

Dynamic or Systematic? Bayesian model selection between dark energy and supernova biases

A.N. Ormondroyd,^{1,2*} W.J. Handley,^{2,3} M.P. Hobson¹ and A.N. Lasenby^{1,2}

¹*Astrophysics Group, Cavendish Laboratory, J.J. Thomson Avenue, Cambridge, CB3 0HE, UK*

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

³*Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA, UK*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

DES-5Y supernovae, combined with DESI BAO, appear to favour Chevallier-Polarski-Linder dynamical dark energy over Λ CDM. [Efstathiou \(2024\)](#) suggested that this is driven by a systematic in the DES pipeline, which particularly affects the low-redshift supernovae brought in from legacy surveys. However, it is difficult to investigate these data in isolation, as the complex supernovae pipelines must properly account for selection effects. This work seeks to determine whether the Bayesian evidence previously found for flexknot dark energy ([Ormondroyd et al. 2025a](#)) is better explained by a systematic difference between the low- and high-redshift supernovae. We also take the opportunity to trial Nested Bridge Sampling with Sequential Monte Carlo as an alternative method for calculating Bayes’ factors.

Key words: methods: statistical – cosmology: dark energy, cosmological parameters

1 INTRODUCTION

[Efstathiou \(2024\)](#) claims that there is a systematic offset between the distance moduli of low- and high-redshift data in the Dark Energy Survey 5-year (DES-5Y) type Ia supernovae [DES Collaboration et al. \(2024\)](#). DES ([Vincenzi et al. 2025](#)) have responded to this, reporting that this claim is unsubstantiated and does not properly account for the complex nature of type Ia supernovae standardisation. [Notari et al. \(2025\)](#) investigated discarding the mutual supernovae between DES-5Y and Pantheon+ ([Brout et al. 2022](#)), then performed an offset between the low- and high-redshift supernovae in DES-5Y, and found that a cosmological constant is less strongly excluded. Of course, we then asked ourselves: is there any *Bayesian* evidence for this?

In this work, we extend the flexknot dark energy reconstructions from [Ormondroyd et al. \(2025b,a\)](#) with an additional low-redshift magnitude offset parameter, and investigate how this additional degree of freedom affects the Bayesian evidence for Λ CDM, the CPL ([Chevallier & Polarski 2001; Linder 2003](#)) parameterisation, and our exotic, free-form, flexknot reconstructions. We will also examine how the tension ratio is affected by this offset. This paper is organised as follows: first, Section 2 will outline the wider context of this work. In Section 3, the additional offset parameter in the DES-5Y supernovae likelihoods will be explained. Section 4 will offer a brief recap of the reconstruction method employed in this work, and demonstrate nested bridge sampling as an alternative method for producing posterior samples and Bayes factors. Results will be discussed in Section 5, finishing with conclusions in Section 6.

2 CONTEXT

Despite the successes of the standard cosmological model, known as Λ CDM, the nature of dark energy has remained an enigma for almost three decades ([Riess et al. 1998; Perlmutter et al. 1999](#)). This has motivated the exploration of alternative phenomenological hypotheses, ranging from a first-order expansion, Gaussian processes, and, of course, our previous flexknot reconstruction. Ultimately, these approaches all seek evidence that the dark energy equation of state parameter, w , has not been -1 for all cosmic time ([Einstein 1917; O’Raifeartaigh et al. 2017](#)).

Type Ia supernovae are a crucial tool in these investigations, and are known as “standard candles” — in fact, *standardisable* candles would be a more appropriate moniker, as meticulous calibration work has to be done by collaborations such as DES to create a standardised dataset suitable for use by others to constrain their favourite alternative cosmologies.

[Efstathiou \(2024\)](#) suggests that the detection of evolving dark energy by the combination of the second release DESI BAO and DES-5Y type Ia supernovae is driven by a systematic in the DES pipeline. In particular, it was suggested that the low-redshift supernovae included in the DES pipeline which are *not* from DES’s own measurements: from the CfA ([Hicken et al. 2009, 2012](#)), the Carnegie Supernova Project ([Krisciunas et al. 2017, 2020](#)) and the Foundation Sample ([Foley et al. 2017](#)), have an apparent magnitude m_B which is systematically 0.04 magnitudes above both the same supernovae (SNe) in Pantheon+, another supernova dataset, and the Planck best fit cosmology. [Vincenzi et al. \(2025\)](#) responded to this claim, citing analysis improvements compared to Pantheon+, and how different selection criteria between the two means it is expected that the datasets contain differences.

The DES-5Y supernova sample represents the largest single-

* E-mail: ano23@cam.ac.uk

survey dataset of its kind, with over 1600 photometrically classified SNe Ia from the DES program. This sample also includes 194 SNe at low redshifts from a number of external surveys to serve as a cosmological anchor. Broadly, the external supernovae are at redshifts of less than 0.1, the DES SNe are from 0.1 to 1.13. In contrast to Pantheon+, DES-5Y presents a stronger preference for evolving dark energy.

Supernova surveys are magnitude-limited, that is, bluer and longer events are more likely to be detected than redder, shorter events, the Malmquist bias. This bias must be corrected for on a supernova-by-supernova basis, which is done using simulations. Of course, there are differences between the Pantheon+ and DES-5Y pipelines, these are covered in detail in the appendix of Vincenzi et al. (2025). For example, it was found that replacing the SALT3 light curve fitting model (Kenworthy et al. 2021; Taylor et al. 2023) of DES-5Y with the older SALT2 (Guy et al. 2007) model used by Pantheon+ would have halved the offset found by Efstathiou (2024). DES-5Y and Pantheon+ also use different selection functions, which means that the bias corrections *should* be different, and equivalence of the “same” supernova event should not be expected, in fact, Pantheon+ contains a strongly biased selection of the DES-3Y supernovae: those with spectroscopic follow-up. Notari et al. (2025) compare DES-5Y and Pantheon+, with the mutual supernovae excluded, of course, this does not properly account for selection effects, if it were to be done properly, the mutual supernovae should be deleted before bias corrections. The approach here suffers a similar limitation, but in lieu of a viable alternative, we proceed.

