The Astrophysical Journal, 827:1 (25pp), 2016 August 10 doi:10.3847/0004-637X/827/1/1

BAHAMAS: NEW ANALYSIS OF TYPE Ia SUPERNOVAE REVEALS INCONSISTENCIES WITH STANDARD COSMOLOGY

Hikmatali Shariff[1,3], Xiyun Jiao[2,3], Roberto Trotta[1,3], and David A. van Dyk[2,3]
[1] Astrophysics Group, Physics Department, Imperial College London, Prince Consort Road, London SW7 2AZ, UK
[2] Statistics Section, Mathematics Department, Huxley Building, South Kensington Campus, Imperial College London, London SW7 2AZ, UK
[3] Imperial Centre for Inference and Cosmology, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK
Received 2015 November 16; accepted 2016 April 4; published 2016 August 3

ABSTRACT

We present results obtained by applying our BAyesian HierArchical Modeling for the Analysis of Supernova cosmology (BAHAMAS) software package to the 740 spectroscopically confirmed supernovae of type Ia (SNe Ia) from the "Joint Light-curve Analysis" (JLA) data set. We simultaneously determine cosmological parameters and standardization parameters, including corrections for host galaxy mass, residual scatter, and object-by-object intrinsic magnitudes. Combining JLA and *Planck* data on the cosmic microwave background, we find significant discrepancies in cosmological parameter constraints with respect to the standard analysis: we find $\Omega_{\rm m} = 0.399 \pm 0.027$, 2.8σ higher than previously reported, and $w = -0.910 \pm 0.045$, 1.6σ higher than the standard analysis. We determine the residual scatter to be $\sigma_{\rm res} = 0.104 \pm 0.005$. We confirm (at the 95% probability level) the existence of two subpopulations segregated by host galaxy mass, separated at $\log_{10}(M/M_{\odot}) = 10$, differing in mean intrinsic magnitude by 0.055 ± 0.022 mag, lower than previously reported. Cosmological parameter constraints, however, are unaffected by the inclusion of corrections for host galaxy mass. We find $\sim 4\sigma$ evidence for a sharp drop in the value of the color correction parameter, $\beta(z)$, at a redshift $z_t = 0.662 \pm 0.055$. We rule out some possible explanations for this behavior, which remains unexplained.

Key words: cosmological parameters – dark energy – dark matter – distance scale – methods: statistical –

1. INTRODUCTION

Supernovae of type Ia (SNe Ia) have been instrumental in establishing the accelerated expansion of the universe, starting with the momentous discovery of the Supernova Cosmology Project and the High-z Supernova Search Team in the late 1990s (Riess et al. 1998; Perlmutter et al. 1999). The accelerated expansion is currently widely attributed to the existence of a "dark energy" component, which is compatible with Einstein's cosmological constant. Over the last decade, the sample of SNe Ia has increased dramatically (e.g., Astier et al. 2006; Wood-Vasey et al. 2007; Bailey et al. 2008; Kowalski et al. 2008; Balland et al. 2009; Freedman et al. 2009; Hicken et al. 2009; Kessler et al. 2009; Amanullah et al. 2010; Contreras et al. 2010; Suzuki et al. 2012; Betoule et al. 2014; Rest et al. 2014), and it now comprises several hundred spectroscopically confirmed SNe Ia. Since SNe Ia probe the low-redshift universe, they are ideal tools to measure the properties of dark energy. Two of the most important tasks required to shed light on the origin of dark energy are to establish whether or not the dark-energy equation of state is evolving with time and whether modified-gravity scenarios might provide a viable alternative explanation.

SNe Ia are observationally characterized by an absence of H in their spectrum, and by the presence of strong Si ii lines. They occur when material from a companion accreting onto a white dwarf triggers carbon fusion, which proceeds until a core of typical mass 0.7$M_{\odot}$ of ^{56}Ni is created. The radioactive decay of ^{56}Ni to ^{56}Co, and subsequently to ^{56}Fe, produces γ-rays that heat up the ejecta, thus powering the light curve (LC). While it is believed that this happens when the mass of the white dwarf approaches (without reaching) the Chandrasekhar limit of $1.4M_{\odot}$, the debate about progenitor scenarios is not settled. There is strong evidence that some systems are likely single-degenerate (Nugent et al. 2011) (where a white dwarf accretes mass from a large, perhaps main-sequence, companion star (Li et al. 2011a)), but studies of SN Ia rates point to the existence of two classes of progenitors (Mannucci et al. 2006). Furthermore, single-degenerate models have been ruled out for the supernova remnant SNR 0509-67 by the lack of an ex-companion star (Schaefer & Pagnotta 2012), and pre-explosion X-ray and optical data for SN2007on are compatible with a single-degenerate model (Voss & Nelemans 2008). Multiple progenitor channels would help to explain the observed variability within the type Ia category (Li et al. 2011b).

Within the more restricted subclass of so-called "normal" SNe Ia, the fundamental assumption underlying their use to measure expansion history is that they can be standardized so that their intrinsic magnitudes (after empirical corrections) are sufficiently homogeneous. This makes them into "standard candles," i.e., objects of almost uniform intrinsic luminosity (within ∼0.1 mag) that can be used to determine the distance–redshift relation. This relies on the empirical observation that intrinsic magnitudes are correlated with decay times of light curves: intrinsically brighter SNe Ia are slower to fade (Phillips 1993; Phillips et al. 1999). It also appears that fainter SNe Ia are redder in color (Riess et al. 1996). Therefore, multi-wavelength observations of light curves can be used to exploit this correlation and reduce the residual scatter in the intrinsic magnitude to typically ∼0.10–0.15 mag. Near-infrared LC data can significantly reduce residual scatter still further (Mandel et al. 2011), as does selecting SNe Ia in young star-forming environments (Kelly et al. 2015).

One of the most widely used frameworks for determining an estimate of the distance modulus from LC data is the SALT2 method (Guy et al. 2005, 2007), which derives color and stretch corrections for the magnitude from the LC fit, and then uses the corrected distance modulus to fit the underlying cosmological parameters. By contrast, the Multi-color Light-curve Shape (Riess et al. 1996; Jha et al. 2007) approach simultaneously infers the Phillips corrections and the cosmological parameters of interest, while explicitly modeling the dust absorption and reddening in the host galaxy. Recently, a fully Bayesian, hierarchical model approach to LC fitting has emerged (Mandel et al. 2009, 2011), but this so-called BAYESN algorithm has not yet been applied for inferring cosmological parameters.

As the size of the SN Ia sample grows, so does the importance of systematic errors relative to statistical errors, to the point where current measurements of the cosmological parameters (including properties of dark energy) are limited by systematics (Betoule et al. 2014). A better understanding of how the properties of SNe Ia correlate with their environment (such as the properties of their host galaxy) will help in improving their usage as standard candles.

In this paper, we introduce BAHAMAS (BAyesian HierArchical Modeling for the Analysis of Supernova cosmology), an extention of the method first introduced by March et al. (2011), and apply it to the SN Ia sample from the "Joint Light-curve Analysis" (JLA, Betoule et al. 2014). Betoule et al. (2014) reanalyzed 740 spectroscopically confirmed SNe Ia obtained by the SDSS-II and SNLS collaboration. March et al. (2011) demonstrated with simulated data that a Bayesian hierarchical model approach of the kind developed here has a reduced posterior uncertainty, smaller mean squared error, and better coverage properties than the standard approach (see also March et al. 2014; Karpenka 2015 for further detailed comparisons). More recently, Rubin et al. (2015) applied a similar method to analyze Union 2.1 data on SNe Ia, extending it to deal with the selection effect and non-Gaussian distribution. Nielsen et al. (2015) adopted the effective likelihood introduced in March et al. (2011) but interpreted the results in terms of profile likelihood (rather than posterior distributions), showing that the profile likelihood in the Ω_Λ, Ω_m plane obtained from JLA data is much wider than what is recovered with the usual χ^2 approach.

This paper re-evaluates the JLA data in the light of the principled statistical analysis made possible by BAHAMAS. As demonstrated in March et al. (2011), the standard χ^2 fitting is an approximation to the Bayesian result in a particular regime, which is usually violated by SALT2 outputs. Therefore we address the question of whether the cosmological constraints obtained from the standard analysis remain unchanged when using a principled likelihood function within a fully Bayesian analysis, as in BAHAMAS. We use our framework to test for evolution with redshift in the properties of SNe Ia, and in particular in their color correction. Finally, we investigate whether the residual scatter around the Hubble law can be further reduced by exploiting correlations between the intrinsic magnitudes of SNe Ia and their host galaxy mass.

This paper is organized as follows: in Section 2 we introduce our notation, the parameters of interest, and our Bayesian hierarchical model. In Section 3 we present results obtained when our approach is applied to the JLA sample; conclusions appear in Section 4. In Appendix A we review our statistical algorithms; in Appendix B we present the full posterior distributions, and in Appendix C we give details of the Gibbs-type samplers that we use to fit our Bayesian models.

2. BAHAMAS: BAYESIAN HIERARCHICAL MODELING FOR THE ANALYSIS OF SUPERNOVA COSMOLOGY

In this section, we review BAHAMAS, an extension of the method introduced by March et al. (2011) for estimating cosmological parameters using the peak magnitudes of SNe Ia adjusted for the stretch and color of their LCs via SALT2. We then discuss features of the model and methods that allow us to adjust for systematic errors, host galaxy mass, and a possible dependence of the color correction on redshift. We also provide a new estimate of the residual scatter in absolute magnitudes of SNe Ia. An outline of our statistical models and methods is presented here. Details of the statistical posterior distributions and the computational techniques we use to explore them appear in Appendix B.

2.1. Distance Modulus in an FRW Cosmology

Our overall modeling strategy leverages the homogeneity of the absolute magnitudes of SNe Ia to allow us to estimate their distance modulus from their apparent magnitudes and thereby estimate the underlying cosmological parameters that govern the relationship between distance modulus and redshift, z. Consider, for example, a sample of n SNe Ia with apparent B-band peak magnitudes, $m_i^\star$. The distance modulus in any passband, $\mu(z; \mathscr{C})$, is the difference between the apparent and the intrinsic magnitudes in that band. Ignoring measurement error for the moment, we can express this relationship statistically via the regression model

$$m_i^\star = \mu(z_i; \mathscr{C}) + M_i, \text{ for } i = 1, \ldots, n, \tag{1}$$

where $M_i \sim \mathcal{N}(M_0, \sigma_{\text{int}}^2)$ is the absolute magnitude of SN Ia i, with M_0 and σ_{int} the mean and intrinsic standard deviation of absolute magnitudes of SNe Ia in the underlying population.[4] Clearly the smaller σ_{int} the better we can estimate $\mu(z; \mathscr{C})$. In Section 2.2 we discuss the inclusion of correlates in Equation (1) that aim to reduce its residual variance, i.e., to make the SNe Ia better standard candles.

The distance modulus is given by

$$\mu(z; \mathscr{C}) = 25 + 5\log_{10}\frac{d_L(z; \mathscr{C})}{\text{Mpc}}, \tag{2}$$

where $\mathscr{C}$ represents a set of underlying cosmological parameters and $d_L(z; \mathscr{C})$ is the luminosity distance to redshift z. In the case of the ΛCDM cosmological model (based on a Friedman–Robertson–Walker (FRW) metric), the luminosity distance is

$$\begin{aligned} d_L(z; \mathscr{C}) = &\frac{c}{H_0}\frac{(1+z)}{\sqrt{|\Omega_\kappa|}}\text{sinn}_{\Omega_\kappa}\left\{\sqrt{|\Omega_\kappa|}\right. \\ &\times \int_0^z dz'[(1+z')^3\Omega_m + \Omega_{DE}(z') \\ &\left. + (1+z')^2\Omega_\kappa]^{-1/2}\right\}, \end{aligned} \tag{3}$$

[4] We use $\mathcal{N}(\mu, \Sigma)$ to denote a (multivariate) Gaussian distribution of mean μ and variance–covariance matrix Σ. For the one-dimensional case, Σ reduces to the variance, σ^2.

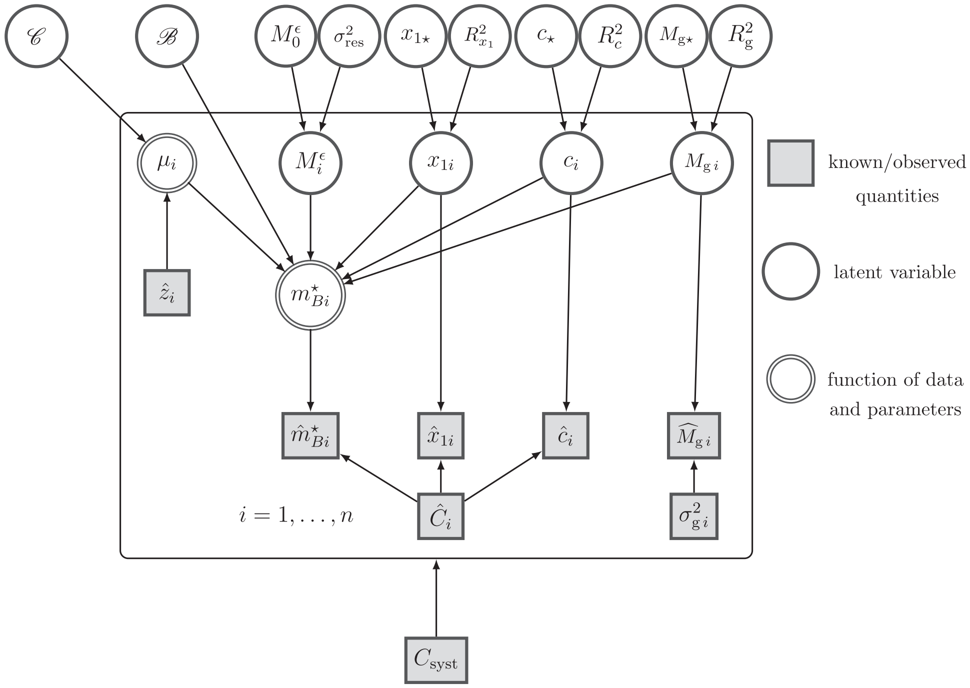

Figure 1. Graphical representation of BAHAMAS. The meaning of the symbols is given in Table 1.

where

$$\text{sinn}_{\Omega_\kappa}(x) = \begin{cases} x, & \text{if } \Omega_\kappa = 0 \\ \sin(x), & \text{if } \Omega_\kappa < 0 \\ \sinh(x), & \text{if } \Omega_\kappa > 0 \end{cases} \quad (4)$$

and $\mathscr{C} = \{\Omega_\kappa, \Omega_{\rm m}, H_0, w\}^T$, with Ω_κ the curvature parameter and $\Omega_{\rm m}$ the total (both baryonic and dark) matter density (in units of the critical density); c is the speed of light, and $H_0 = 100h$ km s^{-1} Mpc^{-1} is the Hubble parameter today, depending on the dimensionless quantity h. For a general dark-energy equation of state as a function of redshift, $w(z)$, we can express

$$\Omega_{\rm DE}(z) = \Omega_\Lambda \exp\left[3\int_0^z \frac{1 + w(x)}{1 + x} dx\right], \quad (5)$$

where Ω_Λ is the dark-energy density parameter. In our analyses, we assume either a flat universe (i.e., $\Omega_\kappa = 0$) with $w(z)$ equal to a constant other than -1 or a curved universe with a cosmological constant (i.e., $w(z) = -1$). In either case, $w(z) = w$ becomes a time-independent constant, and thus

$$\Omega_\Lambda = 1 - \Omega_\kappa - \Omega_{\rm m}. \quad (6)$$

2.2. SALT2 Output and Standardization of SNe Ia

2.2.1. Baseline Model

As described in Guy et al. (2007), the SALT2 fit of the multi-color LC observation of SN Ia i produces measured quantities: $\hat{z}_i$ is the measured heliocentric redshift, $\hat{m}^\star_{Bi}$ the measured B-band apparent magnitude, $\hat{x}_{1i}$ the measured stretch correction parameter, $\hat{c}_i$ the measured color correction parameter, and $\hat{C}_i$ a (3×3) variance–covariance matrix describing the measurement error of $\hat{m}^\star_{Bi}$, $\hat{x}_{1i}$, and $\hat{c}_i$. As shown in March et al. (2011), accounting for observational error in spectroscopically determined redshifts does not lead to any appreciable difference in the results. Thus, after correcting for the translation from heliocentric redshift to the frame of reference of the cosmic microwave background (CMB), we ignore measurement error in the observed redshift and set $\hat{z}_i = z_i$ throughout. Each $\hat{C}_i$ is treated as a known constant, and we denote the SALT2 data by

$$\widehat{\mathscr{D}}_i = \{\hat{m}^\star_{Bi}, \hat{x}_{1i}, \hat{c}_i\}^T, \text{ for } i = 1, \ldots, n. \quad (7)$$

Here we review our Baseline Model that was first introduced by March et al. (2011); extensions appear in Sections 2.2.2 and 2.3. We model $\widehat{\mathscr{D}} = \{\widehat{\mathscr{D}}_1^T, \ldots, \widehat{\mathscr{D}}_n^T\}^T$ via a Bayesian hierarchical model (Kelly 2007); see Figure 1. At the observation level, we model the measured SALT2 fits as independent Gaussian variables centered at their true values,

$$\begin{pmatrix} \hat{m}^\star_{Bi} \\ \hat{x}_{1i} \\ \hat{c}_i \end{pmatrix} \overset{\text{indep}}{\sim} \mathcal{N}\left[\begin{pmatrix} m^\star_{Bi} \\ x_{1i} \\ c_i \end{pmatrix}, \hat{C}_i\right], \text{ for } i = 1, \ldots, n. \quad (8)$$

The true (but unobserved) values, $m^\star_{Bi}$, x_{1i}, and c_i, are treated as latent variables, with x_{1i} and c_i used to predict the intrinsic (absolute) magnitude M_i via the linear regression

$$M_i = -\alpha x_{1i} + \beta c_i + M_i^\epsilon, \quad (9)$$

where $M_i^\epsilon \sim \mathcal{N}(M_0^\epsilon, \sigma^2_{\rm res})$. Here x_{1i} and c_i represent the Phillips stretch and color corrections, respectively, whose predictive strength is controlled by the unknown parameters, α and β, which must be inferred from $\widehat{\mathscr{D}}$; M_i appearing in Equation (1) is the physical absolute magnitude of SN Ia i and M_i^ϵ is the empirically corrected absolute magnitude, after application of the Phillips relations. Substituting Equation (9) into (1) yields

$$m^\star_{Bi} = \mu_i(\hat{z}_i, \mathscr{C}) - \alpha x_{1i} + \beta c_i + M_i^\epsilon. \quad (10)$$

From a statistical point of view Equation (9) is a linear regression model with residuals M_i^ϵ. In principle, including the stretch and color corrections in Equations (9) and (10) should reduce the residual variance, i.e., $\sigma_{\rm res}^2 \leqslant \sigma_{\rm int}^2$, and improve the precision of the estimates of $\mathscr{C}$.[5] In Section 2.3 we investigate whether introducing either host galaxy mass or an interaction[6] between redshift and the color correction as an additional correlated variable in Equation (9) can further reduce the residual variance and increase the precision of the estimate of $\mathscr{C}$.

The population distributions of the latent variables M_i^ϵ, x_{1i}, and c_i are modeled as Gaussian[7], with unknown hyperparameters controlling the mean and variance of each population:

$$M_i^\epsilon | M_0^\epsilon, \sigma_{\rm res} \sim \mathcal{N}\left(M_0^\epsilon, \sigma_{\rm res}^2\right), \quad (11)$$

$$x_{1i} | x_{1\star}, R_{x_1} \sim \mathcal{N}\left(x_{1\star}, R_{x_1}^2\right), \quad (12)$$

$$c_i | c_\star, R_c \sim \mathcal{N}\left(c_\star, R_c^2\right). \quad (13)$$

The distribution in Equation (11) is the model for the residuals in Equation (9).

The prior distributions used for the model parameters are given in Table 1 (along with those for parameters introduced in extensions to the model in Section 2.3). We adopt non-informative proper prior distributions for α, β, and the parameters in $\mathscr{C}$. The value of the Hubble parameter is fixed at $H_0 = 67.3$ km s^{-1} Mpc^{-1} from Planck.[8] Among the population-level parameters, the choice of prior distribution for $\sigma_{\rm res}^2$ is the most subtle. The simple choice of a log-uniform prior, as adopted in March et al. (2011), requires specification of arbitrary bounds to make it proper. Because this might lead to difficulties in interpreting the posterior distribution, we instead adopt a proper inverse Gamma[9] prior distribution, $\sigma_{\rm res}^2 \sim \textsc{InvGamma}(0.003, 0.003)$. We perform a sensitivity analysis for the choice of scale for this distribution and demonstrate that our results (including the posterior distribution of $\sigma_{\rm res}$) are robust to this choice; see Figure 3.

2.2.2. Systematics Covariance Matrix and Selection Effects

In the Baseline Model described in Section 2.2.1, we assume that the SALT2 measurements for each SN Ia are conditionally independent (given their means and variances, see Equation (8)), i.e., the $(3n \times 3n)$ variance–covariance matrix $C_{\rm stat} \equiv {\rm diag}(\hat{C}_1, \ldots, \hat{C}_n)$ is block-diagonal. Betoule et al. (2014) derived a systematic variance–covariance matrix, $C_{\rm syst}$, with correlations among the SNe Ia. The systematic covariance matrix includes contributions from calibration, model uncertainty, bias correction, host, dust, peculiar velocities, and contamination. We account for these systematics by replacing the matrix $C_{\rm stat}$ with $\Sigma_{\hat{\mathscr{D}}} = C_{\rm stat} + C_{\rm syst}$ in the full posterior distribution; see Appendix B.

Betoule et al. (2014) used SNANA simulations to model observational selection effects and corrected for them by shifting the value of $\hat{m}_{Bi}^\star$ accordingly. We adopt the bias-corrected values of $\hat{m}_{Bi}^\star$ and thus do not need to separately account for selection effects. A fully Bayesian approach to forward-modeling of such effects appears in Rubin et al. (2015).

