Optimal Inferences for the Dark Energy Equation of State Using Type Ia Supernova Data

Christopher R. Genovese,[1*], Peter Freeman[1], Larry Wasserman,[1]
Robert C. Nichol[2], Christopher Miller[3]

[1]Department of Statistics, Carnegie Mellon University,
5000 Forbes Ave, Pittsburgh, PA 15213, USA

[2]Institute of Cosmology and Gravitation, University of Portsmouth,
Portsmouth, PO1 2EG, UK

[3]CTIO/NOAO
950 North Cherry Avenue, Tucson, AZ 85719, USA

[*]To whom correspondence should be addressed; E-mail: genovese@cmu.edu.

We present a method for making sharp statistical inferences about the dark energy equation of state from observations of Type Ia Supernovae. The method is based on a new formula for the co-moving distance in terms of the equation of state. This work stands in contrast to current inferential methods that involve estimating derivatives of the co-moving distance as a function of redshift; as we discuss, derivative estimation precludes sharp inference. Using our result, we evaluate the strength of statistical evidence for various competing models of dark energy. We find that with the currently available Type Ia SNe data, it is not possible to distinguish statistically among popular dark-energy mod-

els. In particular, there is no support in the data for rejecting a cosmological constant. A sample size increase by a factor of 10 would likely be sufficient to overcome this problem. Such data should become available with NASA's Joint Dark Energy Mission.

Introduction. Observations of Type Ia supernovae (SNe) provide the strongest available evidence that the expansion of the universe is accelerating (*1*, *2*). One explanation for this apparent acceleration is the presence in the universe of dark energy, energy of unknown composition that has negative pressure $P_{\rm DE}$ and a density $\rho_{\rm DE}$ comprising approximately 73% of the critical density in the present epoch. Support for the existence of dark energy also comes from studies of the Cosmic Microwave Background and Large Scale Structure (*3*).

In this paper, we follow current standard practice and assume a homogeneous, isotropic, and spatially flat universe where matter is non-relativistic and where gravity is described by general relativity with the Friedmann-Robertson-Walker metric. (Specifically, we do not examine an alternative to the dark energy hypothesis, modified gravity, for which some of these assumptions do not hold; see, e.g., *4*.) Under these assumptions, we can express the dark energy pressure $P_{\rm DE}$ and density $\rho_{\rm DE}$ in terms of the co-moving distance r as follows:

$$P_{\rm DE}(z) = -\rho_{\rm crit}\left(\frac{1}{H_0 r'(z)}\right)^2\left(1+(1+z)\frac{2r''(z)}{3r'(z)}\right),$$

and

$$\rho_{\rm DE}(z) = \rho_{\rm crit}\left[\left(\frac{1}{H_0 r'(z)}\right)^2 - \Omega_m(1+z)^3\right],$$

where $\rho_{\rm crit}$ is the current critical density, H_0 is the Hubble constant, Ω_m is the fractional contribution of matter to $\rho_{\rm crit}$ (*5*), and $'$ denotes differentiation of r with respect to z. Taking the ratio of these functions yields the so-called "reconstruction" equation of state parameter

$$w(z) = \frac{H_0^2\Omega_m(1+z)^3 + \frac{2}{3}(1+z)r''(z)/(r'(z))^3}{H_0^2\Omega_m(1+z)^3 - 1/(r'(z))^2} - 1. \qquad (EQ::W)$$

Several important cosmological models can be expressed in terms of w, including the following: the cosmological constant ($w(z) \equiv -1$); frustrated cosmic strings or domain walls ($w(z) \equiv -1/3$ or $w(z) \equiv -2/3$; *6*); various quintessence models (freezing $w'(z) > 0$, thawing $w'(z) < 0$; *7*, *8*), and models which allow $w(0) < -1$, such as Cardassian models (*9*) and phantom dark energy (*10*).

