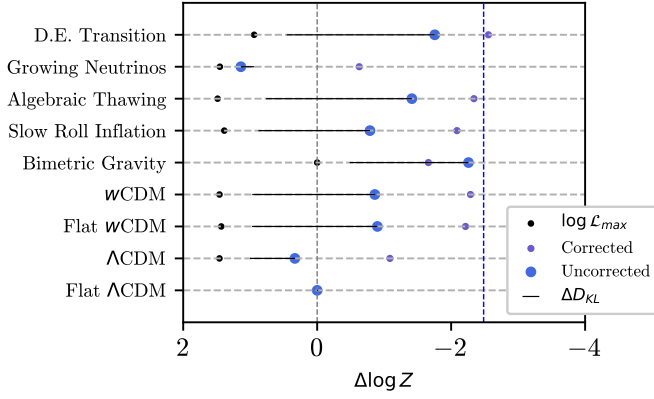


Figure 6. The relative evidences of each model with and without peculiar velocity corrections. Without peculiar velocity corrections no models fall beyond the moderate evidence line, and all models except Bimetric gravity find a better maximum fit relative to Λ CDM than the first analysis figure 5, shown by the black $\log \mathcal{L}_{\max}$ dots.

Finally, A and K are almost independent in their effect on the model comparison. This is shown in the ratio of their evidences:

$$\frac{\mathcal{Z}_K}{\mathcal{Z}_{\text{Gaussian}}} \approx \frac{\mathcal{Z}_{AK}}{\mathcal{Z}_A}. \quad (31)$$

This result is reassuring: A is analogous to the covariances being uniformly smaller/larger, which is independent from the role of K in skewing the distribution (especially since it skews the residuals after normalisation), so including K in a model should produce the same evidence difference whether A is free or fixed.

The t-distribution also finds an improvement over the Gaussian of $\Delta \log \mathcal{Z} = +2.70$, or a Bayes factor of $B = e^{2.70} \approx 15$. Figure 8 shows the 1D posterior for ν on each cosmological model. If the posterior was one sided and increasing towards higher ν , this would suggest that the more Gaussian the distribution, the better the fit, but while the value of ν is not tightly constrained, it peaks at $\nu \sim 150$ on each model. The posterior and evidence show a preference for a non-Gaussian form.

3.4 Model Marginalisation

In the case of having both a selection of cosmological models and a selection of scatter models, we can marginalise out one or both of these choices to obtain either: **a)** Cosmology independent evidences for each scatter model or **b)** Scatter-model independent evidences for each cosmology. Labeling the cosmological model as $\mathcal{M}_c$ and scatter model $\mathcal{M}_s$, we marginalise the evidence (as in equation (35)):

$$p(d|\mathcal{M}_s) = \sum_{\mathcal{M}_c} p(d|\mathcal{M}_c, \mathcal{M}_s) p(\mathcal{M}_c|\mathcal{M}_s), \quad (32)$$

where since $p(\mathcal{M}_c|\mathcal{M}_s)$ is not conditioned on the data it takes the prior values, i.e. is uniform.

This equation also holds with “c” and “s” swapped to marginalise out the scatter model on the cosmological evidences. The results of these marginalisations are shown in figures 9 and 10. Since all of the summing in equation (32) takes place outside of the log, the results are weighted very strongly towards high evidence models, i.e. the marginalised likelihood comparison does not vary appreciably from the likelihood comparison for Λ CDM.

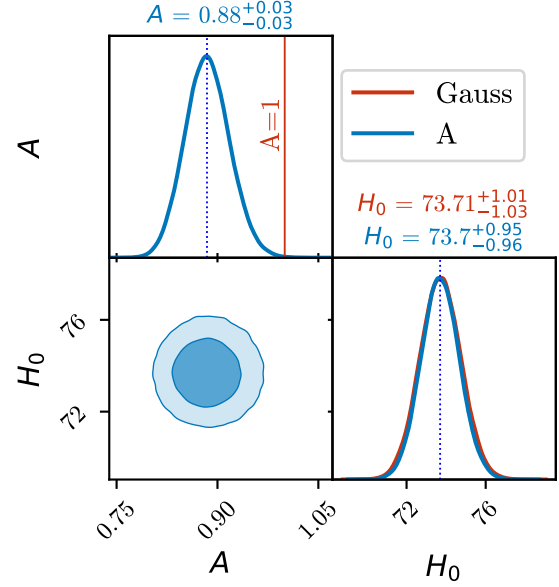


Figure 7. The posterior for A and H_0 in A -scaled Λ CDM (blue), compared with the fiducial analysis (orange). A takes on a distinctly non-null value, and as a result the H_0 estimate is slightly tighter.

