Cosmic Calibration - Statistical Modeling for Dark Energy and the Cosmological Constants

Tracy Holsclaw

September 2009

1 Introduction

It has been assumed for quite some time that the Universe is expanding (de Sitter, 1934). But in recent years, it has been hypothesized that the Universe is expanding at an accelerating rate (Perlmutter et al., 1999). A new entity called dark energy is one possible way to explain this acceleration (Riess et al., 1998). To explore this theory, observations are needed to verify the current cosmological models and estimate the unknown parameters of interest. Type-Ia supernovae data, cosmic microwave background (CMB), and assumptions on the structure of the Universe can be used. One of the best measures of the expanding Universe is the magnitude-redshift relation of type-Ia supernovae (Perlmutter and et al., 1997). Because type-Ia supernovae are standard candles we can measure their distance and we can also measure their redshift; relating these two quantities should give us more information about movement in the Universe. This relationship can further shed light on the dimensionless matter density parameter for the Universe (Ω_m) and the equation of state (EOS, $w(z)$) of dark energy (DE) (Genovese et al., 2009; Huterer and Turner, 2001). These parameters can hopefully shed light on the acceleration of the Universe and about the mysterious dark energy.

The hypothesized dark energy is not directly detectable or measurable, so other means of investigation are employed to learn about this mysterious influence (Genovese et al., 2009). We investigate some of these probes to gain a better understanding of the possible existence and nature of dark energy. These are the distance-redshift relation quantified through abundance of galaxy clusters, baryon acoustic oscillations measured from the size of galaxy clusters, a measure of the background photons passing through hot clusters called the integrated Sachs-Wolfe effect, weak lensing, and type Ia supernovae.

The type-Ia supernovae are one of the best ways to explore the expanding Universe because of their property as a standardizable candle, consisting of measurable luminosity (Leibundgut, 2004; Riess et al., 1996b; Wood-Vasey and et al., 2008). The peak luminosity when a supernova explodes provides information on the distance for the supernova, which can be directly related to the redshift of the object. So, we have two measures from the same object that can be related: redshift (z) and luminosity distance (D_L).

Obtaining luminosity measurements (D_L) is a rather complicated process that we do not discuss here in detail. The luminosity,D_L, is often referred to in terms of μ: $D_L = 10^{\frac{\mu-25}{5}}$. The peak luminosity is of greatest interest and can be obtained from fitting a light curve to each supernova's data (Pskovskii, 1977). The brightness of the supernova is closely related to the shape of its light curve (Leibundgut, 2001). The color, light curve shape, and peak luminosity are related and nearby supernovae data are needed to assist in estimating these values (W. M. Wood-Vasey et al., 2007; Riess et al., 1996a). We assume that the previous work of the astronomer and light curve fitters is correct and provides valid peak luminosity values (D_L) and error bars for these measurements (σ^2). Most of the variability is assumed to be due to differences in the supernovae at low and high redshift and needed color correcting for reddening from intergalactic dust. This includes complex astronomy, filtering, multiple measurements per supernovae, classification of the supernova as a type-Ia, and light curve fitting processes (Guy et al., 2005).

Redshift (z) is one of the measurable quantities and it comes from spectra (Dodelson, 2003). In our paper, this refers to the cosmological redshift of an object. This is not quite the same as the typical redshift and here encompasses the stretching of space between objects (Filippenko, 1997).

Many cosmological assumptions are needed about the Universe and are a result of work done on general relativity with the Friedmann-Robertson-Walker metric. We also need the Hubble parameter and the speed of light for these equations. We assume to know the speed of light (c) precisely. But we need to estimate the Hubble parameter (H_0); there has been previous work to estimate this parameter from other sources which provides good prior information for us.

Finally, we come to our magnitude-redshift relation accounting for the comoving distance ($r(z)$) to an object (Huterer and Turner, 1999; Genovese et al., 2009). This yields:

$$r(z) = \frac{1}{c(1+z)} 10^{\frac{\mu-25}{5}} \tag{1}$$

$$r(z) = \frac{1}{H_0} \int_0^z \left(\Omega_m (1+s)^3 + (1-\Omega_m)(1+s)^3 e^{-3\int_0^s \frac{-w(u)}{1+u} du} \right)^{-1/2} ds \tag{2}$$

As is clear from equations (1) and (2), the data are related to the parameters of interest via a non-linear equation. There are two approaches that could be taken to estimate the unknown parameters Ω_m, $w(z)$, and H_0. One is to use equation (2) directly. This requires a double integration. Another is to take one or both derivatives and work with r' or r'' to relate the data. The data are discrete so it is impossible to differentiate directly without using some sort of smoothing. The second derivative of such curve would be directly related to $w(z)$ (Sahni and Starobinsky, 2006; Saini et al., 2000; Weller and Albrecht, 2002). Following this approach, some authors have fitted different parametric forms to $w(z)$ (Huterer and Turner, 2001; Astier, 2000). We looked into this approach and found that there is loss of information when taking derivatives and poor behavior for low z values. An approach consisting of taking the first derivative, r' case is discussed later in more detail.

There are modeling options we have not explored; this includes a piecewise constant model that requires binning of the supernovae into categories based on their redshift values (Huterer and Starkman, 2003). The principal component analysis (PCA) has some benefits in that it relies on

the data to weight the model and can focus on analyzing specific redshift values, but it produces a non-continuous estimate for $w(z)$ (Krauss et al., 2007; Crittenden and Pogosian, 2005; Simpson and Bridle, 2006). Other approaches in the literature use parametric forms for $w(z)$ or ansatz (trying many parametric forms) based on higher order polynomials.

Ansatz models typically result in more constrained parameter estimates than those from non-parametric models but this comes at a price because they are more rigid models. We have pursued some new non-parametric models in the hope of giving a better alternative to the inflexible parametric approaches or methods relying on binning. Genovese et al. (2009) work with a bases of functions, which they claim cover the polynomial and piecewise constant cases. They employ least squares hypothesis testing of a few of the favored ranges or sets for $w(z)$. Alternately, we propose a Bayesian Gaussian process formulation to estimate $w(z)$, which gives a flexible fit for $w(z)$. Our methods provide smaller probability bands than some of the frequently used ansatz models.

2 Nonlinear Equation

The real data comes in terms of z, μ, and τ. Let $\mu_i = \alpha(z_i) + \epsilon_i$ where τ_i is approximately one standard deviation of uncertainty in μ_i and σ^2 is the estimated variance of the data. We have Normal distributed errors: $\epsilon_i \sim N(0, \tau_i^2\sigma^2)$ or $\mu_i \sim N(\alpha(z_i), \tau_i^2\sigma^2)$.

We use the equation for $r(z)$ found in (Genovese et al., 2009), which is dependent on $w(z)$. $r(z)$ has some useful equalities as well, that will lead to our transformation, T(z), of the data:
$r(z) = \frac{D_L}{c(1+z)} = \frac{1}{c(1+z)} 10^{\frac{\mu-25}{5}}$ or $\mu = 5log_{10}(c(1+z)r(z)) + 25$. This leads to the relationship:

$$r(z) = \frac{1}{H_0}\int_0^z \left(\Omega_m(1+s)^3 + (1-\Omega_m)(1+s)^3 e^{-3\int_0^s \frac{-w(u)}{1+u}du}\right)^{-1/2} ds$$

This leads to a useful transformation of the data. We will write it in the standard form in equation (3) and then we will move H_0 and rewrite it in a more usable form for our analysis in (4).

$$T(z) = 25 + 5log_{10}\left(\frac{c(1+z_i)}{H_0}\int_0^{z_i} \left(\Omega_m(1+s)^3 + (1-\Omega_m)(1+s)^3 e^{-3\int_0^s \frac{-w(u)}{1+u}du}\right)^{-1/2} ds\right) \quad (3)$$

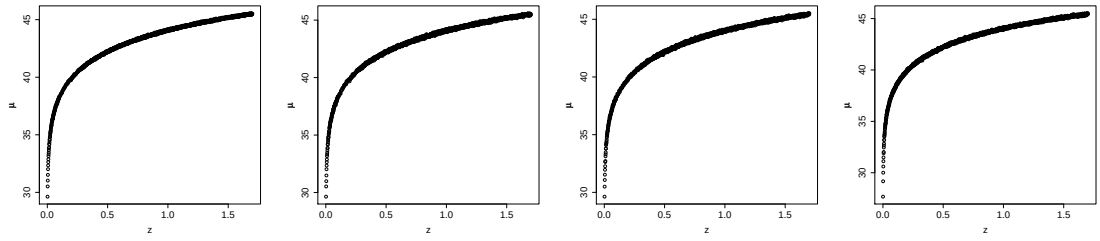
$$T(z) = 25 - 5log_{10}(H_0) + 5log_{10}\left(c(1+z_i)\int_0^{z_i} \left(\Omega_m(1+s)^3 + (1-\Omega_m)(1+s)^3 e^{3\int_0^s \frac{w(u)}{1+u}du}\right)^{-1/2} ds\right) \quad (4)$$

$c = 3 * 10^5$ is the speed of light constant. Ω_{m0} and H_0 can be considered as constants in the simulated data but in reality they are unknown parameters; we examine them as both known and unknown quantities. These are discussed later as we set up priors. We set up our likelihoods for all of our models: $L \propto \left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2}\sum\left(\frac{\epsilon_i}{\tau_i\sigma}\right)^2}$ or $L \propto \left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2}\sum\left(\frac{\mu_i - T(z)}{\tau_i\sigma}\right)^2}$

2.1 Simulated Datasets

We will consider four simulated data sets to test the robustness of our analysis before applying the methods to real data. The data sets include summary statistics for each supernova: a maximum (μ_i) and its associated standard deviation (τ_i). These datasets reflect real data that should soon be available with n=2000 observations. Figure 1 graphs the four data sets; by visual inspection, there seems to be little to no difference but each one is based on a different form of the equation of state, $w(z)$. μ_0, μ_1, μ_2, and μ_3 are simulated and therefore the truth is known.

In these simulated data sets $H_0 = 72$ and $\Omega_m = 0.27$. The weights are set at a constant 0.06 for every observation. In dataset μ_0, the variability is half of the other datasets. The major difference in these four datasets is the underlying $w(z)$, which can be seen in Figure 2. For μ_0 and μ_1 have the simple truth being $w(z) = -1$; the only difference being the uncertainty in μ_0 should be less. In μ_2 the true $w(z)$ has slight curvature and should be fit well by the Chevallier-Polarski-Linder parameterization (Linder, 2003), which we will examine in what we refer to as Model 3. μ_3 is a more complex model with much more curvature and does not correspond to any current models but is a good test for the analysis methods.