Two basic approaches have been used in the literature for distinguishing among competing cosmological models. The first tries to estimate the dimensionless dark energy density $\rho_{\rm DE}(z)/\rho_{\rm crit}$ using polynomial or nonparametric (*23*) models for r and r', typically in the form of piecewise functions where the number of breakpoints acts as a smoothing parameter (*11*). The second approach tries to estimate w using nonparametric models for r, r', and r'' (*12*). The advantage of the former approach is that it requires estimation of only one derivative of r; the advantage of the latter approach is that the resulting estimator can be more easily interpreted in terms of the competing models.

All existing approaches require the estimation of at least one derivative, and derivative estimation is difficult. Because integration is a smoothing operation, it obscures high frequency structure, making detailed inferences about the derivative imprecise. This problem becomes more severe for higher-order derivatives. (See the Supplementary Material for a more detailed explanation.)

The implication for estimating the dark-energy equation of state parameter w using derivatives is the following. If we use parametric models, the estimates are guaranteed to be exquisitely sensitive to the parametric assumptions. If we use nonparametric methods, the estimates are guaranteed to have high variance, high bias, or both.

In this paper, we show how to avoid these difficulties, thus leading to a more precise statistical procedure. Our approach supports both parametric and non-parametric models for the equation of state, enables testing of specific hypotheses, and allows construction of confidence bands for $w(z)$. We use this method to address the following question: are the existing SNe data sufficient to distinguish statistically among currently competing models for $w(z)$ (e.g., to test $w \equiv -1$ versus $w \not\equiv -1$)?

Type Ia SNe Data. We analyze data for 192 SNe Ia from Davis et al. (*14*). The data consist of i) redshifts z, ii) distance moduli $\mu = m - M$, where m and M are the apparent and absolute magnitudes of each supernova, and iii) standard errors $\tau = \sqrt{\tau_\mu^2 + \tau_v^2}$ for 192 SNe Ia, where τ_μ is the intrinsic uncertainty in the distance modulus and τ_v is an estimate of the peculiar velocity of the supernova relative to a local standard of rest. (We ignore the uncertainties of the redshift estimates, which are generally less than 1%.)

Let U_i and z_i denote the observed distance modulus and redshift, respectively, for the ith supernova, $i = 1, \ldots, n$, where $n = 192$. We model U_i as Gaussian with mean $\mu(z_i)$; that is,

$$U_i = \mu(z_i) + \tau_i \epsilon_i, \qquad (EQ :: modulusdata)$$

where the ϵ_is are assumed independent, mean zero, Gaussian noise terms with unit variance and the τ_is are the given standard errors of the distance moduli measurements. We expresss these in terms of co-moving distance (assuming a flat universe) by transforming as follows:

$$\frac{1}{c(1+z_i)} 10^{(U_i - 25)/5} = r(z_i) \cdot 10^{(\tau_i/5)\epsilon_i}, \qquad (EQ :: transform)$$

where c is the speed of light. Thus, letting Y_i denote the $\log_{10}$ of the left hand side of (EQ::transform), we have

$$Y_i = \log_{10} r(z_i) + \sigma_i \epsilon_i, \qquad i = 1, \ldots, n, \qquad (EQ :: DATA)$$

where $\sigma_i = \tau_i/5$. Figure FIG::DATA shows Y plotted against $\log_{10}(z)$ with associated error bars. We thus call r "observable" because it can be directly estimated from the observed data.

Deriving Co-Moving Distance from the Equation of State

In this section, we re-express the relationship between co-moving distance r and the equation of state w to support building more precise statistical models. Specifically, we give explicit analytic expressions for the co-moving distance r (and the deceleration function $q(z) = -1-(1+z)r''(z)/r'(z)$) under various representations for w. With (EQ::DATA), this gives a likelihood directly in w; we show how to use this for parametric or nonparametric inferences in the next section.