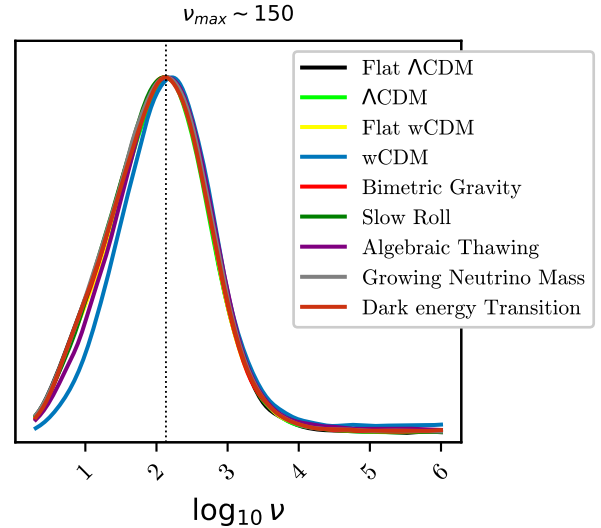


Figure 8. The 1D posterior distribution of ν on each cosmological model. The data show a clear preference for lower ν values, with the posterior dying off where the distribution limits towards a Gaussian at high ν .

We also calculate the total $\mathcal{D}_{\text{KL}}$ for each cosmology/scatter model, by considering the $\mathcal{D}_{\text{KL}}$ of the joint distribution over model and parameter space:

$$p(\theta, \mathcal{M}_s|d, \mathcal{M}_c) = p(\theta|d, \mathcal{M}_c, \mathcal{M}_s) p(\mathcal{M}_s|d, \mathcal{M}_c). \quad (33)$$

From Kroupa et al. (2023), the joint $\mathcal{D}_{\text{KL}}$ can be broken down as:

$$\mathcal{D}_{\text{KL}}(\theta, \mathcal{M}_s|\mathcal{M}_c) = \mathcal{D}_{\text{KL}}(\mathcal{M}_s|\mathcal{M}_c) + \sum_{\mathcal{M}_s} p(\mathcal{M}_s|d, \mathcal{M}_c) \mathcal{D}_{\text{KL}}(\theta|\mathcal{M}_s, \mathcal{M}_c), \quad (34)$$

i.e. the posterior averaged $\mathcal{D}_{\text{KL}}$ from each scatter model, as well as

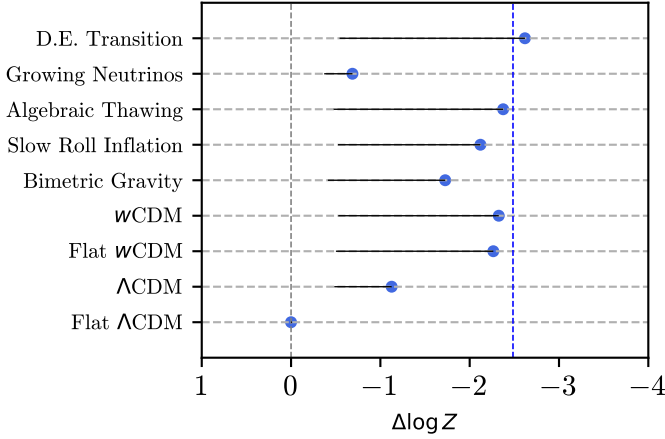


Figure 9. The relative evidences of each cosmology, marginalised over the scatter model. Λ CDM has the highest evidence, providing similar results to the Gaussian analysis (figure 5). The black lines show the joint $\mathcal{D}_{\text{KL}}$ of the scatter model/parameters.

the $\mathcal{D}_{\text{KL}}$ of the (discrete) distribution of scatter models. These $\mathcal{D}_{\text{KL}}$ are plotted on the marginalised evidence plots in figures 9 and 10. For the cosmological models, each model finds the same fit as Λ CDM, and the $\mathcal{D}_{\text{KL}}$ s line up at $\Delta \log \mathcal{Z} \approx -0.5$ just as in the Gaussian case.

Fully marginalising over both cosmology and scatter models should give the best estimate of H_0 that SN are able to offer, using the data to the fullest extent. Marginalising over models gives:

$$p(H_0|d) = \sum_{\mathcal{M}_c, \mathcal{M}_s} p(H_0|d, \mathcal{M}_c, \mathcal{M}_s) p(\mathcal{M}_c, \mathcal{M}_s|d). \quad (35)$$

Over the marginalised posterior, we find a value of $H_0 = 73.67 \pm 0.99 \text{ km s}^{-1} \text{ Mpc}^{-1}$, in agreement with the results of the SH₀ES collaboration [Riess et al. \(2022\)](#), $H_0 = 73.04 \pm 1.04 \text{ km s}^{-1} \text{ Mpc}^{-1}$, although with slightly tighter constraints owing to the more general treatment of the distribution of the residuals.

All other parameters are similarly marginalised, and given in tables 5 and 6, along with the marginalised evidences and odds.

3.5 Absolute Magnitude Evolution Models

The triangle plots for both M_B evolution models are given in figures 11 and 12, as well as the evidence values relative to Λ CDM run on the same dataset.

Both dynamic M_B models come out with lower evidence than Λ CDM, and similarly to the dark energy models have their additional parameters constrained to their null values. The data show no support for a dynamic M_B .

The transition model is ruled out at a higher level than the dark energy transition model, $\Delta \log \mathcal{Z} = -4.0$ relative to Λ CDM. This model has lower evidence than the dark energy transition model, likely because the prior range of $M_{B1} \in [-20, -18]$ has a much larger effect than the dark energy transition parameter $\delta \in [-0.4, 0.6]$.