Two things may be true at the same time. There can be a systematic issue with the low-redshift external supernovae in DES-5Y, even if expecting them to be identical to Pantheon+ is an oversimplification. Therefore, we seek to investigate whether there is any Bayesian evidence for such an offset (Bayes 1763). Deliberately, we do not include the Pantheon+ supernovae in this work. That way, any suggestion we find for this offset is entirely independent of comparisons between the two pipelines.¹

3 DATA

In this work, an agnostic approach is taken. Rather than relying on the -0.04 value as given, we add an additional parameter to our likelihood, Δm_B , which is an offset applied to only the non-DES supernovae:

$$\mathcal{L}(D|\theta) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp -\frac{1}{2}\Delta^T \Sigma^{-1} \Delta, \quad (1)$$

$$\Delta = (m_B + m\Delta m_B - M_B) - \mu(z, \theta).$$

m is a binary mask which is 1 iff the corresponding supernova was not from the DES catalogue itself, and 0 if it was from DES. Setting $\Delta m_B = 0$ is equivalent to the standard supernova likelihood. The distance modulus μ is calculated from the luminosity distance:

$$D_L(z) = (1 + z_{\text{hel}})^c \int_0^{z_{\text{HD}}} \frac{dz'}{H(z')}, \quad \mu(z) = 5 \log_{10} \left(\frac{D_L(z)}{10 \text{ pc}} \right). \quad (2)$$

If Δm_B is unsubstantiated, then it will be Occam-penalised; however if the Bayesian evidence is significantly greater with this parameter,

¹ Of course, if it weren’t for the suggestion of an offset between the two, this work wouldn’t have been carried out. It is left as an exercise to the reader to choose an appropriate prior given they are reading this paper.

Parameter	Prior
Δm_B	$[-0.1, 0.1]$
n	$[1, 20]$
a_{n-1}	0
a_{n-2}, \dots, a_1	sorted($[a_{n-1}, a_0]$)
a_0	1
w_{n-1}, \dots, w_0	$[-3, 1]$
w_a	$[-3, 2], w_0 + w_a < 0$
Ω_m	$[0.01, 0.99]$
$H_0 r_d$ (DESI)	$[3650, 18250]$
H_0 (Ia)	$[20, 100]$

Table 1. Cosmological priors used in this work. Fixed values are indicated by a single number, while uniform priors are denoted by brackets. As BAO only depend on the product $H_0 r_d$, and supernovae depend on H_0 , the former is only sampled when DESI is included, and the latter is analytically marginalised out. Similarly, w_a is only used for the CPL model, with the restriction that the value of w is negative at early times, $w_0 + w_a < 0$, as used by other work.

then it follows that there is a systematic offset between the supernovae from DES.

Since the absolute magnitude M_B is also a parameter to be fitted, an overall offset in m_B has no effect on cosmology, so one would obtain the same results with the opposite mask, albeit Δm_B would have the opposite sign. In fact, it is possible to analytically marginalise out Δm_B , this is discussed further in Appendix A, but all of the results simply sample the parameter.

For a more complete reconstruction of the expansion history, likelihoods with and without the Δm_B offset were combined with the second major data release of DESI DR2 BAO measurements. Since release two is the current state of the art, we will refer to this simply as “DESI BAO”, or just “DESI”, for brevity.

4 METHODS

4.1 Flexknot dark energy reconstructions

The “flexknot” approach of reconstructing the dark energy equation of state parameter is explained in detail in Ormondroyd et al. (2025b) and Ormondroyd et al. (2025a), but we will outline the approach again here. Flexknots are a free-form, model-independent method for reconstructing one-dimensional functions. They consist of a linear spline between n nodes “knots”, whose positions are parameters of the model to be fitted using nested sampling. The horizontal coordinate of the left- and rightmost knots are fixed, in this case, at zero and one respectively, and the remaining coordinates are free to vary within these bounds, with the restriction that they remain sorted. The number of knots n is also a parameter of the model, in practice, separate nested sampling runs are performed for each n using POLYCHORD (Handley et al. 2015a,b), and the posterior of n is proportional to the evidence for each run. This is a well-established technique across many areas of cosmology beyond the dark energy equation of state (Ormondroyd et al. 2025b,a; Hee et al. 2016; Vázquez et al. 2012b), including the primordial power spectrum (Handley et al. 2019; Vázquez et al. 2012a; Aslanyan et al. 2014; Finelli et al. 2018; Planck Collaboration et al. 2014, 2016b), the cosmic reionisation history (Millea & Bouchet 2018; Heimersheim et al. 2022), galaxy cluster profiles (Olamaie et al. 2018), and the 21 cm signal (Heimersheim et al. 2024; Shen et al. 2024).