(a) μ_0 (b) μ_1 (c) μ_2 (d) μ_3

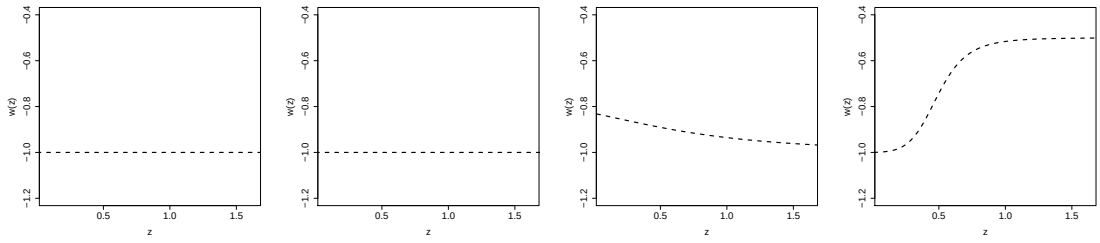
Figure 1: z vs. μ

2.2 MCMC set up: Gibbs and Metropolis-Hasting Steps

2.2.1 Gibbs step for σ^2

All of our model analysis is done using Bayesian methods (Gelman et al., 2004). We use a conjugate prior for σ^2 and this allows a Gibbs step. We let $\pi(\sigma^2) \sim IG(a, b)$; we believe σ^2 should be approximately one because the error bars are accounting for the variation in the data. We will need the posterior for σ^2, where $T(z)$ is the transform given in equation(3):

$$\sigma^2|z, H_0, \Omega_m, w_0 \sim IG\left(\frac{n}{2} + a, \frac{1}{2}\sum\left(\frac{\mu_i - T(z)}{\tau_i}\right)^2 + b\right).$$



(a) μ_0 (b) μ_1 (c) μ_2 (d) μ_3

Figure 2: True $w(z)$

2.2.2 Discussion on H_0

To sample H_0, we use a Gibbs step by transforming it as follows: $h = -5log_{10}(H_0)$. A possible prior for H_0 could be $\pi(H_0) \sim N(c, d^2)$; we found this to be one of the current estimates of H_0 coming from analysis of other types of data sources and it should work as an informative prior. As other data types produce better estimates of H_0, we hope to incorporate this into a more informative prior. This is not a conjugate prior but if we use $-5log_{10}(H_0)$ as our parameter then we can use a transformation of this prior to the form $\pi(h) \sim N(m, s^2)$. Let $h = -5log_{10}(H_0)$ so $\pi(H_0) \sim N(c, d^2)$ is approximately $\pi(h) \sim N\left(m = -5log_{10}(c), s^2 = d^2\left(\frac{-5}{c\ln(10)}\right)^2\right)$ using the delta method, which is now a conjugate prior and we can employ a Gibbs step. The posterior for h is:

$$h|... \sim N\left(\frac{\frac{1}{\sigma^2}\sum\frac{\mu_i - V_i}{\tau_i} + \frac{m}{s^2}}{\frac{1}{\sigma^2}\sum\frac{1}{\tau^2} + \frac{1}{s^2}}, \frac{1}{\frac{1}{\sigma^2}\sum\frac{1}{\tau_i^2} + \frac{1}{s^2}}\right)$$

where $V_i = 25 + 5log_{10}\left(c(1+z_i)\int_0^{z_i}\left(\Omega_m(1+s)^3 + (1-\Omega_m)(1+s)^3 e^{3\int_0^s \frac{w(u)}{1+u}du}\right)^{-1/2} ds\right)$.

This Gibbs step was not employed in much of the analysis because H_0 is highly correlated with other variables. Thus the Gibbs step relies directly on the mixing of Ω_m and also any variables that come from the parameterization of $w(z)$. The method of sampling Ω_m, the variables of $w(z)$, and H_0 together in one Metropolis-Hastings step had superior mixing in most cases.

3 Conventional Parametric Models

Our analysis includes Bayesian estimation for three of the most popular ansatz models for w(u): $w(u) = a$, $w(u) = a + bu$, and $w(u) = a + b\left(\frac{1}{1+u} - 1\right)$. This covers the constant case, linear, and one non-linear case for $w(u)$. First, we discuss a way to distinguish between these models with

Bayesian hypothesis testing and then provide results for each of the three models. We use standard priors in all of these models. These models have been proposed and explored in previous work by Linder (2007) .

Because our data is simulated we have the truth for these models. We will be able to compare our analysis against these truths, as we should get coefficients near these values. For μ_0 and μ_1 things are simple with Model 1 being the truth with $w(z) = -1$. This means that for Model 1 $a = -1$ and both Models 2 and 3 have $a = -1$ and $b = 0$. μ_2 is to come from a model similar to Model 3 with parameters $a = -0.818$ and $b = 0.232$. And Model 1 is much less certain and could be something like $a = -0.892$ and it would have Model 2 parameters near $a = -0.840$ and $b = -0.090$. μ_3 is a more complex model with much more curvature and does not correspond to Models 1, 2, or 3. Model 1 is not able to be fit well but might give something like an average of $a = -0.619$. This would give Model 2 parameters: $a = -0.905$ and $b = -0.324$ and Model 3 may have parameters: $a = -1.083$ and $b = -1.063$.

We used similar priors for all of these models. We did extensive prior sensitivity analysis and found that we could use rather non-informative priors for the parameters in $w(z)$. We will use: $\pi(a) \sim U(-25, 1)$ and $\pi(b) \sim U(-25, 25)$. The prior for σ^2 was chosen because of its properties as a conjugate prior: $\pi(\sigma^2) \sim IG(10, 9)$. And Ω_m and H_0 were given informative priors based on known estimates of these variables: $\pi(H_0) \sim N(72, 1^2)$ and $\pi(\Omega_m) \sim N(0.27, 0.03^2)$.

3.1 Nonlinear Equation Hypothesis Testing

Bayesian hypothesis testing will be used to test the parameters in $w(z)$. This will be done using point masses (using δ functions) in the priors at key levels we wish to test. The δ function is problematic and would typically require just the use of an auxiliary variable to make these into Gibbs steps. However, with our very special non-linear likelihood function it is not possible to use Gibbs steps. So, other methods must be applied here. We introduce an auxiliary variable but must deal with the change in dimensions between spaces. If our $w = -1$ then we are at a lower dimension than if $w \sim U(-25, 1)$, so our prior will be $\pi(w_0|\lambda) \sim \lambda\delta_{w_0=-1} + (1 - \lambda)U(-25, 1)$.

The Algorithm

1. Gibbs step for σ^2: $\sigma^2|z, H_0, \Omega_m, w_0 \sim IG(\frac{n}{2} + 10, \frac{1}{2}\sum\left(\frac{\mu_i - T(z,H_0,\Omega_m,w_0)}{\tau_i}\right)^2 + 9)$

2. Add an auxiliary variable q, which is either: 0 ($w_0 \neq -1$) or 1 ($w_0 = -1$) where $p(q = 1|\lambda) = \lambda$ and $p(q = 0|\lambda) = 1 - \lambda$. Then we re-write the likelihood as follows:

$$L(w, q|x) = L(w = -1|x)I_{q=1} + L(w|x)I_{q=0}$$

Now for the full conditional:

$$\begin{aligned} L(w, q|x)p(q|\lambda)p(\lambda)p(w) &= L(w = -1|x)I_{q=1}p(q = 1|\lambda)p(\lambda)p(w) + L(w|x)I_{q=0}p(q = 0|\lambda)p(\lambda)p(w) \\ &= L(w = -1|x)I_{q=1}\lambda p(\lambda)p(w) + L(w|x)I_{q=0}(1 - \lambda)p(\lambda)p(w) \end{aligned}$$

3. Gibbs step for q:

$$q|w,\lambda... \sim Bern\left(\frac{L(-1|x)\lambda p(\lambda)p(w)}{L(-1|x)\lambda p(\lambda)p(w)+L(w|x)(1-\lambda)p(\lambda)p(w)}\right)$$
$$\sim Bern\left(\frac{L(-1|x)\lambda}{L(-1|x)\lambda+L(w|x)(1-\lambda)}\right)$$

4. Gibbs step for λ:

$$p(w_0,\lambda,q) \propto L(w_0|...) * \lambda^q * 1 * (1-\lambda)^{1-q}$$
$$\lambda|w_0,q... \propto \lambda^{q+1-1}((1-\lambda))^{1-q+1-1} \sim Beta(q+1,2-q)$$

5. Steps for w_0: $w|.... \propto p(w)L(w,q|x) \propto p(w)\left(L(-1|x)\lambda Iq=1+L(w|x)(1-\lambda)I_{q=0}\right)$
   (a) If q=1 then assign $w_{j+1}=w_j$
   (b) If q=0 then propose a w^* from a symmetric distribution such as $w* \sim U(w_j-d,w_j+d)$

$$w|.... \propto L(w|x)(1-\lambda)p(w)I_{q=0}$$
$$\alpha_{MH}=\frac{L(w^*|x)(1-\lambda)\pi(w^*)I_{q=0}}{L(w_j|x)(1-\lambda)\pi(w_j)I_{q=0}}=\frac{L(w^*|x)}{L(w_j|x)}$$

6. Posterior for H_0:

$$p(H_0|...) \sim L(w,q|x)p(q|\lambda)p(\lambda)p(w)p(\Omega_m)p(H_0) \sim L(w,q|x)p(H_0)$$

   (a) when q=1 then $L(w=-1)$; so just let $H_{0,j+1}=H_{0,j}$
   (b) when q=0 then $L(w_{j+1})p(H_0)$ which is a Metropolis step.

7. Posterior for Ω_m:

$$p(\Omega_m|...) \sim L(w,q|x)p(q|\lambda)p(\lambda)p(w)p(\Omega_m)p(H_0) \sim L(w,q|x)p(\Omega_m)$$

   (a) when q=1 then $L(w=-1)$ and let $\Omega_{m,j+1}=\Omega_{m,j}$
   (b) when q=0 then $L(w_{j+1})p(\Omega_m)$ which is a Metropolis step.

Just to complicate things further the steps in 4,5, and 6 need to be combined into one correlated proposal. The three parameters are correlated with one another, so in larger data sets one must take this into account in order to get good mixing in the posterior chains.

Note that in the larger datasets (n=2000) there are issues with the likelihood being zero, which poses problems in step 3. Usually, this can be handled by taking the log of the likelihood and then

later exponentiating but because of the form in 2 this is not straightforward and the computation needs to be done as follows: $L(-1|x) = e^{p_1}$ and $L(w|x) = e^{p_2}$ which leads to:

$$\frac{L(-1|x)\lambda}{L(-1|x)\lambda + L(w|x)\frac{(1-\lambda)}{27}} = \frac{e^{p_1}\lambda}{e^{p_1}\lambda + e^{p_2}\frac{(1-\lambda)}{27}} = \frac{1}{1 + \frac{(1-\lambda)e^{p_2}}{27\lambda e^{p_1}}} = \frac{1}{1 + \frac{(1-\lambda)e^{p_2-p_1}}{27\lambda}}$$

The hypothesis test for the null hypothesis is evaluated as the sum of the qs, not the number of points where $w = -1$ (w is not equal negative one in this algorithm.) In our case $q = 1$ when $w_0 = -1$; so we can simply sum the vector of q and divide by its length to get the proportion associated with the null hypothesis.

3.2 Model 1 - Assume $w(z)$ is a Constant

This model works well with traditional hypothesis testing and results in a single value for $w(z)$ with a probability interval. The downside to this model is that it is very rigid and does not allow for $w(z)$ to change as redshift increases. Many are interested in determining whether $w(z)$ is a a cosmological constant whether it be -1, -2/3, -1/3, or something else (Genovese et al., 2009). If it could be shown that $w(z)$ is constant then this model would be sufficient. In this model we have $w(z) = a$. This leads to a simplified form for $r(z)$:

$$r(z) = \frac{1}{H_0}\int_0^z \left(\Omega_m(1+s)^3 + (1-\Omega_m)(1+s)^3(1+s)^{3a}\right)^{-1/2}.$$

We use a flat rather non-informative prior for a: $\pi(a) \sim Unif(-25, 1)$. H_0 and Ω_m can be fixed or variable; we include a variable Ω_m in the table and graphs. a and Ω_m are highly correlated and they can be sampled jointly from a multivariate Normal distribution with a covariance structure that is obtained after running the process for some time. If they are sampled independently, it affects the mixing negatively.

