MNRAS 000, 1–9 (2017) Preprint 16 January 2018 Compiled using MNRAS LaTeX style file v3.0

SN-BHM: A hierarchical Bayesian model for Supernova Cosmology

Samuel R. Hinton,[1,2]⋆ Alex G. Kim,[3] Tamara M. Davis[1,2]
[1] *School of Mathematics and Physics, The University of Queensland, Brisbane, QLD 4072, Australia*
[2] *ARC Centre of Excellence for All-sky Astrophysics (CAASTRO)*
[3] *Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
Abstract Abstract

Key words: keyword1 – keyword2 – keyword3

1 INTRODUCTION

Almost two decades have passed since the discovery of the accelerating universe (Riess et al. 1998; Perlmutter et al. 1999). Since that time, of the number of observed Type Ia supernovae (SN Ia) have increased by more than an order of magnitude thanks to modern surveys at both low redshift (Bailey et al. 2008; Freedman et al. 2009; Hicken et al. 2009; Contreras et al. 2010; Conley et al. 2011), and higher redshift (Astier et al. 2006; Wood-Vasey et al. 2007; Balland et al. 2009; Amanullah et al. 2010). Cosmological analysis of these supernova samples (Kowalski et al. 2008; Conley et al. 2011; Suzuki et al. 2012; Betoule et al. 2014; Rest et al. 2014) have been combined with complimentary probes of large scale structure (Alam et al. 2017) and the CMB (Hinshaw et al. 2013; Planck Collaboration et al. 2013), and yet, despite these prodigious efforts, the nature of dark energy remains an unsolved mystery.

In attempts to tease out the nature of dark energy, currently running and planned surveys are once again ramping up their statistical power. The Dark Energy Survey (DES, Bernstein et al. 2012; Abbott et al. 2016) will be observing thousands of Type Ia supernova, attaining both spectroscopic and photometric confirmation. The Large Synoptic Survey Telescope (LSST, Ivezic et al. 2008; LSST Science Collaboration et al. 2009) will produce scores of thousands of photometrically classified supernovae. Such increased statistical power demands a similarly increased fidelity and flexibility in modelling the supernovae for cosmological purposes, as systematic uncertainty will prove to be the limiting factor in our analyses.

As such, staggering effort is being put into developing more sophisticated supernovae analyses. Scolnic & Kessler (2016) and Kessler & Scolnic (2017) explore sophisticated simulation corrections to traditional analyses. Approximate Bayesian computation methods also make use of simulations, trading traditional likelihoods and analytic approximations for more robust models with only the cost of increased computational time (Weyant et al. 2013; Jennings et al. 2016). Hierarchical Bayesian Models abound (Mandel et al. 2009; March et al. 2011, 2014; Rubin et al. 2015; Shariff et al. 2016; Roberts et al. 2017), however often face difficulties finding sufficient analytic approximations for complicated effects such as Malmquist bias.

In this paper, we lay out a new hierarchical model that advances the past work of Rubin et al. (2015). Section 2 is dedicated to a quick review of the supernovae cosmology. In Section 3 we outline our methodology. Model verification on simulated datasets is given in Section 4. Forecasts for the impending DES three year spectroscopic supernova survey are contained in Section 5. Section **??** investigates the effect of various systematics on our model, and Section **??** provides details on potential areas of improvement and unsuccessful methodologies.

2 REVIEW

Whilst supernova observations take the form of time-series photometric measurements of brightness in many photometric bands, most analyses do not work from these measurements of apparent magnitude and colour. Instead, most tech-

⋆ E-mail: samuelreay@gmail.com

niques fit these observations of magnitude (along with redshift) to a supernova model, with the most widely used being that of the empirical SALT2 model (Guy et al. 2007, 2010). This model is trained separately before fitting the supernovae light curves for the cosmology selected supernova sample (Guy et al. 2010; Mosher et al. 2014). The resulting output from the model is, for each supernova, a characterised amplitude x_0 (which can be converted into apparent magnitude $m_B = -2.5\log(x_0)$), a stretch term x_1 and colour term c, along with a covariance matrix describing the uncertainty on these summary statistics. As such, the product at the end is a (redshift-dependent) population of m_B, x_1 and c.

The underlying actual supernova population is not as clear cut, and indeed accurately characterising this population, its evolution over redshift and effects from environment is one of the challenges of supernova cosmology. However, given some modelled underlying population that lives in the redshift-dependent space M_B, x_1 and c, the introduction of cosmology into the model is simple – it is encoded in the functional map between those two populations, from apparent magnitude space to absolute magnitude. Specifically, for any given supernova our functional map may take the traditional form:

$$M_B = m_B + \alpha x_1 - \beta c - \mu(z) + \text{corrections}, \tag{1}$$

where α is the stretch correction (Phillips 1993), and β is the colour correction (Tripp 1998) that respectively encapsulate the empirical relation that broader and bluer supernovae are brighter. The corrections term at the end often includes corrections for host galaxy environment, as this has statistically significant effects on supernova properties (Kelly et al. 2010; Lampeitl et al. 2010; Sullivan et al. 2010; D'Andrea et al. 2011; Gupta et al. 2011; Johansson et al. 2013; Rigault et al. 2013; Uddin et al. 2017). The cosmological term, $\mu(z)$ represents the distance modulus, and is precisely known given cosmological parameters and an input redshift.

2.1 Traditional Analyses

Traditional χ^2 analyses such as that found in Kowalski et al. (2008); Conley et al. (2011); Betoule et al. (2014), minimise the difference in distance modulus between the cosmologically predicted values μ_C and the observed distance modulus $\mu_{\rm obs}$, shown respectively below:

$$\mu_C = 5\log\left[\frac{(1+z)r}{10}\right] \tag{2}$$

$$r = \frac{c}{H_0}\int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + \Omega_k(1+z')^2 + \Omega_\Lambda(1+z')^{3(1+w)}}} \tag{3}$$

$$\mu_{\rm obs} = m_B + \alpha x_1 - \beta c - M_B \tag{4}$$

The minimising function is then given as

$$\chi^2 = (\mu_{\rm obs} - \mu_C)^\dagger C^{-1}(\mu_{\rm obs} - \mu_C) \tag{5}$$

where C^{-1} is an uncertainty matrix that combined the uncertainty from the SALT2 fits, intrinsic dispersion, calibration, dust, peculiar velocity and many other factors (see Betoule et al. (2014) for a review). The benefit this analysis methodology provides is speed - for samples of hundreds of supernovae, efficient matrix inversion algorithms allow the likelihood to be evaluated quickly. The speed comes with two costs. Firstly, formulating a χ^2 likelihood requires a loss of model flexibility by building into the model assumptions of uncertainty Gaussianity. Secondly, the computational efficiency is dependent on inverting a covariance matrix with dimensionality linearly proportional to the number of supernovae. As this number increases, the cost of inversion rises quickly, and is not viable for samples with thousands of supernovae.

Selection efficiency, such as the well known Malmquist bias (Malmquist K. G. 1922) is accounted for by correcting data. Simulations following survey observational strategies and geometry are used to calculate the expected bias in distance modulus, which is then added onto the observational data. As these effects are not built into the likelihood, their influence on the error budget is not captured fully in the χ^2 distribution.