Flexknots also have the advantage that $n = 1$ and $n = 2$ knots correspond to the w CDM and CPL models respectively, though, in the latter case, separate sampling runs are also performed with priors consistent with other works for completeness. In response to com-

ments from presenting our previous work, one small change has been made which was previously shown in an appendix of Ormondroyd et al. (2025a), that is, the upper limit of the prior of the w -coordinates of the knots is now 1. This is more consistent with typical priors used for CPL (e.g. Planck Collaboration et al. (2016a, 2021); DESI Collaboration et al. (2024, 2025b,a,c)), and means that Λ CDM is in the centre of the prior, though of course this is of no consequence for a uniform prior. Under each reconstruction, the Kullback-Leibler divergence (KL divergence) is shown as a function of scale factor or redshift as appropriate. This is a more robust way of assessing the constraining power of each dataset throughout cosmic history than simply comparing their contours.

Figure 1 shows prior samples from the usual CPL prior used in most analyses, and the equivalent flexknot prior, and the corresponding posteriors from DES-5Y combined with DESI as kernel density estimates (KDEs). The CPL prior includes the constraint that the value of w at early times, $w_0 + w_a$, is less than zero, to ensure a period of matter domination. Clearly, the two priors are different, but the posteriors are so similar that it is difficult to see that there are two KDEs overlaid. From this, we conclude that it is reasonable to call CPL and two knots the same model. The priors used in this work are listed in Table 1.

4.2 Nested bridge sampling

The nested sampling approach used previously and in this work computes the Bayesian evidence for each combination of data and model. However, in isolation, evidences are impossible to interpret, and it is only with a pair that one can determine a Bayes factor, to which one may apply Jeffreys' scale (Jeffreys 1998) to determine how to feel about their result, or combine with a model prior to determine the posterior odds. An alternative method to compute the Bayes factor, beside two normal nested sampling runs, is to use nested bridge sampling.

Nested bridge sampling (NBS, Chen et al. (2000); Gronau et al. (2017), YALLUP²) makes use of the inevitable similarity between the posteriors for the shared parameters between sampling runs with nested models. This work consists entirely of nested models. Λ CDM is nested within w CDM with $w = 0$, which are both nested within CPL, which are all nested within flexknot models. Also, each cosmological model with $\Delta m_B = 0$ is nested within the same model with Δm_B varying. In theory, one could bridge sample from vanilla Λ CDM to an $n = 20$ flexknot with Δm_B . As a demonstration, let us outline nested bridge sampling to add Δm_B only to an existing run.

YALLUP explains this method in detail and its application to toy and cosmological examples; let us recap the methodology here: The recipe is as follows: first, produce a set of samples with $\Delta m_B = 0$. Then, sample the likelihood ratio, $\tilde{\mathcal{L}}$, using the posterior from the first run as part of an effective prior:

$$\tilde{\mathcal{L}} = \frac{\mathcal{L}(D|\theta, \Delta m_B)}{\mathcal{L}(D|\theta, \Delta m_B = 0)}, \quad \tilde{\pi} = \pi(\Delta m_B|\theta) \frac{\pi(\theta) \mathcal{L}(\theta)}{Z_1}, \quad (3)$$

where D is the data (DES-5Y alone or with DESI BAO), and θ are the appropriate other parameters for the cosmology in question. These

Comparison of CPL and FlexKnot $n=2$ priors

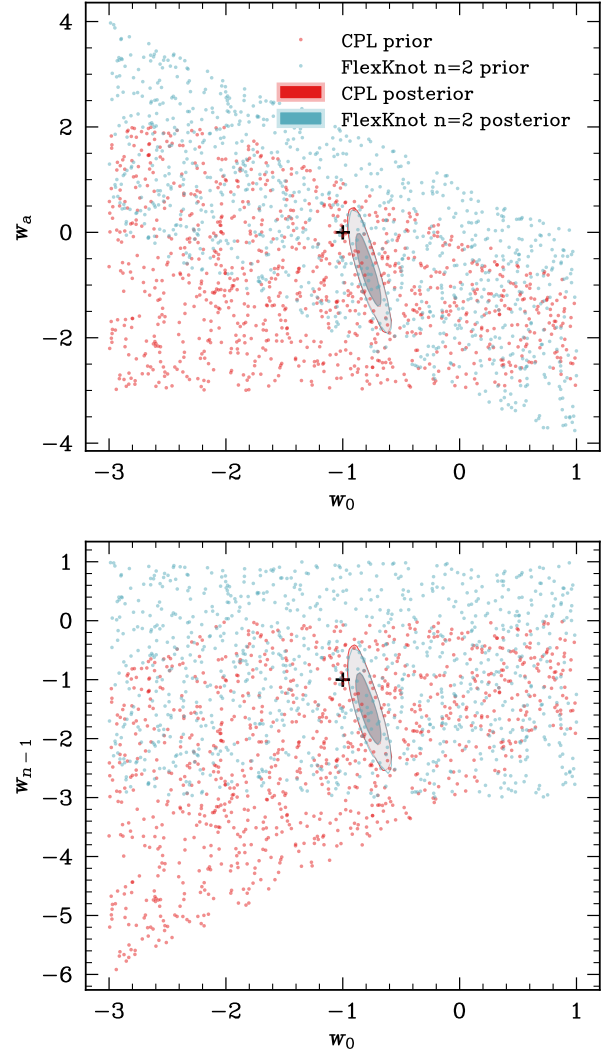


Figure 1. Prior and posterior of CPL and $n = 2$ flexknot, using DES-5Y combined with DESI BAO. The top panel shows the (w_0, w_a) projection, the lower panel shows (w_0, w_{n-1}) . Prior samples are shown as a scatter, the posteriors are shown as kernel density estimates. For reference, the cross marks Λ CDM. The two posteriors are so similar that it is challenging to see one on top of the other!