Equation (EQ::W) produces a differential equation for r in terms of w that has a two-parameter family of solutions. The boundary conditions $r(0) = 0$ and $r'(0) = 1/H_0$ produce a unique solution

$$r(z) = H_0^{-1} \int_0^z ds \left[\Omega_m(1+s)^3 \,+\, (1-\Omega_m)(1+s)^3\, e^{-3\int_0^s \frac{-w(u)}{1+u}\,du}\right]^{-\frac{1}{2}}. \qquad (EQ::R)$$

From this expression, we can also solve for q:

$$q(z) = \frac{1}{2} + \frac{3}{2}\frac{(1-\Omega_m)w(z)}{(1-\Omega_m)+\Omega_m e^{3\int_0^z \frac{-w(s)}{1+s}\,ds}} \qquad (EQ::Q)$$

(A related expressions for the decleration parameter r in terms of q is given in equations EQ::R::Q in the Supplementary Material.) The main use of these expressions is to translate a model for the unobservable w into the observable r.

Equations (EQ::R) and (EQ::Q) have several properties that are valuable for statistical inference as we will show below. First, note that for any w, $r'(z) > 0$, so r is a monotone increasing function of z with $r(0) = 0$. In fact, $(1+z)^{-3/2}/H_0 \le r'(z) \le (1+z)^{-3/2}/\sqrt{H_0^2\Omega_m}$. Second, r is monotone decreasing in w for each fixed value of H_0 and Ω_m. Specifically, if w_1 and w_2 are two candidate equations of state with corresponding co-moving distance fucntions r_1 and r_2, and if $w_2(z) \ge w_1(z)$ for all $z \ge 0$, then $r_2(z) \le r_1(z)$ for all $z \ge 0$. Similarly, q is monotone increasing in w; $w_2(z) \ge w_1(z)$ for all $z \ge 0$ implies that $q_2(z) \ge q_1(z)$ for all

$z \geq 0$. Third, as shown in the Supplementary material, for r to be concave it is sufficient that $w(z) \geq -1/(1-\Omega_m)$ for all $z \geq 0$. Under mild smoothness assumptions on w, the concavity of r holds more broadly. Fourth, in both equations, taking $w \equiv 0$ is equivalent to taking $\Omega_m = 1$.

For any specific choice of w, Ω_m, and H_0, it is straightforward to evaluate r numerically. ATTN specific parameterizations For instance, under a constant w model $w \equiv w_0$, equation (EQ::R) reduces to

$$r(z) = H_0^{-1} \int_0^z ds \, \left[\Omega_m (1+s)^3 + (1-\Omega_m)(1+s)^{3(1+w_0)}\right]^{-\frac{1}{2}}. \qquad (EQ :: RWmone)$$

More generally, expanding $w(z) = -\sum_j \beta_j \psi_j(z)$ in a (not-necessarily orthonormal) basis $\psi_0, \psi_1, \ldots$ yields

$$r(z) = H_0^{-1} \int_0^z ds \, \left[\Omega_m (1+s)^3 \, + \, (1-\Omega_m)(1+s)^3 \, e^{-3 \sum_j \beta_j \tilde{\psi}_j(s)}\right]^{-\frac{1}{2}}, \quad (EQ :: R :: Basis)$$

where $\tilde{\psi}_j(s) = \int_0^s \psi_j(u)/(1+u) du$. Taking the expansion to be finite gives three important special cases:

1. Polynomial in z: $\psi_j(z) = z^j$, $j = 0, \ldots, d$, giving

$$r(z) = H_0^{-1} \int_0^z ds \, \left[\Omega_m (1+s)^3 + (1-\Omega_m)(1+s)^{3(1-\alpha_0)} e^{-3 \sum_{j=1}^d (-1)^j \alpha_j \, s^j / j}\right]^{-\frac{1}{2}}. \qquad (EQ :: R :: POLY)$$

   where $\alpha_k = \sum_{j=k}^d (-1)^j \beta_j$ for $k = 0, \ldots, d$.