2.2 Approximate Bayesian Computation

To try and escape the limitations of the traditional analysis methodology, several recent methods have adopted Approximate Bayesian Computation, where supernova samples are forward modelled in parameter space and compared to observed distributions. Weyant et al. (2013) provides an introduction into ABC methods for supernova cosmology in the context of the SDSS-II results (Sako et al. 2014) and Flat ΛCDM cosmology, whilst Jennings et al. (2016) demonstrates their *superABC* method on simulated first season Dark Energy Survey samples, described in Kessler et al. (2015). In both examples, the supernova simulation package SNANA (Kessler et al. 2009) is used to forward model the data at each point in parameter space.

By building the systematic uncertainties and selection effects into the simulation package, there is vastly more freedom in how to treat and model those effects. Data does not need to be corrected, analytic approximations do not need to be applied, we are free to incorporate algorithms that simply cannot be expressed in a tractable likelihood. This freedom comes with a cost – computation. The classical χ^2 method's most computationally expensive step in a fit is matrix inversion. For ABC methods, we must instead simulate an entire supernova population in its entirety – drawing from underlying supernova populations, modelling light curves, applying selection effects, fitting light curves and applying data cuts. This is an intensive process. Luckily, efficient sampling algorithms that have walkers that rely on the Markov properties of groups instead of individual walkers, such as Ensemble sampling (Foreman-Mackey et al. 2013) allow easy parallelisation of parameter fits and can help ease the computational burden.

One final benefit of ABC methods is that they can move past the traditional treatment of supernovae with summary statistics (m_B, x_1 and c). Jennings et al. (2016) presents both a metric used to compare forward modelled summary statistic populations (denoted the 'Tripp' metric) and a metric directly applicable to the observed supernova light curves themselves, however evaluation of systematic uncertainty was only performed using the Tripp metric.

2.3 Hierarchical Bayesian Models

Sitting comfortably between the traditional models simplicity and the complexity of forward modelling lies Hierarchical Bayesian Models. With the introduction of multiple layers in our model, we can add far more flexibility than a traditional analysis whilst still maintaining most of the computational benefits that come from having a tractable likelihood. Mandel et al. (2009) and Mandel et al. (2011) construct a hierarchical model that they apply to the light curve fitting for supernova. March et al. (2011, 2014); Karpenka (2015) derive a hierarchical model and simplify it by analytically marginalising over nuisance parameters to provide a model that offers increased flexibility with reduced uncertainty over the traditional method. The recent BAHAMAS model (Shariff et al. 2016) builds off this and reanalyses the JLA dataset, whilst including extra freedom in the correction factors α and β, finding evidence for redshift dependence on β. Ma et al. (2016) also performed a reanalysis of the JLA dataset, finding significant differences in α and β values from the original as well. Notably, these methods rely on data that is bias corrected, however the UNITY framework given by Rubin et al. (2015) incorporates selection efficiency analytically in the model, and is applied to the Union 2.1 dataset (Suzuki et al. 2012). The well known BEAMS (Bayesian estimation applied to multiple species) methodology from Kunz et al. (2007) has been extended and applied in several works (Hlozek et al. 2012), mostly lately to include redshift uncertainty for photometric redshift application as zBEAMS (Roberts et al. 2017) and to include simulated bias corrections in Kessler & Scolnic (2017).

The flexibility afforded by a hierarchical model allows for investigations into different treatments of underlying populations, rates, redshift distributions, mass corrections and redshift evolution, each of will will be discussed further in the outline of our model below.

3 OUR METHOD

We construct our Bayesian Hierarchical Model with several goals in mind: creation of a redshift-dependent correlated underlying supernova population, increased treatment of systematics, and analytic correction of selection effects. As this is closest to the UNITY method from Rubin et al. (2015, hereafter denoted R15), we follow a similar model scaffold, and construct the model in Stan (Carpenter et al. 2017; Stan Development Team 2017) which uses automatic differentiation and the no-U-turn Sampler (NUTS), which is a variant of Hamiltonian Monte Carlo, to efficiently sample high dimensional parameter space.

At the most fundamental level, a supernova analysis is simply a mapping from an underlying population onto an observed population, where cosmology is encoded directly in the mapping function. The difficulty arises both in adequately describing the biases in the mapping function, and in adding sufficient, physically motivated flexibility in both of these populations whilst not adding *too* much flexibility, such that model fitting becomes pathological due to increasing parameter degeneracies within the model.

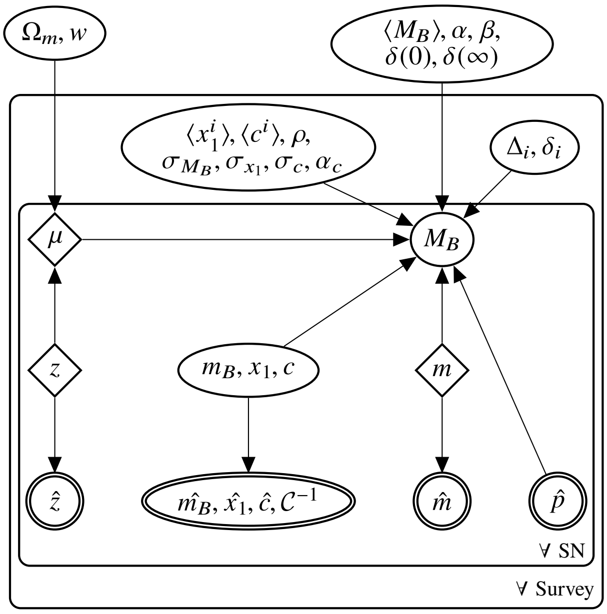

Figure 1. Probabilistic graphical model for our likelihood. Double-lined nodes represent observed variables, diamond nodes represent deterministic variables, and standard nodes represent fit variables.

3.1 Observables Nature Provides: Populations

Like most of the BHM methods introduced previously, we work from the summary statistics as well, where each observed supernova has an apparent magnitude $\hat{m_B}$, stretch $\hat{x_1}$ and colour $\hat{c}$, with uncertainty $\mathcal{C}$ on those values. Additionally, each supernova has an observed redshift $\hat{z}$ and a host galaxy mass associated with it, $\hat{m}$. Given machine learning classifiers to type supernovae, we will also have a probability of being a Type Ia, $\hat{p}$. Our set of observables is therefore given as $\{\hat{m_B}, \hat{x_1}, \hat{c}, \hat{z}, \hat{p}, \hat{m}, \mathcal{C}\}$, as shown in the probabilistic graphical model (PGM) in Figure 1.

The first layer of the hierarchy represents the latent population of m_B, x_1 and c. This can be thought of as the 'true' values for the supernova, and is normally distributed around the observed values such that

$$P(\hat{m_B}, \hat{x_1}, \hat{c} | m_B, x_1, c, \mathcal{C}) = \mathcal{N}(\{\hat{m_B}, \hat{x_1}, \hat{c}\} | \{m_B, x_1, c\}, \mathcal{C}). \quad (6)$$

As we are focused on the spectroscopically confirmed supernovae for this iteration of the method, we assume the observed redshift $\hat{z}$ is the true redshift z such that $P(\hat{z}|z) = \delta(\hat{z} - z)$. Similarly, we assume that the observed mass is the true mass m, such that Whilst the latter is not physically motivated like the redshift measurement, the relatively small contribution of the host galaxy mass to the model makes this an acceptable approximation.