arise from rearranging the Bayes' factor:

$$\begin{aligned} \frac{Z_2}{Z_1} &= \frac{\int \mathcal{L}(\theta, \Delta m_B) \pi(\theta, \Delta m_B) d\theta d(\Delta m_B)}{Z_1} \\ &= \int \underbrace{\frac{\mathcal{L}(\theta, \Delta m_B)}{\mathcal{L}(\theta)}}_{\tilde{\mathcal{L}}} \underbrace{\pi(\Delta m_B|\theta) \frac{\pi(\theta) \mathcal{L}(\theta)}{Z_1}}_{\tilde{\pi}} d\theta d(\Delta m_B), \end{aligned} \quad (4)$$

where the data D is suppressed from the likelihood for clarity, and the probability product rule has been used to factorise the prior $\pi(\theta, \Delta m_B) = \pi(\Delta m_B|\theta)\pi(\theta)$. The word “nested” should not be confused with the same in nested sampling, which refers to the union of likelihood contours. However, nested sampling may indeed be used to perform nested bridge sampling, but one may substitute Sequential Monte Carlo (Naesseth et al. 2019), which is what is used here, for

² in prep

Table 2. Values of Δm_B for DES-5Y, along with the Bayes factor between the standard version ($\Delta m_B = 0$) and with the offset. The Bayes factor is calculated in two ways, first by taking the ratio of the evidences from the nested sampling runs, and also by bridge sampling from $\Delta m_B = 0$ to the offset version. Bridge sampling is not attempted for the flexknot models, as more work is required to determine if this is viable. The shown nested sampling Bayes factors are from the JAX pipeline, which were very consistent with the POLYCHORD pipeline, which are not reported here.

Both Λ CDM values of Δm_B are consistent with -0.04 , and inconsistent with zero to over 2σ . In contrast, both CPL and flexknot dark energy do not exclude zero, this is reflected in the Bayes' factors. Of particular note is the large positive Bayes factor for Λ CDM with DES-5Y + DESI BAO in favour of a low-redshift supernova offset. The Bayes factors from SMC-NSB are beyond error of those computed from two nested sampling runs, but not so much to affect any conclusions.

DES-5Y			
	Δm_B	Bayes factor	SMC-NBS
Λ CDM	-0.035 ± 0.016	0.755 ± 0.098	0.665 ± 0.030
CPL	-0.042 ± 0.028	-0.067 ± 0.127	0.086 ± 0.109
flexknot	-0.018 ± 0.040	-0.525 ± 0.021	N/A
DES-5Y + DESI BAO			
Λ CDM	-0.045 ± 0.012	4.140 ± 0.182	4.488 ± 0.037
CPL	-0.022 ± 0.026	-0.859 ± 0.217	-0.738 ± 0.018
flexknot	-0.017 ± 0.038	-0.396 ± 0.055	N/A

both the initial sampling run and bridging, abbreviated as NBS-SMC. A possible disadvantage of SMC compared to nested sampling is that it does not report an error bar, the error bars here are from ten repeats with different initial random seeds.

4.3 JAX reimplementaion

In addition to the pipeline from previous work, the likelihoods have also been ported to JAX. This allows them to be sampled using BLACKJAX nested slice sampling, which makes use of GPU parallelisation (Yallup et al. 2025; Prathanan et al. 2025; Cabezas et al. 2024). In fact, the $n = 2$ flexknot prior and posterior shown in Figure 1 is from the POLYCHORD pipeline, which uses NUMPY and SCIPY and 64-bit floating point precision, while the CPL result uses the JAX pipeline, with 32-bit floating point precision. Reimplementation in JAX posed some additional challenges, for example, the adaptive QUADPACK method used to compute the integral over $\frac{H_0}{H(z)}$ is not available in JAX.SCIPY, so trapezoidal integration was substituted. Also, it was found that the precision setting for the Mahalanobis distance matrix multiplication had to be increased from its default level for consistency with the other pipeline. Therefore, it is very reassuring that the results, which differ in hardware, sampler, 1-dimensional integration technique for proper motion distance, and floating point representation, are so similar.

One feature of POLYCHORD currently omitted from the BLACKJAX sampler is live point clustering. Clustering is typically considered necessary for nested sampling with multi-modal posteriors, however, the BLACKJAX sampler has proven surprisingly effective on other problems. It is the subject of ongoing research whether no clustering is genuinely a limitation, therefore, we restrict the application of the JAX pipeline to the unimodal Λ CDM and CPL likelihoods only in this work.