2. Polynomial in the scale factor a: $\psi_j(z) = (1+z)^{-j}$, $j = 0, \ldots, d$, giving

$$r(z) = H_0^{-1} \int_0^z ds \, \left[\Omega_m (1+s)^3 + (1-\Omega_m)(1+s)^{3(1-\beta_0)} e^{3 \sum_{j=1}^d \beta_j ((1+z)^{-j} - 1)/j}\right]^{-\frac{1}{2}}. \qquad (EQ :: R :: APOLY)$$

3. Piecewise constant: $\psi_j(z) = 1_{(s_j, s_{j+1}]}(z)$ for $j = 0, \ldots, K-1$, where $0 = s_0 < s_1 < \cdots < s_K$ are breakpoints for K fixed bins and where $1_{(s_j, s_{j+1}]}(z)$ is 1 if $s_j < z \leq s_{j+1}$

and 0 otherwise. In this case, equation (EQ::R) becomes

$$r(z) = \int_0^z ds \left[H_0^2 \Omega_m (1+s)^3 \left(1 - e^{-3\sum_{j=1}^{J(s)} \beta_j \log\left(\frac{1+s_j}{1+s_{j-1}}\right)} e^{-3\beta_{J(s)+1} \log\left(\frac{1+s}{1+s_{J(s)}}\right)} \right) + \frac{1}{\theta^2} \right]^{-\frac{1}{2}}, \qquad (EQ :: R :: PC)$$

where $J(s) = \max\{0 \le j \le K : \; s_j \le s\}$. Despite the discontinuities in w, this expression is a smooth function of the β parameters.

Extension to other bases – such as B-splines, orthogonal polynomials, and wavelets – is straightforward.

Combined with equation (EQ::DATA), each of these expressions produces a likelihood for w, H_0, and Ω_m. Although nonlinear, these likelihoods are well-behaved for optimiztion purposes, and weighted, nonlinear least-squares is computationally efficient in practice. Good estimates of the coefficients can be obtained for a wide variety of models, which in turn supports both parametric and nonparametric inferences about w.

Optimal Statistical Procedures

We distinguish two uses of the word *model* in this paper. A cosmological model for dark energy is a set of assumptions about the underlying physics that gives rise to a particular form of the equation of state. A statistical model for w (or q) is a family of probability distributions for the data indexed (at least) by a parameterization of w (or q). If this parameterization is finite dimensional, the model is *parametric*; otherwise, the model is *nonparametric* (*23*). We consider statistical models whose stochastic component is specified by equation (EQ::DATA); each such model is then determined by the parameters H_0 and Ω_m and by a specific representation of w.

A particular cosmological model can be analyzed under a particular statistical model, but the scope of the inferences is limited by the viability and flexibility of the assumptions made.

Hypothesis Testing. A common method for distinguishing among models of dark energy is to first estimate the equation of state and use this estimate to test hypotheses about cosmological models. This approach has several disadvantages, including that the power of the test depends on having a good estimator and that it requires accurate standard errors for the entire function. Equation (EQ::R), however, allows a broad range of tests to be performed directly, without the unnecessary baggage of a preliminary estimator.

Here we consider null hypotheses of the following forms

A. Simple equalities for w: $w = w_0$,

B. Inequalities for w: $w_0 \leq w \leq w_1$,

C. Inequalities for w': $w'_0 \leq w' \leq w'_1$,

D. Inclusion: $w \in V$ for a linear space V, and

and various intersections of these, where w_0, w_1, w'_0, and w'_1 denote various fixed functions, not necessarily constant. (We use the inequality $w \leq w_0$ to mean that $w(z) \leq w_0(z)$ for all z, and similarly for other inequalities between functions.)

Testing such hypotheses gives direct tests of various cosmological models. The null hypothesis that the cosmological constant model holds, for example, translates to a simple null hypothesis with $w_0 = -1$. Quintessence solutions lead to a variety of constraints on w and w' that can be tested by combining hypotheses that are inequalities for w and for w'. For instance, thawing solutions satisfy

$$1 + w \leq \frac{dw}{d \ln a} \leq 3(1 + w), \qquad (EQ :: THAW :: CON :: lna)$$

and freezing solutions satisfy

$$3w(1 + w) \leq \frac{dw}{d \ln a} \leq 0.2w(1 + w), \qquad (EQ :: FREEZE :: CON :: lna)$$

