Type Ia Supernovae, Evolution, and the Cosmological Constant

Persis S. Drell,[1] Thomas J. Loredo,[2] and Ira Wasserman[1,2]

ABSTRACT

We explore the possible role of evolution in the analysis of data on SNe Ia at cosmological distances. First, using a variety of simple sleuthing techniques, we find evidence that the properties of the high and low redshift SNe Ia observed so far differ from one another. Next, we examine the effects of including simple phenomenological models for evolution in the analysis. The result is that cosmological models and evolution are highly degenerate with one another, so that the incorporation of even very simple models for evolution makes it virtually impossible to pin down the values of Ω_M and Ω_Λ, the density parameters for nonrelativistic matter and for the cosmological constant, respectively. We evaluate the Bayes factor for models with evolution versus models without evolution, which, if one has no prior predilection for or against evolution, is the odds ratio for these two classes of models. The resulting values are always of order one, in spite of the fact that the models that include evolution have additional parameters; thus, the data alone cannot discriminate between the two possibilities. Simulations show that simply acquiring more data of the same type as are available now will not alleviate the difficulty of separating evolution from cosmology in the analysis. What is needed is a better physical understanding of the SN Ia process, and the connections among the maximum luminosity, rate of decline and initial conditions, so that physical models for evolution may be constructed, and confronted with the data. Moreover, we show that if SNe Ia evolve with time, but evolution is neglected in analyzing data, then, given enough SNe Ia, the analysis hones in on values of Ω_M and Ω_Λ which are incorrect. Using Bayesian methods, we show that the probability that the cosmological constant is nonzero (rather than zero) is unchanged by the SNe Ia data when one accounts for the possibility of evolution, provided that we do not discriminate among open, closed and flat cosmologies a priori. The case for nonzero cosmological constant is stronger if the Universe is presumed to be flat, but still depends sensitively on the degree to which the peak luminosities of SNe Ia evolve as a function of redshift.

[1]Floyd R. Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, NY 14853-5001

[2]Center for Radiophysics and Space Research, Cornell University, Ithaca, NY 14853-6801

Subject headings: cosmology: observations — distance scale — statistics — supernovae: general

1. Introduction

The realization that the rates of decline of the brightnesses of Type Ia supernovae (SNe Ia) are correlated with their peak luminosities (Phillips 1992) has led to renewed efforts to use them as cosmological distance markers (Hamuy et al. 1995, 1996a, 1996b; Riess, Press and Kirshner 1995, 1996). Ongoing searches for high redshift ($z \sim 0.5 - 1$) SNe Ia have employed phenomenological models for these correlations to constrain the variation of luminosity distance $D_L(z)$ with redshift; the results have been interpreted to imply the existence of a nonzero cosmological constant (Perlmutter et al. 1998 hereafter P98, Riess et al. 1998 hereafter R98). Moreover, the results appear to rule out the simplest version of a flat cosmology, in which the density parameter for "ordinary" matter (including as yet unidentified nonbaryonic material) $\Omega_M = 1$ and the density parameter for the cosmological constant $\Omega_\Lambda = 0$.

Although the logical possibility that $\Omega_\Lambda \sim 1$ today has long been recognized (Einstein 1917), it is anathema to many theorists, since the associated vacuum energy density must be $\rho_{vac} \sim 10^{-122} M_{\rm P}^4$, where $M_{\rm P}$ is the Planck mass. (Theoretical and conceptual problems with a nonzero cosmological constant so small compared with its "natural" scale have been reviewed by Weinberg 1989 and Carroll, Press & Turner 1992.) On the other hand, there is some evidence from large scale structure simulations that $\Omega_\Lambda \neq 0$ in a flat Universe fits the observations well (Cen 1998). Conceivably, what we interpret as a cosmological "constant" might be an evolving field (e.g., Caldwell, Dave & Steinhardt 1998, Garnavich et al. 1998, Perlmutter, Turner & White 1999). In any case, a convincing demonstration that the expansion rate of the Universe is increasing would have a revolutionary impact on our understanding of fundamental physics.

In view of the importance of the potential discovery of a nonzero cosmological constant, we have undertaken an independent study of the published data in an effort to understand their implications better. Our motivations are both phenomenological and astrophysical (and may end up being related ultimately). On the phenomenological side, we note that three different analysis methods are used to compute distances, the multicolor light curve shape (MLCS) method (Riess, Press & Kirshner 1995, 1996), the M15 or template fitting (TF) method (Phillips 1992; Hamuy et al. 1995, 1996a, 1996b), and the stretch factor (SF) method (Perlmutter et al. 1997). None of these methods is a perfect description of reality. As we will show, they are not always in agreement, and there seems to be no physical or phenomenological reason to prefer one to the other.

On the astrophysical side, we note that there are processes such as evolution of the supernovae sample that can mimic the effects of cosmology at high redshifts and which are extremely difficult to constrain convincingly with the current data. Therefore, it is useful to ask at what level the current data are able to distinguish the effects of cosmology from these other processes. We find that a redshift-dependent shift in SNe Ia peak magnitudes (such that the most distant observed SNe Ia are dimmed by ≈ 0.2 to 0.3 mag) renders $\Omega_M = 1$ and $\Omega_\Lambda = 0$ acceptable, and that this is true for a variety of phenomenological models for the evolution. We also present results of simulations that show that if SNe Ia luminosities evolve with redshift, but evolution is neglected in analyzing the data, then, given enough data, the analyses will settle on precisely determined, but incorrect, values for Ω_M and Ω_Λ, and that the incorrectness of the model will not be detectable with a standard χ^2 goodness-of-fit test.

We believe that it is unjustifiable to try to determine cosmological parameters alone from data on "standardized" candles, such as the peak luminosities of SNe Ia, without taking source evolution into account. Our attitude is that evolution must be assumed to occur, as a default. Hopefully, one can demonstrate from the data that source evolution is absent or negligible. [3] If that turns out to be untrue, one might hope instead to constrain the parameters in an evolutionary model along with the cosmological parameters. Optimistically, one would anticipate that this might be accomplished once enough data are acquired. We argue, from simulations employing simple, phenomenological models, that such optimism may be unrealistic. What is needed is a better physical understanding of the SN Ia process and its evolution with redshift, before cosmological parameters can be determined reliably from SNe Ia catalogues. Such an understanding is currently being sought by theorists (see, e.g., Höflich, Wheeler, and Thielemann 1998; Domínguez et al. 1999).

In using observations of SNe Ia to determine cosmological parameters, the raw data are combined by any of the three methods mentioned above to derive single parameter summaries – the distance moduli to the sources. When a single catalogue of data is

---

[3]The observers have employed supplementary measurements—such as source spectra—to argue that there is no compelling evidence *for* evolution in the SNe Ia samples. If the physical connection between the additional data and the estimators of peak luminosity can be understood quantitatively, then the supplementary measurements can be incorporated usefully into analyses of models with evolution. Without such physical understanding, these additional data cannot be marshalled to argue *against* evolution.

subjected to different types of analysis, each of which derives one quantity per source, the results of the individual analyses need not agree with one another entirely, and there is information contained in the degree to which the answers derived by the different methods differ[4]. In the example under consideration, the different analysis methods might probe slightly different physical aspects of the SN Ia mechanism, and their relationships to reality and to one another may be different at high and low redshift. Indeed, one clue that evolutionary effects are important would be a systematic drift with increasing redshift between the distance moduli implied by the MLCS and TF methods for the SNe Ia observed and analyzed by R98 where identical SNe Ia data are subjected to two different analysis methods.

In order to set the scale of interest for our investigations of potential systematic effects in these data we plot, in Figure 1, the joint credible regions for Ω_M and Ω_Λ for the largest available data set (P98), analyzed as published (Figure 1a) and after introducing a systematic offset to all the high redshift distance moduli ($z > 0.15$) of -0.1 magnitudes (Figure 1b). We see that a correlated systematic shift of this size would have a major impact on the interpretation of the data. A nonzero cosmological constant would still be favored, but the statistical significance of the result would be much reduced.

In Section 2 we review the published data that we employ in our study, as well as the salient features of the light curve fitting methods. In Section 3 we compare the different fitting methods on a supernova by supernova basis where possible. In Section 4 we explore whether the data have sufficient shape information to distinguish effects of cosmology from other cosmological effects such as evolution. We summarize our conclusions in Section 5.

2. Measurements of Ω_M and Ω_Λ Using Type Ia Supernovae

The traditional measure of distance to a SN is its distance modulus, $\mu \equiv m_{\rm bol} - M_{\rm bol}$, the difference between its bolometric apparent magnitude, $m_{\rm bol}$, and its bolometric absolute magnitude, $M_{\rm bol}$. In the Friedman-Robertson-Walker (FRW) cosmology, when the (relative) peculiar velocity of the source is negligible, the distance modulus is determined by the source's redshift, z, according to

$$\mu = 5 \log \left[\frac{D_L(z; \Omega_M, \Omega_\Lambda, H_0)}{1 \ {\rm Mpc}} \right] + 25$$

[4]A simple example might involve computing mean values of data using similar, but not identical, weighting functions. Each weighted mean would then be a slightly different superposition of all of the moments of the data computed with one fiducial weighting function.

$$\equiv \quad f(z; \Omega_M, \Omega_\Lambda, H_0). \tag{1}$$

Here the luminosity distance $D_L(z; \Omega_M, \Omega_\Lambda, H_0) = cH_0^{-1} d_L(z; \Omega_M, \Omega_\Lambda)$, where c is the velocity of light, H_0 is Hubble's constant at the present epoch, and the dimensionless luminosity distance from redshift z is

$$d_L(z; \Omega_M, \Omega_\Lambda) = (1+z)|\Omega_k|^{-1/2} \operatorname{sinn}\{|\Omega_k|^{1/2} \int_0^z dz[(1+z)^2(1+\Omega_M z) - z(2+z)\Omega_\Lambda]^{-1/2}\}, \tag{2}$$

with $\Omega_k = 1 - \Omega_M - \Omega_\Lambda$, and $\operatorname{sinn}(x) = \sinh(x)$ for $\Omega_k \geq 0$ and $\sin(x)$ for $\Omega_k \leq 0$ (e.g., Carroll, Press & Turner 1992). In principle, one could infer the cosmological parameters H_0, Ω_M, and Ω_Λ from the distribution of measured distance moduli of sources at a variety of redshifts.

Several factors complicate implementation of such an analysis. In reality, bolometric data are not available, and one must infer μ using magnitudes m_X and M_X in some bandpass, X. The bandpass maps to a different region of the spectrum as a function of redshift z, so μ cannot be calculated simply by taking the difference between band-limited magnitudes; a K-correction term must be added whose value depends not only on the source's redshift, but also on its spectrum. In addition, extinction along the line of sight increases the apparent magnitude by some amount A_X not due to the cosmological effects modelled in equation (1). Further, the absolute magnitude of the source—bolometric or band-limited—is not directly measured, but must instead be inferred from other source properties. Finally, the inevitable presence of statistical uncertainties and peculiar velocities further complicates straightforward use of equation (1).

Of these complications, the need to infer the absolute magnitude indirectly is the most troublesome. Ideally, one seeks a population of "standard candles" such that all members of the population have the same M (for convenience we henceforth drop the subscript X). If such a population could be identified, the parameters Ω_M and Ω_Λ could be inferred even if the actual value of M for the population were unknown (the remaining parameter, H_0, would remain undetermined). Historically, all attempts to identify such a population have failed. Particularly worrisome is the possibility that some classes of objects that appear to be approximately standard candles locally (at low redshift, where they can be studied in detail) have evolved significantly, so that their younger counterparts at high redshifts have different absolute magnitudes, thwarting their use as cosmological distance indicators.

SNe Ia were briefly considered promising candidates for standard candles, but observers quickly discovered that SNe Ia are not all identically bright (Branch, 1987; Barbon, Rosino, & Iijima 1989; Phillips et al. 1987, 1992; Filippenko, et al. 1992a, 1992b; Leibundgut, et al. 1993). The intrinsic dispersion in the peak absolute magnitudes of SNe Ia, determined from studies of nearby events, is approximately 0.3 - 0.5 mag (Schmidt et al. 1998). However, there is an apparent empirical correlation between the rate of decline of the light curve of a given SN Ia and its luminosity at maximum brightness that was first quantified by Phillips (1992). Various techniques have been developed to take advantage of this correlation to determine the absolute magnitudes of individual supernovae using their light curves (Phillips 1992; Hamuy et al. 1995, 1996a, 1996b; Riess, Press & Kirshner 1995, 1996; Perlmutter et al. 1997); the relationships used in these analyses have come to be known generically as "Phillips relations." When applied to nearby SNe Ia, these methods reduce the dispersion of the distance moduli about the low-z FRW distance modulus vs. redshift relation to ≈ 0.15 (Hamuy et al. 1996a, Riess, Press & Kirshner 1996).

The goal of the high redshift supernovae searches is to observe a large sample of supernovae at relatively large z, and understand their properties well enough to infer reliable distance moduli for them, allowing accurate determination of cosmological parameters. Two experimental groups have recently announced and published results from their independent programs to discover and study high redshift supernovae for this purpose (Perlmutter 1997 and P98; R98). The resulting two data sets share many low redshift SNe discovered by previous surveys, but include different high redshift SNe, and differ in their analysis methods. We take advantage of the similarities and differences among the data and methods used to assess the consistency or inconsistency of the assumptions underlying the analyses.