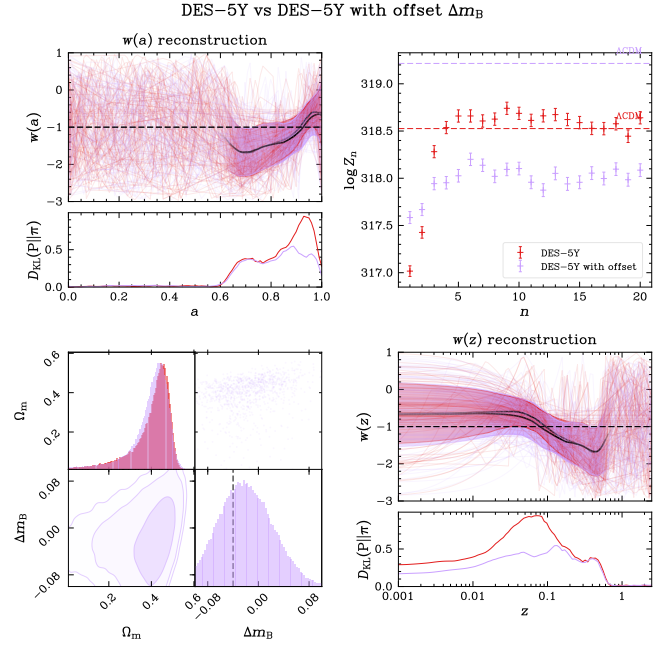


Figure 2. Flexknot reconstruction of the dark energy equation of state parameter using DES-5Y supernovae only. In red is the standard likelihood, in lilac, the version with the Δm_B offset for the low-redshift supernovae. The overall shape of the reconstructions are very similar, but the functional KL divergence and model evidences tell quite different stories. Firstly, note that the high- a /low-redshift KL divergence lacks the peak just below $z = 0.1$, which is to be expected as allowing those magnitudes to float up and down will naturally reduce their constraining power. Second, note that the evidence for Λ CDM has increased with the offset, meanwhile, it has fallen for all flexknots with more than three knots. The Bayes factor between Λ CDM and w CDM is similar between the two likelihoods, but $n = 2$ is more disfavoured with the offset likelihood. Crucially, the evidence for Λ CDM with the offset is greater than the flexknots without the offset. This suggests that the complexity demanded by the flexknot model is just as well, if not better, met by including this additional degree of freedom.

Please note that the posteriors shown in the bottom-left panels do not include Λ CDM, which is shown separately in Figure 5. With flexknots, Δm_B is not well constrained.

5 RESULTS

5.1 Flexknot reconstructions

First, begin with DES-5Y supernovae alone. Figure 2 shows the flexknot reconstruction of $w(a)$ both with and without the low-redshift offset. From the KL divergence panels, it can be seen that the constraining power at low redshifts is reduced with the additional parameter, this is to be expected. The model with the greatest evidence is Λ CDM with the offset. This suggests that the complexity picked up by the flexknot model, which brings models with four or more knots in line with Λ CDM, is better explained by Δm_B . Only Λ CDM, w CDM and CPL ($n = 1$ and $n = 2$ respectively) have greater evidence with the introduction of the offset, the reverse is true for three knots and up. However, at least with DES-5Y alone, these suggestions must be caveated by the relatively small Bayes factors between evidences with the same n , which are at most around 0.7. In fact, Jeffreys' scale suggests that these differences are "barely worth mentioning". We will return to this in Section 5.2.

Now, also include the DESI data, the combined reconstruction is shown in Figure 3. With the BAOs, Λ CDM with Δm_B is still the

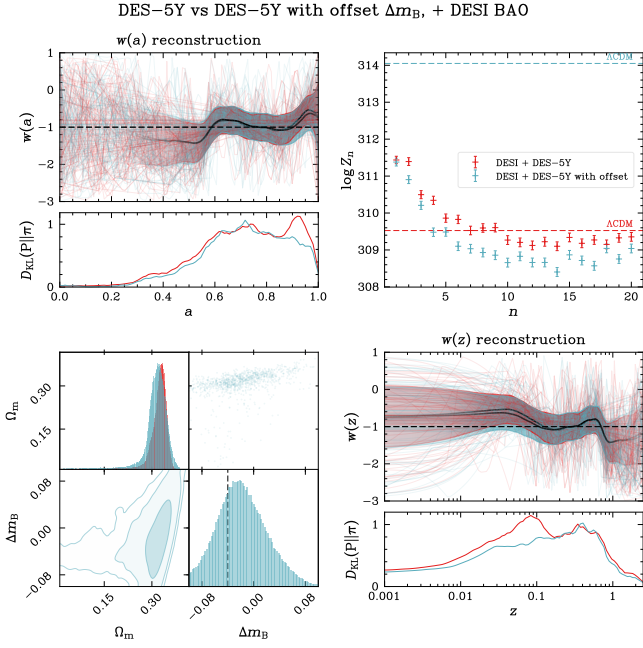


Figure 3. Similar to Figure 2, this time with the addition of DESI BAO. This time, it is even clearer that Λ CDM with the Δm_B offset is the favoured model. Once again, the low-redshift KL divergence peak is lost with the additional parameter. Unlike the results with DES-5Y alone, the Bayes factor for Λ CDM with Δm_B over almost any other model is “decisive”. Again, please note that the posteriors shown in the bottom-left panels do not include Λ CDM, these are shown in Figure 5.

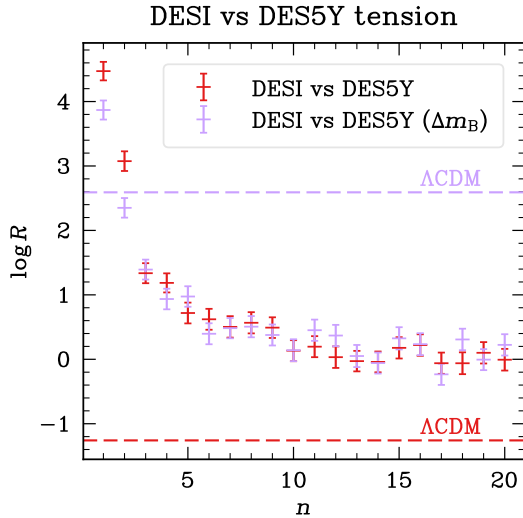


Figure 4. Tension values between DES-5Y and DESI BAO, with and without the low-redshift offset. Λ CDM is shown as horizontal dashed lines, for easy comparison with the other points. For Λ CDM the tension has been reduced (more positive) significantly, while for w CDM and CPL, it has increased slightly. Λ CDM with the offset is now on-par with CPL, though w CDM remains the model with the best dataset concordance. In contrast, without the offset, Λ CDM is the most discrepant model of all. For three knots and above, the tension is similar with or without the offset.