3.2 Underlying Population

The underlying supernova population is often treated with two components - a population distribution in colour and stretch, and intrinsic dispersion. Analytic models often treat the colour and stretch populations with a skew normal and normal, respectively, and have them as independent. Intrinsic dispersion is treated in a variety of manners in other models, from representing it simply as (Gaussian) scatter on the absolute magnitude of the supernova population, to correlated multivariate normal scatter on the combined magnitude, stretch and colour distribution. Additionally, redshift drift of populations' mean colour and stretch will introduce a cosmological bias in our fits unless the population possesses similar ability to change as a function of redshift.

We model the underlying population and intrinsic dispersion together as a redshift-dependent multivariate skew normal for each survey. Following R15 we allow the mean colour and stretch to vary over redshift, anchoring four equally spaced redshift nodes spanning the redshift range of each survey, linearly interpolating between the nodes. These nodes are represented as $\langle x_1^i \rangle$ and $\langle c^i \rangle$. Both the colour and stretch means are modelled with normal priors. The correlation between the magnitude, stretch and colour populations is represented in a correlation matrix ρ, which is treated with an LKJ prior. Skewness is fixed to 0 for the absolute magnitude and stretch, but left free for colour. We parameterise the standard skew normal skewness α_c by sampling $\delta_c = \alpha_c/\sqrt{1+\alpha_c^2}$ which itself is given a uniform prior $\mathcal{U}(0, 0.98)$.

The width of the population, represented by the vector $\{\sigma_{M_B}, \sigma_{x_1}, \sigma_c\}$ is subject to Cauchy priors, however are sampled in log space for efficiency in sampling close to zero.

As such, the only constant between survey populations is the absolute magnitude M_B, with the skewness, redshift-dependent means, width and correlations fit individually for each survey.

On top of the Ia populations, as described above, we also include a simplistic outlier population that also follows R15 (and therefore Kunz et al. 2007) as a Gaussian mixture; where the mean of the population is fixed to the Ia population, but the population width is set to a width of $\sigma^{\rm outl} = 1$ in M_B, x_1 and c. With the spectroscopic DES sample, the contamination rate is expected to be far too low to actually fit contamination population, however in future works with photometric samples that will suffer from significantly more contamination it will be required that extra degrees of freedom are afforded the outlier population. Proof of concept simulation fits show that an acceptable parametrisation is to represent the typically brighter contaminant population as $\langle M_B^{\rm outl} \rangle = \langle M_B \rangle - \delta^{\rm outl}_{M_B}$, where $\delta^{\rm outl}_{M_B}$ is constrained to be positive, or even to be greater than a small positive number to reduce degeneracy between the two populations. For the purposes of the DES spectroscopic sample, which will be dominated by confirmed Type Ia supernovae, $\delta^{\rm outl}_{M_B} = 0$. We assume that supernovae fall into either population as determined by their observed classification probability $\hat{p}$.

3.3 Population Map

3.3.1 Cosmology

We formulate our model with three different cosmological parameterisations; Flat ΛCDM, Flat wCDM and standard ΛCDM. Ω_m is given the prior $\mathcal{U}(0.05, 0.99)$, Ω_Λ was treated with $\mathcal{U}(0, 1.5)$ and the equation of state w was similarly set to a flat prior $\mathcal{U}(-0.4, -2.0)$. For calculating the distance modulus, we fix $H_0 = 70 {\rm km\,s^{-1}\,Mpc^{-1}}$.

3.3.2 Standardisation Parameters

With increasingly large datasets and more nuanced analyses, the choice of how to handle α and β becomes an important consideration when constructing a model. R15 employs a broken linear relationship for both colour and stretch, where different values of α and β are adopted depending on whether x_1 and c are respectively positive or negative (although the cut could be placed at a location other than 0). Shariff et al. (2016) instead of employ a colour-dependent β, model β as redshift-dependent, testing two phenomenological models; $\beta(z) = \beta_0 + \beta_1 z$ and $\beta(z) = \beta_0 + \Delta\beta\,(0.5 + \arctan(100(z - z_t))/\pi)$, where the later effects a rapid but smooth change in β at a turnover redshift z_t.

We tested two models against simulated supernova sets; $\beta(c) = \beta_0 + \beta_1 c$ and $\beta(z) = \beta_0 + \beta_1 z$. See Section 4.2 for details on simulation generation. We found for both models that non-zero values for β_1 are preferred (even with constant β used in simulation) due to severe degeneracy with selection effects. This degeneracy resulted in a significant bias in recovered cosmology, and so in our final model we continue to adopt the constant α and β found in traditional analyses.

3.3.3 Host Galaxy Environment

It is now well known that host galaxy environment has a significant effect on supernova properties. The latest sample of over 1300 spectroscopically confirmed Type Ia supernovae show $> 5\sigma$ evidence for correlation between host mass and luminosity (Uddin et al. 2017). The traditional correction, as employed in analyses such as Suzuki et al. (2012) and Betoule et al. (2014) invoke a step function such that $\Delta M = 0.08\mathcal{H}(\log(M) - 10))$, where $\mathcal{H}$ is the Heaviside step function and M is the galaxy mass in solar masses. The scale of this step function varies from analysis to analysis, with the 0.08 value shown previously sourced from Sullivan et al. (2010) and used in Betoule et al. (2014). In this work we adopt the model used in R15, which follows the work from Rigault et al. (2013), such that we introduce two parameters to incorporate a redshift-dependent host galaxy mass correction:

$$\Delta M = \delta(0)\left[\frac{1.9\left(1 - \frac{\delta(0)}{\delta(\infty)}\right)}{0.9 + 10^{0.95z}} + \frac{\delta(0)}{\delta(\infty)}\right] \tag{7}$$

We also take flat priors on the parametrisation $\delta(0)$, $\delta(0)/\delta(\infty)$.

3.3.4 Uncertainty Propagation

The chief difficulty with including systematic uncertainties in supernova analyses is that they generally occur during the observational pipeline, and have difficult-to-model effects on the output observations. As such, the normal treatment for systematics is to compute their effect on the supernova summary statistics – computing the numerical derivatives $\frac{\partial \hat{m}_B}{\partial \mathcal{Z}_i}$, $\frac{\partial \hat{x}_1}{\partial \mathcal{Z}_i}$, $\frac{\partial \hat{c}}{\partial \mathcal{Z}_i}$, where $\mathcal{Z}_i$ represents the $i^{\rm th}$ systematic.

Assuming that the gradients can be linearly extrapolated – which is a reasonable approximation for modern surveys with high quality control of systematics – we can incorporate into our model a deviation from the observed original values by constructing a $(3 \times N_{\rm sys})$ matrix containing the numerical derivatives for the $N_{\rm sys}$ systematics and multiplying it with the row vector containing the offset for each systematic. By scaling the gradient matrix to represent the shift over 1σ of systematic uncertainty, we can simply enforce a unit normal prior on the systematic row vector to increase computational efficiency.