P98 have published data on 60 SNe Ia. Of these, 18 were discovered and measured in the Calán-Tololo survey (all at low redshift; Hamuy et al. 1996c), and this group discovered 42 new SNe Ia at redshifts between 0.17 and 0.83. The μ values are inferred using the SF light curve fitting method and are typically uncertain to ± 0.2 magnitudes ("1σ"). The SF method (Perlmutter et al. 1997; P98) is based on fitting a time-stretched version of a single standard template to the observed light curves. The stretch factor, s, is then used to estimate the absolute magnitude of the SNe Ia via a linear relationship that is determined jointly with the cosmological parameters. The quoted μ values include a correction for extinction in the Galaxy based on the detailed model of Burstein & Heiles (1982).

R98 have published results based on 50 SNe Ia. Of these, 37, including 27 at low redshift ($z < 0.15$) and ten at high redshift ($z > 0.15$) have well-sampled light curves in addition to spectroscopic information; the quoted "1σ" uncertainties for μ for these SNe Ia are typically smaller than ± 0.2 mag at high z for determinations by either MLCS or TF light curve fitting method. The data for 17 of the SNe Ia at low redshift come from the Calán-Tololo survey (Hamuy et al. 1996c). We focus our attention on these 37 best-observed SNe Ia, which dominate the analysis in R98. These authors extensively tabulate their reduced data and provide detailed information about their fitting techniques, thus facilitating independent analysis of their conclusions.

R98 employ two different methods to estimate the distance modulus based on information from the light curves. The TF method (Hamuy et al. 1996a) fits a set of light curve templates with different values of Δm_{15}, the total decline in brightness from peak to 15 days afterward, to observations of a particular SN Ia. By interpolating between the values of χ^2 for the fits to the various templates, a minimum χ^2 value of Δm_{15} for the SN Ia is estimated. The peak absolute magnitude is deduced from the independently calibrated linear relationship between M and Δm_{15}. The MLCS method consists of fitting an observed light curve to a superposition of a standard light curve and weighted additional templates that parametrize the differences among SNe Ia (e.g., Riess, Press & Kirshner 1996); the outcome of the fits for a particular SN Ia consists of the weights associated with the deviations from the standard, which in turn determine the difference between its peak absolute magnitude and the standard's. The fits are done for more than one color, and reddening and extinction are inferred from color dependences (Riess, Press & Kirshner 1996, R98). Originally the MLCS method used a rather small training set to determine the requisite templates (Riess, Press & Kirshner 1996), but R98 now train on a considerably larger set of nearby SNe Ia to find them. Both the MLCS and TF methods are calibrated on nearby SNe Ia in the Hubble flow ($z < 0.15$) and then applied to the SNe Ia discovered at high redshift. The quoted μ values include a correction for local extinction derived from the Burstein and Heiles model, and in addition the MLCS method uses color dependence to estimate corrections for the extinction and reddening due to absorbing material in the host galaxy.

Schematically, we can consider a lightcurve fitting method to estimate the distance modulus for supernova number i according to the following model:

$$\mu_i = m_i - (M_0 + \Delta_i). \tag{3}$$

Here m_i is the peak apparent magnitude for the SN, M_0 is a fiducial absolute magnitude (a single constant for a particular method), and Δ_i is a shift so that the peak absolute magnitude for the SN is given by $(M_0 + \Delta_i)$. For the purposes of this paper, we have ignored K-corrections and extinction in equation (3) (one can consider them to have been accounted for in the m_i estimates and their uncertainties). These corrections could potentially be important sources of systematic error; the observing teams have gone to some lengths to constrain the sizes of such errors. Here we concentrate on the possibility that systematic errors are introduced by the lightcurve fitting algorithms entering the analyses via the shifts Δ_i. We will seek information about such errors by comparing the shifts *across* methods, rather than through analysis of the internal consistency of a particular method.

The various fitting methods differ in how m_i is interpolated from the observed (incompletely sampled) lightcurve, in the choice for M_0, and in how the shifts Δ_i are determined from the (multicolor) lightcurve shapes. The MLCS method provides Δ_i directly from fitting to a family of templates parameterized by Δ_i. For the TF method, $\Delta_i = \beta_{\mathrm{TF}}(\Delta m_{15} - 1.1)$, with β_{TF} a constant determined by fits. For both the MLCS and TF methods, M_0 is inferred through the use of SNe Ia that have Cepheid distances, and the various parameters specifying the shift as a function of the lightcurve shape are set by analyses of low redshift SNe. For the SF method, $\Delta_i = \alpha(s-1)$, where s is the above-mentioned stretch factor and α is a constant estimated jointly with cosmological parameters in fits to the entire survey. M_0 is simply set at an arbitrary value; accordingly, no attempt is made to infer H_0 using SNe analyzed with the SF method. In principle, each of the quantities on the right hand side of equation (3) has uncertainty associated with it, and the resulting errors in the estimates for these quantities are correlated. But only the combination given by equation (3) appears in cosmological fits, so the lightcurve fitting results can be summarized by the best-fit absolute magnitude $\hat{\mu}_i$ and the total μ_i uncertainty σ_i for each SN. These quantities, and the shifts Δ_i, are the focus of our study.

Figure 2 shows histograms of the shifts deduced from the MLCS (R98), TF (R98) and SF (P98) methods for the observed SNe Ia. Since the choice of M_0 can vary from method to method, we do not expect the histograms to be aligned. However, differences in histogram shape would indicate that the various methods are correcting SNe in different and possibly inconsistent ways. While the three methods claim to reduce the dispersion in the magnitude-redshift relationship at low z, it is clear from the figure that they produce rather different distributions of shifts. Although the SF method has been applied to a different set of SNe Ia than the other methods, this alone cannot explain the obvious differences between the shapes of the histograms (we note that 14 SNe are common to all three methods). Most striking is that the distribution is extremely narrow for the SF method, indicating that, by this measure, the P98 SNe Ia sample consists nearly entirely of standard candles, or that for this sample of SNe Ia, the adopted brightness-decline rate relation is not valid. This suggests to us that these methods may be sensitive to different aspects of the SN Ia phenomenon. A consequence of this is that if the properties of SNe Ia change with redshift, the relationships between the μ estimates produced by the three methods could be z-dependent. A search for such a dependence could thus provide information about redshift dependence of SNe Ia properties. In the following section we use exploratory methods to search for evidence of this and other kinds of dependences.

3. Sleuthing

Our approach in this section is driven by our belief that it is not sufficient to settle for the consistency of the final cosmological inferences of the MLCS, TF and SF analyses. We should expect consistency between them (statistically) on a supernova-by-supernova basis where such a comparison is possible.

Since R98 use two different methods to compute distance moduli for their sample of 37 SNe Ia, we can compare the results and search for systematic differences between them. Both the TF and MLCS techniques are calibrated using the same set of nearby SNe Ia in the Hubble flow and there is only one set of observational data for each SN Ia; consequently, the uncertainties for the two methods are highly correlated. Another comparison set is the group of 14 supernovae from the Calán Tololo survey that are included in both R98 and P98. Since all fitting methods make use of the same published light curves for this sample of 14 events, the inferred quantities for this sample will also be highly correlated.

3.1. Pointwise Consistency

In Figure 3, we compare the distance moduli measured with the different techniques on common samples of SNe Ia. In Figure 3a, we show $\mu_{\rm MLCS}$ vs $\mu_{\rm TF}$ for the 10 high redshift SNe Ia analyzed in R98, and in Figure 3b, we show $\mu_{\rm MLCS}$ vs $\mu_{\rm SF}$ for the 14 common Calán-Tololo SNe Ia analyzed by both SF and MLCS methods. The error bars for μ are derived from the uncertainties in the individual distance moduli for each supernova (R98,P98), except that we have removed the contribution associated with the intrinsic dispersion of the SNe Ia sample, estimated to be $\sigma_{\rm int} = 0.10$ at low z, $\sigma_{\rm int} = 0.15$ at high z, [5] and we have removed the contribution associated with the peculiar velocity of the SNe Ia ($\sigma_v = 300$ in P98, $\sigma_v = 200$ in R98). Both the errors in the distance modulus from intrinsic dispersion in the sample and from the peculiar velocity of the SNe are completely correlated among the different methods. We have not removed the correlations due to K-corrections, photometry and extinction (e.g., Schmidt et al. 1998), because there is insufficient published information for us to do so properly; consequently, we have overestimated the uncorrelated portion of the distance modulus error somewhat.

From Figure 3 it is clear that the estimates for the distance moduli from the different methods are strongly correlated, as they should be. However, there is more dispersion in

---

[5] The error due to intrinsic dispersion in the SNe Ia sample is estimated to be somewhat larger in P98; $\sigma_{int} = 0.17$. For the purposes of comparison we remove the smaller estimated value of the correlated error.

these plots than we would expect based on the quoted errors. A fit of a straight line of slope 1 gives a χ^2/ν (with ν the number of degrees of freedom) of 22.8/9 for Figure 3a and 21.2/13 for Figure 3b indicating that there are errors associated with the analysis methods that have not been accounted for.

We can pursue this type of comparison further with the R98 data where all of the SNe Ia have been fully analyzed with two independent methods. In Figure 4 we compare the MLCS and TF estimates of various quantities that are used in inferring the distance moduli of the SNe Ia events. For the 37 SNe 1a analyzed in R98, Figure 4*a* shows the host galaxy extinction, A, 4*b* shows the correction to the absolute magnitude, Δ, and 4*c* illustrates the peak apparent magnitude, m, calculated with the MLCS and TF analysis methods. (The individual errors for the extinction and Δ are not published but can be crudely estimated to be of order 0.1 magnitudes.) Again, there is more dispersion evident in these plots than might be expected from the quoted or estimated errors except for the correlation plot of $m_{\rm MLCS}$ versus $m_{\rm TF}$. The peak apparent magnitudes inferred via the two methods, which are the quantities most directly related to the raw data, are in excellent agreement.

3.2. Redshift and Luminosity Dependence

In Figure 5, we plot the difference $\Delta\mu \equiv \mu_{\rm MLCS} - \mu_{\rm TF}$ between the distance moduli determined from MLCS and TF respectively, as a function of z. The error bars for $\Delta\mu$ are derived from the uncertainties in the individual distance moduli except that, as described above, we have removed the contribution associated with the intrinsic dispersion of the SNe Ia sample. Formally, we use $\sigma^2_{\Delta\mu} = \sigma^2_{\rm MLCS} + \sigma^2_{\rm TF} - 2\sigma^2_{\rm corr}$ to calculate the error bars shown in Figure 5. Although the data are somewhat scattered at both high and low z, Figure 5 shows that the MLCS and TF methods agree rather well at low z, apart from significant dispersion ($\sigma \sim 0.2$ mag) but there are hints of disagreement at large z, where the dispersion, at least, appears larger, and the mean may also be shifted.

While it is possible that the appearance of Figure 5 at large z merely reflects small number statistics, Figure 6 suggests that the incompatibility between TF and MLCS could be systematic. In Figure 6, we plot $\Delta\mu$ versus $M_B^{\rm AV}$, an estimated absolute magnitude, defined by

$$M_B^{\rm AV} = (M_B^{\rm MLCS} + M_B^{\rm TF})/2 \tag{4}$$

where

$$M_B^{\rm MLCS} = m_B^{\rm MLCS} - \mu_{\rm MLCS} - A_B^{\rm MLCS}$$

$$M_B^{\rm TF} \;=\; m_B^{\rm TF} - \mu_{\rm TF} - A_B^{\rm TF} \tag{5}$$

and $A_B^{\rm MLCS(TF)}$ is an estimate of the extinction due to the host galaxy in the MLCS (TF) correction scheme. For $z > 0.15$, R98 provided all of the information necessary to calculate $M_B^{\rm MLCS}$, $M_B^{\rm TF}$ and hence $M_B^{\rm AV}$, but at low z, M_B was only given for a subset of the SNe Ia in Hamuy et al. (1996a). [6]

According to Figure 6, the difference between $\mu_{\rm MLCS}$ and $\mu_{\rm TF}$ appears to be correlated with the estimated intrinsic brightness, $M_B^{\rm AV}$, at high z, but not at low z. (Recall that the error bars on $\Delta\mu$ are overestimates, as explained above.) A similar correlation is evident if $\Delta\mu$ is plotted against $\Delta_{\rm MLCS}(\Delta_{\rm TF})$, the difference in maximum absolute magnitudes for the observed SNe Ia and the fiducial SNe Ia according to the MLCS(TF) method. (R98 tabulates $\Delta_{\rm MLCS}$ and $\Delta_{\rm TF}$ for all SNe Ia in their sample.) Figure 6 suggests that, at high z, one of the analysis schemes, MLCS or TF, either under-corrects or over-corrects for the luminosity variations in the SNe Ia sample. Since no such systematic trend is evident at low z, we cannot know which method, if either, yields the more accurate value of distance modulus. It is relatively uncontroversial to say that the two methods are not identical, either at low or high z, and hence must probe SNe Ia physics in slightly different, and as yet ill-understood ways (Höflich and Kholkov 1996, Höflich, Wheeler and Thielemann 1998). Thus, the indications of z-dependence implied by Figures 5 and 6, while still based on relatively few events, are not especially surprising.

A worrisome feature of Figures 5 and 6 is that imperfect corrections for luminosity variations can alter the conclusions about Ω_M and Ω_Λ that we draw from these data. To illustrate this point, we computed 1σ confidence contours in $(\Omega_\Lambda, \Omega_M)$ space with separate fits to intrinsically bright and intrinsically dim SNe Ia; the results are shown in Figure 7. [7] The separation into 'bright' and 'dim' was somewhat arbitrary, and we have verified that making different choices does not affect the overall conclusion. [8] The 1σ confidence level contours for the combined data (all M_B^{AV}) are also shown as dashed curves. Figure

---

[6]The zero point reference for M_B may be somewhat different for the high redshift and low redshift data which may account for the fact that in Figure 6, the low redshift supernovae seem to be, on average, less luminous by about 0.5 magnitudes. Our conclusions are robust against a shift in the zero point of the magnitude scale for the low redshift supernovae.