The lack of an M_B transition shown in figure 11 can be attributed to the same reason a dark energy transition is not favoured by the evidence; the effect may not occur at $z_t = 0.1$. However, due to the high peculiar velocity error in the lower end of this data set, we are unable to test the suggested lower value of $z_t = 0.02$. For the linear model, the M_B gradient is not tightly constrained, as its effect is reduced on lower redshifts (where the majority of the data lie), and

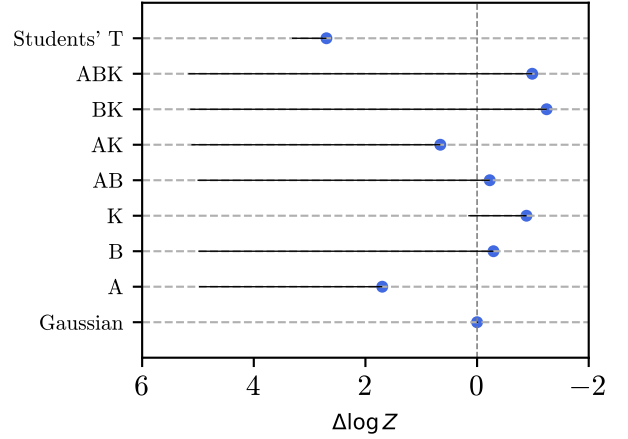


Figure 10. The relative evidences of each scatter model, marginalised over the cosmological model. Two likelihood functions, the A -scaled Gaussian and the Student's t -distribution are favoured over the Gaussian. The black lines show the joint $\mathcal{D}_{\text{KL}}$ of the cosmological model/parameters.

shows some degeneracy with Ω_m in figure 12. This suggests that neglecting the effects of a magnitude gradient could lead to over or underestimating the matter density. However, the SNe Ia evolution modelled by this gradient is likely to be reflected in the light curve of the SN, which is precisely what the Pantheon+ data have been calibrated to standardise. Therefore if a first order $M_B - z$ gradient does exist in SN, the Pantheon+ catalogue may have already been corrected for it, via the light-curve fitting parameters. The negative $\Delta \log \mathcal{Z} = -1.1$ of this model reflects the lack of improvement of fit, and the weakness of the evidence difference reflects the weakness of the constraint on the M_B gradient (as in the model comparison discussion).

Importantly for both of these models the H_0 estimation remains the same, and does not vary drastically towards the Planck value (although the uncertainty in H_0 is doubled by the inclusion of the M_{B2} parameter). This paper finds no evidence to suggest that the Hubble Tension can be resolved via a step-like or linear $M_B - z$ relationship.

4 DISCUSSION

4.1 Alternate Likelihood Distributions

From the forms of the various likelihood functions in equations (15), (25), (27) and (29), each distribution is (naturally) peaked by minimising the residuals $\Delta^T \mathbf{C}^{-1} \Delta$ (except for the skewed distributions $K \neq 0$ which have their peak shifted only by a small factor). As a result the choice of likelihood function does not change the central value of the parameter estimation, and therefore the conclusions of the model comparison do not change over different scatter models. Additionally, since all alternate cosmologies are converging to their Λ CDM limit, scatter model parameters are not sensitive to the specific cosmological model. This leads to an independence in this analysis between the scatter and cosmological models.

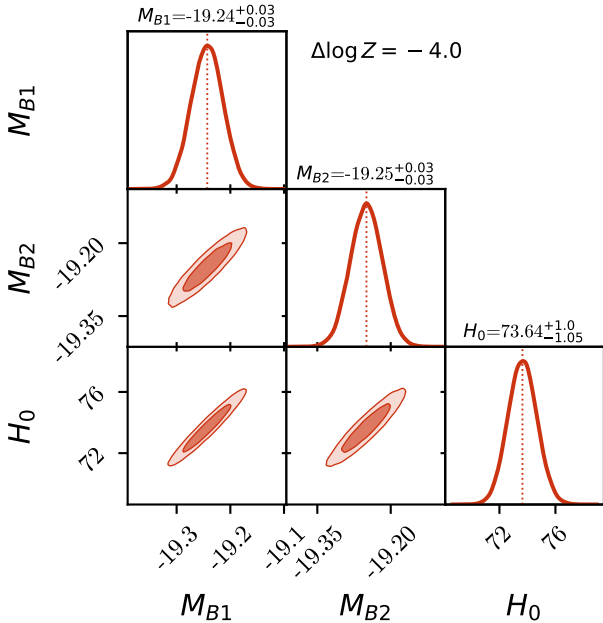
The two scatter models with the greatest evidence - a scaled Gaussian with only A as a free parameter and the student's- t distribution - appear to succeed for unrelated reasons, suggesting there could be more than one factor contributing to the improvement of fit in each case.

Table 4. Evidences for each cosmological/likelihood pair model. The marginalised evidences are given for a cosmology/scatter independent model comparison, as described in section 3.4.