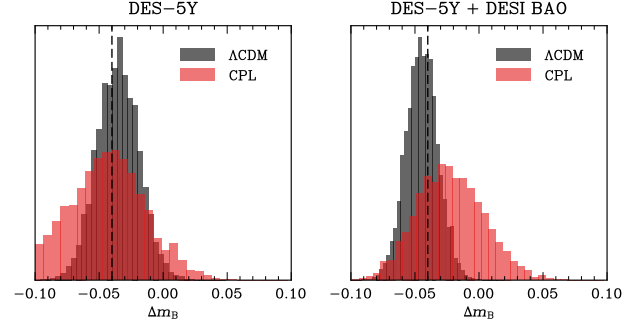


Figure 5. Posterior histograms of Δm_B for Λ CDM and CPL. The left panel uses DES-5Y only, the right also includes DESI BAO. The prior is uniform over the domain of the plot. As predicted by [Efstathiou \(2024\)](#), its value is centred on -0.04 . Note that the Λ CDM posteriors (and the right CPL posterior) are well contained within the prior, therefore, the evidence which would have been found had a wider prior been used can be easily be computed with the ratio of the prior volumes.

favoured model, this time, with a log-Bayes factor of around four over almost all other models, crucially including standard Λ CDM. The offset has also removed the preference for low numbers of flexknots. This strongly suggests that the original preference for dynamical dark energy can be better accounted for by this offset than it can by w CDM, CPL, or even flexknots.

The JAX pipeline produced Λ CDM and CPL evidences in perfect agreement with those from POLYCHORD, in fact, it is those reported in Table 2. These nested sampling runs completed in less than a second on a Google Compute Engine Nvidia L4 GPU, plus a few seconds of compile time. This compares very well to tens of minutes on a 76-core CSD3 CPU, plus submission delay, and has the potential to transform the average Bayesian’s workflow.

It is important to assess how Δm_B may affect the possibility of tension between DES-5Y and DESI BAO. Once again, we follow [Ormondroyd et al. \(2025b\)](#), using the techniques developed in [Handley & Lemos \(2019\)](#); [Hergt et al. \(2021\)](#); [Ormondroyd et al. \(2023\)](#) via the log R statistic. The results are shown in Figure 4. Without the offset, Λ CDM is the most discrepant model, with the only negative log R . With the offset, Λ CDM is now on par with CPL, while w CDM remains the model with the best concordance between the two datasets. For three knots and above, the tension is similar with or without the offset. This is reassuring, as it suggests that the offset is better able to explain the discrepancy between the datasets than flexknot dark energy.

5.2 Effect of Δm_B prior width

From Figure 5, it can be seen that, for Λ CDM, the posterior for Δm_B is reasonably well contained within our chosen prior. This allows us to examine the Bayes factors in more detail.

The uniform prior contains a factor of the reciprocal of the prior volume V , which, in the well contained case, persists into the Bayesian evidence. This means that the evidence that *would* have been found for a different, wider, uniform prior can be calculated as so:

$$\log Z \rightarrow \log \left(Z \times \frac{V_{\text{narrow}}}{V_{\text{wide}}} \right) = \log Z - \log \frac{V_{\text{wide}}}{V_{\text{narrow}}}. \quad (5)$$

Our prior of $\Delta m_B \in [-0.1, 0.1]$ was chosen to be relatively narrow for two reasons: firstly, the nested sampling runtime is proportional

to the KL divergence from prior to posterior; second, if the low-redshift systematic offset does exist and had a value way beyond this range, it is unlikely that it would fall to third parties to investigate it, and frankly this should impact on our model prior. For example, let us examine our conclusions if a prior ten times wider, $[-1.0, 1.0]$, had been chosen. This would reduce Bayes factors by approximately $\log 10 = 2.30$. Consider DES-5Y alone: the only positive Bayes factor is Λ CDM at around 0.76, this would be reversed with this more liberal prior. In fact, we may reverse-engineer this argument to note that, had a prior $e^{0.76} \approx 2.14$ times larger been chosen, the Bayes factor would have been precisely naught. One may rightly criticise that post-hoc revisions to the prior are unwise, however, if this were all the data we had, it seems this would be enough to suggest that there is no practical evidence for the low-redshift offset, since it is so vulnerable to somewhat reasonable alternative priors.

However, now reintroduce the DESI BAO, with a Bayes factor of $\log Z = 4.140 \pm 0.127$ in favour of Δm_B . This time, one would have to have had a prior at least sixty-two times wider to reverse the conclusion — an offset of such magnitude is completely ridiculous. Therefore, we will conclude that there is only evidence for the low-redshift supernova offset in DES-5Y if, and only if, it is combined with baryon acoustic oscillations, and Λ CDM is correct.

5.3 Nested bridge sampling Bayes' factors

Table 2 also includes the SMC-NBS Bayes' factors. Error bars were estimated by running SMC-NBS ten times with different random seeds, while those from nested sampling were computed using ANESTHETIC, which samples possible nested sampling volume compression histories (Handley 2019). It is interesting that all four of the NBS error bars are tighter than those from nested sampling. The different approaches are, in some cases, beyond error of each other, though not so significantly as to affect any conclusions. For example, it is not surprising that the very small Bayes factors with CPL DES-5Y have the opposite signs. Crucially, the Bayes' factor of the most interest, which happens to be the largest, of Δm_B or no Δm_B with DESI in Λ CDM, is similarly large with NBS. The SMC implementation also uses BLACKJAX, and similarly takes mere seconds to run.