Table 1
Summary of the Parameters, Notation, and Prior Distributions Used in Our Hierarchical Model

Parameter	Notation and Prior Distribution
Cosmological Parameters	
Matter density parameter	$\Omega_{\rm m} \sim \textsc{Uniform}(0, 2)$
Cosmological constant density parameter	$\Omega_\Lambda \sim \textsc{Uniform}(0, 2)$
Dark-energy equation of state	$w \sim \textsc{Uniform}(-2, 0)$
Hubble parameter	$H_0 = 67.3$ km s^{-1} Mpc^{-1}
Covariates	
Coefficient of stretch covariate	$\alpha \sim \textsc{Uniform}(0, 1)$
Coefficient of color covariate	β (or β_0) $\sim \textsc{Uniform}(0, 4)$
Coefficient of interaction of color correction and z	$\beta_1 \sim \textsc{Uniform}(-4, 4)$
Jump in coefficient of color covariate	$\Delta\beta \sim \textsc{Uniform}(-1.5, 1.5)$
Redshift of jump in color covariate	$z_t \sim \textsc{Uniform}(0.2, 1)$
Coefficient of host galaxy mass covariate	$\gamma \sim \textsc{Uniform}(-4, 4)$
Population-level Distributions	
Mean of absolute magnitude	$M_0^\epsilon \sim \mathcal{N}(-19.3, 2^2)$
Residual scatter after corrections	$\sigma_{\rm res}^2 \sim \textsc{InvGamma}(0.003, 0.003)$
Mean of absolute magnitude, low galaxy mass	$M_0^{\rm lo} \sim \mathcal{N}(-19.3, 2^2)$
SD of absolute magnitude, low galaxy mass	${\sigma_{\rm res}^{\rm lo}}^2 \sim \textsc{InvGamma}(0.003, 0.003)$
Mean of absolute magnitude, high galaxy mass	$M_0^{\rm hi} \sim \mathcal{N}(-19.3, 2^2)$
SD of absolute magnitude, high galaxy mass	${\sigma_{\rm res}^{\rm hi}}^2 \sim \textsc{InvGamma}(0.003, 0.003)$
Mean of stretch	$x_{1\star} \sim \mathcal{N}(0, 10^2)$
SD of stretch	$R_{x_1} \sim \textsc{LogUniform}(-5, 2)$
Mean of color	$c_\star \sim \mathcal{N}(0, 1^2)$
SD of color	$R_c \sim \textsc{LogUniform}(-5, 2)$
Mean of host galaxy mass	$M_{g\star} \sim \mathcal{N}(10, 100^2)$
SD of host galaxy mass	$R_g \sim \textsc{LogUniform}(-5, 2)$

Note. These include parameters in the Baseline Model described in Sections 2.1–2.2 and its extensions described in Section 2.3. "SD" stands for "standard deviation."

2.3. Generalizing the Phillips Corrections

The advantage of the Phillips corrections is that they are expected to reduce the residual variance in Equation (10) and thus increase the precision in the estimates of $\mathscr{C}$. Introducing additional correlates may further improve precision. In the

[5] This intuition stems from standard linear regression where the dependent variables (here the $m_{Bi}^\star$) and independent variables (here the x_{1i} and c_i) are observed directly. The situation is more complicated when these variables are observed with error.

[6] In statistical terms, an *interaction* between two variables means that the effect of one variable depends on the values of the other. In Section 2.3 we allow the effect of the color correction to vary with redshift.

[7] We assume a single underlying population, but it would be simple to extend our model to multiple populations by drawing M_i^ϵ from a mixture of Gaussians, for example to account for different progenitor scenarios, or contamination from sources other than SNe Ia.

[8] The Hubble parameter is perfectly degenerate with the mean absolute magnitude M_0, hence data from SNe Ia constrain only the degenerate combination $M_0 - 5\log_{10}h$. Therefore changing the value of h amounts to a shift in the mean absolute magnitude.

[9] We parameterize the inverse Gamma distribution so that $X \sim \textsc{InvGamma}(u, v)$ means that $2v/X$ follows a χ^2 distribution with $2u$ degrees of freedom.

context of BAHAMAS, it is straightforward to generalize the Phillips corrections to include additional covariates. To formalize this, we replace x_{1i} and c_i in Equation (9) with a set of p covariates and substitute into Equation (1) to obtain

$$m_{Bi}^{\star} = \mu_i(\hat{z}_i, \mathscr{C}) + X_i^T \mathscr{B} + M_i^{\epsilon}, \tag{14}$$

where X_i is a $(p \times 1)$ vector of covariates and $\mathscr{B}$ is a $(p \times 1)$ vector of regression coefficients. The usual case, given in Equation (10), is a special case of Equation (14) in which only the stretch and color covariates are included ($p = 2$), and it can be recovered by setting $X_i = \{x_{1i}, c_i\}^T$ and $\mathscr{B} = \{-\alpha, \beta\}^T$. If the covariate vector depends nonlinearly on a set of parameters τ, Equation (14) can be further generalized to

$$m_{Bi}^{\star} = \mu_i(\hat{z}_i, \mathscr{C}) + X_i(\tau)^T \mathscr{B} + M_i^{\epsilon}. \tag{15}$$

Equation (15) allows for both linear and nonlinear covariate adjustment.

We consider various instances of Equation (15). First, we investigate the effect of the environment by including the host galaxy mass as a covariate in the correction. The host mass is a (relatively easy to measure) proxy for more fundamental changes in the environment, such as evolution of metallicity. Second, we are interested in testing for possible redshift dependence of the color correction. This could have a physical origin (e.g., dust environments in a high-redshift galaxy being different) or be a reflection of systematic differences between low- and high-redshift surveys.

Future work will aim at investigating the dependence on environmental properties, such as star formation rates and metallicities, a topic of active investigation (Childress et al. 2013a; Rigault et al. 2013, 2015; Jones et al. 2015; Kelly et al. 2015).

2.3.1. Dependence on Host Galaxy Mass

There is strong evidence that the absolute magnitude (after corrections) of SNe Ia correlates with host galaxy mass (e.g., Sullivan et al. 2006; Meyers et al. 2012). Current results indicate that more massive galaxies ($\log_{10}(M/M_{\odot}) > 10$) host brighter SNe Ia, with their average absolute magnitude being of order ∼0.1 mag smaller than in less massive hosts (Kelly et al. 2010; Sullivan et al. 2010; Campbell et al. 2016). This could be a reflection of dust, age, and/or metallicity in the progenitor systems (Childress et al. 2013b).

We investigate three formulations that incorporate host galaxy mass as a covariate in Equation (15) and study how they affect inference for $\mathscr{C}$. In particular, we consider models that (i) divide the SNe Ia into two populations using a hard threshold for host galaxy mass ("Hard Classification Model"), (ii) divide the SNe Ia into two populations using soft probabilistic classification ("Soft Classification Model"), and (iii) adjust for host galaxy mass as a covariate in the regression, analogously to the stretch and color corrections ("Covariate Adjustment Model"). Specifically, we model the observed host galaxy masses (on the $\log_{10}$ scale) as

$$\widehat{M}_{g\,i} \overset{\text{indep}}{\sim} \mathcal{N}(M_{g\,i}, \sigma_{g\,i}^2), \text{ for } i = 1, \ldots, n, \tag{16}$$

where $M_{g\,i}$ is the (true) host galaxy mass of SN Ia i (in $\log_{10}$ solar masses) and $\sigma_{g\,i}$ is the (known) standard deviation of measurement error.

In the "Hard Classification Model," we divide the SNe Ia into two classes using the observed mass: high host galaxy mass if $\widehat{M}_{g\,i} \geqslant 10$ and low host galaxy mass if $\widehat{M}_{g\,i} < 10$. (In this way, we ignore measurement errors in $\widehat{M}_{g\,i}$.) The two classes are allowed to have their own population-level values for the mean absolute magnitude of SNe Ia and residual standard deviation, i.e., $(M_0^{\text{hi}}, \sigma_{\text{res}}^{\text{hi}})$ for high-mass hosts and $(M_0^{\text{lo}}, \sigma_{\text{res}}^{\text{lo}})$ for low-mass hosts. Common values are used for α and β (and of course for $\mathscr{C}$) for both classes. We do not assume a redshift dependence for the color correction. We fix the classification of host galaxy mass at 10^{10} solar masses, analogous to the location of the step function used for the host galaxy mass by Betoule et al. (2014) to enable a direct comparison with their results.

The "Soft Classification Model" is identical to the Hard Classification Model except that measurement errors in the observed masses are accounted for by probabilistically classifying each SN Ia; these errors can be quite significant. Specifically, we let Z_i be an indicator variable that equals one for high host galaxy masses and equals zero for low host galaxy masses, that is,

$$Z_i = \begin{cases} 0, & \text{if } M_{g\,i} < 10 \\ 1, & \text{if } M_{g\,i} \geqslant 10. \end{cases} \tag{17}$$

We treat $\{Z_1, \ldots, Z_n\}^T$ as a vector of unknown latent variables that are estimated along with the other model parameters and latent variables via Bayesian model fitting. This requires specification of a prior distribution on each $M_{g\,i}$. We choose a flat prior so that $M_{g\,i} | \widehat{M}_{g\,i} \overset{\text{indep}}{\sim} \mathcal{N}(\widehat{M}_{g\,i}, \sigma_{g\,i}^2)$; details appear in Appendix B.

The "Covariate Adjustment Model" introduces $M_{g\,i}$ as a covariate in the regression in Equation (14) rather than classifying the SNe Ia by galaxy mass. In particular, we use Equation (14), but with $p = 3$, $X_i = \{x_{1i}, c_i, M_{g\,i}\}^T$, and $\mathscr{B} = \{-\alpha, \beta, \gamma\}^T$ with $\mathscr{B}$ being Bayesianly estimated from the data. The population model for the latent variables M_i^{ϵ}, x_{1i}, and c_i given in Equations (11)–(13) is also expanded to include host galaxy mass:

$$M_{g\,i} | M_{g\star}, R_g \sim \mathcal{N}(M_{g\star}, R_g^2), \tag{18}$$

where $M_{g\star}$ and R_g are hyperparameters analogous to, e.g., $x_{1\star}$ and R_{x_1}; their prior distributions are given in Table 1.

2.3.2. Redshift Evolution of the Color Correction

The SALT2 color correction gives the offset with respect to the average color at maximum B-band luminosity, $c_i = (B - V)_i - \langle B - V \rangle$. This time-independent color variation encompasses both intrinsic color differences and those due to dust in the host galaxy. It is possible that the color correction varies with redshift, as a consequence of evolution of the progenitor and/or changes in the environment, for example, variation in the dust composition with galactic evolution (Childress et al. 2013b). Redshift-dependent dust extinction can lead to biased estimates of cosmological parameters (Menard et al. 2010a, 2010b). This is not captured by the SALT2 fits, since they use a training sample that is distributed over a large redshift range ($0.002 \leqslant z \lesssim 1$) (Guy et al. 2007), and thus the color correction to the training sample is averaged across redshift. It is therefore important to check for a redshift dependence in the color correction by allowing β, which

Table 2
Summary of Extensions to the Baseline Model

Models that adjust for host galaxy mass	
Hard Classification	$(M_0^\epsilon, \sigma_{\rm res})$ split for low/high host galaxy mass at $\widehat{M}_{g\,i} = 10$.
Soft Classification	$(M_0^\epsilon, \sigma_{\rm res})$ split for low/high host galaxy mass at $M_{g\,i} = 10$.
Covariate Adjustment	Host galaxy mass included in linear regression with coefficient γ, see Equation (14).
Models that allow the color adjustment to depend on redshift	
z-linear Color Correction	Color correction given by $\beta_0 + \beta_1 z$, see Equation (19).
z-jump Color Correction	Color correction changes smoothly by $\Delta\beta$ near $z = z_t$, see Equation (20).

controls the amplitude of the linear correction to the distance modulus, to vary with z.

We consider two phenomenological models that allow the color correction to depend on z. In the first, the dependence is linear: we replace the constant β in Equation (10) with the z-dependent $\beta_0 + \beta_1 \hat{z}_i$. This is expressed by setting $X_i = \{x_{1i}, c_i, c_i\hat{z}_i\}^T$ and $\mathscr{B} = \{-\alpha, \beta_0, \beta_1\}^T$ in Equation (14), leading to

$$M_i = -\alpha x_{1i} + \beta_0 c_i + \beta_1 c_i \hat{z}_i + M_i^\epsilon. \tag{19}$$

We refer to this as the "z-linear Color Correction Model."

The second model allows for a sharp transition from a high-redshift to a low-redshift regime: we replace the constant β in Equation (10) with $\beta_0 + \Delta\beta\left(\frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\hat{z}_i - z_t}{0.01}\right)\right)$, where β_0, $\Delta\beta$, and z_t are parameters. This can be viewed as a smoothed step function in that it approaches β_0 as $z \to 0$ and approaches $\beta_0 + \Delta\beta$ as $z \to \infty$, with a smooth monotonic local transition centered at $z = z_t$. Substituting into Equation (9), we have

$$M_i = -\alpha x_{1i} + \beta_0 c_i + \Delta\beta\left(\frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\hat{z}_i - z_t}{0.01}\right)\right)c_i + M_i^\epsilon, \tag{20}$$

where the covariate associated with $\Delta\beta$ depends nonlinearly on z_t as described in Equation (15). We refer to this model as the "z-jump Color Correction Model."

The several model extensions we consider are summarized in Table 2.

2.4. Posterior Sampling

To significantly reduce the dimension of the parameter space under the Baseline Model, March et al. (2011) marginalized out the $3n$ latent variables, $\{M_i^\epsilon, x_{1i}, c_i\}$, from the posterior distribution. This relies on the Gaussian population distributions for analytic tractability. A consequence is that the posterior distributions of the latent variables of each SN Ia are inaccessible.

BAHAMAS improves on March et al. (2011) by using a Gibbs-type sampler to sample from the joint posterior distribution of the parameters and latent variables. This has the advantage that we can present object-by-object posterior distributions for the values of latent color, stretch, and intrinsic magnitude. These can also be mapped onto posterior distributions for the residuals of the Hubble diagram; see Figure 4.

Furthermore, BAHAMAS does not require Gaussian population distributions, as the posterior sampling is fully numerical; Rubin et al. (2015) took a similar approach. Although we do not use non-Gaussian distributions here, BAHAMAS opens the door to a fully Bayesian treatment of non-Gaussianities and selection effects. This will be investigated in a future work.

We present the algorithmic details of our Gibbs-type sampler in Appendix C. We have cross-checked the results obtained with the Gibbs-type sampler with those obtained with the Metropolis–Hastings (MH) sampler of March et al. (2011) and with the MultiNest sampler (a nested sampling algorithm, see Feroz et al. 2009). This comparison is carried out for the Baseline Model as well as for the extensions in Table 2. The main difference is that the Gibbs sampler directly simulates the latent variables while the other two algorithms do not. The marginal distributions obtainable with the latter two methods match within the numerical sampling margin of error with the output from the Gibbs-type sampler. We use the stopping criterion of Gelman & Rubin (1992) and require their scale reduction factor, $\hat{R}$, to be less than 1.1 for all the components of $\mathscr{C}$ and $\mathscr{B}$. This leads to a chain of typically ~3300 samples, with an effective sample size[10] of around 200 for $\mathscr{C}$ components, and 400 for $\mathscr{B}$ components. This requires a CPU time of order 2.0×10^5 s, where the cost of evaluating a single likelihood is of the order of 5–10 s on a single CPU.

3. RESULTS

Here we present the BAHAMAS fits to the JLA data, as well as in combination with *Planck* CMB temperature data, complemented by WMAP9 polarization data (Planck Collaboration et al. 2015).

3.1. Baseline Model

We begin by displaying in Figure 2 the 1D and 2D marginal posterior distributions for the cosmological parameters, and color and stretch correction parameters from the JLA sample of SNe Ia analyzed with BAHAMAS (black contours). We also show the combination with *Planck* CMB data, which we obtain via importance sampling (red contours). We consider either a universe containing a cosmological constant, $w = -1$ (ΛCDM), or a flat universe with a dark-energy component with redshift-independent $w \neq -1$ (wCDM).

Table 3 (ΛCDM) and Table 4 (wCDM) report the corresponding marginal posterior credible intervals. For the

[10] The effective sample size of the parameter ψ is defined as

$$\text{ESS}(\psi) = \frac{T}{1 + 2\sum_{t=1}^{\infty}\rho_t(\psi)}, \tag{21}$$

where T is the total posterior sample size and $\rho_t(\psi)$ is the lag-t autocorrelation of ψ in the MCMC sample. ESS(ψ) approximates the size of an independent posterior sample that would be required to obtain the same Monte Carlo variance of the posterior mean of ψ; see Kass et al. (1998) and Liu (2001). ESS(ψ) is an indicator of how well the MCMC chain for ψ mixes; ESS(ψ) is necessarily less than T and larger values of ESS(ψ) are preferred.

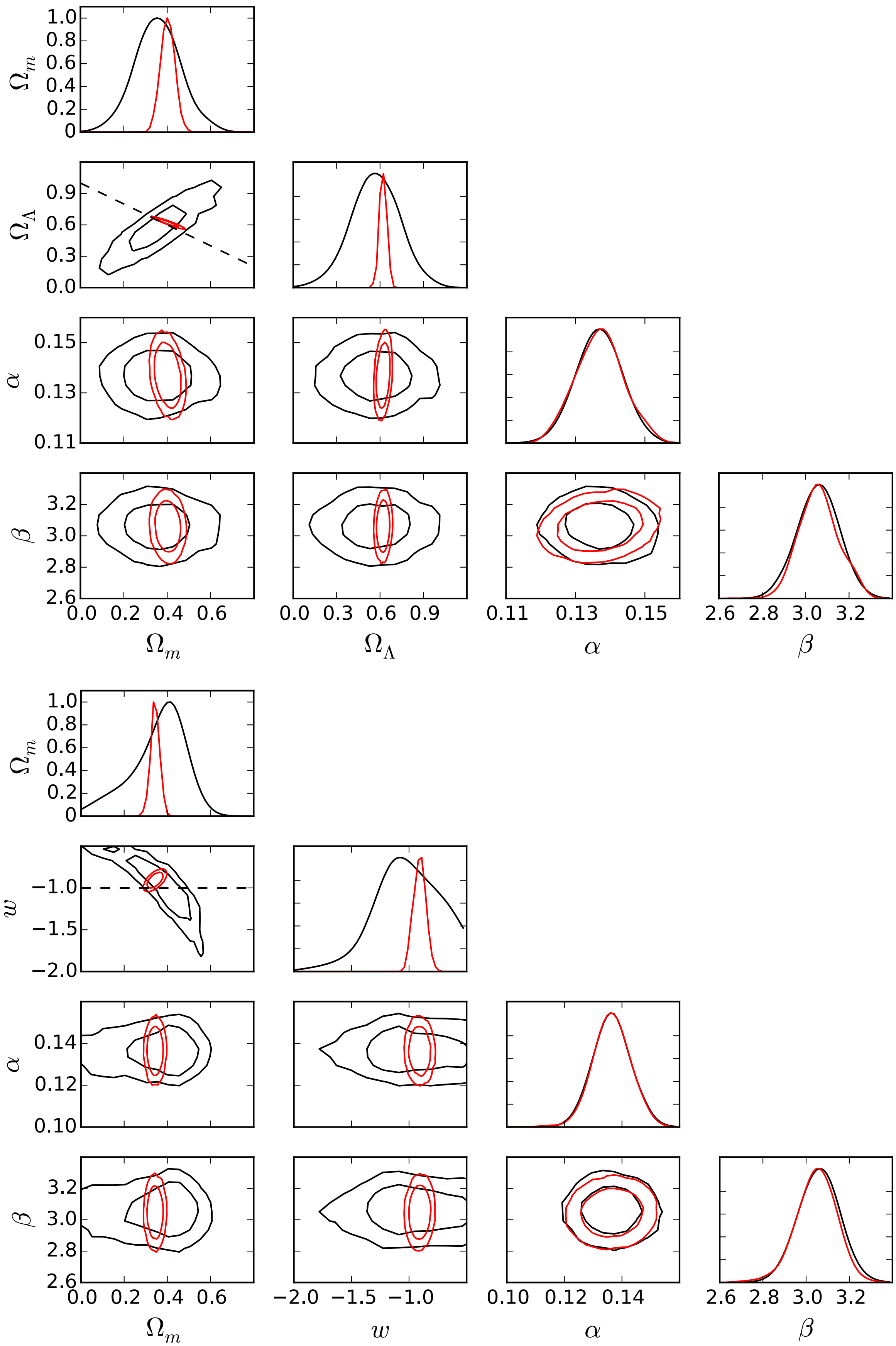

Figure 2. 1D and 2D marginal posterior distributions for the cosmological parameters, and the color and stretch correction parameters under the Baseline Model. Black (red) contours show 68% and 95% highest posterior density regions for JLA SNe Ia data only (JLA combined with *Planck*). The top (bottom) panels display results for the ΛCDM (wCDM) model.