[7]In preparing Figure 7, we have included a contribution to the uncertainty arising from dispersion in galaxy redshift using the technique described in R98.

[8]For the plots shown, we have chosen $M_B^{AV} < -19.45$ as intrinsically bright and $M_B^{AV} > -19.45$ as intrinsically dim for MLCS and TF data. For the SF data, we separated intrinsically bright from intrinsically dim using $\alpha(s-1) > 0$ and $\alpha(s-1) < 0$, respectively. Note that $\alpha(s-1)$ can be calculated from the information in Tables 1 and 2 of P98.

7 indicates a systematic difference between the cosmology favored by intrinsically bright versus intrinsically dim SNe Ia when the MLCS method is used; the effect is much less pronounced for TF and seems to be of the opposite sign for the SF method. The trend may be understood if the MLCS (SF) method tends to overestimate (underestimate) the luminosities of intrinsically bright SNe Ia at high redshift. Such a trend for the MLCS data is also consistent with Figure 6.

The set of plots in Figures 2 through 7 indicate that the analysis methods disagree in their inferences of μ, A, and Δ at a level that is not covered by the quoted errors. We can only speculate on the sources of the discrepancies. However until these methods are understood more systematically, it will be difficult to avoid assigning additional systematic errors to the measured distance moduli with sizes that reflect the systematic differences between the methods, and this will weaken the statistical significance of the results substantially.

3.3. Validity of Phillips Relations

So far we have been investigating the light curve fitting methods as possible sources of systematic error. Potentially, there are other effects that can mimic cosmology that are extremely difficult to constrain with the present data. The most pernicious, discussed at some length by the observers themselves, is evolution of the SN Ia population. It is extremely difficult to put reliable quantitative limits on evolution, and it cannot be excluded conclusively using the currently available spectral and color information. Furthermore, there is already some evidence in the current data that the high redshift sample does not have the same properties as the low redshift sample.

The strongest evidence that the lightcurve corrections improve our knowledge of the SN absolute magnitude would be a demonstration that they reduce the dispersion of the SN distance moduli about the best-fit cosmology. (For example, Riess, Press & Kirshner 1996 showed that MLCS reduces the dispersion about Hubble's Law for low z SNe Ia.) To test this, we have compared the dispersion between the data and the predictions of the best-fit cosmology with and without the corrections, Δ_i, inferred from the light curve fitting. [9] We adopt the quantity

$$D^2 = \frac{1}{N}\sum_i [\hat{\mu}_i - f(z_i; H_0, \Omega_m, \Omega_\lambda)]^2 \tag{6}$$

as a measure of the dispersion, where $\hat{\mu}_i$ is the estimated distance modulus for SN i, z_i is

[9]We redetermine the best fit cosmology when we remove the corrections.

its redshift, the function $f(z)$ is defined by equation (1), and N is the number of SNe Ia in the sample. We compute D separately for high and low z; the results are given in Table 1.

For both MLCS and TF, which are calibrated on low z SNe Ia, we see that the dispersion of the low redshift data is reduced substantially by incorporating the corrections derived from the relation between light curve shape and luminosity at maximum brightness. At high redshift, no such improvement is seen. The dispersion of the high z data about the best fit cosmology is virtually unchanged by the incorporation of either the MLCS or TF corrections.

For SF, there is little evidence from Table I that the corrections reduce the dispersion in the data at all. Recall that in the SF parameterization, the relation between light curve width and luminosity corrections is parameterized by $\Delta_{SF} = \alpha(s-1)$ where α is inferred from a global fit to the data at all redshift. As was shown in Figure 2, the corrections Δ_{SF} are quite small so it is not surprising that they do little to reduce the dispersion in the data. As stated before, the SF method finds little, if any, correlation between light curve width and absolute luminosity when averaged over all redshift. What is startling is that the low redshift sample used in P98 is almost identical to the "peak subsample" of Hamuy et al. (Hamuy et al. 1996a). As detailed in that reference, that low redshift sample does show a significant correlation between light curve width and peak luminosity. If a strong correlation is present in the low redshift sample and only a very weak correlation is evident in the full sample, one is led to suspect that the correlation is not present in the high redshift sample; the large number of high-redshift SNe leverage the joint fit.

4. Accounting For Possible Evolution

Both R98 and P98 assume implicitly that the same light curve fitting methods may be applied at all redshifts sampled. This assumption is only valid if the light curve shape is correlated with peak luminosity in the same way at both high and low redshift. Given the evidence we have presented that this may not be true, which indicates, at least circumstantially, that the SNe Ia population evolves, we feel it is necessary to explore whether the data published so far actually are able to actually distinguish the effects of evolution from those of cosmology.

Such effects fall under the rubric of "systematic errors"—because they are not "random," their effects on one's final inferences are difficult to account for in the conventional frequentist approach to statistical inference. However, both teams have adopted the Bayesian approach for their final analyses (though not for all intermediate stages of their analyses). As noted by Jeffreys (1961), the Bayesian approach is particularly apt for studying the effects of systematic error because of its broader notion of uncertainty. A Bayesian probability density describes how probability is distributed among the possible values of a parameter, rather than how values of the parameter are distributed among some hypothetical population. This permits statistical calculations with quantities that are not "random" in the frequentist sense. In particular, as Jeffreys noted, systematic errors are treatable simply by introducing parameterized models for the errors and marginalizing (integrating over) the extra parameters to obtain one's final inferences.

This procedure, when followed blindly, has the potential to weaken one's conclusions unjustifiably. For example, one could simply introduce a systematic dependence that is identical to the physical dependence one is studying, but with a duplicated set of parameters. This duplication would prevent useful constraints from being placed on the parameters, since any measured effect could be "blamed" on the duplicated systematic dependence. Thus Jeffreys emphasized the need to compare models with and without systematic error terms using the ratio of the model probabilities, the *odds* favoring one model over another. The odds can be written as the product of the prior odds (expressing information from other data, or possibly a subjective comparison of the models) and a *Bayes factor* determined entirely by the data, the models, and the sizes of the model parameter spaces. If we know or strongly believe a systematic effect to be present without consideration of the new data before us, then obviously the systematic error model should be used; the prior odds would lead us to this conclusion even if the Bayes factor is indecisive. If we have no strong prior evidence for a systematic error, one takes the prior odds to be unity and relies on the data alone for determining if the effect is present, taking the Bayes factor to be the odds. An appealing aspect of Bayesian model comparison is that the Bayes factor implements an automatic "Ockham's razor" that penalizes models for the sizes of their parameter spaces. Thus model complexity is accounted for by the Bayes factor. Except in unusual cases, needlessly increasing a model's complexity by simply duplicating terms prevents the Bayes factor from favoring the more complicated model. We provide a brief review of Bayes factors in Appendix A; standard references reviewing their use are Kass and Raftery (1995) and Wasserman (1997).

4.1. Systematic Error in H_0

To illustrate this approach, we show how it can be used to quantitatively account for systematic error introduced by the uncertain Cepheid distances used to infer M_0 in the MLCS and TF methods. (The SF analysis used a "Hubble-constant-free" parameterization and thus could avoid explicit treatment of M_0 and the Hubble constant.) We write the true value of M_0 as $(\hat{M}_0 + \delta)$, where $\hat{M}_0$ is the estimate used for calculating $\hat{\mu}_i$, and the new term, δ, represents the constant (but unknown) error introduced by using Cepheid data to calculate $\hat{M}_0$. We describe the likelihood function for analyzing the SNe Ia data in some detail in Appendix B. The final (approximate) likelihood is equivalent to what one would find from modelling the tabulated $\hat{\mu}_i$ estimates according to,

$$\begin{aligned} \hat{\mu}_i &= f(z_i) + \delta + n_i & (7)\\ &= g(z_i) - \eta + \delta + n_i & (8) \end{aligned}$$

where $f(z_i)$ is the cosmological distance modulus relation defined in equation (1), and n_i is a random error term whose probability distribution is a Gaussian with zero mean and standard deviation σ_i. In the second line, we have separated out the H_0 dependence of $f(z_i)$ into

$$\eta \equiv 5 \log \left(\frac{h}{c_2} \right) - 25, \tag{9}$$

where $H_0 = h \times 100 \mathrm{km\,s^{-1}\,Mpc^{-1}}$, and c_2 is the speed of light in units of $10^2 \mathrm{km\,s^{-1}}$; $g(z_i)$ contains the remaining Ω_M- and Ω_Λ-dependent part of $f(z_i)$.

It is clear from equation (8) that η (and thus H_0) is degenerate with δ; we cannot hope to learn about one without independent knowledge of the other. But δ is constrained by our knowledge of the uncertainty of the Cepheid distance scale. In particular, R98 summarize the uncertainties as introducing an error with a standard deviation of $d = 0.21$ magnitudes (corresponding to $\approx 10\%$ uncertainty in H_0). We account for this by introducing a prior distribution for δ that is a zero-mean Gaussian with standard deviation d.

Our model now has four parameters, δ, h, Ω_M, and Ω_Λ. The likelihood function for the data is the product of N Gaussian distributions specified by equation (8) and is proportional to the exponential of a familiar χ^2 statistic. The full joint posterior distribution is the product of this and priors for the parameters, including the informative prior for δ.[10] We can summarize our inferences for the cosmological parameters by integrating over δ; this can be done analytically and is described in Appendix B. If we want to focus on the conclusions for h, we numerically integrate over Ω_M and Ω_Λ. The result, for the MLCS data, is the marginal distribution for h shown as the rightmost solid curve in Figure 8. The best-fit value is $h = 0.645$, and a 68.3% credible region has a half-width $\sigma_h = 0.063$. This is approximately equal to the "total uncertainty" on H_0 estimated by R98 using standard

---

[10]The priors for Ω_M, and Ω_Λ we take to be flat over the region shown in our plots, excluding the "No Big Bang" region; the prior for h we take to be flat in the logarithm.

"rules of thumb" for accounting for systematic error; we have shown how this estimate could be justified by a formal calculation. For the TF data, the marginal posterior is plotted as the leftmost solid curve in Figure 8, and $h = 0.627 \pm 0.062$.

Of greater current interest are the implications for the density parameters. The marginal distribution for Ω_M and Ω_Λ is found by integrating out δ and h. This can be done analytically (see Appendix B). Contours of the resulting distributions, found using both the MLCS and TF data, appear in Figure 9. They are identical to contours found using a model without δ, and essentially reproduce the results reported in R98 (minor differences result from our omission of the "snapshot" SNe).

In this case, we know that M_0 has been estimated using the Cepheid data, and that this estimate has systematic error. Formally, the prior odds favoring the model with δ over one with $\delta = 0$ is thus infinite. The Bayes factor comparing these models is exactly equal to one (this is because the SNe Ia data can tell us nothing about δ; see Appendix A for discussion of this property of Bayes factors), so the posterior odds is equal to the prior odds. Since we know these errors to be present, we take this δ model to be our "default" model when calculating subsequent Bayes factors in this section.

We conclude our discussion of this model by summarizing the evidence in the data for a nonzero cosmological constant, presuming the δ model to be true. In R98 and P98, the marginal posterior probability that $\Omega_\Lambda > 0$ was presented as such a summary; this probability was found to equal 99.6% (2.9σ), 99.99% (3.9σ), and 99.8% (3.1σ) in the MLCS, TF, and SF analyses, respectively, apparently indicating strong evidence that Ω_Λ is nonzero. But this quantity is not a correct measure of the strength of the evidence that $\Omega_\Lambda \neq 0$. This probability would equal unity if negative values of Ω_Λ were considered unreasonable a priori, yet presumably even in this case one would not consider the data to demand a nonzero cosmological constant with absolute certainty. The correct quantity to calculate is the odds in favor of a model with $\Omega_\Lambda \neq 0$ versus a model with $\Omega_\Lambda = 0$. Considering such models to be equally probable a priori, this is given by the Bayes factor comparing these models. [11] We find $B = 5.4$ using the MLCS data and $B = 6.8$ using the SF data, each indicating positive but not strong evidence for a nonzero cosmological constant (presuming there is no evolution). The TF data give $B = 86$, indicating strong evidence for a nonzero Ω_Λ (again, presuming no evolution). Without clear criteria identifying one method as superior to the others, the data are equivocal about a nonzero cosmological constant, even

[11] These Bayes factor calculations can also be viewed as providing the posterior probability that $\Omega_\Lambda = 0$ by putting a prior probability of 0.5 on the $\Omega_\Lambda = 0$ line; in the calculations reported in R98 and P98, this line has zero prior probability (only finite intervals in $(\Omega_M, \Omega_\Lambda)$ have nonzero prior probability in their analyses).

without accounting for the effects of possible evolution.