Table 1 and Figure 3 contain the results of this analysis. In Figure 3 we have the mean fit of $w(z)$ as a black line, the 68% probability interval in dark blue, and the 95% probability interval in light blue. Table 2 contains the results of hypothesis testing for this model.

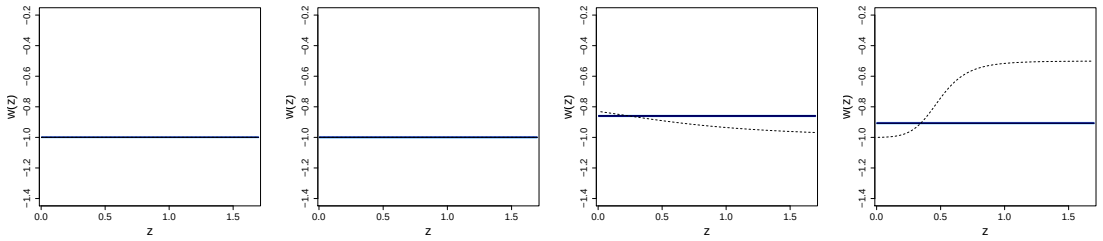
3.2.1 Analysis

The simulations were all run 10,000 times and all acceptance rates for $w(z)$ were around 20% and definately within 10-40%. The mixing of the Metropolis algorithm was acceptable in all cases. As we can see, we do not gain information about Ω_m from the data. Instead this variable is included in the model to show the uncertainty it adds to the fit of $w(z)$.

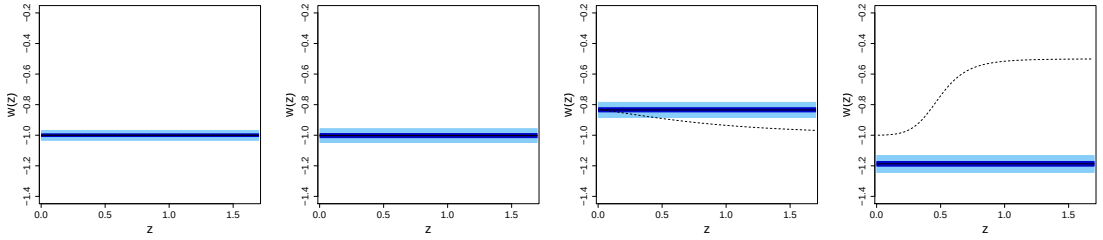
Here we see that Model 1 fits the dataset μ_1 quite well in Tables 1 and 2 and Figure 3 parts (a) and (d). This is expected as the truth for μ_1 is a straight line. However, Model 1 does not fit either dataset μ_2 or μ_3 that well because the truth for $w(z)$ in those datasets is curved. When Ω_m is added to the model it is poorly estimated in the case where the true $w(z)$ is curved.

Table 1: Model 1 - 95% PIs

Dataset	a	Ω_{m0}	H_0	σ^2
μ_0	(-1.003,-0.996)	0.27	72	(0.43,0.47)
μ_1	(-1.005,-0.994)	0.27	72	(0.95,1.05)
μ_2	(-0.865,-0.855)	0.27	72	(0.97,1.07)
μ_3	(-0.912,-0.902)	0.27	72	(1.07,1.18)
μ_0	(-1.033,-0.969)	(0.264,0.278)	(71.74, 72.18)	(0.43,0.47)
μ_1	(-1.048,-0.957)	(0.262,0.282)	(71.63, 72.27)	(0.95,1.05)
μ_2	(-0.884,-0.785)	(0.239,0.271)	(71.77, 72.43)	(0.97,1.08)
μ_3	(-1.243,-1.133)	(0.343,0.359)	(71.84, 72.41)	(0.97,1.07)



(a) μ_0 (b) μ_1 (c) μ_2 (d) μ_3



(e) μ_0 - fit Ω_m and H_0 (f) μ_1 - fit Ω_m and H_0 (g) μ_2 - fit Ω_m and H_0 (h) μ_3 - fit Ω_m and H_0

Figure 3: Model 1

Table 2: Model 1 - Hypothesis Test

Dataset	Percent $a = -1$	a	σ^2
μ_0	0.99	(-1.000,-1.000)	(0.43,0.48)
μ_1	0.98	(-1.000,-1.000)	(0.95,1.05)
μ_2	0.00	(-0.865,-0.855)	(0.97,1.07)
μ_3	0.00	(-0.912,-0.902)	(1.07,1.19)

3.3 Model 2 - Assume $w(u) = a + bu$

Model 2 is fitting a simple linear model with two coefficients to $w(z)$. We let $w(u) = a + bu$ in the $r(z)$ equation. We use a flat rather non-informative prior for a: $\pi(a) \sim Unif(-25, 1)$ and b: $\pi(b) \sim Unif(-25, 25)$. This leads to a simplified version of r(z):

$$r(z) = \frac{1}{H_0} \int_0^z \left(\Omega_m (1+s)^3 + (1-\Omega_m)(1+s)^{3(a-b+1)} e^{3bs} \right)^{-1/2} ds.$$

The general results can be seen in Table 3; we ran the same analysis for $w(z)$ two times: once with Ω_m and H_0 fixed and once with both variable. Table 4 contains the results from performing the hypothesis testing. Graphical fits are in Figure 4 and show a mean line in black, 68% probability bands in dark blue, and 95% probability bands in light blue. We have the truth because these are simulated datasets, so the truth for $w(z)$ is the dashed black line.

Table 3: Model 2 - 95% PIs

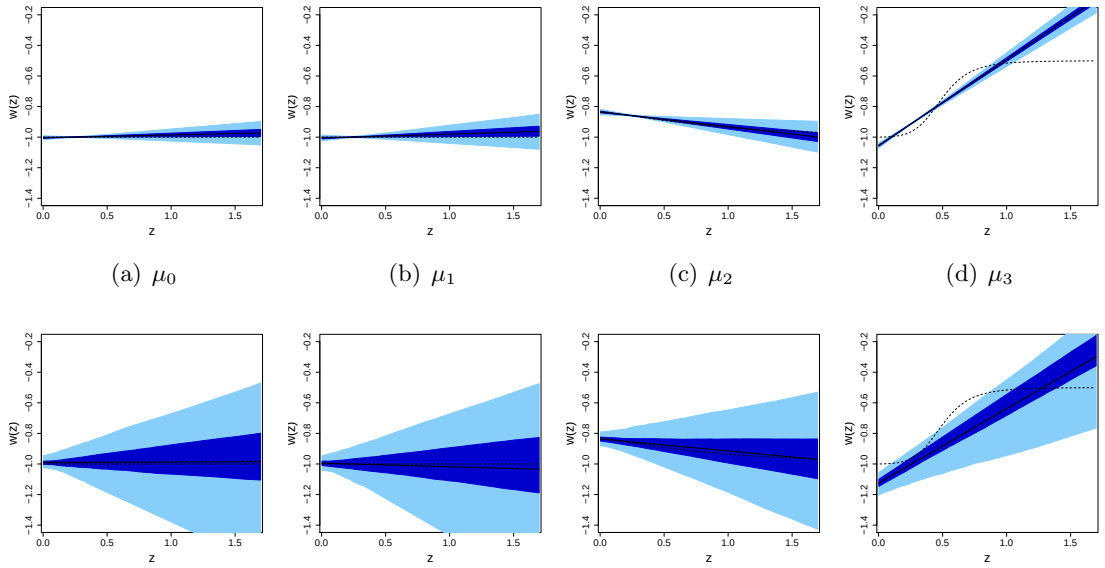
Dataset	a	b	Ω_{m0}	H_0	σ^2
μ_0	(-1.017,-0.991)	(-0.035,0.071)	0.27	72	(0.43,0.47)
μ_1	(-1.025,-0.986)	(-0.055,0.102)	0.27	72	(0.95,1.05)
μ_2	(-0.854,-0.817)	(-0.165,-0.026)	0.27	72	(0.97,1.08)
μ_3	(-1.072,-1.039)	(0.501,0.622)	0.27	72	(0.97,1.08)
μ_0	(-1.024,-0.946)	(-0.382,0.293)	(0.223,0.301)	(71.66, 72.20)	(0.43,0.47)
μ_1	(-1.042,-0.945)	(-0.470,0.288)	(0.226,0.306)	(71.57, 72.24)	(0.95,1.05)
μ_2	(-0.880,-0.790)	(-0.344,0.164)	(0.214,0.301)	(71.69, 72.40)	(0.97,1.08)
μ_3	(-1.204,-1.057)	(0.238,0.653)	(0.267,0.335)	(71.92, 72.50)	(0.97,1.08)

Table 4: Model 2 - Hypothesis Test

Dataset	Percent $a = -1$ and $b = 0$	a	b	σ^2
μ_0	0.99	(-1.000,-1.000)	(0.000,0.000)	(0.43,0.47)
μ_1	0.99	(-1.000,-1.000)	(0.000,0.000)	(0.95,1.05)
μ_2	0.00	(-0.853,-0.817)	(-0.168,-0.031)	(0.97,1.07)
μ_3	0.00	(-1.072,-1.038)	(0.500,0.622)	(0.97,1.08)

3.3.1 Analysis

The simulations were all run 10,000 times for the data sets. All acceptance rates were around 10-40%. Model 2 does an adequate job of fitting datasets μ_1 and μ_2; the flat dataset and the slightly curved one. The hypothesis testing indicates that μ_1 has a truth of $w(z) = -1$, which means this analysis is working as it should. However, Model 2 does not do a superb job of capturing the true



(a) μ_0 (b) μ_1 (c) μ_2 (d) μ_3

(e) μ_0 - fit Ω_m and H_0 (f) μ_1 - fit Ω_m and H_0 (g) μ_2 - fit Ω_m and H_0 (h) μ_3 - fit Ω_m and H_0

Figure 4: Model 2

$w(z)$ for μ_3. This can be seen in Figure 4 parts (c) and (f); the dashed true line for $w(z)$ is outside the probability bands.

Model 2 does estimate the value for Ω_m and H_0 fairly well in all cases, which is an improvement from Model 1. When Ω_m and H_0 are added to the model the probability bands are much wider and our fit of $w(z)$ more uncertain. The parameters a and b are more highly correlated when Ω_m and H_0 are fixed.

3.4 Model 3 - Assume $w(u) = a + b(\frac{1}{1+u} - 1)$

Model 3 was recommended by the cosmologists as a form of interest for $w(u)$. Linder proposes this model as a robust form for an equation of state (EOS) with monotonic behavior (Linder, 2003). The parameters were sampled jointly because they are correlated to one another. Linder advocates using the more robust parameterization $w(u) = a^* + b^*(1 - \frac{1}{1+u})$ to avoid any bias with the cosmological constant (Linder, 2006). We are including two-way plots of the parameters Ω_m, a, and b. He also examines higher order polynomial fits and find they do not estimate the EOS correctly and oscillate heavily and must be truncated for high z values (Linder, 2007). So, here we let $w(u) = a + b(\frac{1}{1+u} - 1) = a + \frac{-bu}{1+u}$ in the $r(z)$ equation and not include any higher order terms.