Table 3
Marginalized Posterior Constraints on Cosmological and SNe Ia Correction Parameters for the ΛCDM Model, Assuming $H_0 = 67.3$ km s^{-1} Mpc^{-1}

	JLA SNe Ia Only			JLA SNe Ia + *Planck* 2015		
	Baseline	z-linear Color Corr.	z-jump Color Corr.	Baseline	z-linear Color Corr.	z-jump Color Corr.
			Baseline Model Parameters			
Ω_{m}	0.340 ± 0.101	0.362 ± 0.094	0.429 ± 0.097	0.399 ± 0.027	0.420 ± 0.031	0.425 ± 0.025
Ω_Λ	0.542 ± 0.157	0.557 ± 0.145	0.632 ± 0.155	0.625 ± 0.020	0.609 ± 0.025	0.604 ± 0.019
Ω_κ	0.119 ± 0.249	0.081 ± 0.230	−0.061 ± 0.244	−0.024 ± 0.008	−0.028 ± 0.008	−0.029 ± 0.007
α	0.137 ± 0.006	0.136 ± 0.006	0.136 ± 0.006	0.137 ± 0.006	0.135 ± 0.007	0.136 ± 0.006
β	3.058 ± 0.085	n/a	n/a	3.068 ± 0.097	n/a	n/a
			Redshift Evolution of Color Correction Parameters			
β_0	n/a	3.211 ± 0.120	3.137 ± 0.092	n/a	3.219 ± 0.119	3.136 ± 0.096
β_1	n/a	−0.622 ± 0.342	n/a	n/a	−0.732 ± 0.360	n/a
$\Delta\beta$	n/a	n/a	−1.120 ± 0.240	n/a	n/a	−1.145 ± 0.243
z_t	n/a	n/a	0.662 ± 0.055	n/a	n/a	0.670 ± 0.056
			Intrinsic Magnitude and Residual Dispersion Parameters			
M_0^ϵ	−19.140 ± 0.022	−19.140 ± 0.020	−19.144 ± 0.021	−19.140 ± 0.018	−19.138 ± 0.018	−19.140 ± 0.016
σ_{res}	0.104 ± 0.005	0.104 ± 0.005	0.103 ± 0.005	0.105 ± 0.005	0.105 ± 0.004	0.103 ± 0.005

Table 4
As in Table 3, but for wCDM

	JLA SNe Ia Only			JLA SNe Ia + *Planck* 2015		
	Baseline	z-linear Color Corr.	z-jump Color Corr.	Baseline	z-linear Color Corr.	z-jump Color Corr.
Ω_{m}	0.355 ± 0.117	0.366 ± 0.119	0.422 ± 0.097	0.343 ± 0.019	0.349 ± 0.015	0.353 ± 0.018
Ω_Λ	0.645 ± 0.117	0.634 ± 0.119	0.578 ± 0.097	0.657 ± 0.019	0.651 ± 0.015	0.647 ± 0.018
w	$-0.995^{+0.418}_{-0.275}$	$-1.022^{+0.425}_{-0.227}$	$-1.145^{+0.394}_{-0.293}$	−0.910 ± 0.045	−0.905 ± 0.050	−0.883 ± 0.043
α	0.136 ± 0.006	0.136 ± 0.006	0.136 ± 0.006	0.136 ± 0.006	0.137 ± 0.006	0.136 ± 0.005
β	3.060 ± 0.088	n/a	n/a	3.047 ± 0.087	n/a	n/a
β_0	n/a	3.206 ± 0.358	3.137 ± 0.090	n/a	3.199 ± 0.109	3.128 ± 0.082
β_1	n/a	−0.629 ± 0.358	n/a	n/a	−0.603 ± 0.320	n/a
$\Delta\beta$	n/a	n/a	−1.116 ± 0.240	n/a	n/a	−1.083 ± 0.237
z_t	n/a	n/a	0.661 ± 0.055	n/a	n/a	0.655 ± 0.055
M_0^ϵ	−19.146 ± 0.024	−19.142 ± 0.022	−19.145 ± 0.021	−19.148 ± 0.024	−19.144 ± 0.020	−19.143 ± 0.020
σ_{res}	0.103 ± 0.005	0.104 ± 0.005	0.103 ± 0.005	0.103 ± 0.007	0.104 ± 0.005	0.102 ± 0.005

$w = -1$ case (i.e., ΛCDM), we find $\Omega_{\mathrm{m}} = 0.340 \pm 0.101$ and $\Omega_\kappa = 0.119 \pm 0.249$ (JLA alone). Including *Planck* data results in[11] $\Omega_{\mathrm{m}} = 0.399 \pm 0.027$, a significantly higher value of the matter content than reported in the standard analysis. (More detailed comparisons are given below.) The curvature parameter is $\Omega_\kappa = -0.024 \pm 0.008$, excluding a flat universe, $\Omega_\kappa = 0$, at the $\sim 3\sigma$ level. For the case of a flat universe (i.e., wCDM, Table 4), we find from JLA and *Planck*, $\Omega_{\mathrm{m}} = 0.343 \pm 0.019$ and $w = -0.910 \pm 0.045$. The contours of the posterior distribution of Ω_{m} and Ω_Λ based on the JLA data only are similar to those obtained by Nielsen et al. (2015). These authors marginalized latent variables out of their effective likelihood, in an approach similar to our own, although with a number of detailed differences.[12] In particular, the 1σ (marginal posterior) contour obtained with BAHAMAS overlaps closely with the 1σ (profile likelihood) contour in Nielsen et al. (2015), while the 2σ contour from BAHAMAS shows a degree of asymmetry that is not present in Nielsen et al. (2015). (Recall that the analysis of Nielsen et al. (2015) relies on approximating the confidence regions using Gaussians, while the numerical sampling of BAHAMAS does not.)

The residual intrinsic dispersion is in all cases close to $\sigma_{\mathrm{res}} = 0.104 \pm 0.005$. This value is to be understood as the average residual scatter in the (post-correction) intrinsic magnitudes across all surveys that make up the JLA data set.

[11] We summarize marginal posterior distributions with their posterior mean and approximate 68% (1σ) posterior credible intervals. We report highest posterior density (HPD) intervals, which are the shortest intervals that capture 68% of the posterior probability. In most cases, the marginal posterior distributions are symmetric and approximately Gaussian, in which case the reported error bar is the posterior standard deviation. The exceptions are the intervals reported for w, which are reported with asymmetric positive and negative errors due to the non-Gaussian shape of the posterior distribution.

[12] Nielsen et al. (2015) adopted implicit uniform priors on the population variances, as well as on σ_{int}. They also maximized the likelihood to obtain confidence intervals on cosmological parameters (after marginalization of the latent variables), rather than integrating the posterior to obtain marginalized credible regions (as in this work). Because BAHAMAS is a nonlinear, non-Gaussian model there is no reason to expect a priori that our results ought to be similar to those obtained by Nielsen et al. (2015).

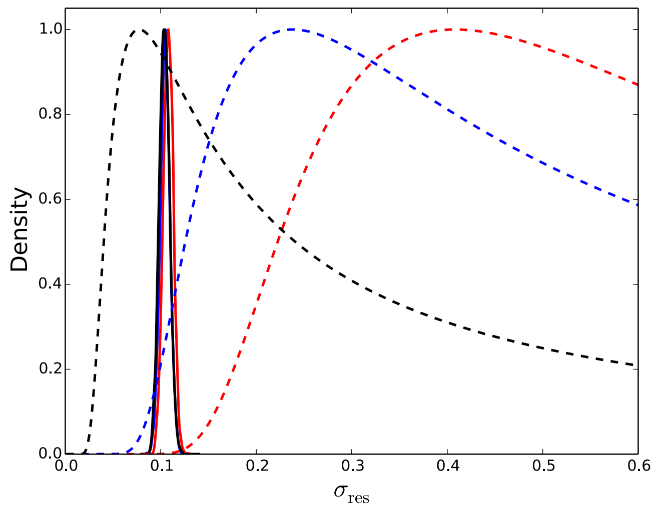

Figure 3. Robustness of the posterior distribution for $\sigma_{\rm res}$ (solid lines) with respect to three different prior specifications (dashed lines). Black: $\sigma_{\rm res}^2 \sim \textsc{InvGamma}(0.003, 0.003)$; blue: $\sigma_{\rm res}^2 \sim \textsc{InvGamma}(0.03, 0.03)$; red: $\sigma_{\rm res}^2 \sim \textsc{InvGamma}(0.1, 0.1)$. Since the three posterior distributions are very similar, we conclude that the posterior distribution of $\sigma_{\rm res}$ is largely insensitive to its prior specification (assuming ΛCDM). Densities have been normalized to their peak for ease of comparison. In the rest of this paper, we use $\sigma_{\rm res}^2 \sim \textsc{InvGamma}(0.003, 0.003)$.

This value should be compared with the parameter $\sigma_{\rm coh}$ in Betoule et al. (2014) (which in that work has an approximately equivalent meaning to our $\sigma_{\rm res}$), ranging from 0.08 (for SNLS) to 0.12 (low-z). It would be easy to extend our analysis to allow for a different value of $\sigma_{\rm res}$ for each of the data sets (SNLS, SDSS, low-z, and *Hubble Space Telescope* (*HST*)) comprising JLA.

The posterior constraints on the residual intrinsic dispersion are, in principle, sensitive to the choice of scale in its inverse Gamma prior distribution. To test the robustness of our posterior inference on $\sigma_{\rm res}$ with respect to its prior specification, we have compared the posterior distributions obtained with three very different prior distributions; each is an inverse Gamma distribution, but with parameters $u = v = 0.003, 0.03, 0.1$. The resulting posterior distributions (alongside their prior distributions) are shown in Figure 3. Despite the widely differing prior distributions, the posteriors are nearly identical, demonstrating the prior-independence of our result. We have verified that all constraints on the other parameters are similarly insensitive to the choice of prior for $\sigma_{\rm res}$.

BAHAMAS allows us to compute the posterior distribution for all latent variables, and for the Hubble residuals. It is instructive to compare the posterior distribution to the standard best-fit estimate to illustrate the phenomenon of "shrinkage": the hierarchical regression structure of the Bayesian model allows estimators to "borrow strength" across the SNe Ia and thus reduces their residual scatter around the regression plane.

We illustrate the shrinkage effect using the Baseline Model. We divide the SNe Ia into four bins using the quartiles of $\hat{x}_1$. For each bin, in the four panels of the first row of Figure 4, we plot in blue $\hat{M}_i \equiv \hat{m}^{\star}_{Bi} - \mu_i(\hat{z}_i, \overline{\mathscr{C}})$ versus $\hat{c}_i$. Here, $\overline{\mathscr{C}}$ is the posterior mean of the cosmological parameters, and $\hat{M}_i$ is a plug-in estimate of the intrinsic magnitude of SN Ia i before corrections. This is equivalent to the standard "best-fit" estimate of the intrinsic magnitude. In red we plot the posterior means, i.e., $\bar{m}^{\star}_{Bi} - \mu_i(\hat{z}_i, \overline{\mathscr{C}})$ versus $\bar{c}_i$. The regression line in each bin (black) has slope $\bar{\beta}$ and intercept $\bar{M}^{\varepsilon}_0 - \bar{\alpha}x_1$, where the bar represents the average with respect to the posterior distribution while x_1 is the mean of $\hat{x}_1$ in that bin.

In each bin, we observe the expected positive correlation between intrinsic magnitude and color (top panels), and negative correlation between intrinsic magnitude and stretch (bottom panels). The most striking feature is that the posterior estimates are dramatically shrunk toward the regression line, when compared with the plug-in estimates. This is because BAHAMAS accounts for the uncertainty in the measured values of $\{\hat{m}^{\star}_{Bi}, \hat{x}_i, \hat{c}_i\}$, and adjusts their fitted values (i.e., their posterior distributions) by "shrinking" them toward their fitted population means and the fitted regression line.

3.2. Including Corrections for Host Galaxy Mass

We now investigate the impact of including information on the host galaxy mass. Marginalized posterior constraints on our model parameters when the host galaxy mass is used as a predictor or a covariate are reported in Tables 5 (ΛCDM) and 6 (wCDM). The posterior distributions are shown in Figure 5, where they are compared with the case when no mass correction is used.

The Hard Classification Model matches exactly the procedure to correct for host galaxy mass adopted in Betoule et al. (2014), hence our results are directly comparable. The only difference is the statistical method adopted in inferring the cosmological parameters from the SALT2 fits. For the matter density parameter (assuming ΛCDM and using JLA data only), we find $\Omega_{\rm m} = 0.343 \pm 0.096$ compared to $\Omega_{\rm m} = 0.295 \pm 0.034$ in Betoule et al. (2014). Our posterior uncertainty is about a factor of $\sim$3 larger, despite the shrinkage effect described above, and the central value is higher by $\sim 0.5\sigma$. We find $w = -0.943^{+0.363}_{-0.255}$. When compared with the Baseline Model, our cosmological parameter constraints hardly change (see Figure 5).[13]

Despite this, we do detect significant difference (with 95% probability) between the mean intrinsic magnitude of SNe Ia in host galaxies of low and high mass. Specifically, we define

$$\Delta M_0 \equiv M_0^{\rm hi} - M_0^{\rm lo} \tag{22}$$

as the difference in intrinsic magnitude between the two subclasses. The posterior interval for ΔM_0 is

$$-0.10 < \Delta M_0 < 0.00 \quad \text{(95\% equal-tail posterior interval)} \tag{23}$$

with $\Delta M_0 = 0$ excluded with 95% probability. The posterior distribution for ΔM_0 is shown in Figure 6, where the result for the Hard Classification Model is compared with the Soft Classification Model. There is no appreciable difference in ΔM_0 between the Hard Classification Model and the Soft Classification Model. In accordance with previous results (Kelly et al. 2010; Sullivan et al. 2010; Campbell et al. 2016), we find that SNe Ia in more massive galaxies are intrinsically brighter, with our posterior estimate of the magnitude difference being $\Delta M_0 = -0.055 \pm 0.022$. However, the size of the effect in our study is smaller than previously reported.

[13] Our treatment in the Baseline Model is not fully consistent. While we ignore any dependence on host galaxy mass we do include the "host relation" term in the systematics covariance matrix. This, however, is likely to have a negligible effect, since Table 11 in Betoule et al. (2014) quantifies the contribution to the error budget on $\Omega_{\rm m}$ from the uncertainty in host relation as a mere 1.3%.

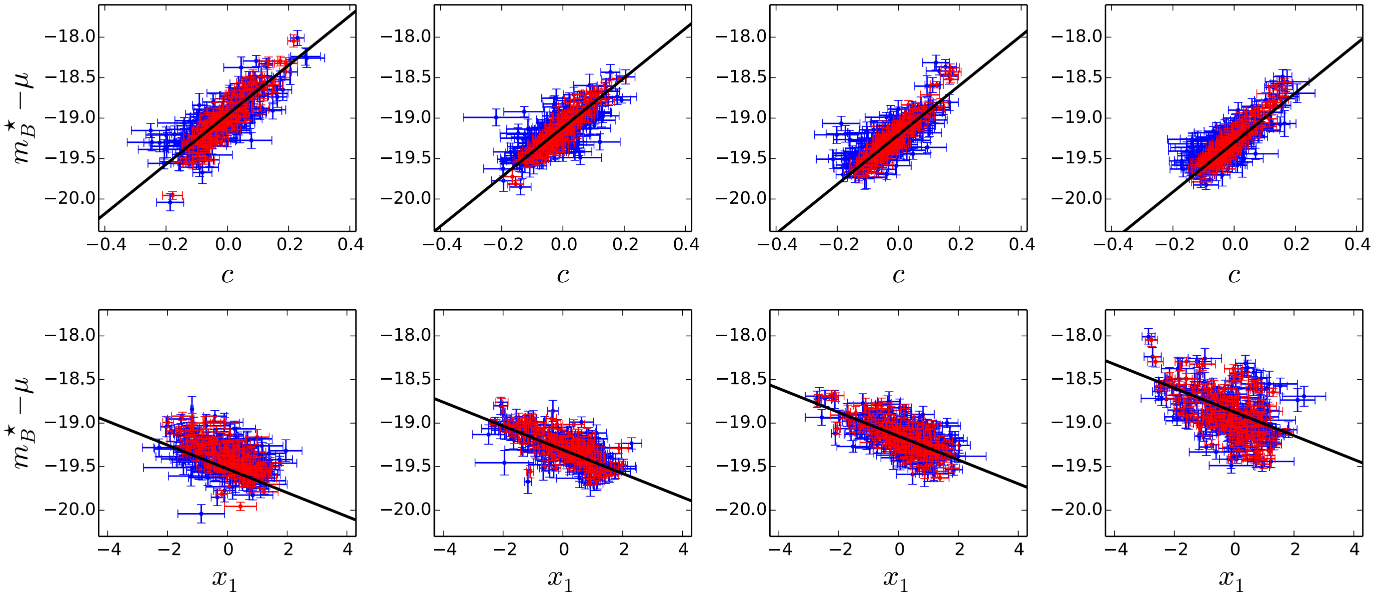

Figure 4. Shrinkage of posterior estimates in BAHAMAS: plug-in estimates of the intrinsic magnitude (blue) and posterior mean (red). The four panels in the first row correspond to quartiles of $\hat{x}_1$; we plot the regression line as a function of the color parameter in each. The horizontal axis plots $\hat{c}_i$ (blue) and the posterior mean of c_i (red). The four panels in the bottom row correspond to quartiles of $\hat{c}$; we plot the regression line as a function of the stretch parameter $\hat{x}_1$ in each. The horizontal axis plots $\hat{x}_{1i}$ (blue) and the posterior mean of x_{1i} (red). The regression lines use the posterior means of the parameters and the mean of the observed covariates in each quartile. The posterior estimates shrink from the plug-in estimates toward the regression line and thus reduce scatter around the regression plane. This is a consequence of the hierarchical regression in the model (this plot is for the ΛCDM case).

Table 5

Posterior Constraints on Our Model Parameters when the Host Galaxy Mass is Used as a Predictor or a Covariate (ΛCDM Case)

	JLA SNe Ia Only			JLA SNe Ia + *Planck* 2015		
	Hard Classification	Soft Classification	Covariate Adjustment	Hard Classification	Soft Classification	Covariate Adjustment
	Baseline Model Parameters					
Ω_m	0.343 ± 0.096	0.338 ± 0.107	0.361 ± 0.100	0.423 ± 0.030	0.400 ± 0.025	0.403 ± 0.031
Ω_Λ	0.523 ± 0.144	0.522 ± 0.165	0.559 ± 0.151	0.603 ± 0.020	0.622 ± 0.019	0.621 ± 0.023
Ω_κ	0.134 ± 0.232	0.140 ± 0.263	0.080 ± 0.244	−0.026 ± 0.011	−0.022 ± 0.008	−0.025 ± 0.010
α	0.141 ± 0.006	0.140 ± 0.006	0.143 ± 0.006	0.142 ± 0.006	0.142 ± 0.007	0.143 ± 0.005
β	3.058 ± 0.095	3.014 ± 0.086	3.068 ± 0.089	3.053 ± 0.068	3.034 ± 0.060	3.031 ± 0.086
M_0^ε	n/a	n/a	−18.837 ± 0.100	n/a	n/a	−18.860 ± 0.096
M_0^{lo}	−19.114 ± 0.023	−19.110 ± 0.023	n/a	−19.111 ± 0.019	−19.110 ± 0.021	n/a
σ_{res}^{lo}	0.110 ± 0.009	0.114 ± 0.009	n/a	0.108 ± 0.006	0.113 ± 0.009	n/a
ΔM_0	−0.055 ± 0.022	−0.049 ± 0.022	n/a	−0.062 ± 0.022	−0.049 ± 0.019	n/a
σ_{res}^{hi}	0.097 ± 0.007	0.096 ± 0.007	n/a	0.095 ± 0.006	0.094 ± 0.006	n/a
γ	n/a	n/a	−0.030 ± 0.010	n/a	n/a	−0.028 ± 0.010
σ_{res}	n/a	n/a	0.101 ± 0.005	n/a	n/a	0.102 ± 0.005

Note. Hard Classification adopts a mass-step correction by splitting the SNe Ia according to host galaxy mass into "Low" ($M_{g\,i} < 10$) and "High" ($M_{g\,i} \geqslant 10$) subclasses. Soft Classification further accounts for uncertainty due to the error in measurement of the host galaxy mass. Covariate Adjustment uses the host galaxy mass as a linear covariate. The quantity ΔM_0 is the difference between the mean peak intrinsic magnitudes of the two populations: $\Delta M_0 \equiv M_0^{hi} - M_0^{lo}$.

For example, Kelly et al. (2010) found (in our notation) $\Delta M_0 = -0.11$, and Sullivan et al. (2010) $\Delta M_0 = -0.08$, while Campbell et al. (2016) reported $\Delta M_0 = -0.091 \pm 0.045$.

The residual intrinsic dispersion of the two subpopulations is marginally smaller for the SNe Ia residing in more massive hosts: $\sigma_{res}^{hi} = 0.097 \pm 0.007$; for the lower mass group the residual dispersion is $\sigma_{res}^{lo} = 0.110 \pm 0.009$. (Those values are for the ΛCDM case, but *w*CDM is similar.)

Figure 7 shows the posterior estimates of the empirically corrected intrinsic magnitudes of SNe Ia, M_i^ϵ, as a function of the measured host galaxy mass. Histograms on either side of the graph show the distributions of the posterior mean estimates of M_i^ϵ for the two populations. The average measurement error of the host galaxy mass is fairly large, especially for low-mass hosts. Therefore, galaxies whose mass is close to the cut-off of $M_{g\,i} = 10$ are of uncertain classification, once the measurement error is taken into account. This could influence the estimate of ΔM_0 and the ensuing cosmological constraints.