This method of adjusting the observed summary statistics is used throughout the traditional and BHM analyses, however it is normally constrained to band systematics. That is, each band for each survey has two systematics associated with it - the calibration uncertainty and the filter wavelength uncertainty. We include these in our approach, in addition to including HST Calspec calibration uncertainty, 10 SALT2 model systematic uncertainties, a dust systematic, a global redshift bias systematic and also the systematic peculiar velocity uncertainty. This gives thirteen global systematics shared by all surveys, plus two systematics per band in each survey. Denoting $\eta \equiv \{m_B, x_1, c\}$, our initial conditional likelihood for our observed summary statistics shown in Equation (6) becomes

$$P(\hat{\eta}, \frac{\partial\hat{\eta}}{\partial\mathcal{Z}_i}|\eta, \delta\mathcal{Z}_i, C) = \mathcal{N}(\hat{\eta} + \delta\mathcal{Z}_i\frac{\partial\hat{\eta}}{\partial\mathcal{Z}_i}|\eta, \mathcal{C}). \tag{8}$$

3.3.5 Selection Effects

Our treatment of selection effects is to incorporate selection efficiency into our model, rather than relying on simulation-driven data corrections. As such, we need to describe the probability that the events we observe are both drawn from the distribution predicted by the underlying theoretical model *and* that those events, given they happened, are subsequently successfully observed. To make this extra conditional explicit, we can write the likelihood of the data given an underlying model, θ, *and* that the data are included in our sample, denoted by S, as

$$\mathcal{L} = P(\mathrm{data}|\theta, S). \tag{9}$$

As the model so far described in previous sections describes $P(D|\theta)$, and we wish to formulate a function $P(S|\mathrm{data}, \theta)$ that describes the chance of an event being successfully observed, we rearrange the likelihood in terms of those functions and find

$$\mathcal{L} = \frac{P(S|\mathrm{data}, \theta)P(\mathrm{data}|\theta)}{\int P(S|D, \theta)P(D|\theta)\, dD}, \tag{10}$$

where the denominator represents an integral over all potential data. Full derivation of this can be found in Appendix A. As θ represents the vector of all model parameters, and D represents a vector of all observed variables, this is not a trivial integral. Techniques to approximate this integral, such as Monte-Carlo integration or high dimensional Gaussian processes failed to give tractable posterior surfaces that could be sampled efficiently by HMC. We therefore simplify the integral and approximate the selection effects in apparent magnitude and redshift space independently, such that the denominator, denoted now w for simplicity, is given as

$$w = \int \left[\int P(S|m_B)P(m_B|z, \theta)\, dm_B \right] P(S|z)P(z|\theta)\, dz. \tag{11}$$

We apply two further approximations similar to those made in R15 – that the redshift distribution of the observed supernova reasonably well sample the $P(S|z)P(z|\theta)$ distribution, and that the survey colour and stretch populations can be treated as Gaussian for the purposes of this integral. The population $P(m_B|z, \theta)$ thus becomes $\mathcal{N}(m_B|m_B^*(z), \sigma^*_{m_B})$, where

$$m_B^*(z) = \langle M_B \rangle + \mu(z) - \alpha\langle x_1(z)\rangle + \beta\langle c(z)\rangle \tag{12}$$

$$\sigma^*_{m_B} = \sigma^2_{M_B} + (\alpha\sigma_{x_1})^2 + (\beta\sigma_c)^2 + 2(\beta\sigma_{M_B,c} - \alpha\sigma m_B, x_1 - \alpha\beta\sigma_{x_1,c}) \tag{13}$$

What then remains is determining the functional form of $P(S|m_B)$. For the treatment of the DES, we find that the error function which smoothly transitions from some constant efficiency down to zero is sufficient. This similar result has been found by many past surveys (Dilday et al. 2008; Barbary et al. 2010; Perrett et al. 2012; Graur et al. 2013; Rodney et al. 2014) THESE CITATIONS EXACTLY MIRROR RUBIN2015, NOT SURE IF I SHOULD CHANGE THEM. For our treatment of the low redshift surveys included Introduce which lowz surveys earlier on we simplify the model fitting and group all low redshift surveys together. This requires a more complicated selection function than the error function used for DES, however is reasonably well fit with a skew normal.

The selection functions were fit to apparent magnitude efficiency ratios calculated from SNANA simulations by calculating an efficiency ratio as a function of apparent magnitude. Uncertainty of the Malmquist bias (entering both through statistical uncertainty from finite sized simulations in the efficiency ratio and the discrepancy between the analytic approximation and non-analytic simulation results) is incorporated into the fitting for the analytic approximation. Uncertainty is uniformly added to the efficiency ratio until the reduced χ^2 of the analytic fit reached 1, allowing us to extract an uncertainty covariance matrix for our analytic fits.

The DES survey is thus characterised by mean fit values $\mu_{\rm DES} = 22.3$, $\sigma_{\rm DES} = 0.7$, with the combined low redshift surveys characterised by $\mu_{\rm LowZ} = 14.1$, $\sigma_{\rm LowZ} = 1.36$, $\alpha_{\rm LowZ} = 3.8$.

With the well sampled approximation as specified previously, we can remove the redshift integral in Eq (11) and replace it with a correction for each observed supernova. For the error function (DES) and skew normal selection (LowZ) functions respectively, this correction becomes

$$w_{\rm DES} = \Phi^c\left(\frac{m_B^* - \mu_{\rm DES}}{\sqrt{\sigma^{*2}_{m_B} + \sigma^2_{\rm DES}}}\right) \tag{14}$$

$$w_{\rm LowZ} = 2\mathcal{N}\left(\frac{m_B^* - \mu_{\rm LowZ}}{\sqrt{\sigma^{*2}_{m_B} + \sigma^2_{\rm LowZ}}}\right)\times \Phi\left(\frac{\mathrm{sign}(\alpha_{\rm LowZ})(m_B^* - \mu_{\rm LowZ})}{\frac{\sigma^{*2}_{m_B} + \sigma^2_{\rm LowZ}}{\sigma^2_{\rm LowZ}}\sqrt{\frac{\sigma^2_{\rm LowZ}}{\alpha^2_{\rm LowZ}} + \frac{\sigma^{*2}_{m_B}\sigma^2_{\rm LowZ}}{\sigma^{*2}_{m_B} + \sigma^2_{\rm LowZ}}}}\right), \tag{15}$$

where Φ represents the normal CDF function, and Φ^c the complimentary CDF.