Similarly, R98 report the number of standard deviations that the $\Omega_M = 1$, $\Omega_\Lambda = 0$ point is away from the best-fit $(\Omega_M, \Omega_\Lambda)$ as a measure of the evidence against the hypothesis that matter provides the closure density; they state this hypothesis is ruled out at the 7σ and 9σ levels using the MLCS and TF methods, respectively. Again, a proper assessment of the hypothesis that $\Omega_M = 1$ and $\Omega_\Lambda = 0$ requires that one give this hypothesis a finite, nonzero prior probability. For the MLCS method, the Bayes factor favoring a model with any $(\Omega_M, \Omega_\Lambda)$ over one with $\Omega_M = 1$ and $\Omega_\Lambda = 0$ is $B = 2.3 \times 10^4$, so that with equal prior odds the probability for the latter model is $p = 1/(1 + B) \approx 5 \times 10^{-5}$ ($\approx 4.5\sigma$). For the TF method, we find $B = 2.1 \times 10^7$, so that $p \approx 5 \times 10^{-8}$ ($\approx 5.8\sigma$). These are small probabilities and indicate very strong evidence against the simpler model, but they are much larger than the probabilities associated with 7σ and 9σ significances ($\approx 2 \times 10^{-11}$ and 3×10^{-18}, respectively, for two degrees of freedom). The incorrect summary statistics used in the previous analyses have exaggerated the evidence for a nonzero cosmological constant, irrespective of whether or not one considers the effects of evolution.

4.2. Models With Evolution

Without a detailed physical idea of the cause of evolution, we cannot explore truly realistic models. Instead we consider two illustrative examples. We first consider a model (Model I) that generalizes the δ model just described by adding an additional offset, ϵ, for the high redshift SNe; we apply this model only to data from R98. As a model of physical evolution, this is certainly too simple, but it is illustrative since for the R98 sample, the observed high redshift SNe Ia are all in a fairly narrow band in redshift near $z \sim 0.5$. Essentially this model merely tells us by how much the luminosities might differ between $z \lesssim 0.1$ and $z \sim 0.5$ as a consequence of evolution. On a purely phenomenological level, the model might be considered more realistic because the low and high redshift SNe are not treated equally in the R98 analysis: the MLCS and TF relations are calibrated using only low redshift SNe. Thus this model can be understood as allowing for a systematic offset when extrapolating the methods beyond the training set. For this model, we use the same prior for δ as in our default model (zero-mean Gaussian with standard deviation $d = 0.21$ mag). The prior for ϵ we also take to be a zero-mean Gaussian but with a different width e. The prior width, e, can be viewed as a description of the scale of errors we might expect from extrapolating low redshift properties to high redshift.

Physically, we might expect evolution to lead to continuous variation of SN Ia properties with redshift. Also, the P98 analysis uses low and high redshift SNe Ia together in calibrating the luminosity/decline rate relation, so there is no clear separation of their data into low and high z subsamples. Thus, Model I is inappropriate for phenomenological modeling of systematic effects from lightcurve fitting of P98 data. Consequently, we consider a second model (Model II) which assumes that the intrinsic luminosities of SNe Ia scale like a power of $1 + z$ as a result of evolution. This second model corresponds to replacing equation (8) with

$$\hat{\mu}_i = g(z_i) - \eta + \delta + \beta \ln(1 + z) + n_i, \tag{10}$$

where δ again represents Cepheid uncertainty in M_0 (relevant only when we apply this model to MLCS or TF data), and β parameterizes the evolution. We use a Gaussian prior for β with zero mean and standard deviation b.

For both models, we explore the dependence of the results on the prior width (e and b) to see how external constraints on evolution (presently unknown) could affect the analysis. We examine values that allow evolutionary changes of up to a few tenths of a magnitude for sources with $z \sim 1$. These characteristic magnitude shifts are comparable to the intrinsic dispersion seen in low redshift SNe Ia (Schmidt et al. 1998), which may be taken as a rough indication of the range of variation of peak magnitude with physical conditions in the explosions. Some current theoretical studies of possible sources of z-dependent variations in SNe Ia luminosities also find magnitude changes of this size to be reasonable (see, e.g., Höflich, Wheeler, and Thielemann 1998; Domínguez et al. 1999).

The new parameters in both models appear linearly in the model equations and can thus be marginalized analytically. Appendix B describes the calculations. Reality could be and probably is far more complicated than either model, but the sparsity of the present data do not justify consideration of more sophisticated models.

4.2.1. Model I

Figure 10*a* shows contours of the marginal density for Ω_M and Ω_Λ using Model I to analyze the MLCS data and taking the prior width for ϵ to be $e = 0.1$ mag. Figure 10*b* shows similar results using the TF data. It is clear from these figures that the presence of a redshift-dependent shift of order e greatly weakens our ability to constrain Ω_M and Ω_Λ from the SNe Ia data. The Bayes factor for this model over the default δ model is 1.1 for both MLCS and TF; the data alone are indecisive about whether such a redshift-dependent error is present. [12] Presuming it is present, the Bayes factor favoring nonzero Ω_Λ over $\Omega_\Lambda = 0$ is reduced significantly from what is found using the default model; it is 1.1 for MLCS and 3.8 for TF.

These results are sensitive to our knowledge of the evolution. Figures 10c and 10d show the MLCS and TF results again, but this time with $e = 0.2$ mag; the credible regions have grown even larger in size. Now the Bayes factor for Model I over the default δ model is 1.3 for MLCS and 1.2 for TF; the data remain indecisive about the presence of an evolutionary offset. Presuming it is present, the Bayes factor favoring nonzero Ω_Λ over $\Omega_\Lambda = 0$ is 0.7 for MLCS (i.e., slightly favoring $\Omega_\Lambda = 0$) and 1.2 for TF.

As one might expect from these results, the most likely value of ϵ tends to be positive, making the more distant sources dimmer than the nearer ones due to evolution rather than cosmology. For example, in Figures 10c and 10d, when $\Omega_M = 1$ and $\Omega_\Lambda = 0$ (a point within the 95.4% credible regions), $\epsilon = 0.31 \pm 0.06$ for MLCS and $\epsilon = 0.17 \pm 0.06$ for TF.

The constraints placed on H_0 by the SNe Ia data arise mostly from the low redshift objects, so one would not expect allowance for evolution to drastically affect the H_0 inferences. The analysis bears this out. The dashed curves in Figure 8 show the marginal distributions for h based on the MLCS and TF data using Model I with $e = 0.2$; they differ little from the distributions found using the reference model with no evolution.

4.2.2. Model II

Figure 11 shows results from analysis of the SF data using Model II. Figure 11a shows contours of the marginal density for Ω_M and Ω_Λ for the $\beta = 0$ case as a reference; these contours essentially duplicate the results of Fit C in P98. Figure 11b shows similar contours, but allowing for a nonzero β; the prior standard deviation for β was $b = 0.25$. Figure 11c repeats the analysis with $b = 0.5$. Again we find that the possibility of evolution significantly weakens the constraints on the density parameters, but if the amount of evolution can be bounded, useful limits might result. The Bayes factor for evolution vs. no evolution is 1.0 for $b = 0.25$ and 1.1 for $b = 0.5$, so the data alone are indecisive about the presence or absence of this type of evolution. We find similar results when using this model to analyze the MLCS and TF data.

[12] The χ^2 values for the default fits are already acceptable, so one might worry that the more complicated models are "overfitting." But the maximum likelihoods for models I and II are only slightly greater than those found with the default model. Bayes factors account for overfitting and it is not playing a role here. The operation of Bayes factors is discussed further in Appendix A.

Figure 12 shows how these findings depend on the prior uncertainty for β. The solid curve shows how the Bayes factor favoring a nonzero Ω_Λ over $\Omega_\Lambda = 0$ depends on b; only for $b \lesssim 0.1$ does the Bayes factor remain near the value of 6.8 found assuming there is no evolution. The dashed curve shows the Bayes factor for Model II versus the default model with no evolution; for no value of b in the range of the plot can the data clearly distinguish evolution from cosmology. This emphasizes the need to use information independent of the μ-z relation to constrain evolution.

4.2.3. Flat Cosmologies

So far we have assessed the evidence for nonzero Ω_Λ by comparing models with $\Omega_\Lambda = 0$ to models with arbitrary Ω_Λ, as was done in the R98 and P98 analyses. However, many cosmologists would consider flat models, with $\Omega_\Lambda = 1 - \Omega_M$, to be of special relevance (e.g., because of inflationary arguments). We have thus analyzed models constrained in this way, using our default model and models I and II.

Figure 13 shows the marginal posterior distributions for Ω_M (and, equivalently, for $\Omega_\Lambda = 1 - \Omega_M$) presuming a flat cosmology. The three panels show analyses of the distance moduli reported using the MLCS and TF data with Model I (top and middle, respectively), and using the SF data with Model II (bottom). The solid curves show results based on the default model; the short-dashed curves show results with a small amount of evolution allowed ($e = 0.1$ or $b = 0.25$), and the long-dashed curves show results with a larger amount of evolution allowed ($e = 0.2$ or $b = 0.5$). As in the previous cases, the Bayes factors comparing models with evolution to the default model are all nearly equal to one. Also, as was found before, accounting for the possibility of evolution significantly weakens the evidence for nonzero Ω_Λ. However, if one restricts attention to flat models, the evidence for nonzero Ω_Λ is stronger than it is if one allows nonflat cosmologies. For the default model presuming no evolution, the Bayes factors favoring nonzero Ω_Λ over a flat model with $\Omega_M = 1$ are 2.1×10^4 (MLCS), 2.5×10^6 (TF) and 5.0×10^3 (SF), much larger values than were found in the comparison using nonflat Ω_Λ models discussed above. But these values fall dramatically if one allows for evolution. For models with a small amount of evolution allowed, they decrease to 20, 48, and 14, respectively, indicating positive but not compelling evidence for nonzero Ω_Λ. For models with a larger amount of evolution allowed, they decrease further to 2.4, 2.5, and 2.3, indicating no significant evidence for nonzero Ω_Λ.

4.3. Simulations

Figure 14 elucidates why introducing the possibility of evolution so greatly weakens our ability to constrain Ω_M and Ω_Λ. The thick solid curve shows $g(z)$ for the best-fit cosmology from fits to the SF data presuming no evolution ($\Omega_M = 0.75$, $\Omega_\Lambda = 1.34$). The dotted curve shows the same function for the flat $\Omega_M = 1$ ($\Omega_\Lambda = 0$) cosmology (which lies within the 68.3% credible region in Figure 11c), and the dashed curve shows the evolutionary component $\beta \ln(1+z)$ for the best-fit value of β given this cosmology ($\beta = 0.83$). The thin solid curve shows the sum of the dotted and dashed curve. Over the range of redshift covered by the data ($z \lesssim 1$) and even beyond, $(\Omega_M, \Omega_\Lambda, \beta) = (0.75, 1.34, 0)$ and $(\Omega_M, \Omega_\Lambda, \beta) = (1, 0, 0.83)$ are indistinguishable if one allows evolution of this type, unless one can determine μ to significantly better than the $\sim 20\%$ accuracy currently obtained at $z \lesssim 1$. However, we note that the models differ substantially at larger redshift (by about 0.8 mag), which offers some hope of discerning evolution. We caution, though, that the best fit values of $(\Omega_M, \Omega_\Lambda)$ are likely to be different for data extending to $z \approx 2$, either with or without evolution, so the comparison is not truly apt, and moreover $(\Omega_M, \Omega_\Lambda) = (0.75, 1.34)$ might be considered implausible intrinsically by many cosmologists. To amplify this point, we also compare $(\Omega_M, \Omega_\Lambda, \beta) = (1, 0, 0.83)$ with another cosmology, $(\Omega_M, \Omega_\Lambda) = (0.3, 0.7)$, within the 68.3% credible region of the no-evolution analysis. As is shown by the bold, long-dashed curve in Figure 14, μ accuracies better than 10% out to $z \approx 2$ would be needed to distinguish these cosmologies from one another out. We have systematically explored a wide range of cosmologies and found similar results: simple power law evolution can make widely disparate cosmologies appear remarkably similar. Put another way, the differences between cosmologies with various Ω_M and Ω_Λ are well mimicked by power-law evolution to redshifts beyond those currently accessible in supernova surveys. We emphasize that we did not choose the form of the evolution to produce this degeneracy; this is a standard phenomenological model for evolution. We have found similar behavior with another simple model for evolution consisting of a power law in lookback time.

To assess how well evolution can mimic cosmology, we analyzed simulated data consisting of μ values with added Gaussian noise at redshifts that themselves had added Gaussian noise. The simulations were designed to roughly mimic possible future data like that reported in P98 (this simplified the analysis since H_0 need not be explicitly accounted for as it would have for data like that reported in R98). The redshifts of the first 16 SNe in our simulated data sets were chosen to be similar to those of the 16 low-redshift SNe in P98 ($z \lesssim 0.1$); the redshifts for the remaining simulated data were spread uniformly randomly over some specified interval. We added redshift errors with a standard deviation of 0.002, and μ errors with standard deviations equal to those reported by P98 for the 16 low-z SNe, and equal to 0.25 magnitudes for the high-z SNe (a typical value for μ values reported in

P98).

Figure 15 shows typical results. Here we simulated data from a flat, $\Omega_M = 1$ ($\Omega_\Lambda = 0$) cosmology with evolution described by Model II with $\beta = 0.5$. Figure 15*a* shows the results of an analysis assuming no evolution, with 38 simulated SNe redshifts in the interval $[0.3, 1]$ (54 total simulated points). This corresponds to a sample size equal to that used in the P98 analysis and extending over a similar range of redshift. The cross indicates the best-fit parameter values, the dot indicates the true values, and the contours bound credible regions of various sizes. One would reject the true model as being improbable if evolution is ignored. Figure 15*b* shows a similar plot, with the number of high-z SNe increased so the total sample size is now 200, with the high-z points now spread over $[0.3, 1.5]$. The contours have shrunk considerably, converging around a point well away from the truth. In both figures, the best-fit point has an excellent χ^2/ν (53.6/52 for the small data set, 201/198 for the large one). Evolution mimics cosmology so well that standard "goodness of fit" reasoning can lead one to conclude, mistakenly, that pure cosmology (with no evolution) is an adequate description of data of this quality even when substantial evolution is present.