To investigate the importance of errors in mass measurement, we fit the Soft Classification Model, which includes

Table 6
As in Table 5, but for wCDM

	JLA SNe Ia only			JLA SNe Ia + *Planck* 2015		
	Hard Classification	Soft Classification	Covariate Adjustment	Hard Classification	Soft Classification	Covariate Adjustment
Ω_m	0.342 ± 0.119	0.343 ± 0.116	0.348 ± 0.114	0.343 ± 0.017	0.350 ± 0.018	0.347 ± 0.015
Ω_Λ	0.658 ± 0.119	0.657 ± 0.116	0.652 ± 0.114	0.657 ± 0.017	0.650 ± 0.018	0.653 ± 0.015
w	$-0.943^{+0.363}_{-0.255}$	$-0.937^{+0.341}_{-0.213}$	$-0.958^{+0.364}_{-0.271}$	−0.906 ± 0.043	−0.902 ± 0.049	−0.898 ± 0.051
α	0.141 ± 0.006	0.141 ± 0.007	0.142 ± 0.007	0.135 ± 0.007	0.142 ± 0.006	0.141 ± 0.005
β	3.034 ± 0.078	3.049 ± 0.085	3.066 ± 0.087	2.917 ± 0.092	3.054 ± 0.085	3.057 ± 0.086
M_0^ε	n/a	n/a	−18.838 ± 0.098	n/a	n/a	−18.846 ± 0.090
M_0^{lo}	−19.117 ± 0.024	−19.111 ± 0.024	n/a	−19.126 ± 0.021	−19.116 ± 0.020	n/a
σ_{res}^{lo}	0.111 ± 0.008	0.112 ± 0.009	n/a	0.110 ± 0.008	0.112 ± 0.009	n/a
ΔM_0	−0.056 ± 0.021	−0.060 ± 0.020	n/a	−0.047 ± 0.025	−0.058 ± 0.020	n/a
σ_{res}^{hi}	0.098 ± 0.006	0.094 ± 0.007	n/a	0.098 ± 0.007	0.094 ± 0.006	n/a
γ	n/a	n/a	−0.030 ± 0.009	n/a	n/a	−0.030 ± 0.009
σ_{res}	n/a	n/a	0.101 ± 0.005	n/a	n/a	0.100 ± 0.005

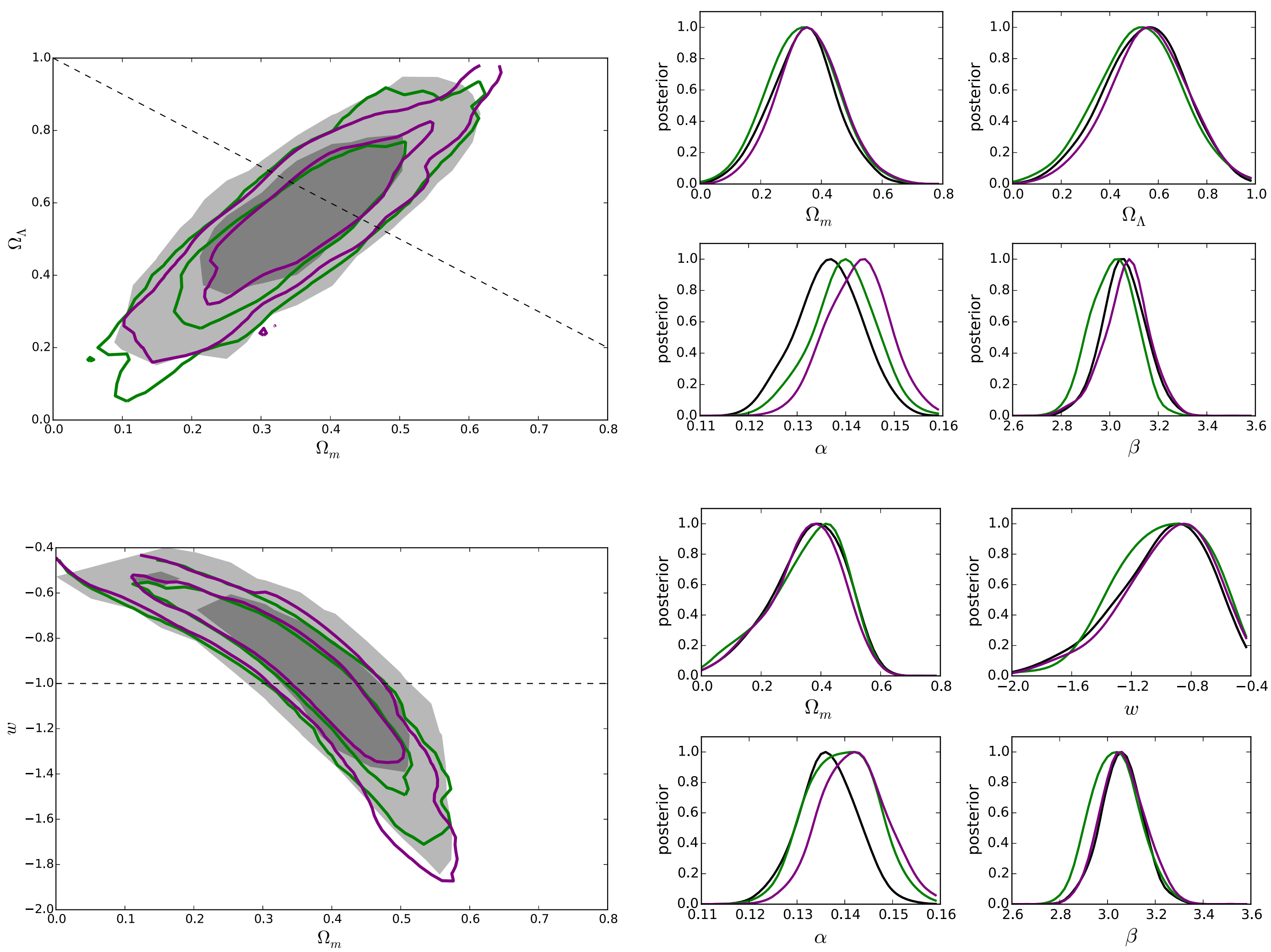

Figure 5. Comparison of cosmological parameters and standardization parameters with and without correction for host galaxy mass (black/shaded: Baseline Model; green: Soft Classification Model; purple: Covariate Adjustment Model). The result of the Hard Classification Model is similar to that of the Soft Classification Model and is not shown. Top panels are fit under the ΛCDM, while the bottom panels are fit under the wCDM. We do not find a significant difference in cosmology when mass information is included in the fit.

indicator variables for each SN Ia; recall that Z_i is one if SN Ia i belongs to the high-mass host class and zero if it does not. Treating Z_i as an unknown variable allows us to assess the posterior probability that each SN Ia belongs to the high-mass host class. In Figure 8 we plot the posterior means and standard deviations for each Z_i. The posterior mean of Z_i is the posterior

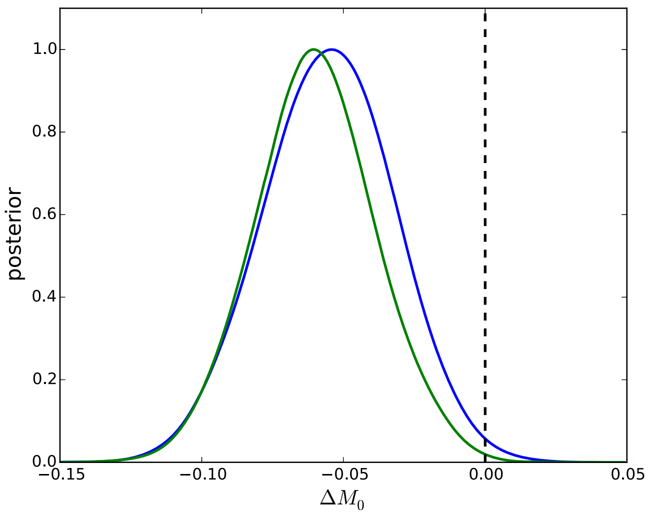

Figure 6. Posterior distribution of ΔM_0, the difference between mean intrinsic magnitudes of SNe Ia in high-mass host galaxies ($M_{g\,i} \geqslant 10$) and low-mass hosts ($M_{g\,i} < 10$). The blue and green curves correspond to the Hard Classification and Soft Classification Models, respectively. Under both models, the posterior probability that $\Delta M_0 < 0$ is greater than 95%, meaning that SNe Ia in more massive hosts are most probably intrinsically brighter ($\Delta M_0 < 0$). (This plot assumes a ΛCDM universe.)

probability that SN Ia i belongs to the high-mass class. Although measurement errors in the host galaxy mass are suppressed for clarity in Figure 8, the fitted model fully accounts for them.

The posterior constraints for the Soft Classification Model are compared with those under the Baseline Model in Figure 5. There is no significant difference between the cosmological fits or in the fitted nuisance parameters of the Baseline, Hard Classification, or Soft Classification Models.

Finally, the Covariate Adjustment Model includes host galaxy mass as a covariate; the fitted regression line under this model is plotted as a solid purple line in Figure 7. The fitted regression line can be expressed as $\hat{m}^{\star}_{Bi} - \mu_i =$ intercept $+ \bar{\gamma}\, \widehat{M}_{g\,i}$, where $\bar{\gamma}$ is the posterior mean of γ and the intercept is $(M_0^{\varepsilon} - \alpha x_1 + \beta c)$ with M_0^{ε}, α, and β replaced by their posterior means, $\bar{M}_0^{\varepsilon}$, $\bar{\alpha}$, and $\bar{\beta}$; $\hat{x}_{1i}$ replaced by $\frac{1}{n}\sum_{i=1}^{n} x_{1i}$; and c replaced by $\frac{1}{n}\sum_{i=1}^{n} \hat{c}_i$. The shaded purple region corresponds to a 68% posterior credible interval of γ (with the intercept fixed as described above). Figure 9 plots the posterior distribution for the slope γ. We find that the posterior probability that $\gamma < 0$ is 99%. The posterior 68% credible interval for γ is -0.030 ± 0.010. This is qualitatively consistent with previous work, but our slope is shallower. Previous analyses (Lampeitl et al. 2010; Gupta et al. 2011; Childress et al. 2013a; Pan et al. 2014; Campbell et al. 2016) (using various samples of SNe Ia) found values of the slope in the range $\gamma = -0.08$ to $\gamma = -0.04$.

Posterior constraints under the Covariate Adjustment Model are compared with those under the Baseline Model in Figure 5. Despite the fact that the posterior probability that $\gamma < 0$ is 99%, there is no significant shift in the cosmological parameters or the residual standard deviation, $\sigma_{\rm res}$. Although intuition stemming from standard linear regression suggests that adding a significant covariate should reduce residual variance, the situation is more complicated in Equation (14) owing to the measurement errors in both the independent and the dependent variables. While the variances of the left and right sides of (14) must be equal, there are numerous random quantities whose variances and covariances can be altered by adding a covariate to the model.

3.3. Redshift Evolution of the Color Correction

We now examine possible redshift evolution of the color correction parameter. The posterior distributions of the cosmological parameters under the Baseline, z-linear Color Correction, and z-jump Color Correction Models are compared in Figure 10 (ΛCDM) and Figure 11 (wCDM). The corresponding marginal posterior constraints are reported alongside the Baseline Model in Tables 3 and 4.

When evolution that is linear in redshift is allowed (as in the z-linear Color Correction Model), we find that a non-zero, negative linear term β_1 is preferred with ∼95% probability, $\beta_1 = -0.622 \pm 0.342$ (JLA data only). Because the standard deviation of $\hat{c}_i$ is of the order of ∼0.1, high-redshift SNe Ia (at $z \sim 1$) are typically ∼0.06 mag brighter than those nearby. However, there is no significant shift in the ensuing distributions of the cosmological parameters when compared with the Baseline Model.

When a sharp transition with redshift is allowed (as in the z-jump Color Correction Model), there is strong evidence for a significant drop in β at $z_t = 0.662 \pm 0.055$. At this redshift, β drops from its low-redshift value, $\beta_0 = 3.137 \pm 0.092$, by $\Delta\beta = -1.120 \pm 0.240$, with a nominal significance of approximately 4.6σ. This represents a correction of typically ∼0.11 mag for SNe Ia at $z > z_t$. The mean value and 1σ uncertainty band in the redshift-dependent $\beta(z)$ are shown in Figure 12. This trend is qualitatively similar to what is reported in Kessler et al. (2009), which attributed the shift to an unexplained effect in the first-year SNLS data. Wang et al. (2014) also found evidence for evolution of β with redshift in the SNLS three-year data. The drop, however, disappears in Betoule et al. (2014), after their reanalysis of the (three-year) SNLS data. The present work, however, uses identical data to Betoule et al. (2014). This is discussed further at the end of this section.

Despite significant evidence for redshift evolution of the color correction, the cosmological parameters are only mildly affected with respect to the Baseline Model. (Differences between the two fits are one standard deviation or less.) The posterior distribution of the residual intrinsic scatter also remains unchanged, giving $\sigma_{\rm res} = 0.103 \pm 0.005$.

In order to quantify the residual scatter around the Hubble diagram, we consider the difference between the theoretical distance modulus, $\mu(\hat{z}_i; \mathscr{C})$, and an estimate based on the observables, $\hat{\mu}_i(M_0^{\varepsilon}, \alpha, \beta) = \hat{m}_{Bi} - M_0^{\varepsilon} + \alpha\hat{x}_i - \beta\hat{c}_i$; that is, we define

$$\Delta\mu_i = \hat{\mu}_i(M_0^{\varepsilon}, \alpha, \beta) - \mu(\hat{z}_i; \mathscr{C}) \tag{24}$$

and its sample variance,

$$\sigma^2_{\Delta\mu} = \frac{1}{n-1}\sum_{i=1}^{n}(\Delta\mu_i - \Delta\bar{\mu})^2, \tag{25}$$

where $\Delta\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n}\Delta\mu_i$. Notice that both $\hat{\mu}_i(M_0^{\varepsilon}, \alpha, \beta)$ and $\mu(\hat{z}_i; \mathscr{C})$ depend on model parameters, and thus for fixed $\widehat{\mathscr{D}}$ we can view $\Delta\mu_i$ and $\sigma^2_{\Delta\mu}$ as functions of the parameters having their own posterior distributions.

We compare the Hubble diagram residuals, $\Delta\mu_i$, for the Baseline Model with those for the z-jump Color Correction

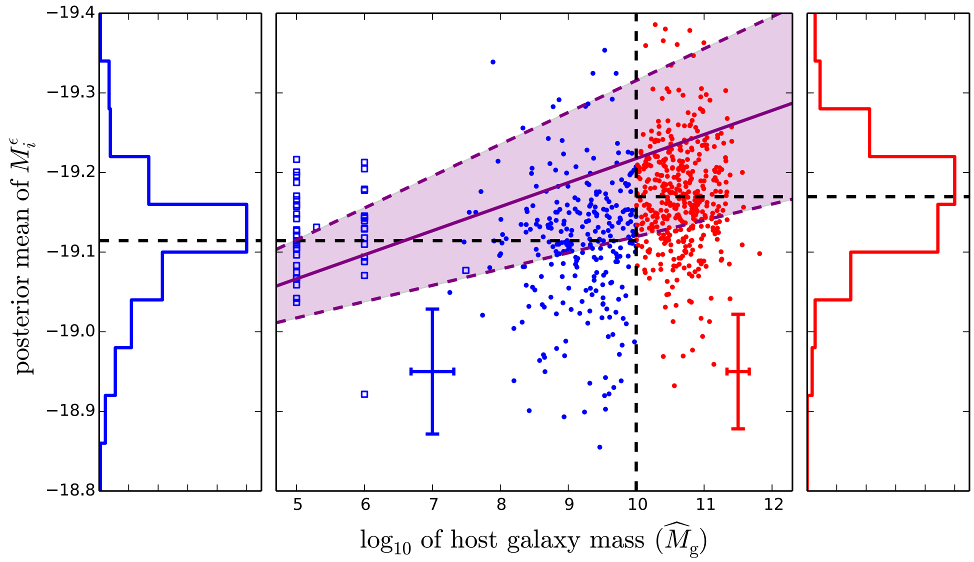

Figure 7. Posterior means and standard deviations for the empirically corrected intrinsic magnitudes of SNe Ia in the JLA sample vs. measured host galaxy mass. The sample has been divided into two populations, with $M_{g\,i} \geqslant 10$ (< 10) depicted in blue (red). A hollow square represents an SN Ia whose nominal measurement error on $M_{g\,i}$ is equal to or larger than 5. The population means of the intrinsic magnitudes are $M_0^{\rm lo} = -19.114 \pm 0.023$ and $M_0^{\rm hi} = -19.169 \pm 0.022$ (horizontal dashed lines) respectively for the classes of low and high host mass. The blue and red vertical error bars represent the average posterior standard deviations of the intrinsic magnitudes in the classes of low and high host mass, respectively. The horizontal error bars represent the average measurement errors of $M_{g\,i}$ in the two classes. The average error bars exclude the SNe Ia represented by hollow squares. The slope of the purple regression line is the posterior mean of γ under the Covariate Adjustment Model, while the purple shaded area represents the 1σ credible region for γ. (The regression line is computed under ΛCDM.)

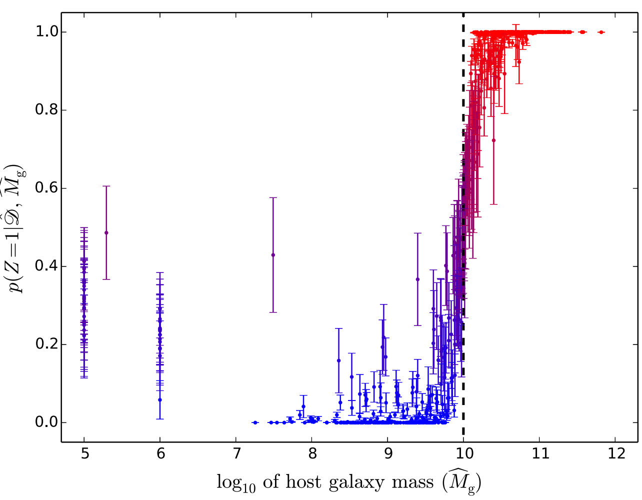

Figure 8. Posterior means and standard deviations of Z_i, the indicator variables for each SN Ia belonging to the high-mass host class ($Z_i = 1$, red) vs. measured host galaxy mass. If $Z_i = 0$ (blue), the SN Ia belongs to the low-mass host class. The posterior mean of Z_i is the posterior probability that an SN Ia belongs to the high-mass host class. Although the horizontal error bars are suppressed for clarity, the model fully accounts for measurement errors in the host galaxy mass. (This plot assumes the ΛCDM.)

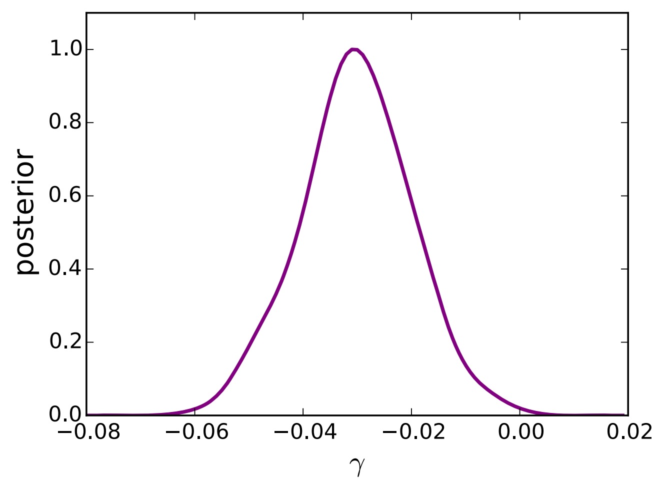

Figure 9. Marginal posterior distribution for γ, the regression coefficient for $M_{g\,i}$ in the Covariate Adjustment Model. (The model is fit assuming a ΛCDM universe.) The probability that γ is less than zero is 99%.

Model in Figure 13. The unknown parameters in $\Delta\mu_i$ are replaced with their posterior means. We only plot SNe Ia with $\hat{z} > 0.6$, because the residuals for low-redshift SNe Ia are very similar for the two models since the β value for $\hat{z} \leqslant 0.6$ is similar. The left panel of Figure 13 shows the Hubble residuals under the Baseline Model, the central panel shows them under the z-jump Color Correction Model, and the right panel compares the two by plotting residuals under the Baseline Model versus residuals under the z-jump Color Correction Model. The scatter is reduced under the z-jump Color Correction Model; it is nearer to zero. This indicates that allowing for a sharp transition in $\beta(z)$ improves the standardization of SNe Ia.

We define the cumulative (i.e., summed over redshift) Hubble residual as

$$s_i = \sum_{\hat{z}_j \leqslant \hat{z}_i} |\Delta\mu_j| \quad (1 \leqslant i \leqslant n). \tag{26}$$

In Figure 14 we use the cumulative residual to highlight the difference in the fit between the Baseline, z-linear Color Correction and z-jump Color Correction Models. Figure 14 shows the cumulative residual as a function of redshift, where at each redshift the Baseline Model residual has been subtracted to facilitate comparison. For $\hat{z} \lesssim 0.7$, the Baseline Model offers a slightly better fit than either of the $\beta(z)$ models. But above $\hat{z} \sim 0.8$ both the z-linear Color Correction Model

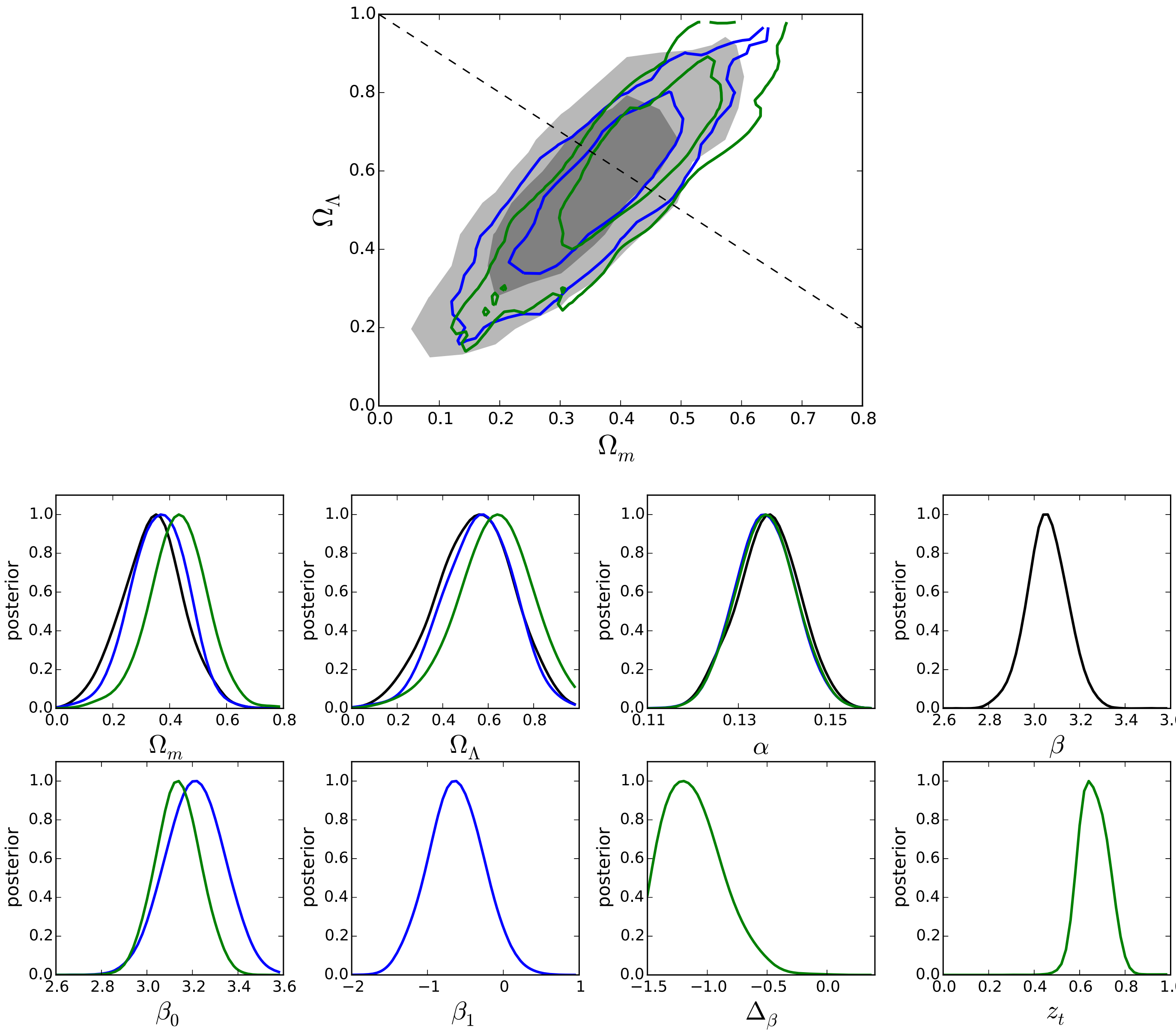

Figure 10. Comparisons of the posterior distributions for the cosmological parameters and the standardization parameters under different models for the color correction parameter: black: Baseline Model (no evolution); blue: z-linear Color Correction Model; green: z-jump Color Correction Model. Posteriors are normalized to the peak. (All models are fit assuming a ΛCDM universe).

and especially the z-jump Color Correction Model provide improved residuals with respect to the Baseline Model. This is shown by their negative values for the relative residual with respect to the Baseline Model. In other words, Figure 14 shows that either of the $\beta(z)$ models improves the fit for high-redshift SNe Ia. Although it is beyond the scope of this paper and the subject of future investigation, formal model comparison should be deployed to weigh the evidence for the evolving color correction model relative to the Baseline Model.