So now we have:

$$r(z) = \frac{1}{H_0}\int_0^z \left(\Omega_m(1+s)^3 + (1-\Omega_m)(1+s)^{3(a-b+1)}e^{\frac{3bs}{1+s}}\right)^{-1/2} ds.$$

We use a flat rather non-informative prior for a: $\pi(a) \sim Unif(-25, 1)$ and b: $\pi(b) \sim Unif(-25, 25)$. The simulations were all run 10,000 times. As with Model 1, the parameters were sampled jointly because they are correlated to one another; this includes a, b, H_0, and Ω_m. The results of this analysis can be seen in Tables 5 and 6 and Figure 5.

Table 5: Model 3 - 95% PIs

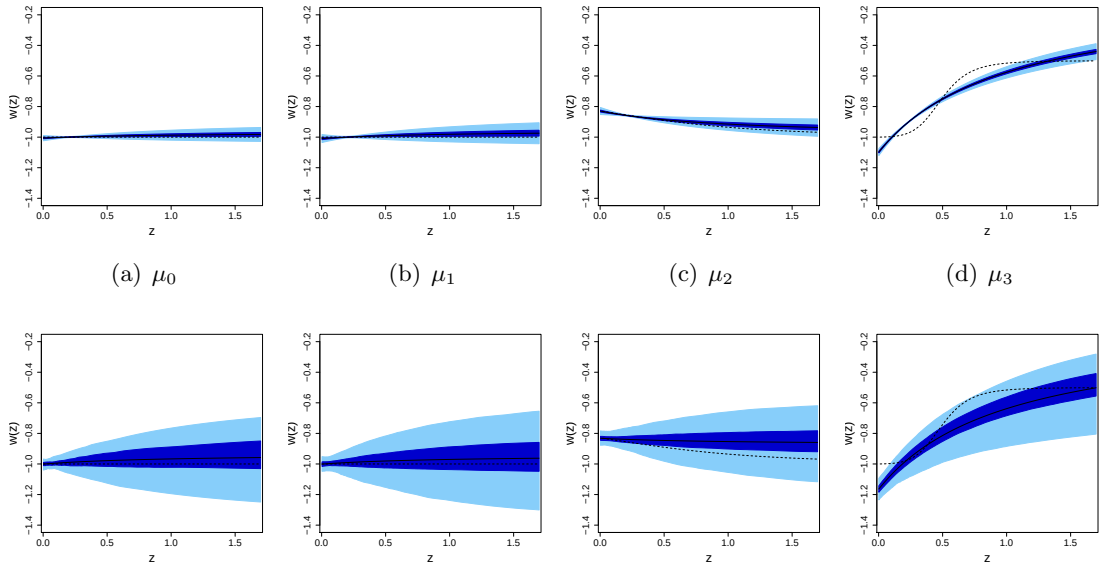
Dataset	a	b	Ω_{m0}	H_0	σ^2
μ_0	(-1.024,-0.989)	(-0.138,0.061)	0.27	72	(0.43,0.47)
μ_1	(-1.036,-0.983)	(-0.205,0.094)	0.27	72	(0.95,1.05)
μ_2	(-0.852,-0.805)	(0.047,0.299)	0.27	72	(0.97,1.07)
μ_3	(-1.122,-1.077)	(-1.164,-0.929)	0.27	72	(0.97,1.08)
μ_0	(-1.040,-0.967)	(-0.473,0.421)	(0.237,0.293)	(71.72, 72.25)	(0.43,0.47)
μ_1	(-1.048,-0.951)	(-0.545,0.510)	(0.229,0.297)	(71.61, 72.34)	(0.95,1.05)
μ_2	(-0.878,-0.784)	(-0.322,0.473)	(0.215,0.291)	(71.72, 72.45)	(0.97,1.08)
μ_3	(-1.237,-1.096)	(-1.368,-0.613)	(0.243,0.324)	(72.05, 72.65)	(0.97,1.08)

Table 6: Model 3 - Hypothesis Test

Dataset	Percent $a = -1$ and $b = 0$	a	b	σ^2
μ_0	0.99	(1.000,1.000)	(0.000,0.000)	(0.43,0.47)
μ_1	0.98	(1.000,1.000)	(0.000,0.000)	(0.95,1.05)
μ_2	0.00	(-0.852,-0.804)	(0.044,0.305)	(0.97,1.08)
μ_3	0.00	(-1.121,-1.078)	(-1.159,-0.936)	(0.97,1.08)

3.4.1 Analysis

The simulations were all run 10,000 times for the data sets. All acceptance rates were around 10-40%. Model 3 is made to pick up the slight curvature in dataset μ_2. Model 3 is capable of fitting μ_1 and μ_2, however, it is not flexible enough to capture the true $w(z)$ in dataset μ_3, as seen in Figure 5. Model 3 does an adequate job of estimating the unknown Ω_m and H_0 parameters in almost all cases (Table 5). When these two parameters are unknown, the results have much wider probability intervals. But overall Model 3 did fairly well and the hypothesis test in Table 6 also chooses μ_1 as having the truth be $w(z) = -1$ which is correct.



(a) μ_0 (b) μ_1 (c) μ_2 (d) μ_3

(e) μ_0 - fit Ω_m and H_0 (f) μ_1 - fit Ω_m and H_0 (g) μ_2 - fit Ω_m and H_0 (h) μ_3 - fit Ω_m and H_0

Figure 5: Model 3

4 Non-parametric Modeling of $w(z)$

We now employ non-parametric models to do inference on $w(z)$. In Model 4, we assume $w(z)$ is a Gaussian process and use the $r(z)$ equation as our relationship for the data (Banerjee et al., 2004). A Gaussian process is a stochastic process such that when sampled at any finite collection of points, the values jointly follow a multivariate Normal distribution. Thus the process can be defined by its mean and correlation functions. This model is advantageous in that it allows for a flexible fit to a function, $w(z)$, based on probability theory rather than assuming a parametric form like the previous models. We still assume that the errors of the data follow a Normal distribution as in the previous models and we use the same likelihood.

4.1 Model 4 - Gaussian Process on $w(u)$

For the Gaussian process, we assume that $w(u_1), ..., w(u_n)$ for any collection of $u_1, ..., u_n$ follows a multivariate Gaussian distribution with mean, θ, and powered exponential covariance function written in a non-standard form: $K(u, u') = \kappa^2 \rho^{|u-u'|^\alpha}$. We consider $w(u)$ to be a Gaussian process: $w(u) \sim GP(\theta, \kappa^2 K(u, u'))$.

A standard powered exponential covariance structure suffices in this case but we reparametrize the typical $e^{-\lambda} = \rho$, so $K(u, u') = \rho^{|u-u'|^\alpha}$. α is typically equal to two but this leads to issues in

Cholesky decomposition and matrix inversion for us. To fix this issue we let α be less than two and in most cases something like 1.9999 works. The Gaussian correlation was tried in this manner, as well as, the Matern with smoothness parameter (ν) equal to 1.5. Ultimately, we ended up choosing $\alpha = 1$ for a more flexible model that fit both flat and curved $w(z)$ equations. Other possible options would typically be to add a nugget term or jitter but this is not possible with the interpolation method we are using. Also, Cholesky decomposition with pivoting was considered.

The main parameter of interest is ρ to test if their is enough relationship in the data for a GP to be needed. In most of the simulated data sets, we would expect this to be the case. The mean of the GP is also important, $\theta = -1$ is typically the value of interest for $w(z)$; other values for θ were used and did not change the GP or results. Allowing θ to vary did not produce stable posteriors for this and other parameters. All of the parameters in the model need priors, for our analysis they will be: $\pi(\rho) = Beta(6, 1)$, $\pi(\kappa^2) = IG(25, 9)$, and $\sigma^2 = IG(10, 9)$.

Altered Gaussian process A typical Gaussian process prior is set up to have a mean and correlation function: $w(u) \sim GP(\theta, \Sigma_{\rho,\kappa^2})$. In our case with our likelihood function, this leads to the following posterior:

$$\sigma^2, \rho, \kappa^2 | \mu_i, \tau_i^2, z_i \propto L(z_i, \mu_i, \tau_i | w(u), \sigma^2) GP(w(u) | \rho, \kappa^2) \pi(\rho) \pi(\kappa^2) \pi(\sigma^2)$$

Instead of the usual Gaussian process set up just described, we went with an altered form to allow for slower changes in the GP. We let $w(u) \sim MVN(\theta, \Sigma)$ and $\Sigma^{-1/2}(w(u) - \theta) = w^o(u) \sim MVN(0, I)$. So now our posterior becomes:

$$L(z_i, \mu_i, \tau_i | w^o(u), \rho, \kappa^2, \sigma^2) MVN(w^o(u); 0, I) \pi(\rho) \pi(\kappa^2) \pi(\sigma^2)$$

4.1.1 The Not so Simple Algorithm

This algorithm is not a simple Gaussian process case because of the double integration and the altered GP form previously discussed. The integration could be done using trapezoid integration and setting up partitioning for both integrals. While that type of algorithm is simpler it takes extensive computational power to perform. Instead, we will use more difficult methods where the inner integral of $T(z_i, w(u), \rho, \kappa^2)$ is calculated based on properties of the GP. Once again, $w(u)$ is our GP (as seen in equation(5)) with m points but now we want to evaluate $y(s) = \int_0^s \frac{w(u)}{1+u} du$ and we know based on the properties of the GP that $y(s)$ is a GP (as seen in equation(6)), too. Equation(7, 8) shows the relationship between the two GPs: $w(u)$ and $y(s)$.

$$w(u) \sim GP(\theta, \Sigma_{22} = \kappa^2 \rho^{|u-u'|^\alpha}) \tag{5}$$

$$y(s) \sim GP\left(\theta \ln(1+s), \Sigma_{11} = \kappa^2 \int_0^s \int_0^{s'} \frac{\rho^{|u-u'|^\alpha}}{(1+u)(1+u')} du du'\right) \tag{6}$$

$$\begin{bmatrix} y(s) \\ w(u) \end{bmatrix} \sim MVN\left[\begin{bmatrix} \theta \ln(1+s) \\ \theta \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right] \tag{7}$$

$$y(s)|w(u) = E(y(s)|w(u)) = \theta \ln(1+s) + \Sigma_{12}\Sigma_{22}^{-1}(w(u) - \theta) \tag{8}$$

We can draw more $y(s)$ points than are in original partition of $w(u)$. This will act as an interpolator so we have enough points for a smoother outer integral calculation. This method does not require one large covariance matrix to be computed and inverted. And it does the inner integral and partitioning/smoothing process all in one step. Σ_{11} is never computed so the slowness of computing double integrals for each entry of the covariance matrix is avoided.

However, we need $\Sigma_{12} = \kappa^2 \int_0^{s'} \frac{\rho^{|u-s|^\alpha}}{1+u} du$ which requires a single integral for every entry.

4.1.2 Chebyshev-Gauss quadrature method for solving the single integral

$K(s,s') = \int_0^{s'} \frac{\rho^{|u-s|^\alpha}}{(1+u)} du$ this integral used in the correlation matrix cannot be solved analytically and there is no good approximation because the limits of integration do not go between 0 and infinity. Numerical methods have to be used and Chebyshev-Gauss quadrature provides a good alternative to other forms of slower numerical integration. R uses Gauss-Kronrod quadrature.