Figure 15*c* shows the results of an analysis of the larger data set using a model that includes evolution; the marginal posterior (with β marginalized) is shown. The credible regions now contain the true model, but they are large even for a data set four times the size of the currently published surveys, and extending to significantly higher redshift. The Bayes factor is of order unity, showing that data of this quality is not capable of distinguishing between models with and without evolution. This is further testimony to the approximate degeneracy between cosmology and evolution, at least at $z \lesssim 1.5$.

The extent to which evolution corrupts the results depends both on the true cosmology and on the amount of evolution allowed. Independent constraints on the amount of evolution could thus play an important role in allowing useful contraints to be placed on the cosmology. They would enter the analysis via the prior for β. Comparison of Figures 11*a* through 11*c* shows how constraints on the amount of evolution can affect one's final inferences.

5. Conclusions

As is well known, attempts to use observations of distant galaxies to measure the cosmological deceleration parameter had to confront the reality that not only are galaxies not standard candles, but their luminosities also evolve with time (e.g., Weinberg 1972, Tinsley 1977, Sandage 1988, Yoshii & Takahara 1988, Bruzual 1990, Peebles 1993).

Nearby SNe Ia have a much narrower range of luminosities than galaxies, and it appears possible to "standardize" their peak brightnesses using the decline rate of their light curves (Phillips 1993, Riess, Press & Kirshner 1996). However, we feel it is unwise to ignore the historical lessons taught us by efforts to measure the deceleration of the Universe from observations of galaxies. Observations of standard or "standardized" candles cannot be used to deduce cosmological parameters without taking account of source evolution. Neglecting the possibility that the luminosities of the candles evolve with redshift introduces uncontrolled systematic errors into the analysis.

One focus of this paper has been to see if there are indications that the SNe Ia population has evolved from $z \sim 0.5 - 1$ to $z \ll 1$. We have presented two arguments that this might be so. The first is that a comparison of the peak luminosities estimated for individual SNe Ia by two different methods, MLCS and TF, are not entirely consistent with one another at high redshifts, $z \sim 0.5$. We have asserted, uncontroversially in our opinion, that the two methods very likely sample slightly different aspects of the SN Ia mechanism, and should not be expected to agree completely. If evolution were entirely absent, though, the differences between them should not depend on redshift, contrary to the admittedly sketchy evidence of the data. A second hint that SNe Ia evolve with redshift is that while the three luminosity estimators, SF as well as MLCS and TF, reduce the dispersion of distance moduli about best fit models at low redshift, they do not at high redshift.

These studies were intended to give us impetus to pursue the more fundamental point of this paper, namely that evolution *must* be considered possible, even if there are no "smoking guns" that seem to require it. Ideally, one should attempt to constrain the parameters of an evolutionary model at the same time as determining the parameters of the cosmological model. As we stated at the outset, changes in peak SN Ia magnitude of order 0.1 magnitudes out to $z \sim 1$ would alter the ranges of acceptable cosmological models substantially. We consider it highly improbable that evolution is so minimal. The dispersion of SNe Ia peak magnitudes at low z is approximately 0.3-0.5 mag (Schmidt et al. 1998), which might indicate a plausible range of variation for diverse physical conditions. Using theoretical models, Höflich, Wheeler & Thielemann (1998) argue that a similar range of variation of peak luminosities could arise as a consequence of changes in composition.

To get an idea of how evolution would affect one's ability to constrain cosmological parameters, we considered two different models. In one, we assumed that there is a magnitude shift between low and high redshift. On applying this model to the data of R98, we found that the ability of the data to discriminate among cosmological models was completely eliminated—virtually any cosmological model would do. We also considered a model in which the peak magnitudes of SNe Ia evolve continuously, with

$\delta m(z) = \beta \log(1+z)$; once again, the data could not discriminate among cosmologies in the face of such evolution.

In order to assess the extent to which the data favor these models over ones without evolution, we computed Bayes factors. The Bayes factor between classes of models with and without luminosity evolution are equivalent to the odds ratio between them if there is no *a priori* reason to prefer one over the other. In all cases, we found that the Bayes factors are of order unity, which means that the data themselves do not favor either model. If we accounted for our prejudice that evolution does occur, the odds would disfavor models in which SNe Ia are identical at all redshifts.

The two models we have considered illustrate how well evolution can mimic cosmology. The less realistic model merely shifted the magnitudes of high z SNe Ia relative to low z SNe Ia by a fixed amount, δm. Since the SNe Ia in the R98 sample were predominantly at $z \approx 0.3 - 0.5$, the cosmological magnitude shift (relative to Hubble's law or any other fiducial cosmology) varies little over the entire redshift range they span. Clearly, for the high z SNe Ia in this sample, one only knows that there is a total magnitude shift between $z \lesssim 0.1$ and $z \sim 0.5$, not how much is due to cosmology and how much to evolution. The characteristic magnitude shifts due to evolution needed for a cosmological model with $(\Omega_M, \Omega_\Lambda) = (1,0)$ are $\approx 0.2 - 0.3$. Ultimately, a model in which there is simply a constant magnitude difference between low and high z SNe Ia should fail to model data spanning a lage range of redshifts.

More daunting is the success of models with a continuous magnitude shift, $\delta m(z) = \beta \ln(1+z)$. While it is unsurprising that such models would be approximately degenerate with cosmology at low z, where the combined magnitude shift, relative to Hubble flow, is $[1.086(1 - \Omega_M/2 + \Omega_\Lambda) + \beta]z$ (e.g., Weinberg 1972), it is remarkable that a continuous magnitude shift with this simple form cannot be discerned out to at least $z \approx 1$. Our simulations show that even if there is truly no evolution, so that reality corresponds to certain values of Ω_M and Ω_Λ with $\beta = 0$, models with $\beta \neq 0$ and $(\Omega_M^{\rm eff}, \Omega_\Lambda^{\rm eff}) \neq (\Omega_M, \Omega_\Lambda)$ yield apparent magnitudes that are indistinguishable from the truth within differences in distance modulus ≈ 0.1 mag. Differences between select cosmological models may be larger at higher z, but often remain within ~ 0.1 mag out to $z \approx 2$.

Even more worrisome is that neglecting evolution in the analysis of samples of evolving SNe Ia can lead to a distorted sense of reality. As an example, we simulated a set of 200 SNe Ia distance moduli, including 184 high-z SNe with redshifts uniformly distributed over $0.3 \leq z \leq 1.5$, in a cosmological model with $(\Omega_M, \Omega_\Lambda, \beta) = (1.0, 0.0, 0.5)$. We then analyzed the data with evolution neglected entirely. The result was that given enough SNe Ia, the analysis would pick out a small range of "allowed" values of $(\Omega_M, \Omega_\Lambda)$, but centered around incorrect values. The true cosmology was well outside the 3σ credible region for these simulations, yet the (incorrect) best-fit model would be judged excellent by a standard χ^2 goodness-of-fit test.

What is needed to separate evolution from cosmology is not just detections of greater numbers of SNe Ia at high redshift, but rather a better physical understanding of the SN Ia process. In particular, one would like to be able to link the Phillips relations uniquely to internal conditions in the explosions themselves, to be able to understand how they might evolve with redshift (see, e.g., Höflich, Wheeler, and Thielemann 1998; Domínguez et al. 1999). This would allow construction of realistic, not phenomenological, models for evolution, and one might hope to be able to constrain the parameters of these models along with cosmological ones. The analogue in galactic astronomy is the use of population synthesis models to study the cosmological evolution of the luminosity function, which *might* permit, given enough data, simultaneous fits for cosmological parameters (Yoshii & Takahara 1988, Bruzual 1990). Conceivably, detailed physical modelling might lead to a detailed, quantitative connection between the peak luminosities of SNe Ia and their spectra, which would allow additional information to be useful quantitatively in fitting for Ω_M and Ω_Λ. R98 and P98 have argued, using the spectral data, that there is no compelling evidence *for* evolution, but that does not translate into a convincing argument *against* evolution unless the salient features of the source spectra can be connected unambiguously to peak luminosity.

In the end, what all cosmologists want to know is the probability that the cosmological constant is nonzero. The Bayes factor provides straightforward mathematical machinery for doing this calculation, whether or not evolution is included in the analysis. When evolution is not included in the analysis, and no prior assumptions are made about the spatial geometry of the Universe, the Bayes factor for $\Omega_\Lambda \neq 0$ compared to $\Omega_\Lambda = 0$ is $B = 5.4$ using the MLCS method, $B = 6.8$ using the SF method, and $B = 86$ using the TF method, which if one is not prejudiced either way, only favors nonzero Ω_Λ equivocally. (There may be reasons to be prejudiced one way or the other; see for example Turner 1999 for a theoretical cosmologist's bias.) When evolution is included in the analysis, the values of the analogous Bayes factors depend on one's prior assumptions, but rather conservatively $B \lesssim 1$. Thus, if we do not discriminate among open, closed and flat cosmological models, the data alone do not choose between $\Omega_\Lambda \neq 0$ and $\Omega_\Lambda = 0$ once the possibility of evolution is taken into account. However, if the Universe is presumed to be flat spatially, then the case for $\Omega_\Lambda \neq 0$ is stronger. If evolution is presumed not to occur, we find Bayes factors $B = 2.1 \times 10^4$ (MLCS), 2.5×10^6 (TF) and 5.0×10^3 (SF), decisive odds in favor of nonzero Ω_Λ. Weak evolution ($e = 0.1$ in Model I or $b = 0.25$ in Model II) lowers these values to $B = 20$ (MLCS), 4.8 (TF) and 14 (SF), which still favors $\Omega_\Lambda \neq 0$ positively but not nearly as persuasively. If evolution is somewhat more pronounced, but still at a plausible level ($e = 0.2$ or $b = 0.5$), the Bayes factors fall to $B = 2.4$ (MLCS), 2.5 (TF) and 2.3 (SF), which is scant evidence for a non-vanishing cosmological constant. Once again, the ability of the data to distinguish $\Omega_\Lambda \neq 0$ from $\Omega_\Lambda = 0$ depends sensitively on prior assumptions about evolution of SNe Ia, and underscores the importance of placing independent constraints on the possible range of variation of their peak luminosities with redshift.

We have benefited from conversations and correspondence with Adam Riess, Peter Garnavich and Bill Press. We thank Saul Perlmutter for providing a draft of P98 prior to publication. We thank Margaret Geller, Saul Teukolsky, Éanna Flanagan, David Chernoff, Ed Salpeter, and Sidney Drell for comments on earlier drafts of this paper. PSD gratefully acknowledges support from the John Simon Guggenheim Memorial Foundation. This work was supported by NASA grants NAG5-3097, NAG5-2762, NAG5-3427, and NSF grant PHY-9310764.

A. Bayes Factors

In Bayesian inference, to form a judgement about an hypothesis H_i, we calculate its probability, $p(H_i|D, I)$, conditional on the data (D) and any other relevant information at hand (I). The desired probability $p(H_i|D, I)$ is not usually assignable directly; instead we must calculate it from other simpler probabilities using the rules of probability theory. Prominent among these is Bayes's theorem, expressing this posterior (i.e., after consideration of the data) probability in terms of a prior probability for H_i and a likelihood for H_i,

$$p(H_i|D, I) = p(H_i|I)\,\frac{\mathcal{L}(H_i)}{p(D|I)}, \tag{A1}$$

where the likelihood for H_i, $\mathcal{L}(H_i)$, is a shorthand notation for the sampling probability for D presuming H_i to be true, $p(D|H_i, I)$. The likelihood notation and terminology emphasizes that it is the dependence of the sampling probability on H_i (rather than D) that is of interest for calculating posterior probabilities. The term in the denominator is the prior predictive probability for the data and plays the role of a normalization constant. It can be calculated according to

$$p(D|I) = \sum_i p(H_i|I)\mathcal{L}(H_i). \tag{A2}$$

We see from this equation that the prior predictive probability is the average likelihood for the hypotheses, with the prior being the averaging weight. It is also sometimes called the marginal probability for the data.

For estimating the values of the parameters θ of some model, the background information is the assumption that the parameterized model under consideration is true; we denote this by M (this may include any other information we have about the parameters apart from that provided by D; for example, previously obtained data). The posterior probability for any hypothesis about continuous parameters can be calculated from the posterior probability density function (PDF), which we may calculate with a continuous version of Bayes's theorem:

$$p(\theta|D, M) = p(\theta|M) \, \frac{\mathcal{L}(\theta)}{p(D|M)}. \tag{A3}$$

Both the posterior and the prior are PDFs in this equation; we continue to use the $p(\cdot)$ notation, letting the nature of the argument dictate whether a probability or PDF is meant. In this case, the normalization constant is given by an integral:

$$p(D|M) = \int d\theta \; p(\theta|M) \, \mathcal{L}(\theta). \tag{A4}$$

The normalization constant is now the average likelihood for the model parameters.

When comparing rival models, M_i, each with parameters θ_i, we return to the discrete version of Bayes's theorem in equation (A3), using $H_i = M_i$ for the hypotheses, and taking the background information to be $I = M_1 + M_2 + \cdots$ (denoting the proposition, "Model M_1 is true *or* model M_2 is true *or* ..."). The likelihood for model M_i is $p(D|M_i, I)$; but since the joint proposition (M_i, I) is equivalent to the proposition M_i by itself, we have $\mathcal{L}(M_i) = p(D|M_i)$. Thus the likelihood for a model in a model comparison calculation is equal to the normalization constant we would use when doing parameter estimation for that model, given by an equation like equation (A4). In other words, the likelihood for a model (as a whole) is the average likelihood for its parameters.