It is conceivable that the evidence for a step in the evolution of $\beta(z)$ is a spurious consequence of the mass-step correction, which is not included in the above analysis. Since more massive ($M_{g\,i} > 10$) host galaxies are preferentially found at low redshift, and SNe Ia in those galaxies are brighter (see Section 3.2), it is possible that such galaxies require on average a smaller color correction than SNe Ia in galaxies at high redshift (which are on average less luminous). However, if such a color–mass–redshift interaction were to exist, it could be identified by fitting a model that allows for both a correction for host galaxy mass and evolution in the color correction. To investigate this possibility, we fit a model that includes both a mass-step correction (as parameterized in the Hard Classification Model) and the z-jump Color Correction. The posterior constraints on all the model parameters change negligibly in this fit compared with the fit of the z-jump Color Correction Model without mass-step correction.

Our result is in stark contrast with Betoule et al. (2014), who found no significant departure of β from a constant. The dependence of reconstructions of color correction on the assumptions of the color scatter model used for SALT2 training has been extensively investigated in Mosher et al. (2014). This study found significant bias (up to ~0.6) in the reconstructed

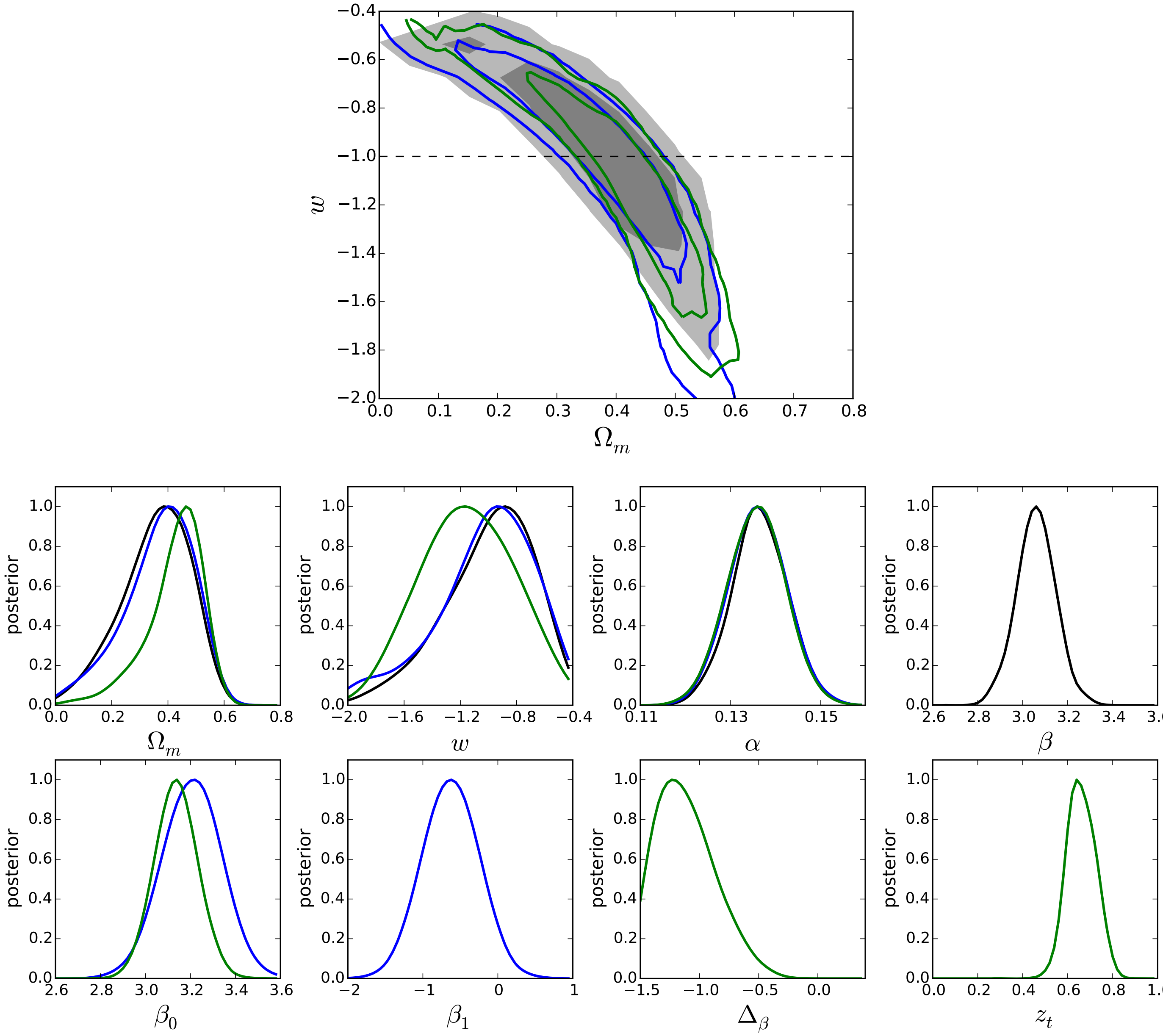

Figure 11. Comparisons of the posterior distributions for the cosmological parameters and the standardization parameters under different models for the color correction parameter: black: Baseline Model (no evolution); blue: z-linear Color Correction Model; green: z-jump Color Correction Model. Posteriors are normalized to the peak. (All models are fit assuming a wCDM universe.)

value for β when the underlying color scatter model was misspecified in the reconstruction. However, Mosher et al. (2014) showed that the reconstructed β (constant with redshift) is biased down (in the cases they considered), that is, in the opposite direction to what we observe. This appears to rule out a misspecification of the color scatter model as an explanation for our result. Mosher et al. (2014) also demonstrated that a color misspecification does not appreciably bias the recovered cosmological parameters. However, they did not investigate a possible z-dependence of the recovered β value. Wang & Wang (2013) analyzed the SNLS3 sample of SNe Ia using different parameterizations of the possible redshift dependence of β, including a linear dependence. They found that β increases significantly with redshift, again in contrast to what is seen in our analysis of the JLA data. Mohlabeng & Ralston (2013) similarly applied a linear z-dependence model for β using the Union 2.1 compilation of SNe Ia. They found a 7σ deviation from a constant β, with a trend to smaller β at larger z, similar to our findings.

The top panel in Figure 11 in Betoule et al. (2014) might suggest that unmodeled selection effects on the color correction at $\hat{z} \gtrsim 0.6$ could lead to our detection of a drop in the value of $\beta(z)$ in that range. To test this possibility, we have artificially corrected the trend to negative colors (as seen in Figure 11 of Betoule et al. 2014) for $\hat{z} > z_t$, and refitted the z-jump Color Correction Model. We find that this correction alters the posterior distributions of the cosmological parameters very significantly, while leaving the strong detection of a jump in the value of $\beta(z)$ largely unchanged. This argues against the existence of unmodeled selection effects due to color correction causing the observed jump in $\beta(z)$ in the z-jump Color Correction Model. By the same token, it is unlikely that our

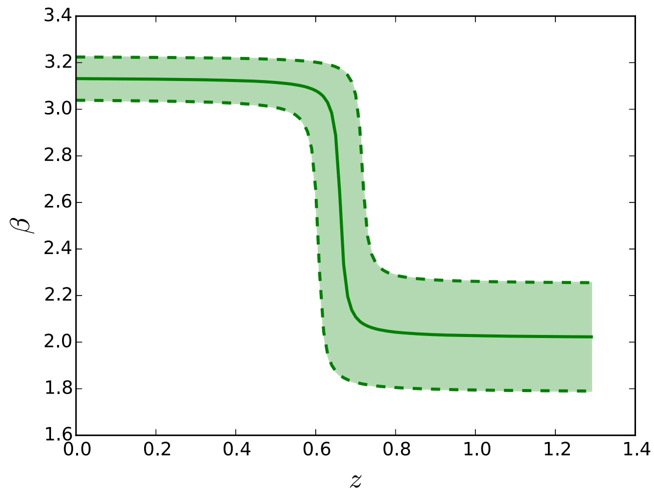

Figure 12. Redshift evolution of the color correction parameter β, assuming the z-jump Color Correction Model. The green line is the posterior mean, while the shaded region represents the 1σ credible region. (ΛCDM case).

result is driven by the redshift evolution of the color (or stretch) correction, as a consequence of selection effects, as seen, e.g., for SNLS one-year data in Astier et al. (2006).

In all of our models above, the population mean and variance of the color and stretch corrections are assumed to be independent of redshift. However, the observed color corrections drift toward the blue near the magnitude limit of a survey (i.e., to larger z). This happens because intrinsically brighter SNe Ia (which are more likely to be observed) are bluer in color. This selection effect thus leads to a z-dependence of the observed color correction, even if the underlying color does not change with redshift. We allow the population mean and variance of the color correction to differ for low-redshift ($z < 0.66$) and high-redshift ($z \geqslant 0.66$) SNe Ia. (The threshold of $z = 0.66$ was chosen because it is the posterior mean of the jump location in the z-jump Color Correction Model.) With this change, we refit both the Baseline Model and the z-jump Color Correction Model. The joint posterior distribution of (Ω_m, Ω_Λ) shifts appreciably toward lower values of matter and cosmological constant, but the evidence for a drop in β persists. This shows that BAHAMAS results are sensitive to the detailed modeling of a potential redshift dependence (induced by selection effects, or otherwise) of the color correction. However, the model for the redshift dependence of color is not what is driving the shift in the posterior distribution of Ω_m toward higher values. We will further investigate this aspect in future work by including an explicit model of selection effects in BAHAMAS.

3.4. Influence of the Systematics Covariance Matrix

To assess the relative importance of the statistical and systematics variance–covariance matrices in our results, we refit the Baseline Model with the statistical covariance matrix only, thus omitting $C_{\rm syst}$. The resulting posterior distributions of (Ω_m, Ω_Λ) (for ΛCDM) and (Ω_m, w) (for wCDM) are shown in Figure 15. Figure 15 compares this fit with the previous Baseline Model that includes the systematics covariance matrix. Adding the systematics covariance matrix not only enlarges the size of the contours—as one expects—but also significantly shifts the mean value of the posterior distribution of Ω_m to larger values, which leads to a smaller Ω_Λ (for ΛCDM) and a larger w (for wCDM). In fact, the posterior means we obtain when neglecting the systematics covariance matrix are broadly compatible with standard results. The Bayesian approach of March et al. (2011) is similar to BAHAMAS and produced results comparable to χ^2 fitting on the data set of Kessler et al. (2009); this analysis did not contain the systematic covariance matrix included in JLA. Thus we are led to conclude that the shift in cosmology is driven by some aspect of the systematics error modeling in JLA. The systematics covariance matrix derived by Betoule et al. (2014) contains contributions from different sources: calibration uncertainty, Milky Way extinction, light-curve model, bias corrections, host relations, contamination, and peculiar velocities. Analysing these individually shows that the main driver shifting Ω_m toward larger values is the calibration uncertainty. The large differences in the fitted values for Ω_m and w between BAHAMAS and the standard χ^2 have been observed previously in simulations by March et al. (2014). These authors showed on simulated SNLS three-year data that the posterior mean of Ω_m tends to be biased high (by ~0.1), while the χ^2 fit tends to be biased low (by a similar amount). However, March et al. (2014) also found that such discrepancies largely disappear when the redshift arm of the sample of SNe Ia is extended to lower and higher z.

In order to further investigate the origin of the observed shift in the fitted cosmological parameters obtained by BAHAMAS, we compute the percentage increase in the variances of $m_B^\star$, x_1, and c when adding the systematics covariance matrix to the statistical covariance matrix, i.e.,

$$F_i = \frac{1}{n}\sum_{j=1}^{n}\frac{\sigma_{i,j}^{2,\rm syst}}{\sigma_{i,j}^{2,\rm stat}} \tag{27}$$

where $i = m_B^\star$, x_1, or c. The quantity $F_i^{1/2}$ is the average percentage increase in the standard deviation for observable i when the systematics covariance matrix is added to the statistical covariance matrix (considering diagonal elements only). We find $F_{m_B^\star}^{1/2} = 2.66$, $F_{x_1}^{1/2} = 0.16$, and $F_c^{1/2} = 0.36$, which shows that the increased error on $m_B^\star$ is by far the dominant contribution from the systematics covariance matrix. This is because the dominant source of systematic error in the JLA data is the flux calibration (Betoule et al. 2014). To check whether the increase in the $m_B^\star$ variance is responsible for the shift in fitted cosmological parameters, we multiply the variance of $m_B^\star$ in the statistical covariance matrix by $(1 + F_{m_B^\star})$ and refit (without adding the systematics covariance matrix) our Baseline Model. The resulting cosmological constraints are shown as purple contours in Figure 15. Comparing with the original Baseline Model fit (black contours), it is clear that most of the shift in the fitted cosmological parameter is due to the large systematic variance of $m_B^\star$. If the model were Gaussian and linear, inflating the errors would only enlarge the uncertainty on the parameters and would not shift the mean of the posterior distribution. Hence we conclude that the shift in cosmology is a reflection of the non-Gaussian, nonlinear nature of our model, something that is only approximately accounted for in the linear propagation of errors used in standard chi-squared analyses.

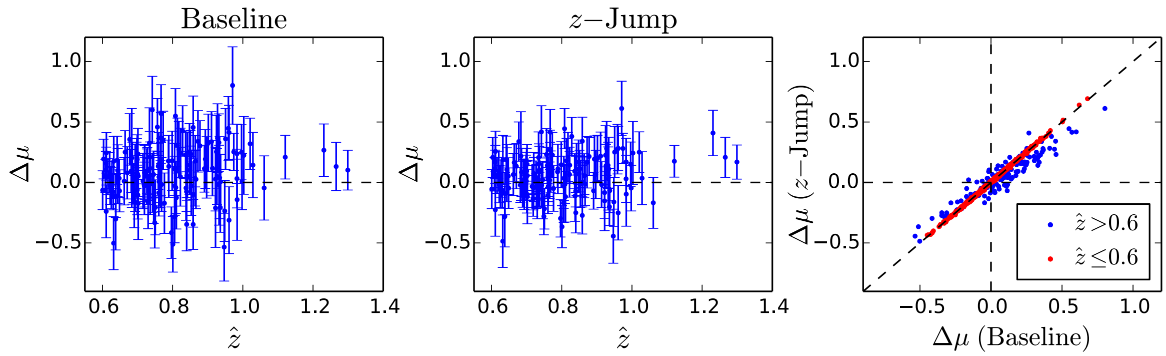

Figure 13. Hubble residuals of the Baseline Model (left, $\beta=$ constant with redshift), z-jump Color Correction Model (center), and comparison between the two (right). In the left and central panels, only SNe Ia with $\hat{z}>0.6$ are plotted to highlight the difference between the two cases. Error bars are the posterior standard deviations of $\Delta\mu_i$. In the right panel, SNe Ia with $\hat{z}\leqslant 0.6$ are plotted in red. This panel shows that the z-jump Color Correction Model reduces the scatter around the Hubble diagram noticeably for $\hat{z}>0.6$, while its Hubble residuals are similar to the Baseline Model for $\hat{z}\leqslant 0.6$ (this plot is for ΛCDM, and the wCDM case is similar).

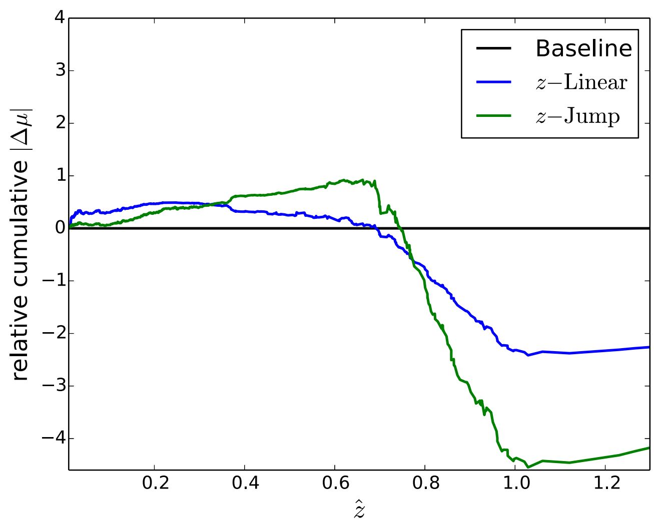

Figure 14. Cumulative Hubble residuals relative to the Baseline Model for the two $\beta(z)$ models considered. For $\hat{z}\gtrsim 0.8$, both the redshift-dependent models improve the fit with respect to the Baseline Model, which has $\beta=$ constant. The z-jump Color Correction model shows the largest improvement in the fit. This plot is for ΛCDM, but the wCDM case is qualitatively similar.

3.5. JLA Subsamples

To further investigate the shift in the fitted cosmological parameters and to check for consistency within the JLA sample of SNe Ia, we split the SNe Ia into a series of subsamples: low-z+SNLS, SDSS+SNLS, low-z+SNLS+*HST*, SDSS+SNLS+*HST* and low-z+SDSS+*HST*. We do not investigate the low-z+SDSS combination in our analysis as this subsample alone does not have a sufficient redshift range to constrain the cosmological parameters. In contrast to Betoule et al. (2014), we vary both Ω_m and Ω_Λ and do not assume flatness (but we do fix $w=-1$). We compare our results against the entire JLA data set in Figure 16. The left panel shows the results when excluding high-z *HST* data, while the right panel includes the nine high-z *HST* SNe Ia.

In contrast to Betoule et al. (2014) (see their Table 10), we find significant shifts in the posterior distributions of the cosmological parameters resulting from the different subsamples. The SNLS sample pushes the cosmology toward a closed universe with higher matter and higher dark-energy content (an effect previously observed in March et al. 2014) while the *HST* sample pulls it in the opposite direction. In particular, for the subsample low-z+SNLS (357 SNe Ia), including just nine extra SNe Ia from *HST* shifts the contours very noticeably to much lower values of both Ω_m and Ω_Λ. If we had assumed flatness, as was done in Betoule et al. (2014), this effect would have been masked. In all cases in Figure 16 if we enforced $\Omega_\kappa=0$, the posterior distribution of Ω_m would be similar to the baseline case. The posterior distributions of all the other parameters for the various subsamples are consistent with each other (hence not shown), except for the low-z+SNLS subsample for which both β and σ_{res} are smaller. This is consistent with the observed redshift dependence of β; see Figure 12.

4. CONCLUSIONS

We have reanalyzed the JLA data on SNe Ia with a principled Bayesian method (BAHAMAS). As shown in March et al. (2011), our approach has better statistical coverage and smaller mean squared errors than the standard χ^2 method. This paper introduces a series of powerful Gibbs-type samplers that allow us to explore the posterior distribution of the latent variables associated with SNe Ia, such as their empirically corrected intrinsic magnitudes. We have presented a general methodology that can easily incorporate additional standardization variables, over and above the usual stretch and color corrections. We have demonstrated this feature by including measurements of host galaxy mass in our fit, fully accounting for the uncertainty in mass measurement.

When the JLA data set is augmented by *Planck* CMB data, we find significant discrepancies with the results from the standard χ^2 fit, in particular in the values of Ω_m and w. We measure the average residual dispersion of the post-correction intrinsic magnitudes in the JLA sample to be $\sigma_{res}=0.104\pm 0.005$. The magnitude of the correction for host galaxy mass is smaller than previously reported. We find significant statistical evidence for a drop in the value of the color correction parameter, β, at a redshift $z_t=0.66$. While we rule out color-dependent selection effects as being responsible for this feature, we cannot trace it back to its origin. Cosmological parameter constraints, however, remain unaffected by marginalization over this non-standard redshift dependence.

Future work will incorporate selection effects into our framework (similarly to Rubin et al. 2015), include additional covariates (such as star formation rate and metallicity) and test their influence on the recovered cosmology, and allow for the

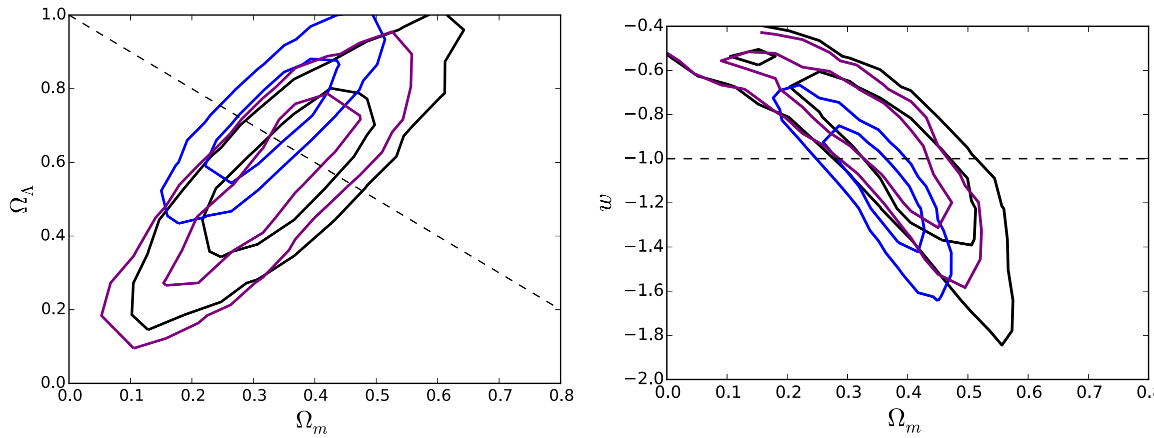

Figure 15. Comparison of posterior distributions when including both statistical and systematic errors (black) with the case when the systematics covariance matrix is neglected (blue). Purple: statistical covariance matrix with diagonal errors on $m_B^\star$ inflated by the average $m_B^\star$ variance from the systematics covariance matrix. Left: ΛCDM; right: wCDM.

possibility of contamination (as in the BEAMS scenario, Kunz et al. 2007; Hlozek et al. 2012; Knights et al. 2013).