when $-1 \leq w \leq -0.8$ (*7*), where a is the scale factor. As we show in the Supplementary Material, these bounds can be re-expressed for w in the same range as

$$\frac{1+w(0)}{(1+z)^3} - 1 \leq w(z) \leq \frac{1+w(0)}{1+z} - 1, \qquad (EQ::THAW::CON)$$

and

$$\frac{w(0)}{(1+z)^3 + w(0)\left((1+z)^3 - 1\right)} \leq w(z) \leq \frac{w(0)}{(1+z)^{0.2} + w(0)\left((1+z)^{0.2} - 1\right)}, \qquad (EQ::FREEZE::CON)$$

where $w(0)$ is a free parameter.

The strategy underlying our testing procedure is to use equation (EQ::R) to translate hypotheses about w into hypotheses about r. Thus, we can test any hypothesis that translates into a manageable form. The procedure is as follows.

0. Select a small $0 < \alpha < 1$.

1. Construct a $1 - \alpha$ confidence set $\mathcal{C}$ for the unknown vector $(r(z_1), \ldots, r(z_n))$.

2. Construct the set R_0 of vectors $(r_0(z_1), \ldots, r_0(z_n))$ where r_0 is a co-moving distance function produced by an equation of state consistent with the null hypothesis

3. Reject the null hypothesis if $\mathcal{C} \cap R_0 = \emptyset$.

In practice, the sets in Steps 1 and 2 need not be constructed explicitly, and the procedure can be made computationally efficient for a broad range of hypotheses. See the Supplementary Material.

We use our procedure to define a test that is independent of a particular parameterization of w. We do take advantage of the prior information in equation (EQ::R) described in the previous section (see equation EQ::R and equation 12 in the Supplementary Material). One way to define the confidence set $\mathcal{C}$ is the set of vectors for which a standard chi-squared goodness-of-fit test (*16*) does not reject the null hypothesis. In light of equation (EQ::DATA), the chi-squared goodness-of-fit ball gives a confidence set for $(\log_{10}(r(z_1)), \ldots, \log_{10}(r(z_1)))$, which is easily transformed into a confidence set for $(r(z_1)), \ldots, r(z_1))$. As the number of data grows, however, the chi-squared confidence sets become unduly conservative, reducing the power of the test. So, we also use alternative confidence set procedures that achieves optimal large-sample performance (*17*). The resulting confidence sets are smaller giving the test higher power. Further details are given in the Supplementary Material.

Our procedure works also under more restrictive assumptions about the form of w, with correspondingly sharper results as the assumptions grow stronger. For this, the confidence set $\mathcal{C}$ is constructed using the assumed parameterization. The resulting test will have higher power than the nonparametric test when the assumed parameterization holds. Note, however, that the validity of *any* inferences under a specific parameterization depends strongly on the parameterization being accurate.

Step 2 of the procedure depends specifically on the hypothesis being tested. We now derive the sets R_0 for null hypotheses of the forms listed above. Let $\mathcal{M}$ denote the set of vectors $(r(z_1), \ldots, r(z_n))$ for functions r that meet the a priori conditions that the co-moving distance must satisfy.

A. Under a simple null hypothesis $w = w_0$ equation (EQ::R) generates a two-parameter family of functions r_0 as H_0 and Ω_m vary; R_0 is the set of vectors $(r_0(z_1), \ldots, r_0(z_n))$ for r_0 in this family.

B. Under the null hypothesis, $w \geq w_0$, equation (EQ::R) shows that, for fixed H_0 and Ω_m, $r \leq r_0$, where r_0 is produced in (EQ::R) by $w = w_0$ for the given value of H_0 and Ω_m. Again, varying H_0 and Ω_m produces a two-parameter family of functions r_0. R_0 is the set of vectors $(r_1, \ldots, r_n) \in \mathcal{M}$ such that $r_1 \geq r_0(z_1)$, …, $r_n \geq r_0(z_n)$ for some r_0 in the family. The restriction to $\mathcal{M}$ sharpens the results. It is not strictly necessary, but because equation (EQ::R) produces functions in $\mathcal{M}$, it is an improvement that is virtually cost free. The other direction of inequality is handled similarly, using the monotonicity of r in w.