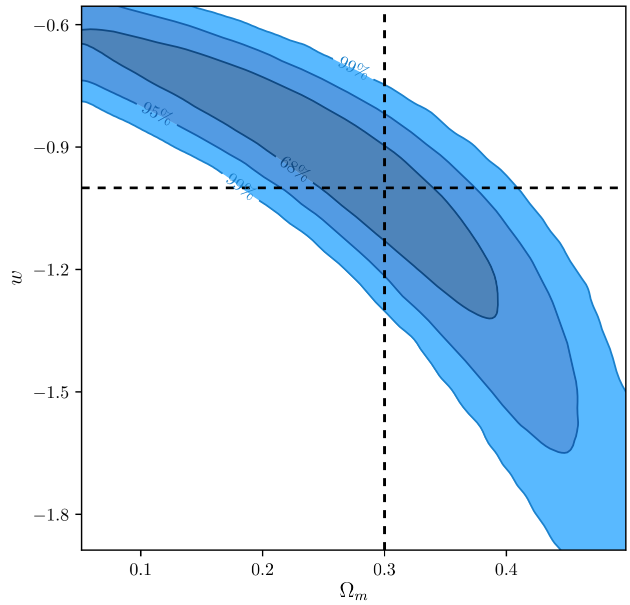

Figure 2. Posterior surfaces for 22 realisations of supernova data with the Flat wCDM model.

4 MODEL VERIFICATION

In order to verify our model we run it through several tests. First, we validate on toy models, verifying that there is not significant cosmological bias in data sets more constraining than the DES spectroscopic sample. We then also generate

4.1 Applied to Toy Spectroscopic Data

We generate simple toy data to validate the basic premise of the model. For both DES and LowZ data we draw from an underlying M_B, x_1, c population and translate into apparent magnitude space using $\Omega_m = 0.3$, $\alpha = 0.14$ and $\beta = 3.1$. Masses are randomly drawn from the interval 0 to 1, and a mass correction with $\delta(0) = 0.08$ and $\delta(0)/\delta(\infty) = 0.5$ included. Independent observational errors of 0.04, 0.2, 0.03 on m_B, x_1 and c (following the mean uncertainty for DES SNANA simulations) are added to create the observables. The results are then passed through the selection effects, where each supernova is only selected based on $P(S|m_B)$, using a skew normal function for the LowZ supernovae and error function for the DES-like supernovae. We draw from each survey simulation until we have 300 LowZ supernovae and 700 DES-like supernovae, representing a statistical sample of greater power than the estimated 250 supernovae for the DES spectroscopic analysis.

The above represents one realisation of data. We test three models (Flat ΛCDM, Flat wCDM, ΛCDM) with 300 realisations and find recovery of underlying cosmology without significant bias. Combined posterior surfaces of 22 realisations for Flat $wCDM$ fits are shown in Figure 2.

4.2 DES SN data validation

Early analyses often treated intrinsic dispersion simply as scatter in the underlying absolute magnitude of the underlying population, but recent analyses require more a more sophisticated approach. In our development of this model and tests of intrinsic dispersion, we analyse the effects of two different scatter models. The first model is the Guy et al. (2010, hereafter denoted the G10 scatter model), which models intrinsic scatter with a 70% contribution from coherent variation and 30% from chromatic variation. The second model, denoted the C11 model is sourced from Chotard et al. (2011) and has variation with 25% contribution from coherent scatter and 75% from chromatic variation. We also test a 'Combined' smear model, which is simply to draw equal samples from the C11 and G10 models. It is against this combined sample that we wish to rigorously validate our model.

Table 1. Tested population distributions, where the SK16 LowZ distribution is formed as sum of two bifurcated Gaussians, with the mean and spread of each component given respectively.

Model	$\langle x_1 \rangle$	σ_{x_1}	$\langle c \rangle$	σ_c
Gaussian LowZ	0.0	1.0	0.0	0.1
Gaussian DES	0.0	1.0	0.0	0.1
SK16 LowZ	0.55 & −1.5	$^{+0.45}_{-1.0}$ & $^{+0.5}_{-0.5}$	−0.055	$^{+0.15}_{-0.023}$
SK16 DES	0.973	$^{+0.222}_{1.472}$	−0.054	$^{+0.101}_{0.043}$

Simulations (using the SNANA package) follow the observational schedule and observing conditions for the DES and LowZ surveys. In addition to the improvements in the scatter models over the simple data, we also include peculiar velocities for the LowZ sample, and now also include our full treatment of systematics. Additionally, we also simulate two different underlying population – a Gaussian distribution in colour and stretch, and skewed colour and stretch populations using population values from Scolnic & Kessler (2016, hereafter SK16).

Each realisation of cosmology fitted contains 300 LowZ supernovae, and 250 DES-like supernovae, such that the uncertainties found when combining chains is representative of the uncertainty in the final DES spectroscopic analysis. Combined posterior surfaces for 150 data realisations are shown in Figure 3. Note that we have combined the posteriors for 150 realisations, and so we should expect the size of the uncertainty to be representative of one realisation, but the statistical spread of the final surface should be $\sqrt{150} \approx 12$ times less than a single realisation. Plot summaries for the fits are shown in Figure 3, and the parameter bounds are listed in Table 2. The results indicate changes in the underlying population can effect significant changes in the recovered standardisation parameters α and β, however do not appear to have a significant effect on recovered cosmology. Whilst the combined sample (on which the selection function was fit) shows no significant bias (as expected), the C11 and G10 only models show opposite biases in Ω_m. Cosmological bias is detected with $\approx 3.0\sigma$ significance for the C11 model and $\approx 1.7\sigma$ significance for the G10 model, indicating that the population and selection effect treatment cannot capture all necessary information to encapsulate the magnitude and chromatic smearing of the supernovae population. However, for the sample size of the DES and LowZ supernova samples (of order 600 supernova), these effects are sub-dominant to the statistical uncertainty, representing at most a deviation

Table 2. Parameter summaries shown for 150 realisations of supernovae data, each comprising of 300 LowZ supernovae and 250 DES-like supernovae. Maximum likelihood statistics, with errors quoted to 1σ (68% confidence) are reported.

Model	Ω_m	α	β
C11 Gaussian	$0.316^{+0.062}_{-0.064}$	$0.143^{+0.028}_{-0.027}$	$2.97^{+0.20}_{-0.19}$
C11 SK16	$0.316^{+0.063}_{-0.062}$	0.155 ± 0.024	$2.77^{+0.26}_{-0.27}$
Combined Gaussian	$0.299^{+0.067}_{-0.060}$	$0.144^{+0.030}_{-0.028}$	$3.11^{+0.21}_{-0.22}$
Combined SK16	$0.297^{+0.075}_{-0.064}$	$0.150^{+0.026}_{-0.025}$	$3.00^{+0.29}_{-0.31}$
G10 Gaussian	$0.290^{+0.070}_{-0.060}$	0.145 ± 0.030	3.27 ± 0.23
G10 SK16	$0.290^{+0.079}_{-0.073}$	$0.147^{+0.026}_{-0.028}$	$3.26^{+0.34}_{-0.33}$

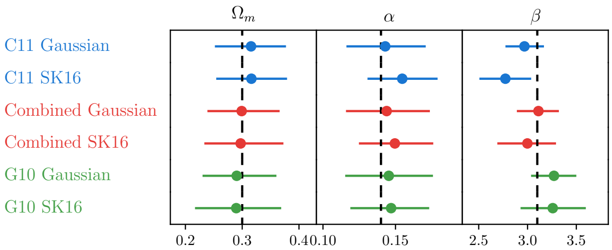

Figure 3. Graphical parameter summaries for 150 realisations of supernova data, following Table 2.