We would like to thank Bruce Bassett, Heather Campbell, Alex Conley, Suhail Dhawan, Josh Frieman, Mike Hobson, Natasha Karpenka, Rick Kessler, Ofer Lahav, Michelle Lochner, Kaisey Mandel, Marisa March, Bob Nichol, Cassio Pigozzo, Mickael Rigault, Mathew Smith, and Mark Sullivan for useful discussions. We thank an anonymous referee for helpful suggestions. We would like to thank Sakina Ali-Khan for help with the color correction tests, and Cassio Pigozzo for help with checks using MultiNest. R.T. acknowledges partial support from an EPSRC "Pathways to Impact" grant and D.v.D. from a Wolfson Research Merit Award (WM110023) provided by the British Royal Society and from Marie-Curie Career Integration (FP7-PEOPLE-2012-CIG-321865) and Marie-Sklodowska-Curie RISE (H2020-MSCA-RISE-2015-691164) Grants both provided by the European Commission.

APPENDIX A
ALGORITHM REVIEW

The Gibbs samplers (Geman & Geman 1984) and Data Augmentation (DA) algorithm (Tanner & Wong 1987), which is a special Gibbs sampler, are widely used Markov chain Monte Carlo (MCMC) methods to sample from highly structured models. Although they are typically easy to implement, they can have slow convergence rates. To improve their convergence, a variety of extensions have been proposed. Among them, the Ancillarity–Sufficiency Interweaving Strategy (ASIS, Yu & Meng 2011) is designed to improve the convergence properties of the DA algorithm, and the Partially Collapsed Gibbs (PCG) sampling (van Dyk & Park 2008) is a useful tool to improve the convergence of Gibbs samplers. In a Gibbs-type sampler, we may also need the help of the MH algorithm (Metropolis et al. 1953; Hastings 1970), when one of the component conditional distributions is not standard.

Consider a generic observed data set, $Y_{\rm obs}$, and model parameters, θ, and suppose we wish to sample from the posterior distribution $p(\theta|Y_{\rm obs})$. When direct sampling is not possible, we may consider introducing a latent variable, $Y_{\rm mis}$, into the model, such that the complete-data model $p(Y_{\rm mis}, Y_{\rm obs}|\theta)$ maintains the target model, $p(Y_{\rm obs}|\theta)$, as its marginal distribution. The DA algorithm proceeds by drawing from $p(Y_{\rm mis}|\theta, Y_{\rm obs})$ and $p(\theta|Y_{\rm mis}, Y_{\rm obs})$ iteratively. This is a useful strategy when these two distributions are easy to sample and the resulting MCMC is relatively quick to converge.

More generally, when the unknown quantity in a model, ψ, consists of two or more components, each of which can be multivariate, that is, $\psi = (\psi_1, \ldots, \psi_N)$ with $N \geqslant 2$, the Gibbs sampler is useful to draw from $p(\psi|Y_{\rm obs})$. In one iteration of a Gibbs sampler, each component of ψ is sampled from its *complete conditional distribution*, i.e., its distribution conditioning on the current values of all the other components. In this paper we consider only systematic-scan Gibbs samplers (Liu et al. 1995), that is, in each complete iteration, the components are updated in a fixed ordering. The DA algorithm is a special case of the Gibbs sampler with two components in ψ, i.e., $\psi = (\theta, Y_{\rm mis})$.

As mentioned above, although they are easy to implement, in some cases the DA algorithm or Gibbs sampler can be slow to converge. We now describe two strategies that can significanlty improve their convergence, ASIS and PCG, along with the MH algorithm.

Ancillarity–Sufficiency Interweaving Strategy. ASIS improves the convergence of a standard DA algorithm by using a pair of special DA schemes. One is the sufficient augmentation $Y_{{\rm mis},S}$, which means the conditional distribution $p(Y_{\rm obs}|Y_{{\rm mis},S}, \theta)$ is free of θ. The other is the ancillary augmentation $Y_{{\rm mis},A}$, for which $p(Y_{{\rm mis},A}|\theta)$ does not depend on θ. Normally, given the parameter, these two augmentation schemes are related via a one-to-one mapping (but see Yu & Meng 2011 for an exception). It is usually the case that if the sampler corresponding to one of these two augmentations is fast, the other is slow. ASIS takes advantage of this "beauty-and-beast" feature of the two DA algorithms by interweaving steps of one into the other (Yu & Meng 2011). The resulting ASIS sampler can substantially outperform both parent DA samplers in terms of convergence, while the additional computational expense is often fairly small.

Partially Collapsed Gibbs Sampling. The PCG sampler can be effective in improving the convergence of Gibbs samplers. It achieves this goal by reducing conditioning, that is, by replacing some of the complete conditional distributions of an ordinary Gibbs sampler with the complete conditionals of marginal distributions of the target joint posterior distribution (van Dyk & Park 2008). This generally leads to larger variance

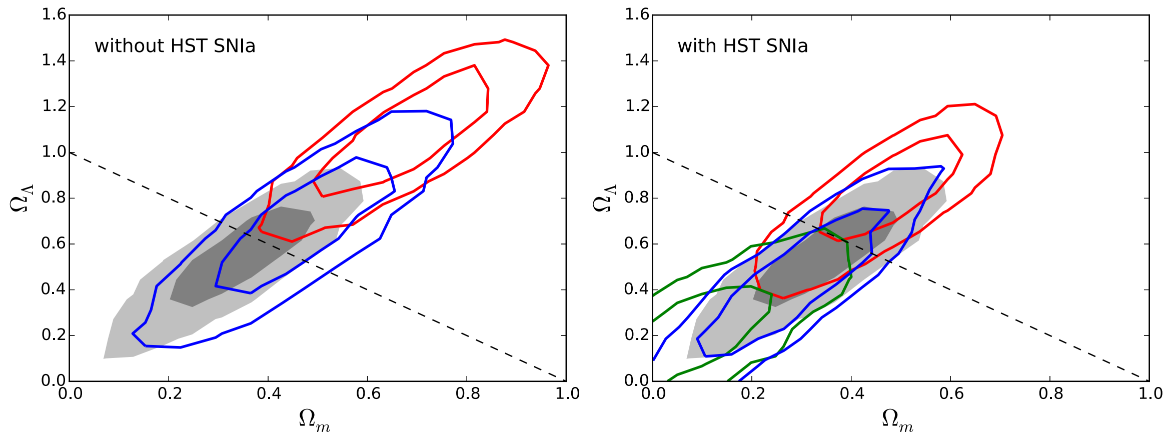

Figure 16. Comparison of posterior distributions for the cosmological parameters in the ΛCDM case when using different subsamples of the JLA data, as compared with the result for the entire JLA data set (black/filled). Blue: SDSS+SNLS (613 SNe Ia, $0.04 < z < 1.06$), red: low-z+SNLS (357 SNe Ia, $0.01 < z < 0.08 \cup 0.13 < z < 1.06$), green: low-$z$+SDSS (492 SNe Ia, $0.01 < z < 0.40$). The left plot does not include the *HST* SNe Ia (nine SNe Ia, $0.84 < z < 1.30$) while the right one does.

of the conditional distribution, and hence bigger jumps. A PCG sampler can be derived from a Gibbs sampler via a three-stage process: (i) *marginalization*, (ii) *permutation*, and (iii) *trimming*. Marginalization can significantly improve the rate of convergence, while permutation typically has a minor effect and trimming has no effect (van Dyk & Park 2008). Thus, we generally expect the PCG sampler to exhibit better, and often much better, convergence properties than its parent Gibbs sampler. In fact, van Dyk & Park (2008) already gave theoretical arguments and Park & van Dyk (2009) gave numerical illustrations of the computational advantage of PCG over ordinary Gibbs samplers. Sometimes, the PCG sampler is simply a blocked or collapsed Gibbs sampler (Liu et al. 1994). However, we are more interested in PCG samplers composed of *incompatible conditional distributions*, that is, there is no joint distribution corresponding to this set of conditional distributions. The incompatibility is introduced by trimming; permuting the order of the steps of a PCG sampler consisting of incompatible conditionals will alter its stationary distribution; see van Dyk & Park (2008).

MH Algorithm. The MH algorithm is frequently used to obtain a correlated sample from a target distribution, $p(\psi|Y_{\rm obs})$, for which direct sampling is difficult. Suppose we have sampled $\psi^{(t)}$ and need to generate $\psi^{(t+1)}$. Instead of sampling from $p(\psi|Y_{\rm obs})$ directly, we generate a candidate value ψ^c from a proposal distribution $g(\psi|\psi^{(t)})$ and accept it as $\psi^{(t+1)}$ with probability $\min(R, 1)$, where $R = \frac{p(\psi^c \mid Y_{\rm obs}) g(\psi^{(t)} \mid \psi^c)}{p(\psi^{(t)} \mid Y_{\rm obs}) g(\psi^c \mid \psi^{(t)})}$. In this way, we construct a reversible Markov chain, $\{\psi^{(t)}, t = 0, 1, ...\}$ with $p(\psi|Y_{\rm obs})$ as its stationary distribution.

To further ease implementation and improve convergence properties, we propose to combine several strategies introduced above into one sampler. Jiao et al. (2015) used a simplified version of the hierarchical model described in Section 2.2 as an example to illustrate the efficiency of both PCG and ASIS in improving the convergence properties of Gibbs-type samplers. They found that combining two strategies into one sampler can produce even more efficient samplers. Thus, we use PCG in each of our samplers to improve the convergence properties of $\mathscr{C}$ or $\mathscr{B}$. In some samplers, we combine PCG and ASIS for better convergence properties. The general method of combining several strategies into one sampler will appear in X. Jiao & D. A. van Dyk (2015, in preparation).

APPENDIX B
THE POSTERIOR DISTRIBUTION

In this section we give explicit expressions for the posterior distributions of the Baseline Model and its extensions listed in Table 2. To this end, we introduce a unified and general notation, see Table 7. We start with an expression that covers all of the models we consider, except the Hard Classification and Soft Classification Models. In particular, this formulation covers the regression model given in Equation (15) with the population distributions given in Equations (11)–(13) and (18) and the systematics covariance matrix described in Section 2.2.2. Under this extended hierarchical model, the posterior distribution is

$$
\begin{aligned}
&p(D, D_\star, \Sigma_D, \mathscr{B}, \mathscr{C} \mid \widehat{\mathscr{D}}) \\
&\quad \propto \frac{\left|\Sigma_{\hat{D}} \Sigma_D \Sigma_{D_\star}\right|^{-\frac{1}{2}}}{R_g^{2\hat{i}} R_c^2 \; R_{x_1}^2} p(\sigma_{\rm res}^2) \\
&\quad \times \exp\left\{-\frac{1}{2}\left[(\hat{D}(\mathscr{C}) - AD)^T \Sigma_{\hat{D}}^{-1} (\hat{D}(\mathscr{C}) - AD)\right.\right. \\
&\quad + (D - JD_\star)^T \Sigma_D^{-1} (D - JD_\star) \\
&\quad \left.\left. + (D_\star - D_{\star\star})^T \Sigma_{D_\star}^{-1} (D_\star - D_{\star\star})\right]\right\}, \qquad (28)
\end{aligned}
$$

where $p(\sigma_{\rm res}^2)$ is the prior distribution of $\sigma_{\rm res}^2$ given in Table 1 and the notation is defined in Table 2. The priors for the cosmological parameters, $\mathscr{C} = \{\Omega_{\rm m}, \Omega_\Lambda, w\}$, the regression coefficients, $\mathscr{B} = \{\alpha, \beta, \beta_0, \beta_1, \Delta\beta, \gamma, z_t\}$, the latent variables, D, their population means, $D_\star$, and their variances, Σ_D, are given in Table 1.

The posterior distribution under the Hard Classification Model is formally identical to that in Equation (28) except that

$$
p(\sigma_{\rm res}^2) \;\text{ is replaced by }\; p({\sigma_{\rm res}^{\rm lo}}^2) p({\sigma_{\rm res}^{\rm hi}}^2), \qquad (29)
$$

with the prior distributions given in Table 1. The (assumed known) indicator variables, Z_i, for low and high host galaxy masses enter through J and Σ_D using the definitions given in Appendix C.1.4.

For the Soft Classification Model, SNe Ia are classified on their true (latent) host galaxy masses (rather than on their

Table 7
Unified General Notation Used in the Posterior Distributions Given in Equations (28)–(32)

Symbol	Description
$\hat{D}(\mathscr{C})$	Column stacked vector of observed quantities, with apparent magnitude corrected for distance modulus, e.g., $\hat{D}(\mathscr{C}) = \{\hat{m}_{B1}^{\star} - \mu_1(\hat{z}_1, \mathscr{C}), \hat{x}_{11}, \hat{c}_1, \ldots, \hat{m}_{Bn}^{\star} - \mu_n(\hat{z}_n, \mathscr{C}), \hat{x}_{1n}, \hat{c}_n\}^T$ in the Baseline Model
D	Column stacked vector of latent variables, e.g., $D = \{M_1^{\epsilon}, x_{11}, c_1, \ldots, M_n^{\epsilon}, x_{1n}, c_n\}^T$ in the Baseline Model
$D_\star$	Vector of population means of the latent variables in D, e.g., $D_\star = \{M_0^{\epsilon}, x_{1\star}, c_\star\}^T$ in the Baseline Model
$D_{\star\star}$	Vector of prior means of quantities in $D_\star$, e.g., with priors given in Table 1, $D_{\star\star} = \{-19.3, 0, 0\}^T$ in the Baseline Model
$\Sigma_{\hat{D}}$	Matrix of variances (uncertainties) of observed quantities in $\hat{D}(\mathscr{C})$, compiled using $\Sigma_{\mathscr{D}} = C_{\rm stat} + C_{\rm syst}$, see Section 2.2.2
Σ_D	Population variance–covariance matrix of latent quantities in D. This is a block-diagonal matrix composed of n blocks, i.e., $\Sigma_D = \mathrm{diag}(S_1, \ldots, S_n)$. For example, each $S_i = \mathrm{diag}(\sigma_{\rm res}^2, R_{x1}^2, R_c^2)$ in the Baseline Model
$\Sigma_{D_\star}$	Prior variance–covariance matrix of quantities in $D_\star$, e.g., with priors given in Table 1, $\Sigma_{D_\star} = \mathrm{diag}(2^2, 10^2, 1^2)$ in the Baseline Model
J	Top-to-bottom stacked matrix of n matrices, i.e., $J = \begin{bmatrix} J_1 \\ \vdots \\ J_n \end{bmatrix}$. In the Hard and Soft Classification Models, $J_i = \begin{bmatrix} 1 - Z_i & Z_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, while in the other models, each J_i is an identity matrix
A	Block-diagonal matrix with n blocks, i.e., $A = \mathrm{diag}(T_1, \ldots, T_n)$. Each block is composed of 0, 1, and elements of $\mathscr{B}$, e.g., each $T_i = \begin{bmatrix} 1 & -\alpha & \beta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ in the Baseline Model
Σ_A	$\Sigma_A^{-1} = A^T \Sigma_{\hat{D}}^{-1} A + \Sigma_D^{-1}$
Σ_K	$\Sigma_K^{-1} = -J^T \Sigma_D^{-1} \Sigma_A \Sigma_D^{-1} J + J^T \Sigma_D^{-1} J + \Sigma_{D_\star}^{-1}$
Δ	$\Delta = A^T \Sigma_{\hat{D}}^{-1} \hat{D}(\mathscr{C})$
$k_\star$	$k_\star = \Sigma_K (J^T \Sigma_D^{-1} \Sigma_A \Delta + \Sigma_{D_\star}^{-1} D_{\star\star})$
λ	Parameter in the prior InvGamma distribution of $\sigma_{\rm res}^2$, i.e., $\sigma_{\rm res}^2 \sim \mathrm{InvGamma}(\lambda, \lambda)$

Note. Here we exemplify the general notation for the Baseline Model in terms of the notation used in Section 2. These details are given for each of the model extensions in Appendix C.1.

observed masses as in the Hard Classification Model). Thus, the indicator variables, Z_i, are treated as unknown and the posterior distribution, $p(D, D_\star, \Sigma_D, \mathscr{B}, \mathscr{C}, Z | \widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g})$, is formally identical to that in Equation (28) except

$$p(\sigma_{\rm res}^2) \quad \text{is replaced by}$$
$$p(\sigma_{\rm res}^{\rm lo\,2}) p(\sigma_{\rm res}^{\rm hi\,2}) \prod_{i=1}^{n} \pi_i^{Z_i} (1 - \pi_i)^{1 - Z_i}, \tag{30}$$

with the prior distributions given in Table 1, $\widehat{\mathscr{D}}_{\rm g} = \{\widehat{M}_{{\rm g}\,i}, i = 1, \ldots, n\}$, and

$$\pi_i = \Pr(Z_i = 1 | \widehat{M}_{{\rm g}\,i}) = \Pr(M_{{\rm g}\,i} \geqslant 10 | \widehat{M}_{{\rm g}\,i})$$
$$= \int_{10}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_{{\rm g}\,i}} \exp[-(M_{{\rm g}\,i} - \widehat{M}_{{\rm g}\,i})^2 / (2\sigma_{{\rm g}\,i}^2)] dM_{{\rm g}\,i}, \tag{31}$$

for $i = 1, \ldots, n$. The specific definitions of the unified notation for the Soft Classification Model are given in Appendix C.1.5.

APPENDIX C
THE MCMC SAMPLERS

To obtain posterior draws of all the variables (including latent variables) of the hierarchical models, we use Gibbs-type samplers, sometimes augmented with an MH step. In order to cross-check our sampling results, we have compared the marginal posteriors for the cosmological parameters, the regression coefficients, and the population variances obtained with Gibbs-type samplers with those obtained from a pure MH algorithm. The MH algorithm has been used to sample from a marginal posterior with latent variables, D, and population mean parameters, $D_\star$, integrated out analytically, akin to what was done in March et al. (2011).

In our Gibbs-type samplers, we make use of PCG to improve convergence. As detailed below, this involves sampling from conditional distributions of the marginal posterior distribution,

$$p(\Sigma_D, \mathscr{B}, \mathscr{C} | \widehat{\mathscr{D}}) \propto \frac{|\Sigma_{\hat{D}} \Sigma_D \Sigma_{D_\star}|^{-\frac{1}{2}} |\Sigma_A \Sigma_K|^{\frac{1}{2}}}{R_c^2\, R_{x_1}^2\, R_g^2} p(\sigma_{\rm res}^2) \times \exp\left\{-\frac{1}{2}\left[\hat{D}(\mathscr{C})^T \Sigma_{\hat{D}}^{-1} \hat{D}(\mathscr{C}) - \Delta^T \Sigma_A \Delta - k_\star^T \Sigma_K^{-1} k_\star + D_{\star\star}^T \Sigma_{D_\star}^{-1} D_{\star\star}\right]\right\}, \tag{32}$$

with notation given in Table 7. The corresponding marginal posterior distributions for the Hard Classification and Soft Classification Models are obtained using the substitutions in Equations (29) and (30), respectively.

This section consists of details of sampling steps of these algorithms.

C.1. Gibbs-type Samplers

We start with Gibbs-type samplers and consider both the Baseline Model and all its extensions discussed in Sections 2.2 and 2.3.

C.1.1. Baseline Model

As stated in Table 7, in the Baseline Model, $\hat{D}(\mathscr{C})$ is the version of $\widehat{\mathscr{D}}$ corrected for distance modulus, that is,

$$\hat{D}(\mathscr{C}) = \left\{\hat{D}(\mathscr{C})_1^T, \ldots, \hat{D}(\mathscr{C})_n^T\right\}^T, \tag{33}$$

where $\hat{D}(\mathscr{C})_i = \{\hat{m}_{Bi}^{\star} - \mu_i(\hat{z}_i, \mathscr{C}), \hat{x}_{1i}, \hat{c}_i\}^T$. Moreover, $D = \{D_1^T, \ldots, D_n^T\}^T$, where $D_i = \{M_i^\epsilon, X_i^T\}^T$ with $X_i = \{x_{1i}, c_i\}^T$; $D_\star = \{M_0^\epsilon, x_{1\star}, c_\star\}^T$; $D_{\star\star} = \{-19.3, 0, 0\}^T$. For the variance–covariance matrices, $\Sigma_{\hat{D}} = C_{\text{stat}} + C_{\text{syst}}$; $\Sigma_D = \text{diag}(S_1, \ldots, S_n)$, where each $S_i = \text{diag}(\sigma_{\text{res}}^2, R_{x_1}^2, R_c^2)$; $\Sigma_{D_\star} = \text{diag}(2^2, 10^2, 1^2)$. In addition, $J_{(3n\times 3)} = \begin{bmatrix} J_1 \\ \vdots \\ J_n \end{bmatrix}$, where each J_i is a (3×3) identity matrix, that is, $J_i = \text{diag}(1, 1, 1)$; $\mathscr{B} = \{-\alpha, \beta\}^T$, and $A_{(3n\times 3n)} = \text{diag}(T_1, \ldots, T_n)$, where each $T_i = \begin{bmatrix} 1 & -\alpha & \beta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

The sampler for the Baseline Model. This is an MH within a PCG sampler, that is, we integrate $(D, D_\star)$ out when updating $\mathscr{C}$ and $\mathscr{B}$. Then the sampling of $\mathscr{C}$ and $\mathscr{B}$ needs the help of the MH algorithm. While using MH in a Gibbs sampler is a standard strategy, embedding MH into a PCG sampler involves more subtleties. We follow exactly the procedure provided by van Dyk & Jiao (2015) when deriving an MH within a PCG sampler. The steps of the sampler are listed below. We use a prime to indicate the current iteration of a parameter, and $\mathcal{M}$ to represent the transition function introduced by the MH algorithm.