C. Null hypotheses of the form $w' \geq w'_0$ can be handled by re-expressing the exponent in equation (EQ::R). Integrating by parts and writing $w(s) = w(0) + \int_0^s w'(u)\, du$ yields that

$$\int_0^s \frac{w(u)}{1+u}\, du = w(0) \log(1+s) + \int_0^s w'(u) \left(\log(1+s) - \log(1+u)\right)\, du.$$

The second term in the right-hand side integrand is non-negative, so $w' \geq w'_0$ implies, for fixed H_0, Ω_m, and $w(0)$, that $r \leq r_0$, where r_0 is the right-hand side of equation (EQ::R) corresponding to $(w(0), w'_0, H_0, \Omega_m)$. Varying H_0, Ω_m, and $w(0)$ produces a three-parameter family of functions, and as before, R_0 is the set of vectors in $\mathcal{M}$ whose compontents are at least as big everywhere as some function in this family. Other inequalities in w' are handled similarly.

D. The null hypothesis that w lies in some linear space of functions V is useful primarily to test the goodness of fit of statistical models for w. We select an arbitrary basis for V and form a $\dim(V) + 1$ dimensional family of functions corresponding to each (H_0, Ω_m) and each vector of coefficients in the basis expansion. R_0 is the set of vectors produced by these functions evaluated at $z_1, \ldots, z_n$. See equation (EQ::R::Basis). This case is handled in practice by numerical optimization and thus works best for low to moderate dimensional spaces. It is not necessary to restrict to a linear space, but that is the best behaved case numerically.

Note that this same approach can be used to test hypotheses about the deceleration function

$q(z) = -\left(1 + (1+z)r''(z)/r'(z)\right)$. In particular, the null hypothesis that the universe is decelerating because of an absence of dark energy can be expressed as $w = 0$ or equivalently, $\Omega_m = 1$ or $q = 1/2$. We call this the weakly non-accelerating hypothesis (no dark energy). In contrast, a non-accelerating universe corresponds to $q \leq 0$; we call this the strongly non-accelerating hypothesis. Using the relationship derived in the Supplementary Material (see equation EQ::R::Q), the weak hypothesis corresponds to the null hypothesis $r = H_0^{-1}(2 - 2(1+z)^{-1/2})$ and the strong hypothesis corresponds to the null hypothesis $r \geq H_0^{-1} \log(1+z)$. Both generate a one-parameter family and corresponding R_0. One sided inequality hypotheses of this form ($r \geq r_0$) require an extra step because by taking H_0 large enough, we can make the lower bound as small as possible and the null trivially true. In this case, we use a priori information about H_0 with a suitable adjustment to the confidence level.

As described, the testing procedure given above is independent of how we parameterize w, which gives a flexible and powerful technique. The resulting inferences are essentially as good as possible without stronger assumptions about the forms of w or r. We turn next to consider inferences for w under more restrictive parameterizations.

Model Fitting and Selection. Two basic methods are commonly used to estimate w from supernova data. In the first, the data are smoothed to estimate r' and r'', and the corresponding estimates of these functions are plugged into equation (EQ::W) to obtain an estimate of w. In the second, w is assumed to be in a parametric family $w(z;\psi)$ and the previous estimator, $\hat{w}$, is computed. Then an estimator of ψ, $\hat{\psi}$, is chosen to minimize the distance (e.g., chi-squared) between $\hat{w}(z)$ and $w(z;\psi)$.

We propose a third method. Given a model for w, use equation (EQ::R) to convert it into an explicit model for r and then fit that to the data. For example, if w is a polynomial with unknown coefficients β, then equation (EQ::R::POLY) shows an explicit nonlinear model for r.