First, we must change the limits of integration from $[0, s']$ to $[-1, 1]$ to be able to use this form of quadrature using the rule: $\int_a^b f(x)dx = \frac{b-a}{2} \int_{-1}^1 f\left(\frac{(b-a)x}{2} + \frac{a+b}{2}\right) dx$. Chebyshev-Gauss quadrature uses Chebyshev polynomials of the first kind as its orthogonal polynomials, $\frac{1}{\sqrt{1-x^2}}$. In total one integral is approximated as such, $\int_{-1}^1 \frac{f(x)}{\sqrt{1-x^2}} dx \sim \sum_{i=1}^n \gamma_i f(x_i)$, $x_i = cos\left(\frac{(2i-1)\pi}{2n}\right)$, and $\gamma_i = \frac{\pi}{n}$. In our case we let: $u_i = cos\left(\frac{(2i-1)\pi}{2n}\right)$, and n=100 so the weights are constant and fully specified. We take our single integral in equation(9) and who how we implement the quadrature in equation(10).

$$K(s,s') = \int_0^{s'} \frac{\rho^{|u-s|^\alpha}}{(1+u)} du = \frac{s'}{2} \int_{-1}^1 \frac{\rho^{|\frac{s'u}{2} + \frac{s'}{2} - s|^\alpha}}{(1 + \frac{s'u}{2} + \frac{s'}{2})} \sqrt{1-u^2} \frac{1}{\sqrt{1-u^2}} du \tag{9}$$

$$K(s,s') = \frac{s'}{2} \sum_{i=1}^n \gamma_i \frac{\sqrt{1-u_i^2}}{(1 + \frac{s'u_i}{2} + \frac{s'}{2})} \rho^{|\frac{s'u_i}{2} + \frac{s'}{2} - s|^\alpha} \tag{10}$$

4.1.3 The Not so Simple Algorithm Continued

1. Initialize all variables: $\rho = \rho_1$, $\kappa^2 = \kappa_1^2$, and $w^o(u) = w_{m,1}^o(u)$. w(u) is a vector with m points in our GP and y(s) has m*h points. We run this algorithm q=1,...,Q times. Set all tuning parameters,$\delta_{1,2,3}$, which needs to be tuned until good mixing occurs. Also, all proposals used are symmetric and do not need a jumping function in α_{MH}. ρ between zero and one, the variance parameters σ^2 and κ^2 must be greater than zero, any proposals that do not fit this criteria will be rejected.

2. Propose $\rho^* = Unif(\rho_1 - \delta_1, \rho_1 + \delta_1)$

   (a) Compute the covariance matrix $K_{22\rho^*} = \rho^{*|u_j - u_i|^\alpha}$

(b) Compute the Cholesky decomposition for $K_{22\rho^*} = U'_{\rho^*}U_{\rho^*}$

(c) Compute the special $K_{12\rho^*} = \int_0^{s'} \frac{\rho^{*|u-s|^\alpha}}{1+u}du$

(d) We want $y_{\rho^*}(s) = \theta\ln(1+s) + [\kappa_{q-1}^2 K_{12*}][\kappa_{q-1}^2 K_{22*}^{-1}](w_{\rho^*}(u) - \theta)$
where: $w_{\rho^*}(u) = [\kappa_{q-1}U'_{\rho^*}]w^o_{m,q-1} + \theta$

$$\begin{aligned} y_{\rho^*}(s) &= \theta\ln(1+s) + [\kappa_{q-1}^2 K_{12*}][\kappa_{q-1}^2 K_{22*}]^{-1}(\big(\kappa_{q-1}U'_{\rho^*}w^o_{m,q-1} + \theta\big) - \theta) \\ &= \theta\ln(1+s) + \kappa_{q-1}K_{12*}[(U'_{\rho^*}U_{\rho^*})^{-1}U'_{\rho^*}]w^o_{m,q-1} \\ &= \theta\ln(1+s) + \kappa_{q-1}K_{12*}[U_{\rho^*}^{-1}]w^o_{m,q-1} \end{aligned}$$

(e) $L(z_i, \mu_i, \tau_i|w_{\rho^*}, \sigma_{q-1}^2) = e^{-\frac{1}{2}\sum\left(\frac{\mu_i - T(z_i, w_{\rho^*}(u))}{\tau_i\sigma_i}\right)^2}$ where the definite integrations in $T(z_i, w_{\rho^*}(u))$ are done numerically through summations of the trapezoid algorithm.

(f) If we accept $\alpha_{MH} = \frac{L_{\rho^*}\pi(\rho^*)}{L_{\rho_{q-1}}\pi(\rho_{q-1})}$ then we let $\rho_q = \rho^*$

3. Draw $\kappa^{2*} = Unif(\kappa_{q-1}^2 - \delta_2, \kappa_{q-1}^2 + \delta_2)$

(a) Compute $y_{\kappa^{2*}}(s) = \theta\ln(1+s) + \kappa^* K_{12\rho_q}[U_{\rho_{q-1}}^{-1}]w^o_{m,q-1}$

(b) $L(z_i, \mu_i, \tau_i|w_{\kappa^{2*}}, \sigma_{q-1}^2) = e^{-\frac{1}{2}\sum\left(\frac{\mu_i - T(z_i, w_{\kappa^{2*}}(u))}{\tau_i\sigma_i}\right)^2}$ where the definite integrations in $T(z_i, w_{\kappa^{2*}}(u))$ are done numerically through summations of the trapezoid algorithm.

(c) If we accept $\alpha_{MH} = \frac{L_{\kappa^{2*}}\pi(\kappa^{2*})}{L_{\kappa_{q-1}^2}\pi(\kappa_{q-1}^2)}$ then we let $\kappa_q^2 = \kappa^{2*}$

4. I want to propose a non-standard w_m^* as my Gaussian Process. I start with drawing a proposal for $w^{o*} \sim MVN(w^o_{q-1}, \delta_3 I_{mxm})$

(a) Compute $y^*(s) = \theta\ln(1+s) + \kappa_q K_{12q}[U_q^{-1}]w^{o*}_{m,}$

(b) $L_{z_i,\mu_i,\tau_i|w^*_{new}(u),\sigma_{q-1}^2} = e^{-\frac{1}{2}\sum\frac{\mu_i - T(z_i, w^*_{new}(u))}{\tau_i\sigma}^2}$

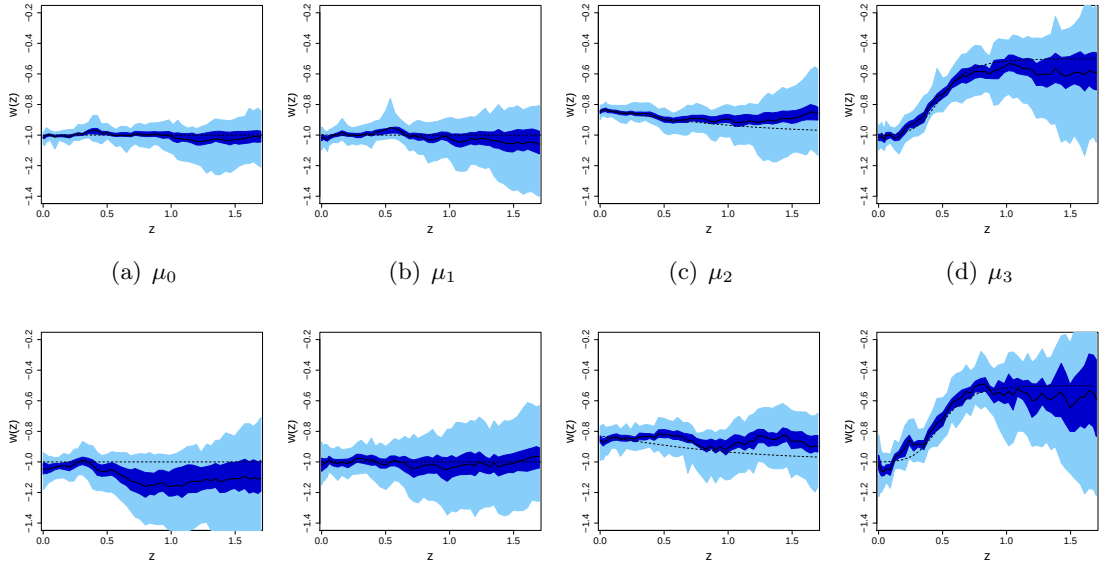
(c) If we accept $\alpha_{MH} = \frac{L_{w^*_{new}(u)}MVN(w_m^{o*}|0,I)}{L_{w_{q-1}}MVN(w_{m,q-1}|0,I)}$ then $w^o_{m,q}(u) = w_m^{o*}(u)$ and the Gaussian process realization is $w_{m,q}(u) = w_m^*(u)$

5. $\sigma_q^2|... \sim IG\left(\frac{n}{2} + 10, \frac{1}{2}\sum\left(\frac{\mu_i - T(z|...)}{\tau_i}\right)^2 + 9\right)$

6. Repeat steps 2-6, Q times and rerun the entire algorithm as needed after resetting the tuning parameters.

4.1.4 Results of Model 4

The best way to summarize the results of the Gaussian process model is with the typical plots. As we see in Figure 6, the Gaussian process model is capable of finding the true $w(z)$ for all three datasets: μ_1, μ_2, and μ_3. The roughness in the plots comes from the exponential correlation function. All other correlation functions over smooth the fit and do allow enough flexibility for fitting a true $w(z)$ like μ_3. We chose a model with more rough edges to be able to fit a more curvaceous $w(z)$.



(a) μ_0 (b) μ_1 (c) μ_2 (d) μ_3

(e) μ_0 - fit Ω_m and H_0 (f) μ_1 - fit Ω_m and H_0 (g) μ_2 - fit Ω_m and H_0 (h) μ_3 - fit Ω_m and H_0

Figure 6: Model 4

The two parameters we can can fit with this model are Ω_m and H_0. And then in the second set of simulations we let Ω_m and H_0 both be parameters to be estimated. We used a Gibbs step for H_0 and it worked well and had better mixing than a Metropolis-Hastings step in this analysis.

Table 7: Model 4

Dataset	Ω_m	H_0
μ_0	(0.265,0.299)	(71.74, 72.37)
μ_1	(0.250,0.290)	(71.53,72.34)
μ_2	(0.243,0.280)	(71.77,72.62)
μ_3	(0.237,0.279)	(71.81,72.68)

Overall, the Gaussian process fits well for all three datasets when Ω_m is known. But we see that the true Ω_m is at the boundary of the estimated interval for μ_0 and μ_3. This is also seen in several of the other models. We could go back and run a smoother fit using a Gaussian correlation function for μ_0, μ_1, and μ_2 because they do not need the extra flexibility of the exponential correlation function.

4.2 Approximating a Gaussian Process with Damped Hermite Polynomials

Linder opposes fitting $w(z)$ with higher order polynomials because of their oscillating behavior and the need to truncate the fit for higher z values, as well as, the fact that they can introduce bias in the fit (Linder, 2007). However, Genovese et al. (2009) use a mixture of basis functions (which need not be orthonormal) to fit $w(z)$ in their analysis. But they find that the $w(z)$ equals a constant is the best while using model selection criteria. Their analysis allows for all three subforms that include polynomial basis, scale factor polynomials, and piecewise constant fits and can be extended to include B-splines, orthogonal polynomials and wavelet analysis (Genovese et al., 2009). However, we believe they may have only tried the first few orders of the basis function in their comparison.