6 CONCLUSIONS

In this work, it has been found that there is substantial Bayesian evidence for a low-redshift supernovae systematic in DES-5Y, when it is combined with DESI BAO. However, without the BAO, the claim is unsubstantiated, though the posterior on the offset value indeed agrees with the -0.04 of Efstathiou (2024) and excludes zero to 2σ . Crucially, the Bayesian evidence favours offset Λ CDM over flexknot dark energy with or even without the offset, so we must conclude that the systematic is a better model than dynamical dark energy. We accept that this approach is limited, and fiddling with a subset of apparent magnitudes post bias correction does not constitute a sensible supernova catalogue. Nevertheless, the options are clear: either there is a systematic issue with the DES-5Y supernovae, or dark energy really is dynamical.

It has also been found that the Λ CDM tension between DES-5Y and DESI BAO is significantly reduced with Δm_B , while w CDM and CPL tensions are slightly increased. w CDM still remains the model with the best agreement.

The alternative JAX pipeline has produced Λ CDM and CPL posterior samples and evidences in excellent agreement with those from our original POLYCHORD-powered approach. The reduced sampling

time to mere seconds has the potential to be transformative for the Bayesian workflow. We leave it for future work to investigate whether these tools are robust for the more challenging multi-modal posterior of flexknot reconstructions. We also find that Bayes factors obtained using nested bridge sampling are sufficiently similar to those obtained by the ratio of nested sampling evidences, therefore, we hope that this approach can be trialled more widely.

ACKNOWLEDGEMENTS

A.N. Ormondroyd and W.J. Handley were supported by the research environment and infrastructure of the Handley Lab at the University of Cambridge. 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. WJH was supported by a Royal Society University Research Fellowship.

The PYTHON pipeline in this work made use of NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), and pandas (The pandas development team 2023; McKinney 2010). The GPU-accelerated pipeline was written in JAX (Bradbury et al. 2018), using Google gcloud compute resources. The nested sampling chains were analysed using ANESTHETIC Handley (2019), plots were produced in matplotlib (Hunter 2007), using the smplotlib template created by Li (2023).

The authors thank David Yallup and Toby Lovick for guidance with SMC-NBS.

DATA AVAILABILITY

TODO

REFERENCES

- Aslanyan G., Price L. C., Abazajian K. N., Easter R., 2014, *J. Cosmology Astropart. Phys.*, 2014, 052
- Bayes T., 1763, *Philosophical Transactions of the Royal Society of London*, 53, 370
- Bradbury J., et al., 2018, JAX: composable transformations of Python+NumPy programs, <http://github.com/google/jax>
- Brout D., et al., 2022, *ApJ*, 938, 110
- Cabezas A., Corenflos A., Lao J., Louf R., 2024, BlackJAX: Composable Bayesian inference in JAX ([arXiv:2402.10797](https://arxiv.org/abs/2402.10797))
- Chen M.-H., Shao Q.-M., Ibrahim J. G., 2000, *Monte Carlo Methods in Bayesian Computation*, 1 edn. Springer Series in Statistics, Springer New York, NY, doi:10.1007/978-1-4612-1276-8
- Chevallier M., Polarski D., 2001, *Int. J. Mod. Phys. D*, 10, 213
- DES Collaboration et al., 2024, *ApJ*, 973, L14
- DESI Collaboration et al., 2024, *arXiv e-prints*, p. [arXiv:2404.03002](https://arxiv.org/abs/2404.03002)
- DESI Collaboration et al., 2025a, *arXiv e-prints*, p. [arXiv:2503.14738](https://arxiv.org/abs/2503.14738)
- DESI Collaboration et al., 2025b, *arXiv e-prints*, p. [arXiv:2503.14739](https://arxiv.org/abs/2503.14739)
- DESI Collaboration et al., 2025c, *arXiv e-prints*, p. [arXiv:2503.14743](https://arxiv.org/abs/2503.14743)
- Efstathiou G., 2024, *arXiv e-prints*, p. [arXiv:2408.07175](https://arxiv.org/abs/2408.07175)
- Einstein A., 1917, *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pp 142–152
- Finelli F., et al., 2018, *J. Cosmology Astropart. Phys.*, 2018, 016
- Foley R. J., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 475, 193
- Gronau Q. F., et al., 2017, *arXiv e-prints*, p. [arXiv:1703.05984](https://arxiv.org/abs/1703.05984)