A similar expression can be derived for any basis expansion (see Supplementary Material). We fit a weighted least-squares nonlinear regression to r, which produces estimates and standard errors for the coefficients in the model, and in turn for w and q.

The latter method has the advantage of not requiring intermediate estimates of the derivatives. And indeed, in terms of statistical accuracy, the latter method is guaranteed to perform better than the other two. This is true because smoothing to get the derivatives introduces unnecessary additional variability into the procedure.

We consider three families of nested statistical models for w.

I. Many investigators have studied the dark energy equation of state using a linear model $w(z) = -(\beta_0 + \beta_1 z)$ (*19*). This family contains the constant models (e.g., cosmological constant has $\beta_0 = 1$) but also allows inferences about more complicated cosmological models listed above. It can be extended to a nested family of polynomial models $w(z) = -\sum_{d=0}^{D} \beta_d P_d(z)$ for $D \geq 0$, where $P_0, \ldots, P_D$ are linearly independent polynomials of degree $0, \ldots, D$ respectively.

II. We also consider models in which w is a polynomial in the scale factor a. These take the form $w(z) = -\sum_{d=0}^{D} \beta_d (1+z)^{-d}$, for $D \geq 0$.

III. It is also common to model w as in equation (EQ::W::PC) with a piecewise constant function (*11*). For large enough K and suitably spread breakpoints, such functions approximate any square integrable function on an interval arbitrarily well. If B_k is the set of breakpoints for the kth order piecewise constant model, then taking $B_1 \subset B_2 \subset \cdots$ yields a nested family of models for w as K increases.

All three families contain the constant w model, corresponding to $D = 0$ or $K = 0$ respectively.

If $r(z; \beta, \theta)$ denotes the co-moving distance predicted by a model with parameters β and θ, then following equation (EQ::DATA), we fit the Y_is to $\log_{10} r(z_i; \beta, \theta)$'s by non-linear least squares. We choose the breakpoints in the piecewise constant model to balance the total weight $\sum 1/\sigma_i^2$ within each bin, although it makes little difference if one instead balances sample size.

Given a family of parameterizations for w, the next question becomes which to use. Models with too many parameters provide a better apparent fit but estimates of the parameters have high variance; models with too few parameters give estimates with low variance but with bias from the misspecified model. With a two nested parameterizations for w, such as a constant model and the linear model, a likelihood ratio test can be used to select between them. For more general inferences, it is useful to select a model from a larger collection. Here the goal is to select a model that best balances bias and variance. Many good methods are available; we use BIC (Bayesian Information Criterion, also called the Schwarz criterion *24*) as it is both simple and effective.

Confidence Bands. For much the same reasons as just discussed, confidence bands for w (and q) based on smoothed estimates of r' and r'' are necessarily wide (*20*). Better confidence bands can be obtained by applying the bootstrap (*21*) to the fitted models. We generate confidence intervals for the parameters in the model by resampling residuals from the model fit, renormalized to have appropriate variance. The Supplementary Material gives further details and shows how to construct confidence bands for w and q under a particular model. It is common practice to use the confidence bands corresponding to the model selected by BIC. These are straightforward and accurate when the selected model holds. But they necessarily optimistic because the bands do not account for the variation in the model selection process or for the potential bias induced by choosing too simple a model. Nonparametric bands for w and q can be constructed using the methods in this paper but requires somewhat intensive nonlinear optimization. The width of these bands depends strongly on the level of smoothness one assumes, but there is little a priori guidance in this choice. As a compromise, we use the confidence bands from an undersmoothed model (moderate order B-spline expansion).