Here we follow the work of Steinberg (2004) on damped Hermite polynomials as the basis functions. We want to fit a series expansion to $W(u) = \frac{w(u)}{1+u}$ in this way: $W(u) = \gamma(u) + \sum_{s=0}^{\infty} \beta_s J_s(u)$ where $\gamma(u) = \frac{-\theta}{1+u}$ is the mean of $W(u)$ where θ is a constant, $J_s(u) = H_s^* \exp -\frac{wu^2}{2(1+m)}$ are the damped polynomials, and $H_s^*(u) = H_s(u/\sqrt(2))/(2^s s!)^{1/2}$ is a physicist Hermite polynomial. The Hermite polynomials, H_s^*, are orthonormal with respect to a standard Normal, and have the property: $E[H_s^*(W)H_t^*(W)] = \int_{-\infty}^{\infty} \exp^{-u^2/2} H_s^*(u)H_t^*(u)du = \delta_{s,t}$. But the damped polynomials $J_s(u)$ are not orthogonal.

We are interested in the Hermite polynomial parameterization because it corresponds to a Gaussian process with Gaussian correlation where range parameter $\lambda = \frac{m}{2(1-m^2)}$ and variance $\sigma^2 = \tau^2(1-m^2)^{-1/2}$ and $0 \leq m \leq 1$. The prior on β also contains m, the damping parameter, $\pi(\beta_s) \sim MVN(0, \sigma^2 m^s)$. Both τ^2 and m are unknown parameters that must be estimated, so we have priors: $\pi(m) \sim Beta(6,1)$ and $\tau^2 \sim IG(25,9)$, and then for our variance in the likelihood we have $\sigma^2 \sim IG(10,9)$.

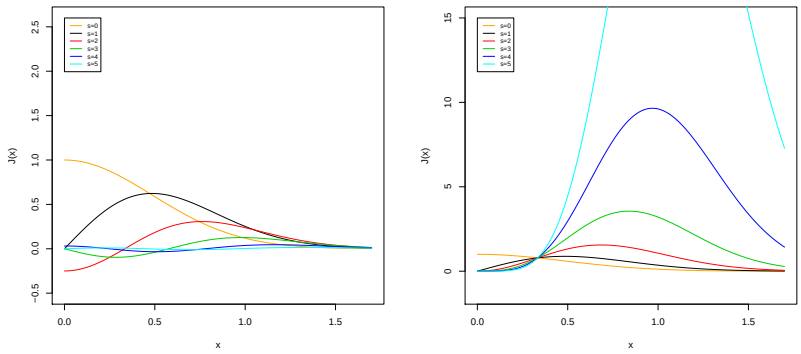
If we let $k = \sqrt{\frac{1+m}{m}}$, we notice that k^2 is a variance term and it $2 \leq k^2 \leq \infty$. We found that because our distance data (z) has range between zero to two the basis of Hermite polynomials needs to be rescaled by a factor of two or three like $J(2z)$ or $J(3z)$ before we begin the analysis. The unscaled polynomials do not dampen until closer to 3.0 on the z scale.

We will write the series expansion up to the first two polynomials with β's and show that if the

polynomials are truncated then they can be re-written with new coefficients,a's.

$$\begin{aligned} W(u) &= \gamma(u) + \sum_{s=0}^{\infty} \beta_s J_s(u) \\ &= \gamma(u) + \exp -\frac{mu^2}{2(1+m)}[\beta_0 - \frac{1}{\sqrt{2}}\beta_2 + \beta_1 u + (\beta_2 \frac{1}{\sqrt{2}})u^2 ...] \\ &= \gamma(u) + \exp\left(-\frac{mu^2}{2(1+m)}\right)\left[a_0 + a_1 u + a_2 u^2 ...\right] \end{aligned}$$

In figure 7, we see the first six damped Hermite polynomials. The first plot is of the H^* polynomials with β_s coefficients and the second with the a_s coefficients. We set $m = 0.9$ for these graphs and use a rescaling factor of 3.



(a) β_s Polynomials (b) a_s Polynomials

Figure 7: Damped Hermite Polynomials

We need to integrate $W(u)$; this does not yield a fully analytic solution. We choose to use $W(u)$ instead of $w(u)$ because we wish to perform the inner integration in our transform using the properties of the Normal distribution which is well approximated by our statistical software.

$$\begin{aligned} \int_0^s W(u)du &= \int_0^s \gamma(u)du + e^{-\frac{mu^2}{2(1+m)}}\left(a_0 + a_1 u + a_2 u^2\right) du \\ &= \int_0^s \frac{-\theta}{1+u} du + a_0 \int_0^s e^{-\frac{u^2}{2k^2}} ds + a_1 \int_0^s u e^{-\frac{u^2}{2k^2}} du + a_2 \int_0^s u^2 e^{-\frac{u^2}{2k^2}} du \\ &= -\theta \ln(1+s) + a_0\sqrt{2\pi k^2}[N(s|0,k^2) - 1/2] + a_1\{-k^2 e^{-s^2/2k^2} + k^2\} \\ &\quad + a_2\{-k^2 e^{-s^2/2k^2}(s) + k^2\sqrt{2\pi k^2}[N(s|0,k^2) - 1/2]\} \end{aligned}$$

4.2.1 Results

The best fit comes from using only the first two terms of the damped Hermite polynomial expansion. If any higher order terms are added the analysis breaks down with our current data. We get poor mixing in the MCMC, wide probability bands, and a poor fitting mean especially for high z values. This is partly expected as the same thing happens with other basis expansions. We would like to continue to do further work with this analysis by adding Ω_m and H_0, and possibly adding θ as a variable in the mean.

4.3 The Real Data

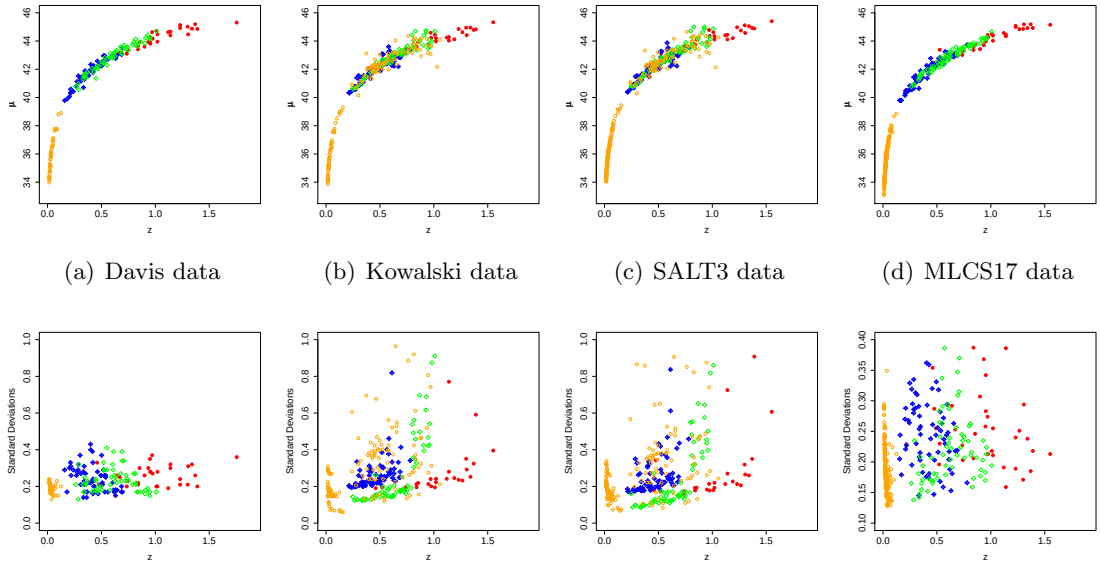
There are currently several sets of real data available for the supernovae Ia (SNe Ia). All four sets contain the same basic set of supernovae data and then a few more points depending on the light curve fitters criteria. Three of these sets are fit using SALT light curve fitter and one uses the MLCS17 light curve fitter. We call these data sets the Davis data set (n=192) (Davis et al., 2007; Riess et al., 2007; W. M. Wood-Vasey et al., 2007), Kowalski data set (n=307) (Kowalski et al., 2008), SALT3 data set (n=397), and MLCS17 data set (n=372). The real data has larger error bars associated with μ than in our simulated data and this produces much larger error bands on our fits of w(z). Also, with the real data both Ω_m and H_0 are currently unknown but have general values found from other data sources, we use as priors.

Figure 8 is of the real data. We have four general types of surveys and each is represented with a separate color and symbol: HST - red filled circles, ESSENCE - blue diamonds with x's, SLOAN-LEGACY (SNLS) - green diamond, and HZSST - orange circles (which could be subdivided further based on the team). These graphs show that the different telescopes focus their search in different ranges of the redshift. Figure 8 (e)-(h) shows the estimated errors for each observation (μ); these are given by our light curve fitter.

4.3.1 Real Data Analysis

First, we fit ansatz Models 1, 2, and 3 to the four real data sets. Ω_m and H_0 are variable in all of these models. The prior for H_0 is changed to $\pi(H_0^*) \sim N(72, 4^2)$. This allows for the extra term, M, that gets incorporated into H_0 from the marginalization of the data that happens during the light curve fitting. The estimates of H_0 may be closer to 65 than to 72 because of this term as seen in Table 8. H_0 and M cannot be distinguished in these equations, so we just call the combined item H_0^*. The results of the analysis for Models 1, 2,and 3 can be best seen in Table 8 and Figure 9.

Analysis Every model and dataset is within the 95% probability of having the truth be $w(z) = -1$. This is partly due to the fact that our bands for most of these models are quite large. More data or less variability data would help to reduce these bands. Also, if we had more certainty about the parameters like H_0 and Ω_m, this could reduce the uncertainty in the cosmological parameter. All of our 95% probablity estimates of Ω_m include the value 0.27, which is of interest. It is also



(a) Davis data (b) Kowalski data (c) SALT3 data (d) MLCS17 data

(e) Davis data (f) Kowalski data (g) SALT3 data (h) MLCS17 data

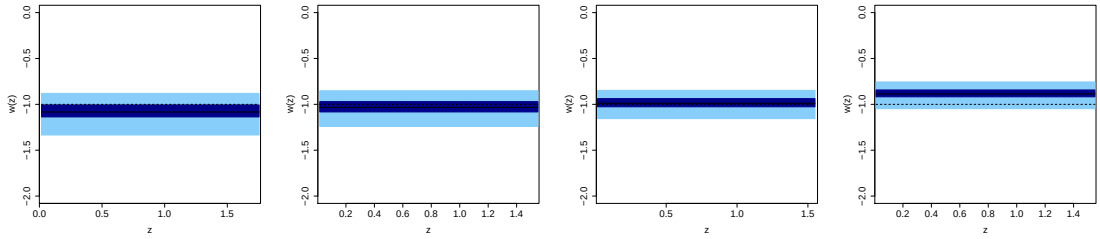
Figure 8: z vs. mu

noted that we are gaining very little information here about Ω_m; what we know prior to running the analysis is fairly similar to the distribution of Ω_m after the analysis. Thus, we can conclude that with this data we are most likely not learning anything about Ω_m and if we could get a better estimate of Ω_m from another data source it may prove to shrink our uncertainty about the EOS curve. We cannot fully comment on H_0^* because its meaning is not fully understood as it is a mixture of H_0 and the marginalization during light curve fitting that results in M.