It is convenient and common to report model probabilities via odds, ratios of probabilities of models. The (posterior) odds for M_i over M_j is

$$\begin{aligned} O_{ij} &\equiv \frac{p(M_i|D,I)}{p(M_j|D,I)} \\ &= \frac{p(M_i|I)}{p(M_j|I)} \times \frac{p(D|M_i)}{p(D|M_j)} \\ &\equiv \frac{p(M_i|I)}{p(M_j|I)} \times B_{ij} \end{aligned} \tag{A5}$$

where the first factor is the prior odds, and the ratio of model likelihoods, B_{ij}, is called the *Bayes factor*. When the prior information does not indicate a preference for one model over another, the prior odds is unity and the odds is equal to the Bayes factor. Kass and Raftery (1995) provide a comprehensive review of Bayes factors, and Wasserman (1997) provides a survey of their use and methods for calculating them. When the prior odds does not strongly favor one model over another, the Bayes factor can be interpreted just as one would interpret an odds in betting; Table 2 summarizes the recommended interpretation of Kass and Raftery.

The Bayes factor is a ratio of prior predictive probabilities; it compares how rival models predicted the observed data. Simple models with no or few parameters have their predictive probability concentrated in a small part of the sample space. The additional parameters of complicated models allow them to assign more probability to other regions of the parameter space, but since the predictive probability must be normalized, this broader explanatory power comes at the expense of reducing the probability for data lying in the regions accessible to simpler models. As a result, model comparison using Bayes factors tends to favor simpler models unless the data are truly difficult to account for with such models. Bayes factors thus implement a kind of automatic and objective "Ockham's razor" (Jaynes 1979; Jefferys and Berger 1992).

This notion of simplicity is somewhat subtle, but in some simple situations it accords well with our intuition that models with more parameters are more complicated and should only be preferred if they account for the data significantly better than a simpler alternative. Because Bayes factors are ratios of *average* likelihoods, rather than the *maximum* likelihoods that are used for model comparison in frequentist statistics, they penalize models for the sizes of their parameter spaces. A simple, approximate calculation of the average parameter likelihood given by equation (A4) elucidates how this comes about.

First, we assume that the data are informative in the sense of producing a likelihood function that is strongly localized compared to the prior. Suppose that the scale of variation of the prior is $\Delta\theta$, and the scale of variation of the likelihood is $\delta\theta \ll \Delta\theta$. If the likelihood is maximized at $\theta = \hat{\theta}$, then we find

$$p(D|M) \approx p(\hat{\theta}|M) \int d\theta \; \mathcal{L}(\theta). \tag{A6}$$

Since the prior is normalized with respect to θ, $p(\hat{\theta}|M)$ will be roughly equal to $1/\Delta\theta$. The integral will be roughly equal to the product of the peak and width of the likelihood, $\mathcal{L}(\hat{\theta})\delta\theta$. Thus,

$$p(D|M) \approx \mathcal{L}(\hat{\theta})\frac{\delta\theta}{\Delta\theta}. \tag{A7}$$

We find that the likelihood for a model is approximately given by the maximum likelihood for its parameters, multiplied by a factor that is always ≤ 1 that is a measure of how the size of the probable part of the parameter space changes when we account for the data. This latter factor is colloquially known as the *Ockham factor*. To see why, consider the case of nested models: M_1 and M_2 share parameters θ, but M_2 has additional parameters ϕ. In such cases, it is not uncommon that the prior and posterior ranges for θ are usually comparable for both models (this is not the case in the present work, however). Then the Bayes factor in favor of the more complicated model is approximately given by

$$B_{21} \approx \frac{\mathcal{L}(\hat{\theta}, \hat{\phi})}{\mathcal{L}(\hat{\theta})} \frac{\delta\phi}{\Delta\phi}. \tag{A8}$$

Thus the data will favor M_2 only if the maximum likelihood ratio is high enough to offset $\frac{\delta\phi}{\Delta\phi}$, which will be < 1 if the data contain any information about ϕ (and cannot be > 1 in any case). This is in contrast to the frequentist approach, where only the ratio of maximum likelihoods is used. This ratio cannot disfavor M_2; one thus requires the likelihood ratio to exceed some critical value before preferring M_2, on the grounds that one should prefer the simpler model a priori. Unfortunately, the critical value is set in a purely subjective and ad hoc manner, and comparisons using likelihood ratios can be inconsistent (in the formal statistical sense of giving the incorrect answer when the amount of data becomes infinite). The Bayesian approach can (and often does) prefer the simpler model even when both models are given equal prior probabilities, and the critical likelihood ratio needed to just prefer M_2 is determined by the likelihood functions and the size of the parameter space searched. The odds is known to be a consistent statistic for choosing between models.

The approximations leading to the simple result of equation (A8) are not valid for the present work, so a simple "Ockham's razor" interpretation of our results is not possible. Although the default model is nested in the models that have z-dependent systematic errors, it is clear from the figures that the addition of the systematic error parameters (corresponding to ϕ in the above analysis) greatly affects inferences of the cosmological parameters (corresponding to θ). Thus the $\delta\theta$ factors (here associated with the cosmological parameters) do not approximately cancel in the Bayes factor. Moreover, inferences for the θ and ϕ parameters are highly correlated in the SNe Ia problem, so it is not possible to identify separate $\delta\theta$ and $\delta\phi$ factors separately quantifying the uncertainties in the nested and additional parameters. We do know that the maximum likelihoods (e.g., minimum χ^2 values) are comparable for models with and without z-dependent systematic errors. The more complicated models are not improving the best fit substantially, but rather the additional parameter allows one to make the fit nearly as good as the best fit in a large region of the parameter space (because of the near-degeneracy of evolution and cosmology). It is this increase of the acceptable volume of parameter space that accounts for the Bayes factors slightly favoring the more complicated models here.

As is clear from equation (A7), the prior ranges for parameters play an important role in Bayesian model comparison. This is in contrast to their role in parameter estimation, where in Bayes's theorem the prior range factor appears in both the numerator (through the prior) and the denominator (through the average likelihood) and thus cancels, typically having a negligible effect on inferences (though the range itself cancels, some effect can remain due to truncation of the tails of the likelihood). In particular, parameter estimation is typically well-behaved when prior ranges become infinite. But model comparison fails when the ranges for any parameters not common to all models become infinite, because the Ockham factors associated with those parameters vanish. This may at first appear to be troubling (or at best a nuisance), but a similar behavior is acknowledged to be necessary even in frequentist treatments of many problems. For example, consider detection of a periodic signal in a noisy time series using a discrete Fourier transform (DFT). This is a model comparison problem (comparing a model without a signal to one with a periodic signal), and in fact the Fourier power is simply related to the likelihood for a periodic (sinusoidal) signal. In frequentist analyses, one cannot simply use the number of standard deviations the DFT peak is above the null expectation to assess the significance of a signal; one must also take into account the number of statistically independent frequencies examined, which depends on the frequency range searched and on the number and locations of frequencies examined within that range. Similar considerations arise in searches for features in spectra, or searches for sources in images—one must take into account the number and locations of points searched in order to properly assess the significance of a detection. The results of the corresponding Bayesian calculations similarly depend on the ranges of parameters searched (but not on the number and locations of the parameter values used). Bayes's theorem indicates that the sizes of parameter spaces (i.e., search ranges) must be taken into account whenever we compare models; such considerations should not be unique to the few applications where they have been recognized to be important in conventional analyses.

B. Statistical Methodology

As in the analyses of R98 and P98, we adopt the Bayesian approach for inferring the cosmological parameters Ω_M and Ω_Λ, extending their analyses to include parameterized systematic and evolutionary components. The additional parameters are dealt with by marginalizing (as the R98 analysis did with H_0 and the P98 analysis did with the SF fitting parameters). Many of the needed marginalizations can be done analytically; this Appendix describes these calculations. Some remaining marginalizations (including calculation of Bayes factors) were done numerically with various methods including straightforward quadrature, adaptive quadrature, and Laplace's method; application of these methods to Bayesian integrals is surveyed in Loredo (1999).

B.1. Basic Framework

Let D_i denote the data associated with SN number i, and D denote all the data associated with the N SNe in a particular survey. Let $\mathcal{C}$ denote the cosmological parameters, $\mathcal{C} = (H_0, \Omega_M, \Omega_\Lambda)$, and $\mathcal{S}$ denote possible extra parameters associated with modelling evolution or other sources of systematic errors. Our task is to find the posterior distribution for these parameters given the data and some model, M. Actually, we are ultimately interested in the marginal distribution for Ω_M and Ω_Λ, found by marginalizing: $p(\Omega_M, \Omega_\Lambda|D, M) = \int dH_0 \int d\mathcal{S} p(\mathcal{C}|D, M)$. Bayes's theorem gives the joint posterior distribution for $\mathcal{C}$ and $\mathcal{S}$,

$$p(\mathcal{C}, \mathcal{S}|D, M) \propto p(\mathcal{C}|M)p(\mathcal{S}|M)\mathcal{L}(\mathcal{C}, \mathcal{S}). \tag{B1}$$

The first factor is the prior for $\mathcal{C}$, which we will take to be flat over the ranges shown in our plots (or flat in the logarithm for H_0; see below). The second factor is the prior for $\mathcal{S}$ which we assume is independent of $\mathcal{C}$; we discuss it further in the context of specific models, below. The last factor is the likelihood for $\mathcal{C}$ and $\mathcal{S}$, which we have abbreviated as $\mathcal{L}(\mathcal{C}, \mathcal{S}) \equiv p(D|\mathcal{C}, \mathcal{S}, M)$. Rigorous calculation of this likelihood is very complicated, requiring introduction and estimation of many additional parameters, including parameters from the lightcurve model and parameters for characteristics of the individual SNe (such as their apparent and absolute magnitudes, redshifts, K-corrections, etc.). With several simplifying assumptions, the final result is relatively simple; it can be written as the product of independent Gaussians for the redshifts and distance moduli of the SNe integrated over the redshift uncertainty, so that

$$\mathcal{L}(\mathcal{C}, \mathcal{S}) \approx \prod_i \int dz_i \exp\left[-\frac{[F(z_i) - \hat{\mu}_i]^2}{2s_i^2}\right] \exp\left[-\frac{(z_i - \hat{z}_i)^2}{2w_i^2}\right]. \tag{B2}$$

Here $\hat{\mu}_i$ is the best-fit distance modulus for SNe number i, s_i is its uncertainty, $\hat{z}_i$ is the best-fit cosmological redshift, and w_i is its uncertainty (mostly due to the source's peculiar velocity). The function $F(z_i)$ gives the true distance modulus for a SN Ia at redshift z_i; in the absence of systematic or evolutionary terms, it is given by $f(z_i)$ in equation (1). For the results reported in P98, two complications appear in the likelihood. First, the factors are not independent; the use of common photometric calibration data for groups of SNe Ia that are studied together introduces correlations. P98 have reported a correlation matrix accounting for these, but the correlations are very small and we have neglected them here. In addition, one of the parameters defining the lightcurve model—the α parameter described in § 2, above—appears explicitly in the P98 likelihood so that it can be estimated jointly with the cosmological parameters. This parameter would appear in the $\hat{\mu}_i$ estimates in equation (B2). The data tabulated in P98 use the best-fit α, however, so we could not account for the uncertainty of α in our analysis. The close similarity between our contours in Figure 11a and those presented in P98 argues that rigorous accounting for the uncertainty in α plays only a minor role in the final results.

As was done in R98 and P98, we approximate the z_i integrals in equation (B2) by linearizing the z_i dependence of $F(z_i)$ about $\hat{z}_i$ and performing the resulting convolution of Gaussians analytically. The result is

$$\mathcal{L}(\mathcal{C}, \mathcal{S}) \approx \prod_i \exp\left[-\frac{[\mu_i - \hat{\mu}_i]^2}{2\sigma_i^2}\right]. \tag{B3}$$

where $\mu_i = F(\hat{z}_i)$ and

$$\sigma_i^2 = s_i^2 + [F'(\hat{z}_i)]^2 w_i^2. \tag{B4}$$

The total variance σ_i^2 depends on $\mathcal{C}$ through $F'(z)$. But this dependence is weak in general, and $F'(z)$ is actually independent of $\mathcal{C}$ at low redshift in the pure cosmology model, with

$$F'(z) = \frac{5 \log e}{z}. \tag{B5}$$

We follow the practice of R98 and simply use this formula for all redshifts. We use the same formula for models with systematic error terms that introduce an additional (weak) dependence on redshift and $\mathcal{S}$; the redshift uncertainties are negligibly small at high redshifts where such dependences might become important, so the dependence of σ_i^2 on redshift is negligible. It is possible to do the z_i integrals in equation (B2) accurately using Gauss-Hermite quadrature. We have done some calculations this way and verified that the final inferences are negligibly affected by the redshift integral approximations.

Equation (B3) is the starting point for the analyses reported in the body of this work. It is of a simple form: -2 times the log-likelihood is of the form of a χ^2 statistic. This is the same likelihood we would have written down had we simply presumed at the outset that the reported $\hat{\mu}_i$ values were equal to some underlying true values given by $F(\hat{z}_i)$;

$$\hat{\mu}_i = F(\hat{z}_i) + n_i, \tag{B6}$$

where the probability distribution for the value of n_i is a zero-mean Gaussian with standard deviation σ_i^2.