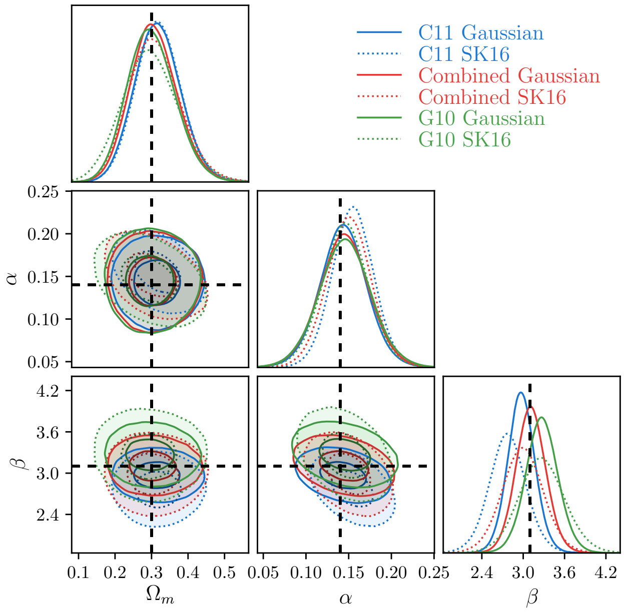

Figure 4. Posterior surfaces for 150 realisations of supernova data. Neither population nor scatter model has a significant effect on posterior surface location or shape. Population skewness has the primary effect of changing the degeneracy direction for the $\alpha - \beta$ contour, whilst the scatter models and skewness have noticeable effects on the value and uncertainty of the recovered parameter β.

of 0.25σ and, accounting for statistical fluctuations up to 2σ, a minimum bias of 0.08σ for the C11 model.

5 DES FORECASTS

To forecast the results for the 3 year DES spectroscopic supernovae sample, we generate 100 realisations of datasets, each comprising 300 low redshift supernovae and 250 DES-like supernovae. We fit these through the three cosmological models - Flat ΛCDM, ΛCDM and wCDM, however focus mostly on the wCDM model as it is the primary scientific concern of the Dark Energy Survey. We run all three cosmological models with systematics enabled and disabled to explicitly show the uncertainty included from systematic contribution.

How to include flat lambdacdm - the plot is boring, just quote the results, or just quote uncertainty, should I round the uncertainty to make it symmetric? Cant do that for wCDM or LambdaCDM because theyre definitely not gaussian

Talk about how these arent even competitive with JLA and we really need to combine data? Talk about how the photometric supernova dataset should be way better? Talk about statistics limited.

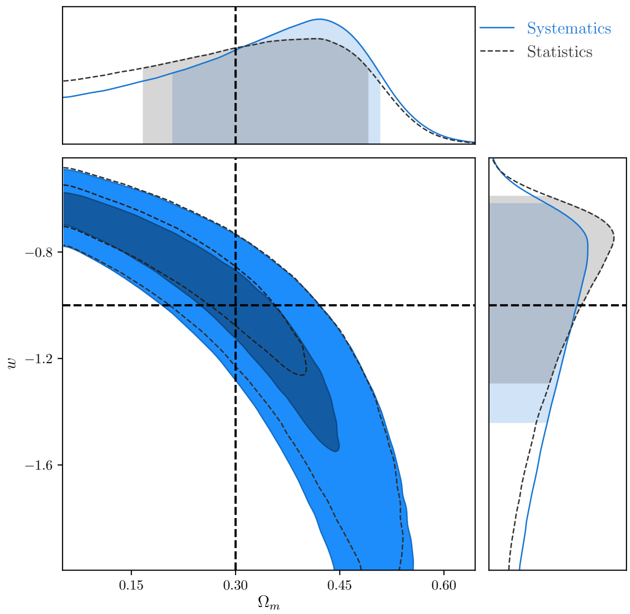

Figure 5. w.

6 CONCLUSIONS

Conclude.

ACKNOWLEDGEMENTS

Plots of posterior surfaces and parameter summaries were created with `ChainConsumer` (Hinton 2016).

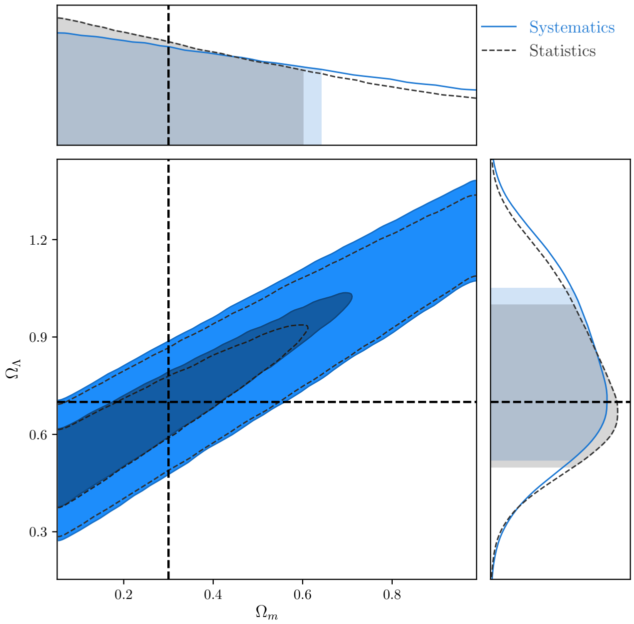

Figure 6. ole.

REFERENCES

Abbott T., et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 1270

Alam S., et al., 2017, Monthly Notices of the Royal Astronomical Society, 470, 2617

Amanullah R., et al., 2010, The Astrophysical Journal, 716, 712

Astier P., et al., 2006, Astronomy and Astrophysics, 447, 31

Bailey S., et al., 2008, eprint arXiv:0810.3499

Balland C., et al., 2009, Astronomy and Astrophysics, 507, 85

Barbary K., et al., 2010, The Astrophysical Journal, 745, 27

Bernstein J. P., et al., 2012, The Astrophysical Journal, 753, 152

Betoule M., et al., 2014, Astronomy & Astrophysics, 568, 32

Carpenter B., et al., 2017, Journal of Statistical Software, 76, 1

Chotard N., et al., 2011, Astronomy & Astrophysics, 529, 6

Conley A., et al., 2011, The Astrophysical Journal Supplement Series, 192, 1

Contreras C., et al., 2010, The Astronomical Journal, 139, 519

D'Andrea C. B., et al., 2011, The Astrophysical Journal, 743, 172

Dilday B., et al., 2008, The Astrophysical Journal, 682, 262

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publications of the Astronomical Society of Pacific, 125, 306

Freedman W. L., et al., 2009, The Astrophysical Journal, 704, 1036

Graur O., et al., 2013, The Astrophysical Journal, 783, 28

Gupta R. R., et al., 2011, ApJ, 740, 92

Guy J., et al., 2007, Astronomy and Astrophysics, 466, 11

Guy J., et al., 2010, Astronomy and Astrophysics, 523, 34

Hicken M., et al., 2009, The Astrophysical Journal, 700, 331

Hinshaw G., et al., 2013, The Astrophysical Journal Supplement Series, 208, 19

Hinton S., 2016, JOSS, 1

Hlozek R., et al., 2012, The Astrophysical Journal, 752, 79

Ivezic Z., et al., 2008, eprint arXiv:0805.2366

Jennings E., Wolf R., Sako M., 2016, eprint arXiv:1611.03087, pp 1–22

Johansson J., et al., 2013, Monthly Notices of the Royal Astronomical Society, 435, 1680