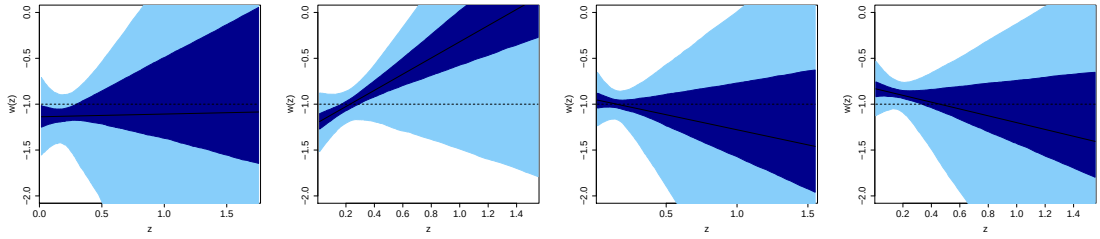
4.3.2 Model 4 - Gaussian Process

The Gaussian process was initially run with an exponential correlation function; the results were mostly flat with some slight curvature as seen in Figure 10. Because of these rather flat results and all of the training sets we had run on simulated data, we felt it safe to use an approximately Gaussian correlation function which yields smoother results. It is approximate only in that we use an exponent of 1.9999 instead of 2 for numerical stability. These results can be seen in Figure 11. Table 9 contains the parameter estimates for both the exponential and Gaussian correlation function simulations.

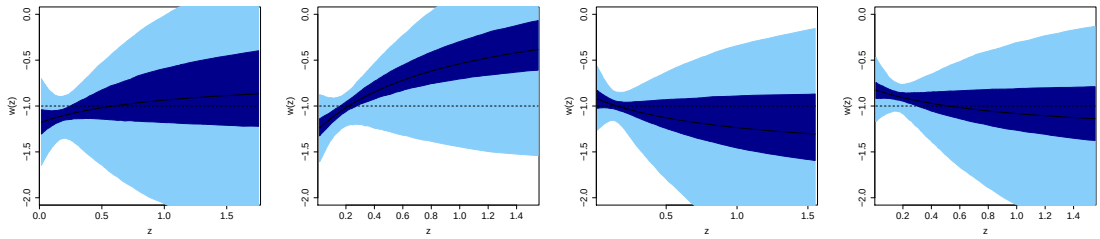
Analysis We noticed first that our non-parametric fit had tighter probability bands than that of some of the parametric models. All of the Gaussian process runs had $w(z) = -1$ within the 95%



(a) Davis (b) Kowalski (c) SALT3 (d) MLCS17



(e) Davis (f) Kowalski (g) SALT3 (h) MLCS17



(i) Davis (j) Kowalski (k) SALT3 (l) MLCS17

Figure 9: Model 1

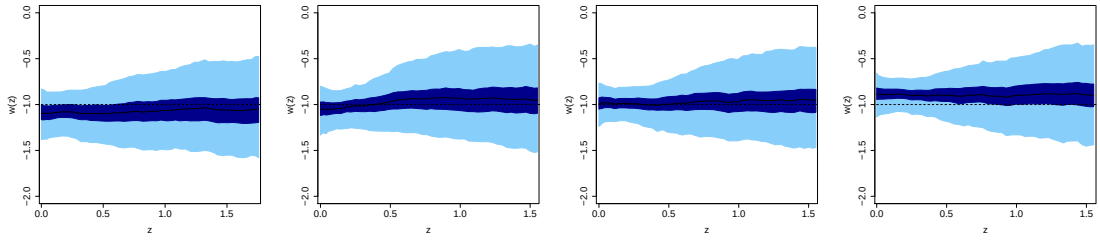
Table 8: Models 1, 2, 3 - 95% PIs

Model	Dataset	a	b	Ω_{m0}	H_0^*	σ^2
Model 1	Davis	(-1.334,-0.881)	N/A	(0.229,0.324)	(64.62, 67.64)	(0.87,1.20)
	Kowalski	(-1.241,-0.852)	N/A	(0.235,0.334)	(69.08, 71.47)	(0.90,1.16)
	SALT3	(-1.156,-0.848)	N/A	(0.233,0.323)	(64.44, 66.10)	(1.04,1.31)
	MLCS17	(-1.048,-0.756)	N/A	(0.226,0.320)	(64.37, 65.76)	(0.95,1.21)
Model 2	Davis	(-1.576,-0.678)	(-0.263,1.889)	(0.228,0.337)	(64.60, 67.99)	(0.88,1.21)
	Kowalski	(-1.545,-0.876)	(-0.549,2.175)	(0.229,0.327)	(69.36, 72.08)	(0.90,1.16)
	SALT3	(-1.251,-0.622)	(-2.460,1.199)	(0.233,0.338)	(64.24, 66.22)	(1.05,1.31)
	MLCS17	(-1.132,-0.516)	(-2.301,0.958)	(0.229,0.333)	(64.20, 65.77)	(0.96,1.22)
Model 3	Davis	(-1.699,-0.666)	(-3.260,2.622)	(0.232,0.328)	(64.54, 68.20)	(0.87,1.21)
	Kowalski	(-1.655,-0.867)	(-3.683,1.024)	(0.231,0.327)	(69.45, 72.28)	(0.90,1.16)
	SALT3	(-1.286,-0.525)	(-1.805,3.699)	(0.232,0.336)	(64.16, 66.19)	(1.04,1.32)
	MLCS17	(-1.184,-0.442)	(-1.707,3.110)	(0.232,0.332)	(64.16, 65.82)	(0.96,1.21)

Table 9: Models 4 - 95% PIs

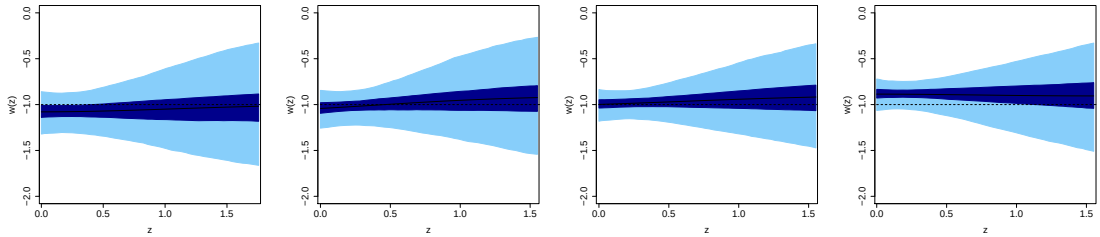
Model	Dataset	Ω_{m0}	H_0	σ^2
Model 4	Davis	(0.232,0.325)	(64.60, 67.63)	(0.87,1.20)
Exponential	Kowalski	(0.231,0.330)	(69.11, 71.58)	(0.89,1.16)
	SALT3	(0.229,0.323)	(64.39, 66.15)	(1.04,1.31)
	MLCS17	(0.226,0.324)	(64.32, 65.83)	(0.96,1.21)
Model 4	Davis	(0.228,0.324)	(64.60, 67.51)	(0.87,1.20)
Gaussian	Kowalski	(0.233,0.328)	(69.10, 71.52)	(0.90,1.16)
	SALT3	(0.231,0.324)	(64.47, 66.16)	(1.04,1.31)
	MLCS17	(0.226,0.323)	(64.35, 65.77)	(0.96,1.21)

probability. The mixing in the algorithm was good and had stable MCMC posteriors. The estimates are very similar in this model to the other parametric ones; we are learning very little about H_0 and Ω_m from the current supernova data. Either more data is needed to reduce uncertainty or more data or more precise data. We were very happy with these results as the non-parameteric flexible fit of $w(z)$ we obtained through our Gaussian process method had smaller probability bands than Model 2 and Model 3. This means not only did we achieve fitting $w(z)$ with a non-parametric form but we also were able to reduce the uncertainity in the estimation of the EOS.



(a) Davis (b) Kowalski (c) SALT3 (d) MLCS17

Figure 10: Model 4- Gaussian Process- Exponential Correlation



(a) Davis (b) Kowalski (c) SALT3 (d) MLCS17

Figure 11: Model 4- Gaussian Process- Gaussian Correlation

5 Future work

5.1 More parameters for the supernovae data

Our first future test is to add more information to the current model by looking at the most recent SALT data. The real data has two more variables that so far we have not used in our analysis: k (the color of the supernovae) and t (the scale of the light curve of the supernovae). Both of these variables are important in the light curve fitting process of the real data and are two of the sources of variability in the fitting process. Up until this point we have not included these in our models but rather have been using μ, a unified response variable that accounts for all types of variability. We now want to break down μ into more components: $\mu_i = m_B - M - \alpha(t-1) - \beta k$. We would like to include the color and scale variables separately in the model as linear terms. It can be useful to split the data into independent variables because the analysis can yield which variable is

introducing the most uncertainty to the model. We have the new model:

$$m_B = \alpha(t-1) + \beta k + 5log_{10}(c(1+z_i)r(z_i)) + 25 + M$$
$$r(z) = \frac{1}{H_0}\int_0^z \left(\Omega_m(1+s)^3 + (1-\Omega_m)(1+s)^3 e^{-3\int_0^s \frac{-w(u)}{1+u}du}\right)^{-1/2} ds$$

Hopefully, this will help observers in their future work toward mitigating some of the measurement error. Whereas, if a factor adds little variability to the model then in reducing its measurement error is not important.

As a note, there is a variable M that appears in the real data it comes from the marginalization in the light curve fitting. As in all the real data both M and H_0 cannot both be estimated. We will set $H_0 = 65$ and then let all of the variability be in the new variable M. Some of the most recent ideas for priors are: $\pi(M) = -19.46 \pm 1.0$, $\pi(\alpha) = -1.34 \pm 0.08$, $\pi(\beta) = 2.59 + 0.12, 2.59 - 0.08$, $\pi(\Omega_m) \sim N(0.27, 0.03^2)$, and $\pi(\sigma^2) \sim IG(10, 9)$.

We plan on fitting $w(z)$ with a Gaussian process with exponential correlation function, as we did with the previous parameterization of this problem. We obtained good results from this non-parametric fit that had smaller probability bands than the Chevallier-Polarski-Linder parametric model. We believe a non-parametric fit of $w(z)$ allows for more flexibility than an ansatz model. We hope to extend our understand of the nature $w(z)$ through this type of analysis.

Another feature we wish to add is estimation for radiation density, Ω_r. Currently, we have been assuming this parameter contributes little to the original transformation and adds instability with the additional parameter. Ω_r can be added in the following way:

$$T(z) = M + \alpha(t-1) + \beta k + 25 + 5log_{10}\frac{c(1+z_i)}{H_0}\int_0^{z_i} H(s)ds \tag{11}$$
$$H(s) = \left(\Omega_r(1+s)^4 + \Omega_m(1+s)^3 + (1-\Omega_m-\Omega_r)(1+s)^3 e^{-3\int_0^s \frac{-w(u)}{1+u}du}\right)^{-1/2} \tag{12}$$

5.2 Experimental Design

Currently, our real data is comprised of several hundred supernova observations. Each of the telescopes search different depths of space for the supernovae and collect their results. To gain a better understanding of the nature of dark energy, through our function $w(z)$, we need more data. We would like to explore which redshift ranges will give us the best information for constraining the cosmological parameter. This would need to account for the any variation in observations as further objects tend to have larger error bars. Ultimately we would like to answer what would be better, more noisy data or less data which is more accurate? These lines of investigation are relevant, as many observers continue to collect this supernova data.