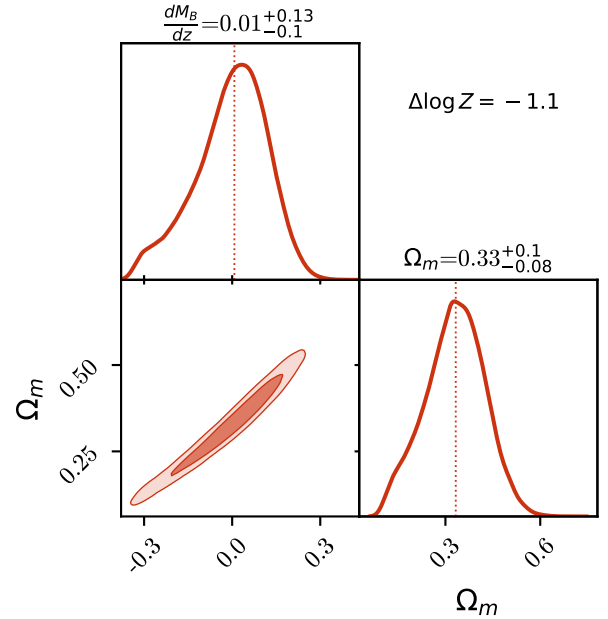
Model	Gaussian	A	B	K	AB	AK	BK	ABK	Student's T	Scatter Independent
Flat Λ CDM	811.3	813.1	811.1	810.4	811.1	812.0	810.1	810.3	814.0	812.4
Λ CDM	810.2	811.9	810.0	809.3	810.0	810.9	809.0	809.3	812.9	811.3
Flat w CDM	809.1	810.8	808.8	808.3	808.8	809.8	807.8	808.1	811.8	810.1
w CDM	809.0	810.7	808.7	808.1	808.8	809.7	807.7	808.1	811.7	810.1
Bimetric Gravity	809.7	811.3	809.4	808.8	809.4	810.3	808.4	808.6	812.3	810.7
Slow Roll Inflation	809.2	810.8	808.9	808.4	809.0	810.0	808.0	808.3	811.9	810.3
Algebraic Thawing	809.0	810.6	808.7	808.2	808.7	809.8	807.8	808.0	811.7	810.0
Growing Neutrino Mass	810.7	812.4	810.3	809.7	810.4	811.3	809.4	809.6	813.4	811.7
Dark Energy Transition	808.8	810.4	808.4	807.9	808.4	809.4	807.4	807.7	811.4	809.8
Cosmology Independent	810.1	811.8	809.8	809.2	809.8	810.7	808.8	809.1	812.8	–

Table 5. Parameter estimations and evidences for each cosmological model, now marginalised over the scatter model. All parameter values are quoted as median $\pm$ Bayesian credibility interval, and again the (marginalised) odds and evidences of each model are listed, with odds given as $Z : Z_0$.

Model	H_0	Ω_m	Parameter 2	Parameter 3	$\Delta \log Z$	Odds
Flat Λ CDM	$73.70^{+0.95}_{-0.99}$	$0.33^{+0.02}_{-0.02}$	–	–	0	–
Λ CDM	$73.67^{+0.95}_{-1.00}$	$0.32^{+0.05}_{-0.05}$	$\Omega_k = 0.03^{+0.12}_{-0.12}$	–	-1.1	<1:3
Flat w CDM	$73.67^{+0.96}_{-1.00}$	$0.32^{+0.07}_{-0.05}$	$w = -0.97^{+0.16}_{-0.14}$	–	-2.3	<1:9
w CDM	$73.69^{+0.96}_{-1.01}$	$0.29^{+0.07}_{-0.06}$	$\Omega_k = 0.08^{+0.33}_{-0.22}$	$w = -1.07^{+0.49}_{-0.28}$	-2.3	<1:10
Bimetric Gravity	$73.78^{+0.99}_{-0.95}$	$0.35^{+0.03}_{-0.04}$	$B_1 = 0.53^{+0.22}_{-0.53}$	–	-1.7	<1:5
Slow Roll Inflation	$73.68^{+0.99}_{-1.00}$	$0.32^{+0.05}_{-0.04}$	$\delta w_0 = 0.02^{+0.15}_{-0.13}$	–	-2.1	<1:8
Algebraic Thawing	$73.69^{+1.00}_{-0.98}$	$0.33^{+0.05}_{-0.05}$	$w_0 = -0.99^{+0.16}_{-0.14}$	$p = 0.54^{+2.62}_{-2.29}$	-2.4	<1:10
Growing Neutrino Mass	$73.58^{+1.01}_{-0.94}$	$0.26^{+0.07}_{-0.05}$	$\Omega_e = 0.13^{+0.12}_{-0.05}$	$\Omega_\nu = 0.10^{+0.04}_{-0.10}$	-0.7	<1:1.9
Dark Energy Transition	$73.69^{+1.37}_{-1.44}$	$0.33^{+0.03}_{-0.04}$	$\delta = 0.00^{+0.03}_{-0.03}$	–	-2.6	<1:13

**Figure 11.** The posterior for an M_B transition model, showing the near and far magnitudes (split about redshift $z=0.1$). Both values are tightly centred on the Cepheid calibrated value.