Karpenka N. V., 2015, The supernova cosmology cookbook: Bayesian numerical recipes. (`arXiv:1503.03844`), `http://arxiv.org/abs/1503.03844`

Kelly P. L., Hicken M., Burke D. L., Mandel K. S., Kirshner R. P., 2010, The Astrophysical Journal, 715, 743

Kessler R., Scolnic D., 2017, The Astrophysical Journal, 836, 56

Kessler R., et al., 2009, Publications of the Astronomical Society of the Pacific, 121, 1028

Kessler R., et al., 2015, The Astronomical Journal, 150, 172

Kowalski M., et al., 2008, The Astrophysical Journal, 686, 749

Kunz M., Bassett B., Hlozek R., 2007, Physical Review D, 75, 1

LSST Science Collaboration et al., 2009, eprint arXiv:0912.0201

Lampeitl H., et al., 2010, The Astrophysical Journal, 722, 566

Ma C., Corasaniti P.-S., Bassett B. A., 2016, Monthly Notices of the Royal Astronomical Society, 463, 1651

Malmquist K. G. 1922, Lund Medd. Ser. I, 100, 1

Mandel K. S., Wood-Vasey W. M., Friedman A. S., Kirshner R. P., 2009, The Astrophysical Journal, 704, 629

Mandel K. S., Narayan G., Kirshner R. P., 2011, The Astrophysical Journal, 731, 120

March M. C., Trotta R., Berkes P., Starkman G. D., Vaudrevange P. M., 2011, Monthly Notices of the Royal Astronomical Society, 418, 2308

March M. C., Karpenka N. V., Feroz F., Hobson M. P., 2014, Monthly Notices of the Royal Astronomical Society, 437, 3298

Mosher J., et al., 2014, The Astrophysical Journal, 793, 16

Perlmutter S., et al., 1999, The Astrophysical Journal, 517, 565

Perrett K., et al., 2012, The Astronomical Journal, 144, 59

Phillips M. M., 1993, The Astrophysical Journal, 413, L105

Planck Collaboration et al., 2013, Astronomy & Astrophysics, 571, 66

Rest A., et al., 2014, The Astrophysical Journal, 795, 44

Riess A. G., et al., 1998, The Astronomical Journal, 116, 1009

Rigault M., et al., 2013, Astronomy & Astrophysics, 560, A66

Roberts E., Lochner M., Fonseca J., Bassett B. A., Lablanche P.-Y., Agarwal S., 2017, eprint arXiv:1704.07830

Rodney S. A., et al., 2014, The Astronomical Journal, 148, 13

Rubin D., et al., 2015, The Astrophysical Journal, 813, 15

Sako M., et al., 2014, eprint arXiv:1401.3317

Scolnic D., Kessler R., 2016, The Astrophysical Journal Letters, 822

Shariff H., Jiao X., Trotta R., van Dyk D. A., 2016, The Astrophysical Journal, 827, 1

Stan Development Team 2017, PyStan: the interface to Stan, `http://mc-stan.org/`

Sullivan M., et al., 2010, Monthly Notices of the Royal Astronomical Society, 406, 782

Suzuki N., et al., 2012, The Astrophysical Journal, 746, 85

Tripp R., 1998, A two-parameter luminosity correction for Type IA supernovae. Vol. 331, EDP Sciences [etc.], `http://adsabs.harvard.edu/abs/1998A{%}26A...331..815T`

Uddin S. A., Mould J., Lidman C., Ruhlmann-Kleider V., Zhang B. R., 2017, eprint arXiv:1709.05830

Weyant A., Schafer C., Wood-Vasey W. M., 2013, The Astrophysical Journal, 764, 116

Wood-Vasey W. M., et al., 2007, The Astrophysical Journal, 666, 694

APPENDIX A: SELECTION EFFECT DERIVATION

A1 General Selection Effects

When formulating and fitting a model using a constraining dataset, we wish to resolve the posterior surface defined by

$$P(\theta|\text{data}) \propto P(\text{data}|\theta)P(\theta), \tag{A1}$$

which gives the probability of the model parameter values (θ) given the data. Prior knowledge of the allowed values of the model parameters is encapsulated in the prior probability $P(\theta)$. Of primary interest to us is the likelihood of observing the data given our parametrised model, $\mathcal{L} \equiv P(\text{data}|\theta)$. When dealing with experiments which have imperfect selection efficiency, our likelihood must take that efficiency into account. We need to describe the probability that the events we observe are both drawn from the distribution predicted by the underlying theoretical model *and* that those events, given they happened, are subsequently successfully observed. To make this extra conditional explicit, we write the likelihood of the data given an underlying model, θ, *and* that the data are included in our sample, denoted by S, as:

$$\mathcal{L} = P(\text{data}|\theta, S). \tag{A2}$$

A variety of selection criteria are possible, and in our method we use our data in combination with the proposed model to determine the probability of particular selection criteria. That is, we characterise a function $P(S|\text{data}, \theta)$, which colloquially can be stated as *the probability of a potential observation passing selection cuts, given our observations and the underlying model.* We can introduce this expression in a few lines due to symmetry of joint probabilities and utilising that $P(x, y, z) = P(x|y, z)P(y, z) = P(y|x, z)P(x, z)$:

$$P(\text{data}|S, \theta)P(S, \theta) = P(S|\text{data}, \theta)P(\text{data}, \theta) \tag{A3}$$

$$P(\text{data}|S, \theta) = \frac{P(S|\text{data}, \theta)P(\text{data}, \theta)}{P(S, \theta)} \tag{A4}$$

$$= \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)P(\theta)}{P(S|\theta)P(\theta)} \tag{A5}$$

$$= \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{P(S|\theta)} \tag{A6}$$

which is equal to the likelihood $\mathcal{L}$. Introducing an integral over all possible events D, so we can evaluate $P(S|\theta)$,

$$\mathcal{L} = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S, D|\theta)\, dD} \tag{A7}$$

$$\mathcal{L} = \frac{P(S|\text{data}, \theta)P(\text{data}|\theta)}{\int P(S|D, \theta)P(D|\theta)\, dD}, \tag{A8}$$

where we define the denominator as w for simplicity in future derivations.