Step 1. $\mathscr{C} \sim \mathcal{M}(\mathscr{C}|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}')$:
Use MH to sample $\mathscr{C}$ from $p(\mathscr{C}|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}')$, which is proportional to $p(\Sigma_D', \mathscr{B}', \mathscr{C}|\widehat{\mathscr{D}})$, under the constraint imposed by the priors[14];

Step 2. $\mathscr{B} \sim \mathcal{M}(\mathscr{B}|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{C})$:
Use MH to sample $\mathscr{B}$ from $p(\mathscr{B}|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{C})$, which is proportional to $p(\Sigma_D', \mathscr{B}, \mathscr{C}|\widehat{\mathscr{D}})$, under the constraint imposed by the priors;

Step 3. $(D, D_\star) \sim p(D, D_\star|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}, \mathscr{C})$:
This step consists of two substeps:
1. Sample $D_\star$ from $\mathcal{N}(k_\star, \Sigma_K)$, where $k_\star$ and Σ_K are defined in Table 7;
2. Sample D from $\mathcal{N}(\mu_A, \Sigma_A)$, where Σ_A is defined in Table 7 and $\mu_A = \Sigma_A(\Delta + \Sigma_D^{-1} J D_\star)$;

Step 4. $\sigma_{\text{res}} \sim p(\sigma_{\text{res}}|\widehat{\mathscr{D}}, D, D_\star, R_{x_1}', R_c', \mathscr{B}, \mathscr{C})$:
Sample σ_{res}^2 from $\text{InvGamma}\left[\frac{n}{2} + \lambda, \frac{\sum_{i=1}^n (M_i^\epsilon - M_0^\epsilon)^2}{2} + \lambda\right]$, and $\sigma_{\text{res}} = \sqrt{\sigma_{\text{res}}^2}$;

Step 5. $R_{x_1} \sim p(R_{x_1}|\widehat{\mathscr{D}}, D, D_\star, \sigma_{\text{res}}, R_c', \mathscr{B}, \mathscr{C})$:
Sample $R_{x_1}^2$ from $\text{InvGamma}\left[\frac{n}{2}, \frac{\sum_{i=1}^n {x_{1i} - x_{1\star}}^2}{2}\right]$ with $\log(R_{x_1}) \in [-5, 2]$, and $R_{x_1} = \sqrt{R_{x_1}^2}$;

Step 6. $R_c \sim p(R_c|\widehat{\mathscr{D}}, D, D_\star, \sigma_{\text{res}}, R_{x_1}, \mathscr{B}, \mathscr{C})$:
Sample R_c^2 from $\text{InvGamma}\left[\frac{n}{2}, \frac{\sum_{i=1}^n (c_i - c_\star)^2}{2}\right]$ with $\log(R_c) \in [-5, 2]$, and $R_c = \sqrt{R_c^2}$.

C.1.2. z-linear Color Correction Model

In the z-linear Color Correction Model, the specification of $\hat{D}(\mathscr{C})$, $D_\star$, $D_{\star\star}$, $\Sigma_{\hat{D}}$, Σ_D, $\Sigma_{D_\star}$ and J is identical to that in the Baseline Model. As above, $D = \{D_1^T, \ldots, D_n^T\}^T$, where $D_i = \{M_i^\epsilon, X_i^T\}^T$, but under this model $X_i = \{x_{1i}, c_i, \hat{z}_i c_i\}^T$. In addition, $\mathscr{B} = \{-\alpha, \beta_0, \beta_1\}^T$; $A_{(3n\times 3n)} = \text{diag}(T_1, \ldots, T_n)$, where $T_i = \begin{bmatrix} 1 & -\alpha & \beta_0 + \beta_1 \hat{z}_i \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

The sampler for the z-linear Color Correction model. In this sampler, we combine ASIS and MH within PCG algorithms. We integrate $(D, D_\star)$ out when updating $\mathscr{C}$, and use the ASIS algorithm to update $\mathscr{B}$. The distribution of D conditioning on $\mathscr{B}$ and other parameters is

$$D|D_\star, \Sigma_D, \mathscr{B}, \mathscr{C} \sim \mathcal{N}(JD_\star, \Sigma_D). \tag{34}$$

Because this distribution is free of $\mathscr{B}$, D is an ancillary augmentation for $\mathscr{B}$ conditioning on other parameters. To derive a sufficient augmentation, we set $\tilde{D} = AD$. The distribution of $\hat{D}(\mathscr{C})$ conditioning on $\tilde{D}$, $\mathscr{B}$, and other parameters is

$$\hat{D}(\mathscr{C})|\tilde{D}, D_\star, \Sigma_D, \mathscr{B}, \mathscr{C} \sim \mathcal{N}(\tilde{D}, \Sigma_{\hat{D}}). \tag{35}$$

Because this distribution is free of $\mathscr{B}$, $\tilde{D}$ is the corresponding sufficient augmentation for $\mathscr{B}$.

We use "I" in the superscript to indicate intermediate draws that are not part of the final output. The steps of the sampler for the z-linear Color Correction model are:

Step 1. $\mathscr{C} \sim \mathcal{M}(\mathscr{C}|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}')$:
Use MH to sample $\mathscr{C}$ from $p(\mathscr{C}|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}')$, which is proportional to $p(\Sigma_D', \mathscr{B}', \mathscr{C}|\widehat{\mathscr{D}})$, under the constraint imposed by the priors;

Step 2. $(D^I, D_\star) \sim p(D, D_\star|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}', \mathscr{C})$:
This step consists of two substeps:
1. Sample $D_\star$ from $\mathcal{N}(k_\star, \Sigma_K)$, where $k_\star$ and Σ_K are defined in Table 7;
2. Sample D^I from $\mathcal{N}(\mu_A, \Sigma_A)$, where Σ_A is defined in Table 7 and $\mu_A = \Sigma_A(\Delta + \Sigma_D^{-1} J D_\star)$;

Step 3. $\mathscr{B}^I \sim p(\mathscr{B}|\widehat{\mathscr{D}}, D^I, D_\star, \Sigma_D', \mathscr{C})$:
Sample $\mathscr{B}^I$ from $\mathcal{N}(\zeta_B, \Sigma_B)$ (details about this distribution are given below) with constraint $\mathscr{B}^I \in [-1, 0] \times [0, 4] \times [-4, 4]$;
Use $\mathscr{B}^I$ to construct A^I. Then set $\tilde{D} = A^I D^I$;

Step 4. $\mathscr{B} \sim p(\mathscr{B}|\widehat{\mathscr{D}}, \tilde{D}, D_\star, \Sigma_D', \mathscr{C})$:
Sample $\mathscr{B}$ from $\mathcal{N}(\tilde{\zeta}_B, \tilde{\Sigma}_B)$ (details about this distribution are given below) with constraint $\mathscr{B} \in [-1, 0] \times [0, 4] \times [-4, 4]$;
Use $\mathscr{B}$ to construct A. Then set $D = A^{-1}\tilde{D}$;

Step 5. $\sigma_{\text{res}} \sim p(\sigma_{\text{res}}|\widehat{\mathscr{D}}, D, D_\star, R_{x_1}', R_c', \mathscr{B}, \mathscr{C})$:
Sample σ_{res}^2 from $\text{InvGamma}\left[\frac{n}{2} + \lambda, \frac{\sum_{i=1}^n (M_i^\epsilon - M_0^\epsilon)^2}{2} + \lambda\right]$, and $\sigma_{\text{res}} = \sqrt{\sigma_{\text{res}}^2}$;

[14] The proportionality is a consequence of the relationship: $p(X|Y) = P(X, Y)/P(Y) \propto P(X, Y)$, as a function of X.

Step 6. $R_{x_1} \sim p(R_{x_1}|\widehat{\mathscr{D}}, D, D_\star, \sigma_{\rm res}, R_c', \mathscr{B}, \mathscr{C})$:
Sample $R_{x_1}^2$ from InvGamma$\left[\frac{n}{2}, \frac{\sum_{i=1}^n (x_{1i} - x_{1\star})^2}{2}\right]$ with $\log(R_{x_1}) \in [-5, 2]$, and $R_{x_1} = \sqrt{R_{x_1}^2}$;

Step 7. $R_c \sim p(R_c|\widehat{\mathscr{D}}, D, D_\star, \sigma_{\rm res}, R_{x_1}, \mathscr{B}, \mathscr{C})$:
Sample R_c^2 from InvGamma$\left[\frac{n}{2}, \frac{\sum_{i=1}^n (c_i - c_\star)^2}{2}\right]$ with $\log(R_c) \in [-5, 2]$, and $R_c = \sqrt{R_c^2}$.

In Step 3, $\Sigma_B^{-1} = E^T V_m^{-1} E$, where V_m is the $(n \times n)$ submatrix of $\Sigma_{\hat{D}}$ after deleting the $(3i-1)^{\rm th} (i = 1, \ldots, n)$ and $(3i)^{\rm th}$ $(i = 1, \ldots, n)$ rows and columns, and $E_{(n\times 3)} = \begin{bmatrix} X_1^T \\ \vdots \\ X_n^T \end{bmatrix}$. Furthermore, $\zeta_B = \Sigma_B E^T V_m^{-1}(\hat{\xi}_m - \xi_m - \Delta\xi)$, where $\hat{\xi}_m = \{\hat{m}_{B1}^\star - \mu_1(\hat{z}_1, \mathscr{C}), \ldots, \hat{m}_{Bn}^\star - \mu_n(\hat{z}_n, \mathscr{C})\}^T$, $\xi_m = \{M_1^\epsilon, \ldots, M_n^\epsilon\}^T$, and $\Delta\xi = V_{m,-m} V_{-m}^{-1}(\hat{\xi}_{-m} - \xi_{-m})$; V_{-m} is the $(2n \times 2n)$ submatrix of $\Sigma_{\hat{D}}$ after deleting the $(3i-2)^{\rm th} (i = 1, \ldots, n)$ rows and columns; $V_{m,-m}$ is the $(n \times 2n)$ submatrix of $\Sigma_{\hat{D}}$ after deleting the $(3i-1)^{\rm th}$ $(i = 1, \ldots, n)$ and $(3i)^{\rm th} (i = 1, \ldots, n)$ rows and the $(3i-2)^{\rm th}$ $(i = 1, \ldots, n)$ columns; $\hat{\xi}_{-m} = \{\hat{x}_{11}, \hat{c}_1, \ldots, \hat{x}_{1n}, \hat{c}_n\}^T$; $\xi_{-m} = \{x_{11}, c_1, \ldots, x_{1n}, c_n\}^T$.

In Step 4, $\tilde{\Sigma}_B^{-1} = (\tilde{E}^T \tilde{E}) / \sigma_{\rm res}^{2'}$, where $\tilde{E}_{(n\times 3)} = \begin{bmatrix} \tilde{E}_1^T \\ \vdots \\ \tilde{E}_n^T \end{bmatrix}$ with $\tilde{E}_i = \{-\tilde{x}_{1i}, -\tilde{c}_i, -\hat{z}_i \tilde{c}_i\}^T$; $\tilde{x}_{1i}$ and $\tilde{c}_i$ are the $(3i-1)^{\rm th}$ and $(3i)^{\rm th}$ components of $\tilde{D}$ respectively. Furthermore, $\tilde{\zeta}_B = \tilde{\Sigma}_B[\tilde{E}^T(\xi_{M_0} - \tilde{\xi}_m)/\sigma_{\rm res}^{2'}]$, where $\xi_{M_0} = \{\underbrace{M_0^\epsilon, \ldots, M_0^\epsilon}_{n}\}^T$ and $\tilde{\xi}_m = \{\tilde{M}_1^\epsilon, \ldots, \tilde{M}_n^\epsilon\}^T$; $\tilde{M}_i^\epsilon$ is the $(3i-2)^{\rm th}$ component of $\tilde{D}$.

C.1.3. z-jump Color Correction Model

In the z-jump Color Correction Model, the specification of $\hat{D}(\mathscr{C})$, $D_\star$, $D_{\star\star}$, $\Sigma_{\hat{D}}$, Σ_D, $\Sigma_{D_\star}$, and J is identical to that in the Baseline Model. As above, $D = \{D_1^T, \ldots, D_n^T\}^T$, where $D_i = \{M_i^\epsilon, X_i^T\}^T$, but under this model $X_i = X_i(z_t) = \left\{x_{1i}, c_i, \left(\frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\hat{z}_i - z_t}{0.01}\right)\right) c_i\right\}^T$. In addition, $\mathscr{B} = \{-\alpha, \beta_0, \Delta\beta\}^T$; $A_{(3n\times 3n)} = {\rm diag}(T_1, \ldots, T_n)$, where $T_i = \begin{bmatrix} 1 & -\alpha & \beta_0 + \Delta\beta\left(\frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\hat{z}_i - z_t}{0.01}\right)\right) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Because we have an additional unknown parameter, z_t, under this model, the complete and marginal posterior distributions should be written as $p(D, D_\star, \Sigma_D, \mathscr{B}, z_t, \mathscr{C}|\widehat{\mathscr{D}})$ and $p(\Sigma_D, \mathscr{B}, z_t, \mathscr{C}|\widehat{\mathscr{D}})$ respectively, although they are formally identical to (28) and (32), respectively.

The sampler for the z-jump Color Correction Model. As in the sampler for the z-linear Color Correction Model, we also combine ASIS and MH within PCG algorithms in this sampler. We integrate $(D, D_\star)$ out when updating both $\mathscr{C}$ and z_t, and use the ASIS algorithm to update $\mathscr{B}$. When implementing ASIS, we also regard D as the ancillary augmentation, and $\tilde{D} = AD$ as the corresponding sufficient augmentation for $\mathscr{B}$, conditioning on other parameters.

The steps of the sampler for the z-jump Color Correction Model are:

Step 1. $\mathscr{C} \sim \mathcal{M}(\mathscr{C}|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}', z_t')$:
Use MH to sample $\mathscr{C}$ from $p(\mathscr{C}|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}', z_t')$, which is proportional to $p(\Sigma_D', \mathscr{B}', z_t', \mathscr{C}|\widehat{\mathscr{D}})$, under the constraint imposed by the priors;

Step 2. $z_t \sim \mathcal{M}(z_t|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}', \mathscr{C})$:
Use MH to sample z_t from $p(z_t|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}', \mathscr{C})$, which is proportional to $p(\Sigma_D', \mathscr{B}', z_t, \mathscr{C}|\widehat{\mathscr{D}})$, under the constraint $z_t \in [0.2, 1]$;

Step 3. $(D^I, D_\star) \sim p(D, D_\star|\widehat{\mathscr{D}}, \Sigma_D', \mathscr{B}', z_t, \mathscr{C})$:
This step consists of two substeps:
1. Sample $D_\star$ from $\mathcal{N}(k_\star, \Sigma_K)$, where $k_\star$ and Σ_K are defined in Table 7;
2. Sample D^I from $\mathcal{N}(\mu_A, \Sigma_A)$, where Σ_A is defined in Table 7 and $\mu_A = \Sigma_A(\Delta + \Sigma_D^{-1} J D_\star)$;

Step 4. $\mathscr{B}^I \sim p(\mathscr{B}|\widehat{\mathscr{D}}, D^I, D_\star, \Sigma_D', z_t, \mathscr{C})$:
Sample $\mathscr{B}^I$ from $\mathcal{N}(\zeta_B, \Sigma_B)$ with constraint $\mathscr{B}^I \in [-1, 0] \times [0, 4] \times [-1.5, 1.5]$. The construction of ζ_B and Σ_B is identical to that in the z-linear Color Correction sampler;
Use $\mathscr{B}^I$ and z_t to construct A^I. Then set $\tilde{D} = A^I D^I$;

Step 5. $\mathscr{B} \sim p(\mathscr{B}|\widehat{\mathscr{D}}, \tilde{D}, D_\star, \Sigma_D', z_t, \mathscr{C})$:
Sample $\mathscr{B}$ from $\mathcal{N}(\tilde{\zeta}_B, \tilde{\Sigma}_B)$ with constraint $\mathscr{B} \in [-1, 0] \times [0, 4] \times [-1.5, 1.5]$. The construction of $\tilde{\zeta}_B$ and $\tilde{\Sigma}_B$ is identical to that in the z-linear Color Correction sampler, except that under this model $\tilde{E}_i = \left\{-\tilde{x}_{1i}, -\tilde{c}_i, -\left(\frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\hat{z}_i - z_t}{0.01}\right)\right)\tilde{c}_i\right\}^T$;
Use $\mathscr{B}$ and z_t to construct A. Then set $D = A^{-1}\tilde{D}$;

Step 6. $\sigma_{\rm res} \sim p(\sigma_{\rm res}|\widehat{\mathscr{D}}, D, D_\star, R_{x_1}', R_c', \mathscr{B}, z_t, \mathscr{C})$:
Sample $\sigma_{\rm res}^2$ from InvGamma$\left[\frac{n}{2} + \lambda, \frac{\sum_{i=1}^n (M_i^\epsilon - M_0^\epsilon)^2}{2} + \lambda\right]$, and $\sigma_{\rm res} = \sqrt{\sigma_{\rm res}^2}$;

Step 7. $R_{x_1} \sim p(R_{x_1}|\widehat{\mathscr{D}}, D, D_\star, \sigma_{\rm res}, R_c', \mathscr{B}, z_t, \mathscr{C})$:
Sample $R_{x_1}^2$ from InvGamma$\left[\frac{n}{2}, \frac{\sum_{i=1}^n (x_{1i} - x_{1\star})^2}{2}\right]$ with $\log(R_{x_1}) \in [-5, 2]$, and $R_{x_1} = \sqrt{R_{x_1}^2}$;

Step 8. $R_c \sim p(R_c|\widehat{\mathscr{D}}, D, D_\star, \sigma_{\rm res}, R_{x_1}, \mathscr{B}, z_t, \mathscr{C})$:
Sample R_c^2 from InvGamma$\left[\frac{n}{2}, \frac{\sum_{i=1}^n (c_i - c_\star)^2}{2}\right]$ with $\log(R_c) \in [-5, 2]$, and $R_c = \sqrt{R_c^2}$.

C.1.4. Hard Classification Model of Host Galaxy Mass

In this model, we divide the population of SNe Ia into two classes according to host galaxy mass. The specification of $\hat{D}(\mathscr{C})$, D, $\Sigma_{\hat{D}}$, and A is identical to that in the Baseline Model. However, the specification of $D_\star$, Σ_D, $\Sigma_{D_\star}$, and J is changed to reflect the existence of two populations according to host galaxy mass. Under this model, $D_\star = \{(M_0^{\rm lo}), (M_0^{\rm hi}), x_{1\star}, c_\star\}^T$; $D_{\star\star} = \{-19.3, -19.3, 0, 0\}^T$; $\Sigma_D = {\rm diag}(S_1, \ldots, S_n)$, where $S_i = {\rm diag}[(1 - Z_i)(\sigma_{\rm res}^{\rm lo})^2 + Z_i(\sigma_{\rm res}^{\rm hi})^2, R_{x_1}^2, R_c^2]$; $\Sigma_{D_\star} = {\rm diag}(2^2, 2^2, 10^2, 1^2)$; $J_{(3n\times 4)} = \begin{bmatrix} J_1 \\ \vdots \\ J_n \end{bmatrix}$, where

$J_i = \begin{bmatrix} 1 - Z_i & Z_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. As stated in Appendix B, under this model, $Z = \{Z_1, ..., Z_n\}$ is assumed known with

$$Z_i = \begin{cases} 1 & \text{if } \widehat{M}_{g\,i} \geqslant 10 \\ 0 & \text{otherwise.} \end{cases} \quad (36)$$

The Sampler for the Hard Classification Model. This is also an MH within a PCG sampler, that is, we integrate $(D, D_\star)$ out when updating $\mathscr{C}$ and $\mathscr{B}$. The steps of the sampler are listed as follows.

Step 1. $\mathscr{C} \sim \mathcal{M}(\mathscr{C}|\widehat{\mathscr{D}}, \Sigma'_D, \mathscr{B}')$:
Use MH to sample $\mathscr{C}$ from $p(\mathscr{C}|\widehat{\mathscr{D}}, \Sigma'_D, \mathscr{B}')$, which is proportional to $p(\Sigma'_D, \mathscr{B}', \mathscr{C}|\widehat{\mathscr{D}})$, under the constraint imposed by the priors;

Step 2. $\mathscr{B} \sim \mathcal{M}(\mathscr{B}|\widehat{\mathscr{D}}, \Sigma'_D, \mathscr{C})$:
Use MH to sample $\mathscr{B}$ from $p(\mathscr{B}|\widehat{\mathscr{D}}, \Sigma'_D, \mathscr{C})$, which is proportional to $p(\Sigma'_D, \mathscr{B}, \mathscr{C}|\widehat{\mathscr{D}})$, under the constraint imposed by the priors;

Step 3. $(D, D_\star) \sim p(D, D_\star|\widehat{\mathscr{D}}, \Sigma'_D, \mathscr{B}, \mathscr{C})$:
This step consists of two substeps:
1. Sample $D_\star$ from $\mathcal{N}(k_\star, \Sigma_K)$, where $k_\star$ and Σ_K are defined in Table 7;
2. Sample D from $\mathcal{N}(\mu_A, \Sigma_A)$, where Σ_A is defined in Table 7 and $\mu_A = \Sigma_A(\Delta + \Sigma_D^{-1} J D_\star)$;

Step 4. $(\sigma_{\text{res}}^{\text{lo}}) \sim p((\sigma_{\text{res}}^{\text{lo}})|\widehat{\mathscr{D}}, D, D_\star, (\sigma_{\text{res}}^{\text{hi}})', R'_{x_1}, R'_c, \mathscr{B}, \mathscr{C})$:
Sample $(\sigma_{\text{res}}^{\text{lo}})^2$ from $\textsc{InvGamma}\left[\frac{\sum_{i=1}^n (1 - Z_i)}{2} + \lambda, \frac{\sum_{i=1}^n 1 - Z_i(M_i^\epsilon - (M_0^{\text{lo}}))^2}{2} + \lambda\right]$, and $(\sigma_{\text{res}}^{\text{lo}}) = \sqrt{(\sigma_{\text{res}}^{\text{lo}})^2}$;

Step 5. $(\sigma_{\text{res}}^{\text{hi}}) \sim p((\sigma_{\text{res}}^{\text{hi}})|\widehat{\mathscr{D}}, D, D_\star, (\sigma_{\text{res}}^{\text{lo}}), R'_{x_1}, R'_c, \mathscr{B}, \mathscr{C})$:
Sample $(\sigma_{\text{res}}^{\text{hi}})^2$ from $\textsc{InvGamma}\left[\frac{\sum_{i=1}^n (Z_i)}{2} + \lambda, \frac{\sum_{i=1}^n Z_i(M_i^\epsilon - (M_0^{\text{hi}}))^2}{2} + \lambda\right]$ and $(\sigma_{\text{res}}^{\text{hi}}) = \sqrt{(\sigma_{\text{res}}^{\text{hi}})^2}$;

Step 6. $R_{x_1} \sim p(R_{x_1}|\widehat{\mathscr{D}}, D, D_\star, (\sigma_{\text{res}}^{\text{lo}}), (\sigma_{\text{res}}^{\text{hi}}), R'_c, \mathscr{B}, \mathscr{C})$:
Sample $R_{x_1}^2$ from $\textsc{InvGamma}\left[\frac{n}{2}, \frac{\sum_{i=1}^n (x_{1i} - x_{1\star})^2}{2}\right]$ with $\log(R_{x_1}) \in [-5, 2]$, and $R_{x_1} = \sqrt{R_{x_1}^2}$;

Step 7. $R_c \sim p(R_c|\widehat{\mathscr{D}}, D, D_\star, (\sigma_{\text{res}}^{\text{lo}}), (\sigma_{\text{res}}^{\text{hi}}), R_{x_1}, \mathscr{B}, \mathscr{C})$:
Sample R_c^2 from $\textsc{InvGamma}\left[\frac{n}{2}, \frac{\sum_{i=1}^n (c_i - c_\star)^2}{2}\right]$ with $\log(R_c) \in [-5, 2]$, and $R_c = \sqrt{R_c^2}$.