It is clear why the A-scaled Gaussian succeeds: from figure 2 the Hubble residuals are packed more tightly than the Pantheon+ covariance predicts, and therefore tightening the distribution with $A = 0.88$ allows the smaller residuals to contribute more to the likelihood and find a better fit. Figure 7 also shows that uniformly

**Figure 12.** The posterior for a linear M_B model. The figure shows a degeneracy with Ω_m , and approximately triples the size of its error bounds compared to Λ CDM.

scaling the Pantheon+ covariances down makes the uncertainty on other model parameters smaller, a natural result considering that for a uniform prior, the likelihood and posterior are proportional. A-scaled Λ CDM achieves the tightest H_0 of any model, at ± 0.95 . This, more clearly than any marginalised result, displays the usefulness of treating the residuals properly in achieving better parameter estimates. As

motivation to look into the root of this result; if we had an a priori reason to insert this scaled covariance matrix (and ignore all of the other tested likelihoods) then we would gain the full benefit of the lower variances and we could, in good faith, report this even tighter H_0 estimate. The t-distribution is less immediately interpreted. In a full Bayesian treatment, instead of using the Pantheon+ covariance matrix as it is, we would incorporate the observations that led to the creation of this covariance matrix (such as SN simulations), as well as a prior for the covariance matrix, into our analysis. This would allow us to marginalise out the true covariance and construct a posterior predictive, i.e. a likelihood that is conditioned on the previously observed SN data. In the case that the prior for the covariance matrix was an uninformative inverse-Wishart distribution, the correct likelihood for the data would be a multivariate t-distribution, where the shape parameter Σ is the sample sum of squares of the input data. In this case the degrees of freedom $\nu = n - k$, where n is the number of observations used to construct the covariance matrix, and k is the size of the covariance matrix (1448). For a full derivation of these conjugate distributions, see Chapter 3.6 of Gelman et al. (2013).

While a 1D t-distribution with degrees of freedom $\nu \geq 10$ is indistinguishable from a Gaussian by eye, in the multivariate case much higher ν values are relevant, due to the higher dimensionality. This can be seen in the $\nu + k$ term in the exponent of equation (29). Over 1448 data points, this difference can be resolved from the standard Gaussian distribution, and a statistical preference for $\nu \sim 150$ can be identified. The construction described above is not exactly how this inference runs, but treating the covariance matrix as a sample covariance matrix estimated as a sum of deviations, and accounting for the uncertainty in this via a t-distribution, yields a slightly higher evidence, in a cosmology independent way, by a Bayes factor of $e^{2.70} \approx 151$. Although the bound is not very tight, the posterior on ν says that the covariance matrix was effectively estimated with $n = k + \nu \sim 2000$ observations.

It should be noted that the multivariate t-distribution has covariances given by $(\nu/(\nu - 2)) \times C$ which is a similar effect to the A parameter. However at $\nu \sim 150$ this corresponds to an A value very close to 1, so is not the main effect at play in changing the evidences.

4.2 Model Comparison

From figure 9, the Occam penalty approximation equation (9) holds; the difference between where the black $\mathcal{D}_{\text{KL}}$ lines end ($\langle \log \mathcal{L} \rangle_{\mathcal{P}}$) and the black log $\mathcal{L}_{\text{max}}$ dots is in all cases $\approx 1/2$, and each model has $\Delta \hat{d} \approx 1$.

In both the Gaussian and the marginalised analysis (tables 3 and 5) all parameters lie within 1σ of their null value, i.e. the value that turns the theory back into Λ CDM, and therefore all models find the same fit as Λ CDM. The alternate dark energy models do not achieve a significantly better fit on the data, and due to the Occam penalty on their expanded prior space have lower evidence than Λ CDM in all cases, where the scale of the evidence difference is determined solely by the $\mathcal{D}_{\text{KL}}$.

With some simplifying assumptions we can calculate the theoretical $\Delta \log \mathcal{Z}$ of a future model comparison. Consider a one parameter expansion to Λ CDM, where the posterior of the additional parameter is still centred around its null value (so that $\Delta \log \mathcal{L}_{\text{max}} \approx 0$), and the additional parameter θ^* is approximately independent of Ω_m (so that $\Delta \mathcal{D}_{\text{KL}} \approx \mathcal{D}_{\text{KL}}(\theta^*)$). For the sake of simplicity, give this parameter a uniform prior width b and Gaussian posterior with variance σ^2 . Additionally assume that the prior is sufficiently wide, so that $\mathcal{D}_{\text{KL}}$ can be effectively evaluated over $\pm\infty$. Calculating $\mathcal{D}_{\text{KL}}(\theta^*)$ analytically, we find:

$$\int P(\theta^*) \log(P(\theta^*)/\pi(\theta^*)) d\theta^* = \log(b/\sigma) - \frac{1}{2} \log(2\pi) - \frac{1}{2}. \quad (36)$$

With $\Delta \hat{d} = 1$ this provides a difference in evidence of

$$\begin{aligned} \Delta \log \mathcal{Z} &= (\langle \log \mathcal{L} \rangle_{\mathcal{P}} - \mathcal{D}_{\text{KL}})_1 - (\langle \log \mathcal{L} \rangle_{\mathcal{P}} - \mathcal{D}_{\text{KL}})_0 \\ &\approx -\mathcal{D}_{\text{KL}}(\theta^*) - \frac{1}{2} = \frac{1}{2} \log(2\pi) - \log(b/\sigma). \end{aligned} \quad (37)$$

So for the well behaved, one-parameter extensions that centre around Λ CDM (FwCDM, Λ CDM, Slow Roll) the Bayes factor has a clear relationship to the constraint on the additional parameter: $B_{01} \propto \sigma^{-1}$. While this result is only exact in a highly idealized case, as a heuristic it informs us how much more precise future surveys need to become to rule out parameters like curvature with SNe alone. The strength of our belief in discarding an extension to our base model is directly proportional to how tightly we can constrain its added parameters to 0.