B.2. FRW Cosmology

Presuming a FRW cosmology and no systematic errors, equation (B6) can be written,

$$\begin{aligned}\hat{\mu}_i &= f_i + n_i \\ &= g_i - \eta + n_i,\end{aligned} \tag{B7}$$

where $f_i = f(\hat{z}_i)$ is the magnitude-redshift relation, which we can separate into a part $g_i = g(\hat{z}_i)$ that depends implicitly only on Ω_M and Ω_Λ, and the H_0 dependence is contained in η (defined in equation (9)). Define the quadratic form Q according to

$$Q = \sum_i \frac{(\hat{\mu}_i - g_i + \eta)^2}{\sigma_i^2}. \tag{B8}$$

This is the χ^2 statistic used in R98; the joint likelihood for h, Ω_M, and Ω_Λ is simply proportional to $e^{-Q/2}$. We can analytically marginalize over h (or equivalently, over η) to find the marginal likelihood for the density parameters. To do so, we must assign a prior for h. We use the standard noninformative "reference" prior for a positive scale parameter, a prior flat in the logarithm and thus scale-invariant (Jeffreys 1961; Jaynes 1968; Yang and Berger 1997). This corresponds to a prior that is flat in η. We bound this prior over some range $\Delta\eta$ (with limits corresponding to $h = 0.1$ and $h = 1$, so $\Delta\eta = \ln[10]$). The prior range has negligible effect on all our results (so long as it contains the peak of the likelihood) because the H_0 parameter is common to all models, so the prior range cancels out of all probability ratios. Thus we could let it become infinite, but it is a good practice in Bayesian calculations to always adopt proper (i.e., normalizable) priors, especially if Bayes factors (ratios of normalization constants) are of interest.

Using the log-flat prior, the marginal likelihood for the density parameters is

$$\mathcal{L}(\Omega_M, \Omega_\Lambda) = \frac{1}{\Delta\eta} \int d\eta e^{-Q/2}. \tag{B9}$$

To do the integral, complete the square in Q as a function of η, writing

$$Q = \frac{(\eta - \hat{\eta})^2}{s^2} - q(\Omega_M, \Omega_\Lambda), \tag{B10}$$

where

$$\frac{1}{s^2} = \sum_i \frac{1}{\sigma_i^2}, \tag{B11}$$

$$\hat{\eta}(\Omega_M, \Omega_\Lambda) = s^2 \sum_i \frac{g_i - \hat{\mu}_i}{\sigma_i^2}, \tag{B12}$$

and the $(\Omega_M, \Omega_\Lambda)$-dependence is isolated in

$$\begin{aligned} q(\Omega_M, \Omega_\Lambda) &= -\frac{\hat{\eta}^2}{s^2} + \sum_i \frac{(\hat{\mu}_i - g_i)^2}{\sigma_i^2} \\ &= \sum_i \frac{(\hat{\mu}_i - g_i + \hat{\eta})^2}{\sigma_i^2}. \end{aligned} \tag{B13}$$

The integral in equation (B9) is thus simply an integral over a Gaussian in η located at $\hat{\eta}$ with standard deviation s; $\hat{\eta}$ is the best-fit (most probable) value of η given Ω_M and Ω_Λ, and s is its conditional uncertainty. As long as η is inside the prior range and $s \ll \Delta\eta$, the value of this integral is well approximated by $s\sqrt{2\pi}$, so that

$$\mathcal{L}(\Omega_M, \Omega_\Lambda) = \frac{s\sqrt{2\pi}}{\Delta\eta} e^{-q/2}. \tag{B14}$$

This is the marginal likelihood one would use to infer the density parameters in the absence of any systematic error terms. Note from equation (B13) that the quadratic form is just what one would obtain by calculating the "profile likelihood" for the density parameters (the likelihood maximized over the nuisance parameters, a frequentist method sometimes used to approximately treat nuisance parameters). Since the uncertainty s is independent of Ω_M and Ω_Λ, it follows from equation (B14) that the marginal likelihood is proportional to the profile likelihood in this problem.

It is also possible to do the calculation analytically using a flat prior for h, spanning a prior range Δh. The corresponding prior for η is exponential;

$$p(\eta|M) = \frac{10^5 c_2}{2a\Delta h} e^{\eta/2a}, \tag{B15}$$

where $a = 2.5 \log e$, a constant known as Pogson's ratio (Pogson 1856). The product of the likelihood and the prior can still be written as $e^{-Q/2}$ with Q quadratic in η; but there is an additional linear term in Q from the prior. Completing the square duplicates equation (B10), but with $\hat{\eta}$ replaced with

$$\hat{\eta} = s^2 \left[-\frac{1}{4a} + \sum_i \frac{g_i - \hat{\mu}_i}{\sigma_i^2} \right]. \tag{B16}$$

The marginal likelihood for Ω_M and Ω_Λ also has a different factor out front; it is given by

$$\mathcal{L}(\Omega_M, \Omega_\Lambda) = \frac{10^5 c_2 s\sqrt{2\pi}}{2a\Delta h} e^{-q/2}. \tag{B17}$$

We present this result for reference only; we use the scale-invariant prior in the body of this work and in the remainder of this Appendix. We have compared calculations with flat and log-flat priors for some models; the resulting marginal likelihoods are negligibly different. Note that equation (B17) is not proportional to the profile likelihood; the proportionality is a special property of the scale-invariant prior.

B.3. Systematic Error in H_0

Among the lightcurve model parameters, regardless of the method, is the fiducial absolute magnitude for SNe Ia, M_0. To obtain definite values for the distance moduli, M_0 must be estimated or at least arbitrarily specified. Let $\hat{M}_0$ denote the value used to calculate the tabulated $\hat{\mu}_i$ estimates. We can write the true value as

$$M_0 = \hat{M}_0 + \delta, \tag{B18}$$

where δ is an uncertain error in our estimate. Since the $\hat{\mu}_i$ estimates are calculated using $\hat{M}_0$, they will have an additive error equal to δ that is systematic (the same for every SNe Ia). To account for this, equation (B7) must be replaced by

$$\hat{\mu}_i = g_i - \eta + \delta + n_i. \tag{B19}$$

Note here the degeneracy between η and δ; since they play identical roles (up to a sign) in the model for the distance moduli, they cannot be individually constrained using only magnitude/redshift data; additional information setting a distance scale to at least one SN is required. Only the quantity $\gamma = \delta - \eta$ can be inferred from the basic data.

P98 arbitrarily specify $\hat{M}_0$, so there is no useful information about δ that can break the degeneracy between δ and η. Recognizing this, they simply forgo any attempt to infer the Hubble constant. Their analysis amounts to replacing η and δ with γ and marginalizing over γ with a flat prior; the resulting marginal likelihood for the density parameters is of the same form as equation (B14), though with an arbitrarily large prior range for γ (which can be ignored since it is common to all models being compared). This is the likelihood we used for the analyses of LBL data (and simulated data) described in § 4 when we assume no evolutionary effects are present.

R98 use Cepheid distances for three SNe Ia to estimate M_0 for use with the MLCS and TF methods. We can consider this extra data to provide a prior distribution for δ; this prior breaks the degeneracy between η and δ in the analysis. R98 report a 10% uncertainty in the Cepheid distance scale for SNe Ia, corresponding to 0.21 magnitude uncertainty in distance moduli. We accordingly adopt a Gaussian prior for δ with zero mean and standard deviation $d = 0.21$, so that

$$p(\delta) = \frac{1}{d\sqrt{2\pi}} e^{-\delta^2/2d^2}. \tag{B20}$$

We can calculate the likelihood for the cosmological parameters by multiplying the joint likelihood for them and δ by this prior, and integrating over δ, as follows.

The quadratic form in the exponential resulting from multiplying this prior by the likelihood resulting from equation (B19) is,

$$\begin{aligned} Q &= \frac{\delta^2}{d^2} + \sum_i \frac{(\hat{\mu}_i - g_i + \eta - \delta)^2}{\sigma_i^2} \\ &= \frac{(\delta - \hat{\delta})^2}{s^2} - q(\Omega_M, \Omega_\Lambda), \end{aligned} \tag{B21}$$

where

$$\frac{1}{s^2} = \frac{1}{d^2} + \sum_i \frac{1}{\sigma_i^2}, \tag{B22}$$

$$\hat{\delta}(\eta, \Omega_M, \Omega_\Lambda) = s^2 \sum_i \frac{\hat{\mu}_i - g_i + \eta}{\sigma_i^2}, \tag{B23}$$

and the $(\eta, \Omega_M, \Omega_\Lambda)$-dependence is isolated in

$$\begin{aligned} q(\Omega_M, \Omega_\Lambda) &= -\frac{\hat{\delta}^2}{s^2} + \sum_i \frac{(\hat{\mu}_i - g_i + \eta)^2}{\sigma_i^2} \\ &= \frac{\hat{\delta}^2}{d^2} + \sum_i \frac{(\hat{\mu}_i - g_i + \eta - \hat{\delta})^2}{\sigma_i^2}. \end{aligned} \tag{B24}$$

As with η in the previous subsection, the integral over δ is a simple Gaussian integral, equal to $s\sqrt{2\pi}$. Thus the marginal likelihood for the cosmology parameters is

$$\mathcal{L}(\eta, \Omega_M, \Omega_\Lambda) = \frac{s}{d} e^{-q/2}. \tag{B25}$$

This is the likelihood used for the analyses of the MLCS and TF data using the default model, as reported in § 4.1.

B.4. Systematic Error From Evolution

The simplest model we considered with a redshift-dependent systematic or evolutionary component is Model I, which adds a shift of size ϵ to the distance moduli of the high redshift SNe Ia. For this model,

$$\hat{\mu}_i = \begin{cases} f_i + \delta + n_i & \text{if } z_i < z_c \\ f_i + \delta + \epsilon + n_i & \text{if } z_i \geq z_c \end{cases}, \tag{B26}$$

with $z_c = 0.15$. We seek the marginal likelihood for the cosmological parameters, requiring us to introduce priors for δ and ϵ and marginalize over them.

The prior for δ is given by equation (B20), and the prior for ϵ is similarly a zero-mean Gaussian, but with a different standard deviation, e;

$$p(\epsilon) = \frac{1}{e\sqrt{2\pi}} \exp\left[-\frac{\epsilon^2}{2e^2}\right]. \tag{B27}$$

The quadratic form associated with the product of these priors and the likelihood function is

$$Q = \frac{\delta^2}{d^2} + \frac{\epsilon^2}{e^2} + \sum_{z_i<z_c} \frac{(\hat{\mu}_i - f_i - \delta)^2}{\sigma_i^2} + \sum_{z_i\geq z_c} \frac{(\hat{\mu}_i - f_i - \delta - \epsilon)^2}{\sigma_i^2}. \tag{B28}$$

To marginalize over ϵ, we complete the square in ϵ by introducing the ϵ uncertainty t, given by

$$\frac{1}{t^2} = \frac{1}{e^2} + \sum_{z_i\geq z_c} \frac{1}{\sigma_i^2}, \tag{B29}$$

and the conditional best-fit value of ϵ,

$$\hat{\epsilon}(\delta, \mathcal{C}) = t^2 \sum_{z_i\geq z_c} \frac{\hat{\mu}_i - f_i - \delta}{\sigma_i^2}. \tag{B30}$$

After completing the square and integrating the resulting Gaussian dependence on ϵ, we find that

$$p(\delta)\mathcal{L}(\delta, \mathcal{C}) = \frac{t}{ed\sqrt{2\pi}} e^{-q/2}, \tag{B31}$$

where

$$q = -\frac{\hat{\epsilon}^2}{t^2} + \frac{\delta^2}{d^2} + \sum_i \frac{(\hat{\mu}_i - f_i - \delta)^2}{\sigma_i^2}. \tag{B32}$$

Note that the sum is over all SNe, and that δ appears in $\hat{\epsilon}$. Completing the square in δ lets us identify the δ uncertainty, s, given by

$$\frac{1}{s^2} = \frac{1}{d^2} + \sum_i \frac{1}{\sigma_i^2} - \frac{t^2}{v^2}, \tag{B33}$$

and the conditional estimate for δ,

$$\hat{\delta}(\mathcal{C}) = s^2 \left[-\frac{t^2}{v^2} F + \sum_i \frac{\hat{\mu}_i - f_i}{\sigma_i^2} \right], \tag{B34}$$

where in these equations we have defined v and F according to

$$\frac{1}{v} = \sum_{z_i\geq z_c} \frac{1}{\sigma_i^2}, \tag{B35}$$

and

$$F = \sum_{z_i\geq z_c} \frac{\hat{\mu}_i - f_i}{\sigma_i^2}. \tag{B36}$$

Using these, we can rewrite q as

$$q = \frac{(\delta - \hat{\delta})^2}{s^2} + q'(\mathcal{C}), \tag{B37}$$

where the dependence on the cosmological parameters is in

$$q'(\mathcal{C}) = -\frac{\hat{\delta}^2}{s^2} - t^2F^2 + \sum_i \frac{(\hat{\mu}_i - f_i)^2}{\sigma_i^2}. \tag{B38}$$

After integrating over the Gaussian dependence on δ, the marginal likelihood is

$$\mathcal{L}(\mathcal{C}) = \frac{ts}{ed}e^{-q'/2}. \tag{B39}$$

This is the likelihood used for analyses of the MLCS and TF data based on Model I.

For Model II, used to model the SF data, the estimated distance moduli are given by

$$\hat{\mu}_i = g_i + \gamma + \beta h_i + n_i, \tag{B40}$$

where as before $\gamma = \delta - \eta$, and $h_i = \ln(1 + \hat{z}_i)$. We will marginalize over γ and β, using a flat prior for γ and a zero-mean Gaussian prior for β with standard deviation b.