This analysis would help guide the observers to collect the most relevant data that would be most efficient to current research. This type of problem is referred to as experimental design in statistics. The experimental design can give guidelines for where the most useful data can be

collected on the redshift range. Typical experimental design would specify exact redshifts to collect data but the Universe probably will not comply with providing a supernova at that exact location, so instead we will alter the standard approach and give redshift targets where data would be most useful. These targeted values will be suggestions on where to look or what telescope may be more useful.

Currently, we have data from four general telescope classifications which tend to concentrate in particular regions of the redshift range. It is possible for astronomers to use different telescopes to probe different redshift ranges but they need to know where to spend their time and money.

It may be useful, too, if we can give some general idea of how the small, medium, and large redshift values contribute differently to constraining the cosmological parameter. This line of research would include employing a utility function based on our posterior of the $w(z)$ and estimate the expected utility of additional observations at various redshift values. As with all of the previous work on this problem we expect this to be computationally intensive due to the highly non-linear relationship in this problem. Currently, the method that is used to weight different regions of the redshift range is principal component analysis (Huterer and Starkman, 2003). This requires that the $w(z)$ be a piecewise discontinuous function and the supernovae are binned according to redshift to perform this analysis. We hope to move away from those kind of assumptions in our experimental design and allow $w(z)$ to be flexible.

5.3 Different Probes

Currently, we have only examined the supernovae data. This has required much work setting up the non-linear relationship and working with the computational intensive algorithms to produce a non-parametric estimate for the equation of state. There are several other types of probes that have similar non-linear relationships and use many of the same parameters we have already studied with the supernovae data. We hope to configure our current algorithms to include these other data sources and hopefully gain more information about the cosmological parameter and reduce our estimation uncertainty.

The current supernovae relationship is given by equations(13) and it can be seen in this section that it shares much of the same form and parameters as the other probes.

$$\frac{1}{c(1+z_i)} 10^{\frac{\mu_i - 25}{5}} = \frac{1}{H_0} \int_0^z H(s) ds \tag{13}$$

5.3.1 Baryon Acoustic Oscillation Data

We would like to include the baryon acoustic oscillation (BAO) data to our model. This data is not derived from supernovae but rather a different probe. We are hoping to combine this small data set with the supernova data and continue with $w(z)$ with our unique Gaussian process method. This method deals with some of the computational demanding features of the non-linear relationship and allows for a flexible fit to $w(z)$ and does not seem to increase the probability bands when compared to other parametric fits.

Adding the BAO data would require assuming this data is independent of the supernovae data and creating a joint likelihood. We would also need to account for the difference in standard error associated with the two different types of data through weights in the likelihood equation. We assume the BAO data is far less noisy than the supernovae data, even though the data set only contains several points. By using more data that is less noisy we would hope to constrain our non-parametric fit of $w(z)$ further and reduce the probability bands. Also, we would hope to gain more understanding about Ω_m and H_0 through this method. We can use equation14 for $H(s)$ or include Ω_r as in equation(12). Our data A relates to z in the following manner:
$A = \frac{\sqrt{\Omega_m}}{h(z_1)^{1/3}} \left(\frac{1}{z_1} \int_0^{z_1} \frac{H_0}{H(s)} ds \right)^{2/3}$ where $A = 0.469 \left(\frac{n_s}{.98} \right)^{-.35} \pm 0.017$.

There are currently two data values from the BAO probe: $z_1 = 0.35$ where $n_s = 0.958 \pm 0.016$ and there may be a second data point at $z_1 = 0.20$ that has certain issue we may also be able to use. Because there is uncertainty in both A and n_s this will require some type of hierarchical model. This way we can include both types of uncertainty into our analysis. We will also be able to include other parameters of interest like: $\Omega_r h^2 = 2.47x10^{-5}$ where $h = \frac{H_0}{100}$.

5.3.2 Cosmic Microwave Background Data

Another type of cosmological probe that could be included in our study is the cosmic microwave background data (CMB). The physics of this data is best understood compared to other types of data. This probe explores very high redshift, many orders of magnitude higher than any of the previous data types discussed. There is only one data point available for this probe near redshift $z = 2000$. We could include this probe similarly by assuming it independent to the current data sets and incorporating its likelihood. We would like to have the fit for $w(z)$ be non-parametric and flexible. We would hope that the extra data would give us more information on the nature of $w(z)$ in another redshift region. Also, we could possible gain more information on the other parameters of interest are Ω_m and H_0 and the data is $R = 1.713 \pm 0.020$ and $z_2 = 1087.9 \pm 1.2$. The relation between this form of the data is given by:

$$R = \sqrt{\Omega_m} \int_0^{z_2} \frac{H_0}{H(s)} ds$$

$$H(s) = \left(\Omega_m (1+s)^3 + (1-\Omega_m)(1+s)^3 e^{-3\int_0^s \frac{-w(u)}{1+u} du} \right)^{-1/2}$$

5.3.3 Combined Data Sources

Overall, we can obtain output from a high accuracy simulator and use an already constructed emulator that allows for five different cosmological parameters (like Ω_m, H_0, w(z)...). We would like to blend these parameters with our current information obtained from the supernova data. This type of analysis should be more robust as it employs multiple data sources (like supernova, CMB, BAO...). A benefit is that this reverse engineers the estimation process and will allow us to

compare the accuracy of different probes. We hope this will be the general framework of design, optimization, and analysis of future work on the dark energy surveys.

6 Timeline

Timeline 2009-2010

	Sept	Oct	Nov	Dec	Jan	Feb	March	April	May	June
Hermite orthogonal polynomial model										
Add more parameters to the supernovae model										
Add BAO data to the models										
Add CMB data to the models using hierarchical modeling										
Experimental design - where is more data needed on the z axis										
Add other types of data to the model										
Finish writing up dissertation										

References

Astier, P. (2000). Can luminosity distance measurements probe the equation of state of dark energy? *Physics Letters B*, **500**(1-2): 8–15.

Banerjee, S., B.P., C., and E., G. A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall, New York.

Crittenden, R. G. and Pogosian, L. (2005). Investigating dark energy experiments with principal components. *arXiv:astro-ph/0510293v1*.

Davis, T. M., Mortsell, E., Sollerman, J., and et al. (2007). Scrutinizing exotic cosmological models using essence supernova data combined with other cosmological probes. *Astrophysics Journal*, **666**: 716.

de Sitter, W. (1934). On distance, magnitude, and related quantities in an expanding universe. *Bulletin of the Astronomical Institutes of the Netherlands*, **7**: 205.

Dodelson, S. (2003). *Modern Cosmology*. Academic Press.

Filippenko, A. V. (1997). Optical spectra of supernovae. *Annual Review of Astronomy and Astrophysics*, **35**: 309–355.

Gelman, A., Carlin, B., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapmand and Hall, New York.

Genovese, C. R., Freeman, P., Wasserman, L., Nichol, R., and Miller, C. (2009). Inference for the dark energy equation of state using type ia supernova data. *Annals of Applied Statistics*, **3**(1): 144–178.

Guy, J., Astier, P., Nobili, S., Regnault, N., and Pain, R. (2005). Salt: a spectral adapitve light curve template for type ia supernovae. *Astrophysics*.

Huterer, D. and Starkman, G. (2003). Parameterizations of dark-energy properties: a principal-component approach. *Phys. Rev. Lett.*, **90**(031301).

Huterer, D. and Turner, M. S. (1999). Prospects for probing the dark energy via supernova distance measurements. *Phys. Rev. D*, **60**(081301).

Huterer, D. and Turner, M. S. (2001). Probing dark energy: methods and strategies. *Phys. Rev. D*, **64**(123527).

Kowalski, M., Rubin, D., and et al. (2008). Improved cosmological constraints from new, old and combined supernova datasets. *Astrophysics*, **686**: 749–778.

Krauss, L. M., Jones-Smith, K., and Huterer, D. (2007). Dark energy, a cosmological constant, and type ia supernovae. *New Journal of Physics*, **9**: 141.

Leibundgut, B. (2001). Cosmological implications from observations of type ia supernovae. *Annual Review of Astronomy and Astrophysics*, **39**: 67–98.

Leibundgut, B. (2004). Are type ia supernovae standard candles? *Astrophysics and Space Science*, **290**(1-2).

Linder, E. V. (2003). Exploring the expansion history of the universe. *PhysRevLett*, **90**(091301).

Linder, E. V. (2006). Biased cosmology: pivots, parameters and figures of merit. *Astroparticle Physics*, **26**:102–110.

Linder, E. V. (2007). The mirage of w=-1. *Astrophysics*.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., and et al. (1999). Measurements of omega and lambda for 42 high-redshift supernovae. *Astrophysics Journal*, **517**:565–586.

Perlmutter, S. and et al. (1997). Cosmology from type ia supernovae. *Bulletin of the American Astronomical Society*, **29**:1351.

Pskovskii, I. P. (1977). Light curves, color curves, and expansion velocity of type i supernovae as functions of the rate of brightness decline. *Soviet Astronomy*, **21**:675.

Riess, A., Press, W. H., and Kirshner, R. P. (1996a). Is the dust obscuring supernovae in distant galaxies the same as dust in the milky way? *The Astronpysical Journal*, **473**:588–594.

Riess, A., Press, W. H., and Kirshner, R. P. (1996b). A precise distance indicator: type ia supernova multicolor light curve shapes. *The Astronpysical Journal*, **473**:88–109.

Riess, A. G., Filippenko, A. V., and et al. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, **116**:1009–1038.

Riess, A. G., Strolger, L.-G., Casertano, S., Ferguson, H. C., Mobasher, B., Gold, B., and et al. (2007). New hubble space telescope discoveries of type ia supernovae at z $>$ 1: narrowing constraints on the early behavior of dark energy. *The Astrophysics Journal*, **659**:98–121.

Sahni, V. and Starobinsky, A. (2006). Reconstructing dark energy. *Int.J.Mod.Phys.D*, **15**:2105.

Saini, T. D., Raychaudhury, S., Sahni, V., and Starobinsky, A. A. (2000). Reconstructing the cosmic equation of state form supernova distances. *Phys.Rev.Lett*, **85**:1162–1165.

Simpson, F. and Bridle, S. (2006). Redshift sensitivities of dark energy surveys. *Phys.Rev.D*, **73**(083001).

W. M. Wood-Vasey, W. M., Miknaitis, G., Stubbs, C. W., Jha, S., Riess, A. G., Garnavich, P. M., Kirshner, R. P., and Aguilera, C. (2007). Observational constraints on the nature of dark energy: first cosmological resutls from the essence supernova survey. *Astrophysics Journal*, **666**.

Weller, J. and Albrecht, A. (2002). Future supernovae observations as a probe of dark energy. *Phys.Rev.D*, **65**(103512).

Wood-Vasey and et al. (2008). Type ia supernovae are good standard candles in the near infrared: evidence from pairitel. *ApJ*, **689**.