4.3 Peculiar Velocity Corrections

From the evidence diagram figure 6 the inference with the CMB frame redshifts does not distinguish between the dark energy models as well as inference after peculiar velocity correction are added to the redshifts and two models, the Growing Neutrino mass and curved Λ CDM now find a higher evidence than the fiducial cosmology. These changes in the evidence differences in the uncorrected analysis could be caused by the data having less constraining power, due to the additional noise. From equation (37) if additional parameters cannot be constrained as tightly, then evidence differences will be smaller. Due to the change in the covariance matrix, as well as the additional scatter, the uncorrected analysis has evidence values ~ 100 lower than the corrected, reflecting that these corrections are a necessary part of the analysis of SN Ia.

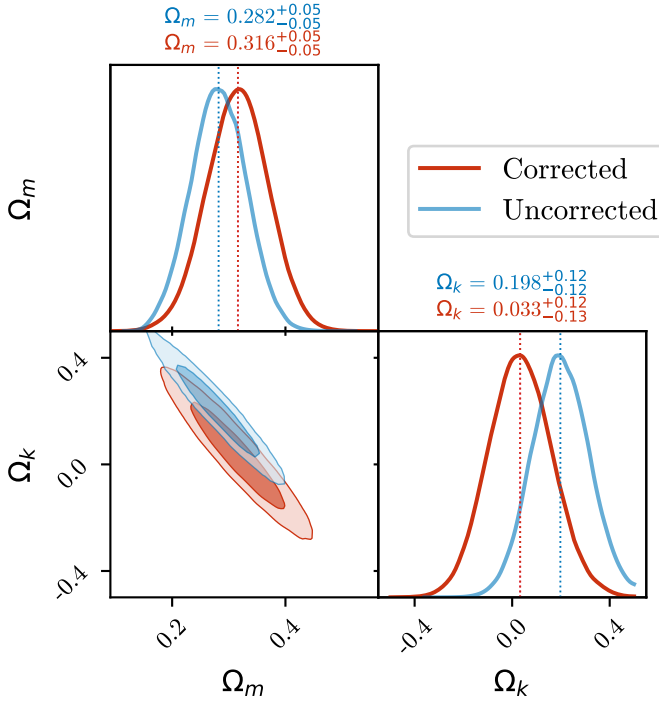
In the first analysis (figure 5), the black log $\mathcal{L}_{\text{max}}$ dots line up with Λ CDM; the alternate models provide no improvement in fit. As discussed before, because all alternate models contain Λ CDM this implies that their additional parameters are all being constrained around their null values. In the case of the uncorrected data however, all models except for Bimetric gravity are now achieving a slightly better fit, and the evidence difference is no longer just the penalty terms. The additional scatter that is added by removing peculiar velocity corrections is fit well by all models except Bimetric gravity, in a way that the scatter in the normal analysis does not favour any model over Λ CDM. As a result, the growing neutrino mass model is now weakly favoured by the evidence, at $\Delta \log \mathcal{Z} = +1.13$, and Λ CDM is now inconclusive against Λ CDM at $\Delta \log \mathcal{Z} = +0.33$. In order for a model to achieve an improvement in fit it needs to have one of its parameters stray from the null value, displayed most clearly with curvature in Λ CDM. Figure 13 shows the posterior of Λ CDM with and without velocity corrections, where the maximum likelihood value of curvature is non-zero in the uncorrected analysis. Unlike the corrected data where introducing curvature doesn't fit the data better, the added scatter from removing peculiar velocity corrections is slightly better fit by a non-zero curvature, by enough to overcome the Occam penalty.

5 CONCLUSIONS

In this paper we have considered nine dark energy models, covering Λ CDM and its extensions, dynamical dark energy theories, and a

Table 6. Parameter estimations and evidences for each scatter model, marginalised over the cosmological model. All parameter values are quoted as median $\pm$ Bayesian credibility interval, and the marginalised odds and evidences of each scatter model are listed, with odds given as $\mathcal{Z} : \mathcal{Z}_0$.