As already noted, the γ marginalization is similar to the η marginalization already treated above. The result is

$$p(\beta)\mathcal{L}(\Omega_M, \Omega_\Lambda) = \frac{s}{\Delta\gamma b}e^{-q/2}, \tag{B41}$$

where s is given by equation (B11),

$$\hat{\gamma}(\beta, \Omega_M, \Omega_\Lambda) = s^2 \sum_i \frac{\hat{\mu}_i - g_i - \beta h_i}{\sigma_i^2}, \tag{B42}$$

and the $(\beta, \Omega_M, \Omega_\Lambda)$-dependence is isolated in

$$\begin{aligned} q(\beta, \Omega_M, \Omega_\Lambda) &= -\frac{\hat{\gamma}^2}{s^2} + \sum_i \frac{(\hat{\mu}_i - g_i - \beta h_i)^2}{\sigma_i^2} \\ &= \sum_i \frac{(\hat{\mu}_i - g_i - \beta h_i - \hat{\gamma})^2}{\sigma_i^2}. \end{aligned} \tag{B43}$$

We assume that the prior range for γ, $\Delta\gamma$, contains the peak of the Gaussian. Since this range is common to all models for this data and thus cancels in all calculations, we do not need to specify it more precisely, and we simply drop it from subsequent calculations.

Note that $\hat{\gamma}$ depends on β; we can isolate this dependence by writing

$$\hat{\gamma} = s^2H - \beta s^2 G, \tag{B44}$$

where

$$H = \sum_i \frac{\hat{\mu}_i - g_i}{\sigma_i^2}, \tag{B45}$$

and

$$G = \sum_i \frac{h_i}{\sigma_i^2}. \tag{B46}$$

This helps us to do the remaining marginalization over β. We now complete the square in β, identifying the β uncertainty, τ, given by

$$\frac{1}{\tau^2} = \frac{1}{b^2} - s^2 G^2 + \sum_i \frac{h_i^2}{\sigma_i^2}, \tag{B47}$$

and the conditional best-fit β,

$$\hat{\beta} = \tau^2 \left[-s^2 GH + \sum_i \frac{h_i(\hat{\mu}_i - g_i)}{\sigma_i^2} \right]. \tag{B48}$$

Integrating over the Gaussian dependence on β gives a factor of $\tau\sqrt{2\pi}$, and the final likelihood for the density parameters is

$$\mathcal{L}(\Omega_M, \Omega_\Lambda) = \frac{s\tau\sqrt{2\pi}}{b} e^{-q'/2}, \tag{B49}$$

where

$$q' = \frac{\hat{\beta}^2}{b^2} + \sum_i \frac{(\hat{\mu}_i - g_i - \hat{\beta} h_i - s^2 H)^2}{\sigma_i^2}. \tag{B50}$$

This is the likelihood used for the calculations with Model II in § 4.

REFERENCES

Barbon, R., Rosino, L. and Iijima, T. 1989, AA, 220, 83

Branch, D., 1987 ApJ, 316, 81L

Bruzual, G. 1990, in *Evolution of the Universe of Galaxies: Proceedings of the Edwin Hubble Centennial Symposium*, ed. R. G. Kron (San Francisco: Astronomical Society of the Pacific), pp. 185-199

Burstein, D., & Heiles, C. 1982, AJ, 87, 1165

Carroll, S. M.,Press, W. H. and Turner, E. L. 1992, ARAA, 30, 499

Caldwell, R. R., Dave, R. and Steinhardt, P. J. 1998, Phys. Rev. Lett., 80, 1582

Cen, R. 1998, ApJ, 509, 16

Domínguez, I., Höflich, P., Straniero, O., Wheeler, C., & Thielemann, F.K. 1999, in *Nuclei in the Cosmos V*, in press (astro-ph/9809292)

Einstein, A. 1917, Sitzugnsber. Preuss. Akad. Wiss. Phys.-Math. K1. 142 (English translation in *The Principle of Relativity* (NY: Dover, 1952), p. 177)

Filippenko, A. V., et al. 1992a, ApJ, 384, L15

Filippenko, A. V., et al. 1992b, AJ, 104, 1543

Garnavich, P. M., et al. 1998, ApJ, in press (astro-ph/9806396)

Hamuy, M., Phillips, M., Maza, J., Suntzeff, N., Schommer, R., and Aviles, R. 1995, AJ, 109, 1

Hamuy, M., Phillips, M., Schommer, R., Suntzeff, N., Maza, J., and Aviles, R. 1996a, AJ, 112, 2391

Hamuy, M., Phillips, M., Suntzeff, N., Schommer, R., Maza, J., and Aviles, R. 1996b, AJ, 112, 2398

Hamuy, M. et al. 1996c, AJ, 112, 2408

Höflich, P. and Khokhlov, A. 1996, AJ, 457, 500

Höflich, P., Wheeler, J.C. & Thielemann, F.K. 1998, ApJ, 495, 617

Jaynes, E. T. 1968, IEEE Trans. Sys. Sci. Cyb., SSC-4, 227

Jaynes, E. T. 1979, J. Am. Stat. Assoc., 74, 740

Jefferys, W. H., & Berger, J. O. 1992, Am. Scientist, 80, 64

Jeffreys, H. 1961, *Theory of Probability* (Oxford: Oxford University Press)

Kass, R. E., & Raftery, A. E. 1995, JASA, 90, 773

Leibundgut, B., et al. 1993, AJ, 105, 301

Loredo, T. J. 1999, in *Astronomical Data Analysis and Software Systems (ADASS) VIII*, ed. R. Crutcher & D. Mehringer, (San Francisco: Astronomical Society of the Pacific), in press

Peebles, P. J. E. 1993, *Principles of Physical Cosmology* (Princeton: Princeton University Press), pp. 202-3, 328-330

Perlmutter, S., et al. 1997, ApJ, 483, 565

Perlmutter, S., et al. 1998, AJ, in press (P98) (astro-ph/9812133)

Perlmutter, S., Turner, M. and White, M 1999, submitted to PRL (astro-ph/9901052)

Phillips, M. M., et al. 1987, PASP, 99, 592

Phillips, M. M., Wells, L., Suntzeff, N., Hamuy, M., Leibundgut, B., Kirshner, R., Foltz, C. 1992, AJ, 103, 1632

Phillips, M. M. 1993, AJ, 413, L105

Pogson, R. 1856, MNRAS, 17, 12

Riess, A. G, Press, W. H. and Kirshner, R. P. 1995, ApJ, 438, L17

Riess, A. G, Press, W. H. and Kirshner, R. P. 1996, ApJ, 473, 88

Riess, A. G., et al. 1998, AJ, in press (R98) (astro-ph/9805201)

Sandage, A. 1988, ARAA, 26, 561

Schmidt, B. P., et al. 1998, ApJ, in press (astro-ph/9805300)

Tinsley, B. 1977, ApJ, 211, 621

Turner, M. 1999, astro-ph/9904049

Wasserman, L. 1997, Carnegie Mellon University Dept. of Statistics Technical Report #666, http://www.stat.cmu.edu/www/cmu-stats/tr/tr666/tr666.html

Weinberg, S. 1972, *Gravitation and Cosmology*, (NY: Wiley), pp. 441-5

Weinberg, S. 1989, Rev. Mod. Phys., 61, 1

Yang, R., & Berger, J.). 1997, ISDS Discussion Paper 97-42 (http://www.isds.duke.edu/papers/)

Yoshii, Y. & Takahara, F., 1988, ApJ, 326, 1

Fig. 1.— Joint credible regions for Ω_Λ versus Ω_m for the data set of P98. The 68.3%, 95.4% and 99.7% confidence level contours are shown for (*a*) the data as published and (*b*) after introducing a systematic offset of -0.1 magnitudes to the high redshift ($z > 0.15$) sample.

Fig. 2.— Histogram of the corrections to the absolute magnitudes of the observed SNe Ia, in magnitudes, deduced from the MLCS, TF, and SF methods.

Fig. 3.— (*a*) μ_{MLCS} versus μ_{TF} for the 10 high redshift SNe Ia of R98. (*b*) μ_{MLCS} versus μ_{SF} for the 14 low reshift SNe Ia from the Calán Tololo survey that are used in both R98 and P98. The dashed lines are straight line fits to the data where the slope of the line is fixed to 1.

Fig. 4.— Scatterplots comparing SNe Ia properties inferred using the MLCS and TF methods for the R98 sample of SNe Ia. Compared are (*a*) the host galaxy extinction, A, for the 37 well measured SNe Ia; (*b*) the correction to the absolute magnitude Δ for the 37 well measured SNe Ia; and (*c*) the peak apparent magnitude, m, of the 10 well measured SNe Ia at high redshift. The errors on A and Δ can be estimated to be of the order of 0.1 magnitudes. The dashed lines each have a slope of 1.

Fig. 5.— The difference between the distance moduli inferred using the MLCS and TF lightcurve fitting methods, $\Delta\mu = \mu_{MLCS} - \mu_{TF}$, as a function of redshift z. The errors on the data points are described in the text.

Fig. 6.— The difference between the distance moduli, $\Delta\mu = \mu_{MLCS} - \mu_{TF}$, is plotted versus an estimate of the absolute magnitude, M_B^{AV}, for SNe Ia at low redshift ($z < 0.15$) in the left hand plot and high redshift ($z > 0.15$) in the right hand plot. The dashed line in the right hand plot is the result of a least squares fit to the data which gives a slope of -0.6 ± 0.15.

Fig. 7.— The 68.3% joint credible regions plotted separately for intrinsically dim and intrinsically bright SNe Ia for the (*a*) MLCS, (*b*) TF, and (*c*) SF analysis methods. The contours for the full data set (all M_B^{AV}) are shown as dashed curves.

Fig. 8.— Marginal posterior distribution for h using the MLCS and TF estimated distance moduli. The labeled solid curves show results using our reference model incorporating systematic uncertainties in the Cepheid distances for the SNe Ia used to set the absolute magnitude scale of SNe Ia. The dotted curves show results using Model I, with the prior uncertainty for the high-z offset $e = 0.2$ mag.

Fig. 9.— The 68.3%, 95.4% and 99.7% joint credible regions for Ω_M and Ω_Λ based on our reference model, using distance moduli calculated with the (a) MLCS and (b) TF lightcurve fitting methods.

Fig. 10.— The 68.3%, 95.4% and 99.7% joint credible regions for for Ω_M and Ω_Λ based on Model I. Panels (a) and (b) are for MLCS and TF, respectively, for $e = 0.1$, and (c) and (d) are for MLCS and TF, respectively, with $e = 0.2$.

Fig. 11.— The 68.3%, 95.4% and 99.7% joint credible regions for Model II applied to distance moduli calculated with the SF lightcurve fitting method. Panel (a) is for no evolution, and shown for reference. Panels (b) and (c) are for $b = 0.25$ and $b = 0.5$, respectively.

Fig. 12.— The solid line shows the Bayes factor for $\Omega_\Lambda \neq 0$ versus $\Omega_\Lambda = 0$ as a function of the parameter b of Model II, and the dashed line shows the Bayes factor for Model II versus the zero-evolution reference model.

Fig. 13.— Marginal posterior distributions for Ω_M (and, equivalently, for $\Omega_\Lambda = 1 - \Omega_M$) presuming a flat cosmology, using data from the MLCS (top), TF (middle), and SF (bottom) methods. Results are shown presuming no evolution (solid curves), allowing a small amount of evolution (short-dashed), and allowing a larger amount of evolution (long-dashed).

Fig. 14.— Comparison of μ in cosmological models with and without evolution. The thick solid line is for the best-fit cosmology for SF presuming no evolution $(\Omega_M, \Omega_\Lambda) = (0.75, 1.34)$. The dotted curve is for $(\Omega_M, \Omega_\Lambda) = (1, 0)$ without evolution; the dashed curve is $\beta \ln(1 + z)$ with the best-fit value $\beta = 0.83$ for this cosmology. The thin solid line is the sum, and depicts the best-fit presuming $(\Omega_M, \Omega_\Lambda) = (1, 0)$ with evolution included. The bold, long-dashed line is for $(\Omega_M, \Omega_\Lambda) = (0.3, 0.7)$ without evolution.

Fig. 15.— Results of analyzing simulated data with $(\Omega_M, \Omega_\Lambda, \beta) = (1, 0, 0.5)$. Panels (a) and (b) are for analyses presuming no evolution using data sets with 38 and 186 high-redshift $(0.3 \leq z \leq 1.5)$ sources, respectively; both data sets had 14 low-redshift sources. Panel (c) repeats the analysis of the larger data set with evolution included in the model. The crosses indicate the best-fit parameter values from the analyses; the dots indicate the true values used to generate the data.

Table 1. Dispersion of the data, in magnitudes, from the best fit cosmology for low z ($z < 0.15$) and high z ($z > 0.15$).

Fitting Method		With Corrections	Without Corrections
MLCS			
	low z	0.18 ± 0.02	0.33 ± 0.05
	high z	0.22 ± 0.05	0.20 ± 0.04
TF			
	low z	0.20 ± 0.03	0.33 ± 0.05
	high z	0.17 ± 0.04	0.20 ± 0.05
SF			
	low z	0.18 ± 0.03	0.18 ± 0.03
	high z	0.30 ± 0.03	0.30 ± 0.03

Table 2. Interpretation of Bayes Factors

$\ln(B_{ij})$	B_{ij}	Evidence against H_j
0 to 1	1 to 3	Not worth more than a bare mention
1 to 3	3 to 20	Positive
3 to 5	20 to 150	Strong
> 5	> 150	Very Strong