A2 Supernova Selection Effects

We assume that our selection effects can be reasonably well encapsulated by independent functions of (actual) apparent magnitude and redshift, such that $P(S|\text{data}, \theta) = P(S|z)P(S|m_B)$. Our denominator then becomes

$$w = \int \cdots \int d\hat{z}\, d\hat{m_B}\, dz\, dm_B P(S|z)P(S|m_B)P(\hat{z}|z)P(\hat{m_B}|m_B)P(z, m_B|\theta), \tag{A9}$$

where for simplicity we have not written out all the integrals which do not interact with the selection effects explicitly. Due to our assumed perfect measurement of redshift, $P(\hat{z}|z) = \delta(\hat{z} - z)$. $P(\hat{m_B}|m_B)$ is a Gaussian due to our Gaussian model of summary statistics m_B, x_1, c, and can be analytically integrated out, collapsing the integral over $\hat{m_B}$. Finally, we can express $P(z, m_B|\theta)$ as $P(m_B|z, \theta)P(z|\theta)$, where the first term requires us to calculate the magnitude distribution of our underlying population at a given redshift, and the second term is dependent on survey geometry and supernovae rates. We can thus state

$$w = \int \left[\int P(S|m_B)P(m_B|z, \theta)\, dm_B \right] P(S|z)P(z|\theta)\, dz. \tag{A10}$$

By assuming that the distribution $P(S|z)P(z|\theta)$ is well sampled by the observed supernovae redshifts, we can approximate the integral over redshift by evaluating

$$\int P(S|m_B)P(m_B|z, \theta)\, dm_B \tag{A11}$$

for each supernova in the dataset – i.e. Monte Carlo integration with assumed perfect importance sampling.

As stated in Section 3.3.5, the underlying population in apparent magnitude, when we discard skewness, can be represented as $\mathcal{N}(m_B|m_B^*(z), \sigma^*_{m_B})$, where

$$m_B^*(z) = \langle M_B \rangle + \mu(z) - \alpha\langle x_1(z)\rangle + \beta\langle c(z)\rangle \tag{A12}$$

$$\sigma^*_{m_B} = \sigma^2_{M_B} + (\alpha\sigma_{x_1})^2 + (\beta\sigma_c)^2 + 2(\beta\sigma_{M_B,c} - \alpha\sigma m_B, x_1 - \alpha\beta\sigma_{x_1,c}). \tag{A13}$$

Then, modelling $P(S|m_B)$ as either a normal or a skew normal, we can analytically perform the integral and reach equations (14) and (15).

A3 Approximate Selection Effects

This paper has been typeset from a TEX/LATEX file prepared by the author.

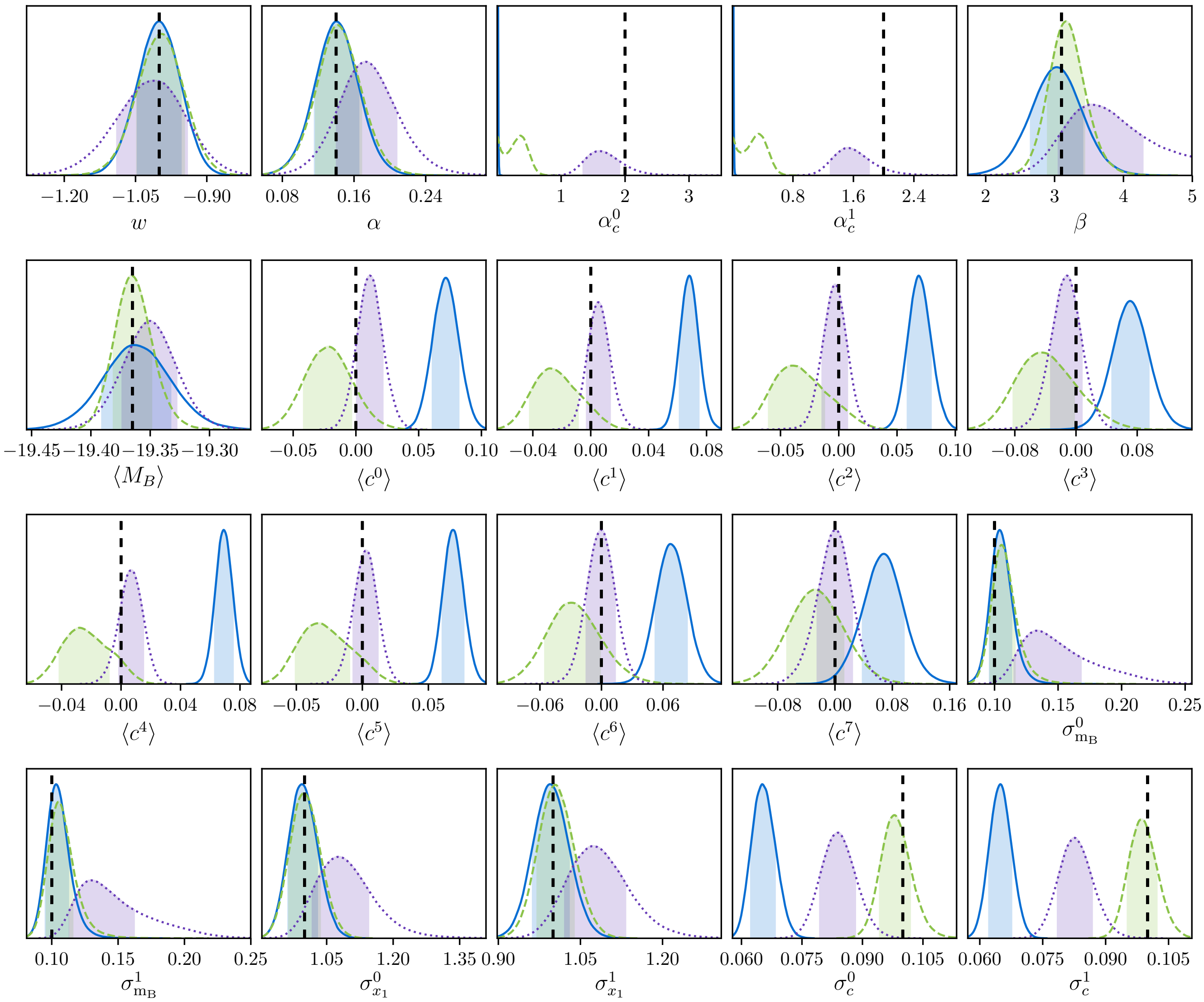

Figure A1. Marginalised distributions for 200 realisations of cosmology, fit to Flat wCDM with prior $\Omega_m \sim \mathcal{N}(0.3, 0.01)$, each containing 1000 simulated high-z and 1000 simulated low-z supernovae. The dashed green surfaces represent a fit to an underlying Gaussian colour population with the unshifted model. The blue solid surface represents fits to a skewed colour population with the unshifted model, and the purple dotted surface represents a fit to a skewed colour population with the shifted model. The superscript 0 and 1 denote the two different surveys (high-z and low-z respectively), and similarly the first four $\langle c^i \rangle$ parameters represent the four redshift nodes in the high-z survey, and the last four represent the nodes for the low-z survey. We can see that the shifted model is far better able to recover skewed input populations than the unshifted, performing better in terms of recovering skewness α_c, mean colour $\langle c \rangle$ and width of the colour distribution σ_c. The unshifted model recovers the correct colour mean and width if you approximate a skew normal as a normal: $\Delta\mu = \sqrt{2/\pi}\sigma_c \delta_c \approx 0.071$, which is approximately the deviation found in fits to the colour population mean. Whilst the performance of the shifted model on skewed data populations is an improvement over the unshifted model, it comes at a cost - significant bias appears in the fits to α, β, σ_{m_B} and w. These biases in cosmological parameters are not found in the unshifted model, and similarly the bias when moving from a Gaussian to skewed colour population is negligible. For these reasons, we adopt the unshifted model.