Model	Parameter 1	Parameter 2	Parameter 3	$\Delta \log \mathcal{Z}$	Odds
Gaussian	–	–	–	0	–
<i>A</i>	$A = 0.88^{+0.03}_{-0.03}$	–	–	+1.70	>5:1
<i>B</i>	$B = 1.02^{+0.005}_{-0.005}$	–	–	-0.29	<1:1.3
<i>K</i>	$K = 0.02^{+0.02}_{-0.02}$	–	–	-0.88	<1:2
<i>AB</i>	$A = 1.31^{+0.68}_{-0.27}$	$B = 1.06^{+0.16}_{-0.12}$	–	-0.23	<1:1.2
<i>AK</i>	$A = 0.89^{+0.03}_{-0.03}$	$K = 0.02^{+0.02}_{-0.02}$	–	+0.65	>2:1
<i>BK</i>	$B = 1.02^{+0.005}_{-0.005}$	$K = 0.02^{+0.02}_{-0.02}$	–	-1.25	<1:3
<i>ABK</i>	$A = 1.42^{+0.58}_{-0.26}$	$B = 1.06^{+0.05}_{-0.03}$	$K = 0.02^{+0.02}_{-0.02}$	-0.98	<1:2
Student's t	$\log_{10}(\nu) = 2.05^{+0.71}_{-0.75}$	–	–	+2.70	>14:1

**Figure 13.** The scatter added by removing peculiar velocity corrections appears to be better fit by a non-zero curvature, as shown in the blue posterior. This is the reason that the evidence for curved Λ CDM is higher than the default cosmology when run on the uncorrected data.

theory of modified gravity. We find that Flat Λ CDM has a higher evidence than all other models (figure 9), and this preference holds independent of the form of the likelihood function.

When removing peculiar velocity redshift corrections, the growing neutrino mass model shifts to have higher evidence than Λ CDM by $\Delta \log \mathcal{Z} = +1.13$, making the model comparison weakly in favour of the growing neutrino mass model and on all other models the degree of the model comparison is weakened without the corrections. This could simply be caused by a lack of constraining power on the uncorrected data, however the values of the $\Delta \log \mathcal{L}_{\max}$ suggest that the model have a different fit on the uncorrected redshifts. This could suggest that the assumed Λ CDM model used in velocity field reconstruction introduces a slight bias towards Λ CDM when performing model comparison.

On the handful of alternate likelihood functions tested, we find

evidence to suggest that the covariance matrix should be uniformly scaled down, or equal evidence that the residuals should be modelled with a Student's t-distribution. The success of these two distributions appear to contradict one-another; the t-distribution succeeds because it correctly models the additional uncertainty in the estimation of the covariance matrix, whereas the A-only test succeeds by uniformly reducing the uncertainty. We believe that a more precise formulation of the t-distribution on this data could clarify the emergence of this non-Gaussian result, and reconcile this apparent contradiction.

As a result of this treatment of the residuals on a more general class of likelihood functions, the uncertainty on the Hubble constant is reduced by 5%. We have tightened the H_0 constraint with no new data by more appropriately modelling the scatter, and regardless of the increase in precision, the scatter-independent parameter estimates of table 5 are a more honest reflection of the data, as they marginalise out our uncertainty about the form of the likelihood function. Future work into likelihood-free inference could result in being able to constraint H_0 , and other parameters, even further. This would improve not only current estimation and model comparison, but set us up well for the next generation of future telescopes, which are expected to increase the size of SN catalogues by an order of magnitude (Rose et al. 2021).

It's worth stressing that all of the conclusions of this paper are based on SNe Ia alone, achieving precision cosmology results without high-redshift constraints, and evaluating the cosmologies through only a single mechanism (luminosity distances). Incorporating other refined probes (e.g. BAOs, CMB) with this improved analysis of SN could lead to a similar increase in precision to that found in this paper, on a much broader set of parameters and cosmological tests.

ACKNOWLEDGEMENTS

This work was performed using the Cambridge Service for Data Driven Discovery (CSD3), part of which is operated by the University of Cambridge Research Computing on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The DiRAC component of CSD3 was funded by BEIS capital funding via STFC capital grants ST/P002307/1 and ST/R002452/1 and STFC operations grant ST/R00689X/1. DiRAC is part of the National e-Infrastructure.

DATA AVAILABILITY

The Pantheon+ dataset is publicly available on the Pantheon+ repository, at <https://github.com/PantheonPlusSH0ES/DataRelease/>

The nested sampling chains underlying this article are available on Zenodo, at <https://doi.org/10.5281/zenodo.10026539>