- Guy J., et al., 2007, *A&A*, **466**, 11
- Handley W., 2019, *The Journal of Open Source Software*, **4**, 1414
- Handley W., Lemos P., 2019, *Phys. Rev. D*, **100**, 043504
- Handley W. J., Hobson M. P., Lasenby A. N., 2015a, *MNRAS*, **450**, L61
- Handley W. J., Hobson M. P., Lasenby A. N., 2015b, *MNRAS*, **453**, 4384
- Handley W. J., Lasenby A. N., Peiris H. V., Hobson M. P., 2019, *Phys. Rev. D*, **100**, 103511
- Harris C. R., et al., 2020, *Nature*, **585**, 357
- Hee S., Handley W. J., Hobson M. P., Lasenby A. N., 2016, *MNRAS*, **455**, 2461
- Heimersheim S., Sartorio N. S., Fialkov A., Lorimer D. R., 2022, *ApJ*, **933**, 57
- Heimersheim S., Rønneberg L., Linton H., Pagani F., Fialkov A., 2024, *MNRAS*, **527**, 11404
- Hergt L. T., Handley W. J., Hobson M. P., Lasenby A. N., 2021, *Phys. Rev. D*, **103**, 123511
- Hicken M., et al., 2009, *The Astrophysical Journal*, **700**, 331
- Hicken M., et al., 2012, *The Astrophysical Journal Supplement Series*, **200**, 12
- Hunter J. D., 2007, *Computing in Science & Engineering*, **9**, 90
- Jeffreys H., 1998, *The Theory of Probability*, 3rd edn. Oxford University Press, Oxford, England
- Kenworthy W. D., et al., 2021, *The Astrophysical Journal*, **923**, 265
- Krisciunas K., et al., 2017, *The Astronomical Journal*, **154**, 278
- Krisciunas K., et al., 2020, *The Astronomical Journal*, **160**, 289
- Li J., 2023, AstroJacobLi/smplotlib: v0.0.9, doi:10.5281/zenodo.8126529, <https://doi.org/10.5281/zenodo.8126529>
- Linder E. V., 2003, *Phys. Rev. Lett.*, **90**, 091301
- McKinney W., 2010, in Stéfan van der Walt Jarrod Millman eds, *Proceedings of the 9th Python in Science Conference*. pp 56 – 61, doi:10.25080/Majora-92bf1922-00a
- Millea M., Bouchet F., 2018, *A&A*, **617**, A96
- Naesseth C. A., Lindsten F., Schön T. B., 2019, *arXiv e-prints*, p. arXiv:1903.04797
- Notari A., Redi M., Tesi A., 2025, *J. Cosmology Astropart. Phys.*, **2025**, 048
- O’Raifeartaigh C., O’Keeffe M., Nahm W., Mitton S., 2017, *European Physical Journal H*, **42**
- Olamaie M., Hobson M. P., Feroz F., Grainge K. J. B., Lasenby A., Perrott Y. C., Rumsey C., Saunders R. D. E., 2018, *MNRAS*, **481**, 3853
- Ormondroyd A. N., Handley W. J., Hobson M. P., Lasenby A. N., 2023, *arXiv e-prints*, p. arXiv:2310.08490
- Ormondroyd A. N., Handley W. J., Hobson M. P., Lasenby A. N., 2025a, *arXiv e-prints*, p. arXiv:2503.17342
- Ormondroyd A. N., Handley W. J., Hobson M. P., Lasenby A. N., 2025b, *MNRAS*, **541**, 3388
- Perlmutter S., et al., 1999, *ApJ*, **517**, 565
- Planck Collaboration et al., 2014, *A&A*, **571**, A22
- Planck Collaboration et al., 2016a, *A&A*, **594**, A13
- Planck Collaboration et al., 2016b, *A&A*, **594**, A20
- Planck Collaboration et al., 2021, *A&A*, **652**, C4
- Prathaban M., Yallup D., Alvey J., Yang M., Templeton W., Handley W., 2025, *arXiv e-prints*, p. arXiv:2509.04336
- Riess A. G., et al., 1998, *AJ*, **116**, 1009
- Shen E., Anstey D., de Lera Acedo E., Fialkov A., 2024, *MNRAS*, **529**, 1642
- Taylor G., et al., 2023, *Monthly Notices of the Royal Astronomical Society*, **520**, 5209
- The pandas development team 2023, pandas-dev/pandas: Pandas, doi:10.5281/zenodo.8092754, <https://doi.org/10.5281/zenodo.8092754>
- Vázquez J. A., Bridges M., Hobson M. P., Lasenby A. N., 2012a, *J. Cosmology Astropart. Phys.*, **2012**, 006
- Vázquez J. A., Bridges M., Hobson M. P., Lasenby A. N., 2012b, *J. Cosmology Astropart. Phys.*, **2012**, 020
- Vincenzi M., et al., 2025, *MNRAS*, **541**, 2585
- Virtanen P., et al., 2020, *Nature Methods*, **17**, 261
- Yallup D., Kroupa N., Handley W., 2025, in *Frontiers in Probabilistic Inference: Learning meets Sampling*, <https://openreview.net/forum?id=ekbkMSuPo4>

APPENDIX A: ANALYTIC MARGINALISATION OVER

Δm_B

This appendix contains an extension to the analytical marginalisation in Ormondroyd et al. (2025b) and Ormondroyd et al. (2025a). In those works, it was demonstrated that the constant offset from M_B could be analytically marginalised from the SNe likelihoods. Since this value affects all terms in the data vector, there is essentially an accompanying mask of all ones. In fact, the algebra is identical for any mask, assuming that the likelihood tends to zero at the limits of the prior, so Δm_B could be marginalised over by replacing the covariance matrix:

$$\Sigma^{-1} \rightarrow \Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{m}^T \mathbf{m} \Sigma^{-1}}{\mathbf{m}^T \Sigma^{-1} \mathbf{m}}, \quad (\text{A1})$$

and adjusting the normalisation by a factor of $\frac{1}{V_m} \sqrt{\frac{2\pi}{\mathbf{m}^T \Sigma \mathbf{m}}}$, where V_m is the prior volume of the offset, Δm_B in this case. However, this introduces two challenges. First, each one of these marginalisations introduces an additional zero eigenvalue to the inverse covariance matrix, reducing its rank by one. This introduces precision challenges, in particular when working in 32-bit floating point. Secondly, the posterior of Δm_B is of interest, and while it is possible to then sample it from the posterior of the other parameters, it is more straightforward to simply include it as a parameter during nested sampling, as this poses no challenge for our nested slice sampling tools.

This paper has been typeset from a \LaTeX file prepared by the author.