C.1.5. Soft Classification Model of Host Galaxy Mass

In this model, the specification of $\hat{D}(\mathscr{C})$, D, $D_\star$, $D_{\star\star}$, Σ_D, $\Sigma_{D_\star}$, J, and A is identical to that in the Hard Classification Model. But here J is stochastic, since Z is stochastic.

The sampler for the Soft Classification Model. This is also an MH within a PCG sampler, that is, we integrate $(D, D_\star)$ out when updating $\mathscr{C}$ and $\mathscr{B}$. The steps of the sampler are:

Step 1. $Z \sim p(Z|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, D', D'_\star, \Sigma'_D, \mathscr{B}', \mathscr{C}')$:
For each i, sample Z_i from Bernoulli$(\tilde{p}_i)$, where

$\tilde{p}_i = \frac{p_{i,\text{high}}}{p_{i,\text{low}} + p_{i,\text{high}}}$, with

$$p_{i,\text{low}} = \frac{1}{(\sigma_{\text{res}}^{\text{lo}})'} \exp\left\{-\frac{[(M_i^\epsilon)' - (M_0^{\text{lo}})']^2}{2(\sigma_{\text{res}}^{\text{lo}})^{2'}}\right\} 1 - \pi_i \quad (37)$$

and

$$p_{i,\text{high}} = \frac{1}{(\sigma_{\text{res}}^{\text{hi}})'} \exp\left\{-\frac{[(M_i^\epsilon)' - (M_0^{\text{hi}})']^2}{2(\sigma_{\text{res}}^{\text{hi}})^{2'}}\right\} \pi_i; \quad (38)$$

π_i is defined in Appendix B;
Use Z to construct J, as in Table 7;

Step 2. $\mathscr{C} \sim \mathcal{M}(\mathscr{C}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, \Sigma'_D, \mathscr{B}', Z)$:
Use MH to sample $\mathscr{C}$ from $p(\mathscr{C}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, \Sigma'_D, \mathscr{B}', Z)$, which is proportional to $p(\Sigma'_D, \mathscr{B}', \mathscr{C}, Z|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g)$, under the constraint imposed by the priors;

Step 3. $\mathscr{B} \sim \mathcal{M}(\mathscr{B}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, \Sigma'_D, \mathscr{C}, Z)$:
Use MH to sample $\mathscr{B}$ from $p(\mathscr{B}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, \Sigma'_D, \mathscr{C}, Z)$, which is proportional to $p(\Sigma'_D, \mathscr{B}, \mathscr{C}, Z|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g)$, under the constraint imposed by the priors;

Step 4. $(D, D_\star) \sim p(D, D_\star|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, \Sigma'_D, \mathscr{B}, \mathscr{C}, Z)$:
This step consists of two substeps:
1. Sample $D_\star$ from $\mathcal{N}(k_\star, \Sigma_K)$, where $k_\star$ and Σ_K are defined in Table 7;
2. Sample D from $\mathcal{N}(\mu_A, \Sigma_A)$, where Σ_A is defined in Table 7 and $\mu_A = \Sigma_A(\Delta + \Sigma_D^{-1} J D_\star)$;

Step 5. $(\sigma_{\text{res}}^{\text{lo}}) \sim p((\sigma_{\text{res}}^{\text{lo}})|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, D, D_\star, (\sigma_{\text{res}}^{\text{hi}})', R'_{x_1}, R'_c, \mathscr{B}, \mathscr{C}, Z)$:
Sample $(\sigma_{\text{res}}^{\text{lo}})^2$ from $\textsc{InvGamma}\left[\frac{\sum_{i=1}^n (1 - Z_i)}{2} + \lambda, \frac{\sum_{i=1}^n 1 - Z_i(M_i^\epsilon - (M_0^{\text{lo}}))^2}{2} + \lambda\right]$, and $(\sigma_{\text{res}}^{\text{lo}}) = \sqrt{(\sigma_{\text{res}}^{\text{lo}})^2}$;

Step 6. $(\sigma_{\text{res}}^{\text{hi}}) \sim p((\sigma_{\text{res}}^{\text{hi}})|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, D, D_\star, (\sigma_{\text{res}}^{\text{lo}}), R'_{x_1}, R'_c, \mathscr{B}, \mathscr{C}, Z)$:
Sample $(\sigma_{\text{res}}^{\text{hi}})^2$ from $\textsc{InvGamma}\left[\frac{\sum_{i=1}^n Z_i}{2} + \lambda, \frac{\sum_{i=1}^n (Z_i)(M_i^\epsilon - (M_0^{\text{hi}}))^2}{2} + \lambda\right]$, and $(\sigma_{\text{res}}^{\text{hi}}) = \sqrt{(\sigma_{\text{res}}^{\text{hi}})^2}$;

Step 7. $R_{x_1} \sim p(R_{x_1}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, D, D_\star, (\sigma_{\text{res}}^{\text{lo}}), (\sigma_{\text{res}}^{\text{hi}}), R'_c, \mathscr{B}, \mathscr{C}, Z)$:
Sample $R_{x_1}^2$ from $\textsc{InvGamma}\left[\frac{n}{2}, \frac{\sum_{i=1}^n (x_{1i} - x_{1\star})^2}{2}\right]$ with $\log(R_{x_1}) \in [-5, 2]$, and $R_{x_1} = \sqrt{R_{x_1}^2}$;

Step 8. $R_c \sim p(R_c|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_g, D, D_\star, (\sigma_{\text{res}}^{\text{lo}}), (\sigma_{\text{res}}^{\text{hi}}), R_{x_1}, \mathscr{B}, \mathscr{C}, Z)$:
Sample R_c^2 from $\textsc{InvGamma}\left[\frac{n}{2}, \frac{\sum_{i=1}^n (c_i - c_\star)^2}{2}\right]$ with $\log(R_c) \in [-5, 2]$, and $R_c = \sqrt{R_c^2}$.

C.1.6. Covariate Adjustment Model of Host Galaxy Mass

In this model, since we include $M_{g\,i}$ as an additional covariate, the specification of quantities in the posterior distribution is different from the Baseline Model. First, $\hat{D}(\mathscr{C})$ is the combination of the $\widehat{\mathscr{D}}$ corrected for distance modulus and the host galaxy mass, $\widehat{\mathscr{D}}_g$, that is, $\hat{D}(\mathscr{C}) = \{\hat{D}(\mathscr{C})_1^T, ..., \hat{D}(\mathscr{C})_n^T\}^T$, where $\hat{D}(\mathscr{C})_i = \{\hat{m}_{Bi}^\star - \mu_i(\hat{z}_i, \mathscr{C}), \hat{x}_{1i}, \hat{c}_i, \widehat{M}_{g\,i}\}^T$. Moreover, $D = \{D_1^T, ..., D_n^T\}^T$, where $D_i = \{M_i^\epsilon, X_i^T\}^T$ with $X_i = \{x_{1i}, c_i, M_{g\,i}\}^T$; $D_\star = \{M_0^\epsilon, x_{1\star}, c_\star, M_{g\star}\}^T$; $D_{\star\star} = \{-19.3, 0, 0, 10\}^T$. For the variance–covariance matrices, $\Sigma_{\hat{D}}$ now has the dimension of $(4n \times 4n)$. The $(3n \times 3n)$ submatrix of $\Sigma_{\hat{D}}$, after deleting the $(4i)^{\text{th}} (i = 1, ..., n)$ rows and columns, is $(C_{\rm stat} + C_{\rm syst})$. The $(4i, 4i)^{\rm th}$ element of $\Sigma_{\hat{D}}$ is $\sigma^2_{{\rm g}\,i}$, while the other elements in the $(4i)^{\rm th}$ rows and columns are all zero, because we ignore correlations between $\widehat{M}_{{\rm g}\,i}$ and other observed quantities; $\Sigma_D = {\rm diag}(S_1,..., S_n)$, where each $S_i = {\rm diag}(\sigma^2_{\rm res}, R^2_{x_1}, R^2_c, R^2_{\rm g})$; $\Sigma_{D_\star} = {\rm diag}(2^2, 10^2, 1^2, 100^2)$. In addition, $J_{(4n\times4)} = \begin{bmatrix} J_1 \\ \vdots \\ J_n \end{bmatrix}$, where each J_i is a (4×4) identity matrix; $\mathscr{B} = \{-\alpha, \beta, \gamma\}^T$ and $A_{(4n\times4n)} = {\rm diag}(T_1,..., T_n)$, where each $T_i = \begin{bmatrix} 1 & -\alpha & \beta & \gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Because we include data on host galaxy mass, $\widehat{\mathscr{D}}_{\rm g}$, under this model, the complete and marginal posterior distributions should be written as $p(D, D_\star, \Sigma_D, \mathscr{B}, \mathscr{C}|\, \widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g})$ and $p(\Sigma_D, \mathscr{B}, \mathscr{C}|\, \widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g})$ respectively. But they are formally identical to (28) and (32), respectively.

The sampler for the Covariate Adjustment Model. This is also an MH within a PCG sampler, that is, we integrate $(D, D_\star)$ out when updating $\mathscr{C}$ and $\mathscr{B}$. Then the sampling of $\mathscr{C}$ and $\mathscr{B}$ needs the help of the MH algorithm. The steps of the sampler are listed below.

Step 1. $\mathscr{C} \sim \mathcal{M}(\mathscr{C}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g}, \Sigma'_D, \mathscr{B}')$:
Use MH to sample $\mathscr{C}$ from $p(\mathscr{C}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g}, \Sigma'_D, \mathscr{B}')$, which is proportional to $p(\Sigma'_D, \mathscr{B}', \mathscr{C}|\, \widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g})$, under the constraint imposed by the priors;

Step 2. $\mathscr{B} \sim \mathcal{M}(\mathscr{B}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g}, \Sigma'_D, \mathscr{C})$:
Use MH to sample $\mathscr{B}$ from $p(\mathscr{B}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g}, \Sigma'_D, \mathscr{C})$, which is proportional to $p(\Sigma'_D, \mathscr{B}, \mathscr{C}|\, \widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g})$, under the constraint imposed by the priors;

Step 3. $(D, D_\star) \sim p(D, D_\star|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g}, \Sigma'_D, \mathscr{B}, \mathscr{C})$:
This step consists of two substeps:
1. Sample $D_\star$ from $\mathcal{N}(k_\star, \Sigma_K)$, where $k_\star$ and Σ_K are defined in Table 7;
2. Sample D from $\mathcal{N}(\mu_A, \Sigma_A)$, where Σ_A is defined in Table 7 and $\mu_A = \Sigma_A(\Delta + \Sigma_D^{-1} J D_\star)$;

Step 4. $\sigma_{\rm res} \sim p(\sigma_{\rm res}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g}, D, D_\star, R'_{x_1}, R'_c, R'_{\rm g}, \mathscr{B}, \mathscr{C})$:
Sample $\sigma^2_{\rm res}$ from $\textsc{InvGamma}\left[\frac{n}{2} + \lambda, \frac{\sum_{i=1}^n (M_i^\epsilon - M_0^\epsilon)^2}{2} + \lambda\right]$, and $\sigma_{\rm res} = \sqrt{\sigma^2_{\rm res}}$;

Step 5. $R_{x_1} \sim p(R_{x_1}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g}, D, D_\star, \sigma_{\rm res}, R'_c, R'_{\rm g}, \mathscr{B}, \mathscr{C})$:
Sample $R^2_{x_1}$ from $\textsc{InvGamma}\left[\frac{n}{2}, \frac{\sum_{i=1}^n (x_{1i} - x_{1\star})^2}{2}\right]$ with $\log(R_{x_1}) \in [-5, 2]$, and $R_{x_1} = \sqrt{R^2_{x_1}}$;

Step 6. $R_c \sim p(R_c|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g}, D, D_\star, \sigma_{\rm res}, R_{x_1}, R'_{\rm g}, \mathscr{B}, \mathscr{C})$:
Sample R^2_c from $\textsc{InvGamma}\left[\frac{n}{2}, \frac{\sum_{i=1}^n (c_i - c_\star)^2}{2}\right]$ with $\log(R_c) \in [-5, 2]$, and $R_c = \sqrt{R^2_c}$;

Step 7. $R_{\rm g} \sim p(R_{\rm g}|\widehat{\mathscr{D}}, \widehat{\mathscr{D}}_{\rm g}, D, D_\star, \sigma_{\rm res}, R_{x_1}, R_c, \mathscr{B}, \mathscr{C})$:
Sample $R^2_{\rm g}$ from $\textsc{InvGamma}\left[\frac{n}{2}, \frac{\sum_{i=1}^n (M_{{\rm g}\,i} - M_{{\rm g}\star})^2}{2}\right]$ with $\log(R_{\rm g}) \in [-5, 2]$, and $R_{\rm g} = \sqrt{R^2_{\rm g}}$.

When MH updates are required in the samplers above, we use truncated normal distributions centered at the current draw with variance–covariance matrix adjusted to obtain an acceptance rate of around 40% (univariate) or 25% (multivariate as proposal distributions). Truncations are applied according to prior constraints.

PCG and ASIS show significant power in improving the convergence properties of $\mathscr{C}$ and $\mathscr{B}$. Although our PCG and ASIS samplers require 30%–50% more CPU time per iteration than our ordinary Gibbs samplers, their correlation lengths are smaller. For example, the effective sample size for the components of $\mathscr{C}$ is 5–6 times larger, and for the components of $\mathscr{B}$ is 3–4 times larger. See Jiao et al. (2015) for more numerical illustrations.

C.2. MH Samplers

We also use the MH algorithm to obtain samples of Σ_D, $\mathscr{B}$, $\mathscr{C}$, and (for the z-jump Color Correction Model) z_t from their joint posterior distribution under all the models, except the Soft Classification one, with the purpose of cross-checking the results obtained under the Gibbs-type samplers described above. The proposal distribution of the MH algorithm is a normal distribution centered at the current draw. For the variance–covariance matrix of the normal proposal distribution, we initially choose a diagonal matrix with randomly chosen entries. We then run a preliminary chain and use it to obtain an estimate of the variance–covariance matrix of the parameters. Finally we replace the variance–covariance matrix in the proposal distribution with this estimate and run the MH sampler to obtain posterior samples (ignoring the initial run when plotting the marginal distributions).

REFERENCES

Amanullah, R., Lidman, C., Rubin, D., et al. 2010, ApJ, 716, 712
Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31
Bailey, S., Aldering, G., Antilogus, P., et al. 2008, arXiv:0810.3499
Balland, C., Baumont, S., Basa, S., et al. 2009, A&A, 507, 85
Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22
Campbell, H., Fraser, M., & Gilmore, G. 2016, MNRAS, 457, 3470
Childress, M., Aldering, G., Antilogus, P., et al. 2013a, ApJ, 770, 107
Childress, M., Aldering, G., Antilogus, P., et al. 2013b, ApJ, 770, 108
Contreras, C., Hamuy, M., Phillips, M. M., et al. 2010, ApJ, 139, 519
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Freedman, W. L., Burns, C. R., Phillips, M., et al. 2009, ApJ, 704, 1036
Gelman, A., & Rubin, D. 1992, StaSc, 7, 457
Geman, S., & Geman, D. 1984, ITPAM, 6, 721
Gupta, R. R., DAndrea, C. B., Sako, M., et al. 2011, ApJ, 740, 92 (Erratum: ApJ 741, 127 (2011))
Guy, J., Astier, P., Nobili, S., Regnault, N., & Pain, R. 2005, A&A, 443, 781
Guy, J., Astier, P., Baumont, S., et al. 2007, A&A, 466, 11
Hastings, W. 1970, Biometrika, 57, 97
Hicken, M., Challis, P., Jha, S., et al. 2009, ApJ, 700, 331
Hlozek, R., Kunz, M., Bassett, B., et al. 2012, ApJ, 752, 79
Jha, S., Riess, A. G., & Kirshner, R. P. 2007, ApJ, 659, 122
Jiao, X., van Dyk, D. A., Trotta, R., & Shariff, H. 2015, in Proc. Joint Applied Statistics Symp. Int. Chinese Statistical Association & Korean Int. Statistical Society 2014 (Berlin: Springer)
Jones, D. O., Riess, A. G., & Scolnic, D. M. 2015, ApJ, 812, 31
Karpenka, N. V. 2015, PhD thesis, Stockholm University
Kass, R. E., Carlin, B. P., Gelman, A., & Neal, R. M. 1998, The American Statistician, 52, 93
Kelly, B. C. 2007, ApJ, 665, 1489
Kelly, P. L., Filippenko, A. V., Burke, D. L., et al. 2015, Sci, 347, 1459
Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, ApJ, 715, 743
Kessler, R., Becker, A. C., Cinabro, D., et al. 2009, ApJS, 185, 32

Knights, M., Bassett, B. A., Varughese, M., et al. 2013, JCAP, 1301, 039
Kowalski, M., Rubin, D., Aldering, G., et al. 2008, ApJ, 686, 749
Kunz, M., Bassett, B. A., & Hlozek, R. 2007, PhRvD, 75, 103508
Lampeitl, H., Nichol, R. C., Seo, H.-J., et al. 2010, MNRAS, 401, 2331
Li, W., Bloom, J. S., Podsiadlowski, P., et al. 2011a, Natur, 480, 348
Li, W., Leaman, J., Chornock, R., et al. 2011b, MNRAS, 412, 1441
Liu, J. S. 2001, Monte Carlo Strategies in Scientific Computing (New York: Springer)
Liu, J. S., Wong, W. H., & Kong, A. 1994, Biometrika, 81, 27
Liu, J. S., Wong, W. H., & Kong, A. 1995, Journal of the Royal Statistical Society. Series B (Methodological), 57, 157
Mandel, K. S., Narayan, G., & Kirshner, R. P. 2011, ApJ, 731, 120
Mandel, K. S., Wood-Vasey, W. M., Friedman, A. S., & Kirshner, R. P. 2009, ApJ, 704, 629
Mannucci, F., Della Valle, M., & Panagia, N. 2006, MNRAS, 370, 773
March, M. C., Karpenka, N. V., Feroz, F., & Hobson, M. P. 2014, MNRAS, 437, 3298
March, M. C., Trotta, R., Berkes, P., Starkman, G. D., & Vaudrevange, P. M. 2011, MNRAS, 418, 2308
Menard, B., Kilbinger, M., & Scranton, R. 2010a, MNRAS, 406, 1815
Menard, B., Scranton, R., Fukugita, M., & Richards, G. 2010b, MNRAS, 405, 1025
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, JChPh, 21, 1087
Meyers, J., Aldering, G., Barbary, K., et al. 2012, ApJ, 750, 1
Mohlabeng, G. M., & Ralston, J. P. 2013, MNRAS, 439, L16
Mosher, J., Guy, J., Kessler, R., et al. 2014, ApJ, 793, 16
Nielsen, J. T., Guffanti, A., & Sarkar, S. 2015, arXiv:1506.01354
Nugent, P. E., Sullivan, M., Cenko, S. B., et al. 2011, Natur, 480, 344
Pan, Y.-C., Sullivan, M., Maguire, K., et al. 2014, MNRAS, 438, 1391
Park, T., & van Dyk, D. A. 2009, Journal of Computational and Graphical Statistics, 18, 283
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Phillips, M. 1993, ApJL, 413, L105
Phillips, M., Lira, P., Suntzeff, N., et al. 1999, AJ, 118, 1766
Planck Collaboration et al. 2015, arXiv:1502.01589
Rest, A., Scolnic, D., Foley, R., et al. 2014, ApJ, 795, 44
Riess, A. G., Press, W. H., & Kirshner, R. P. 1996, ApJ, 473, 88
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Rigault, M., Copin, Y., Aldering, G., et al. 2013, A&A, 560, A66
Rigault, M., Aldering, G., Kowalski, M., et al. 2015, ApJ, 802, 20
Rubin, D., Aldering, G., Barbary, K., et al. 2015, ApJ, 813, 137
Schaefer, B. E., & Pagnotta, A. 2012, arXiv:1201.2195
Sullivan, M., Le Borgne, D., Pritchet, C. J., et al. 2006, ApJ, 648, 868
Sullivan, M., Conley, A., Howell, D. A., et al. 2010, MNRAS, 406, 782
Suzuki, N., Rubin, D., Lidman, C., et al. 2012, ApJ, 746, 85
Tanner, M. A., & Wong, W. H. 1987, J. Am. Stat. Assoc., 82, 528
van Dyk, D. A., & Jiao, X. 2015, Journal of Computational and Graphical Statistics, 24, 301
van Dyk, D. A., & Park, T. 2008, J. Am. Stat. Assoc., 103, 790
Voss, R., & Nelemans, G. 2008, Natur, 451, 802
Wang, S., Li, Y.-H., & Zhang, X. 2014, PhRvD, 89, 063524
Wang, S., & Wang, Y. 2013, PhRvD, 88, 043511
Wood-Vasey, W. M., Miknaitis, G., & Stubbs, C. W. 2007, ApJ, 666, 694
Yu, Y., & Meng, X.-L. 2011, Journal of Computational and Graphical Statistics, 20, 531