REFERENCES

- Bernardo J. M., Smith A. F., 2009, Bayesian theory. Vol. 405, John Wiley & Sons
- Brout D., Hinton S. R., Scolnic D. M., 2020, *The Astrophysical Journal Letters*, 912
- Brout D., et al., 2022, *ApJ*, 938, 110
- Carr A., Davis T. M., Scolnic D., Said K., Brout D., Peterson E. R., Kessler R., 2022, *Publications of the Astronomical Society of Australia*, 39
- Dainotti M. G., Bargiacchi G., Bogdan M., Capozziello S., Nagataki S., 2023, Reduced uncertainties up to 43% on the Hubble constant and the matter density with the SNe Ia with a new statistical analysis ([arXiv:2303.06974](https://arxiv.org/abs/2303.06974))
- Dhawan S., Brout D., Scolnic D., Goobar A., Riess A. G., Miranda V., 2020, *ApJ*, 894, 54
- Ellis G. F. R., van Elst H., 2008, Cosmological models (Cargèse lectures 1998) ([arXiv:gr-qc/9812046](https://arxiv.org/abs/gr-qc/9812046))
- Fardon R., Nelson A. E., Weiner N., 2004, *J. Cosmology Astropart. Phys.*, 2004, 005
- Feeney S. M., Mortlock D. J., Dalmaso N., 2018, *Monthly Notices of the Royal Astronomical Society*, 476, 3861–3882
- Gelman A., Carlin J., Stern H., Dunson D., Vehtari A., Rubin D., 2013, Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis
- Goobar A., Leibundgut B., 2011, *Annual Review of Nuclear and Particle Science*, 61, 251
- Handley W., 2021, *Physical Review D*, 103
- Handley W., Lemos P., 2019, *Phys. Rev. D*, 100, 023512
- Handley W. J., Hobson M. P., Lasenby A. N., 2015a, *Monthly Notices of the Royal Astronomical Society: Letters*, 450, L61
- Handley W. J., Hobson M. P., Lasenby A. N., 2015b, *MNRAS*, 453, 4384
- Hergt L., Handley W., Hobson M., Lasenby A., 2021, *Physical Review D*, 103
- Hobson M. P., Jaffe A. H., Liddle A. R., Mukherjee P., Parkinson D., 2010, Bayesian methods in cosmology. Cambridge University Press
- Hyndman R. J., 1996, *The American Statistician*, 50, 120
- Jeffreys H., 1961, Theory of Probability. International series of monographs on physics, Clarendon Press, <https://books.google.co.uk/books?id=AavQAAAAAAAJ>
- Keeley R., Shafieloo A., L’Huillier B., 2022, An Analysis of Variance of the Pantheon+ Dataset: Systematics in the Covariance Matrix? ([arXiv:2212.07917](https://arxiv.org/abs/2212.07917))
- Kelly P. L., Hicken M., Burke D. L., Mandel K. S., Kirshner R. P., 2010, *ApJ*, 715, 743
- Kistler M. D., Stanek K. Z., Kochanek C. S., Prieto J. L., Thompson T. A., 2013, *The Astrophysical Journal*, 770, 88
- Kroupa N., Yallup D., Handley W., Hobson M., 2023, Kernel-, mean- and noise-marginalised Gaussian processes for exoplanet transits and H_0 inference ([arXiv:2311.04153](https://arxiv.org/abs/2311.04153))
- Kullback S., Leibler R. A., 1951, *The Annals of Mathematical Statistics*, 22, 79
- Kurek A., Szydłowski M., 2008, *The Astrophysical Journal*, 675, 1
- Linder E. V., 2007, *General Relativity and Gravitation*, 40, 329
- Lindley D. V., 2000, *Journal of the Royal Statistical Society Series D: The Statistician*, 49, 293
- Maguire K., 2017, in Alsabti A. W., Murdin P., eds., *Handbook of Supernovae*. Springer International Publishing, p. 293, doi:10.1007/978-3-319-21846-5_36
- Meng X. C., Li Z. M., Yang W. M., 2011, *Publications of the Astronomical Society of Japan*, 63, L31
- Mortonson M., Hu W., Huterer D., 2009, *Physical Review D*, 80
- Mörtsell E., Dhawan S., 2018, *J. Cosmology Astropart. Phys.*, 2018, 025
- Pascal F., Bombrun L., Tourneret J.-Y., Berthoumieu Y., 2013, *IEEE Transactions on Signal Processing*, 61, 5960
- Perivolaropoulos L., Skara F., 2022, *New Astronomy Reviews*, 95, 101659
- Perlmutter S., et al., 1999, *The Astrophysical Journal*, 517, 565
- Peterson E. R., et al., 2022, *The Astrophysical Journal*, 938, 112
- Phillips M. M., 1993, *ApJ*, 413, L105
- Planck Collaboration et al., 2020, *A&A*, 641, A6
- Riess A. G., et al., 1998, *AJ*, 116, 1009
- Riess A. G., et al., 2022, *ApJ*, 934, L7
- Rose B. M., et al., 2021, A Reference Survey for Supernova Cosmology with the Nancy Grace Roman Space Telescope ([arXiv:2111.03081](https://arxiv.org/abs/2111.03081))
- Scheutwinkel K. H., Handley W., de Lera Acedo E., 2023, *Publications of the Astronomical Society of Australia*, 40
- Slepian Z., Gott J. R., Zinn J., 2014, *MNRAS*, 438, 1948
- Sullivan M., et al., 2010, *Monthly Notices of the Royal Astronomical Society*, pp no–no
- Tripp R., 1998, *A&A*, 331, 815
- Trotta R., 2008, *Contemporary Physics*, 49, 71
- Wetterich C., 2007, *Physics Letters B*, 655, 201
- d’Amico G., Gleyzes J., Kokron N., Markovic K., Senatore L., Zhang P., Beutler F., Gil-Marín H., 2020, *Journal of Cosmology and Astroparticle Physics*, 2020, 005
- de Rham C., Gabadadze G., Tolley A. J., 2011, *Phys. Rev. Lett.*, 106, 231101
- du Mas des Bourboux H., et al., 2020, *The Astrophysical Journal*, 901, 153

This paper has been typeset from a $\text{\LaTeX}$ file prepared by the author.