Results

Testing Cosmological Models. We test seven cosmological models using the procedure described earlier, independently of any parameterization for w. Three models (cosmological constant, frustrated cosmic strings, and domain walls), with a simple null hypothesis of the form $w = w_0$. The two quintessence models (thawing and freezing solutions respectively) were tested with inequality null hypotheses given by equations (EQ::THAW::CON) and (EQ::FREEZE::CON) intersected with the condition that $-1 \leq w \leq -0.8$. (We also tested more expansive versions of these hypotheses using i. equations (EQ::THAW::CON) and (EQ::FREEZE::CON) alone and ii. the hypothesis $w' \geq 0$ and $w' \leq 0$ intersected with the condition that $-1 \leq w \leq -0.8$. But being strict supersets of the original null hypotheses these are less likely to reject.) We tested both the weakly (no dark energy) and strongly non-accelerating universe hypotheses. For the latter, $q \leq 0$, we used an apriori confidence interval for H_0 $72 \pm 8\mathrm{km/s/Mpc}$ *(ATTN ref)*. Finally, we tested the inclusion hypothesis that w is a constant, possibly different from -1. Table 1 shows the results of these tests at various significance levels. The no dark energy model is clearly inconsistent with the data (p-value $p \approx 0$), but none of the other models is rejected at the 13% level. Note in particular that the cosmological constant is consistent with the data.

A false null hypotheses might fail to be rejected because the power of the test is too low. Because our procedure has essentially as much power as possible given the available information about w, the only ways to improve power are either to make stronger assumptions about the form of w or to get more data. We argue that the latter is necessary. Table 2 shows the results of the same tests assuming a linear form $w(z) = -(\beta_0 + \beta_1 z)$, which is the simplest nontrivial parameterization and a correspondingly smaller confidence set. The pattern of rejections is basically the same. and in particular, there is insufficient evidence to move away from a cosmological constant. Of course, there is no reason to believe the linear form for w, and if it is false, inferences under that assumption can be misleading. But this shows that strengthening the assumptions is not enough to overcome the lack of information in the data.

Model	Rejected At Level			
	32%	13%	5%	1%
Cosmological Constant	**yes**	no	no	no
Frustrated Cosmic Strings	**yes**	**yes**	**yes**	no
Domain Walls	**yes**	no	no	no
Nonaccelerating	**yes**	no	no	no
Quintessence Thawing	no	no	no	no
Quintessence Freezing	no	no	no	no
Constant w	no	no	no	no

Table 1. Results of nonparametric hypothesis tests for various cosmological models of w. The significance levels correspond to 1, 1.5, 2, and 2.5 standard deviations respectively from a Gaussian mean.

Model	Rejected At Level			
	32%	13%	5%	1%
Cosmological Constant	**yes**	no	no	no
Frustrated Cosmic Strings	**yes**	**yes**	**yes**	no
Domain Walls	**yes**	no	no	no
Nonaccelerating	**yes**	no	no	no
Quintessence Thawing	no	no	no	no
Quintessence Freezing	no	no	no	no
Constant w	no	no	no	no

Table 2. Results of hypothesis tests for various cosmological models of w assuming a linear parameterization for w in z.

Fitted Models for the Equation of State. In all three families of models we consider (I,II, and III above), likelihood ratio tests between the constant model and the higher-order models in the family fails to reject with pvalue $p > 0.85$. In all three cases, BIC is monotone increasing with the constant model the clear choice. A likelihood ratio test of the cosmological constant versus the constant w model fails to reject with pvalue $p = 0.137$. Figure FIG::LLSURF shows the log-likelihood surface for the constant w model fit, with the bootstrap confidence interval for β_0 overlayed. Although the cosmological constant cannot be rejected at the 87% level, it is suggestively on the boundary or rejection.

Figure FIG::W::FITS shows the fitted w for the constant, linear, piecewise constant $K = 2$, and linear in a models along with their associated *in model* confidence bands. Already these confidence bands are wide enough that we cannot rule out many of the competing cosmological models, and these are still potentially optimistic. FIG::W::R overlays on the data the co-moving distances derived from these fits; the differences between the fits are small relative to measurement error. FIG::W::COMP shows the best fit W with compromise bands from the undersmoothed model, which are wider still.

ATTN FIGURE

ATTN FIGURE

Possible Outliers. We note that two data points – supernovae d083 and 04D3gt – appear suspect, although they cannot be conclusively eliminated. These two points consistently show large residuals (< -4 in standard units), while the remaining residuals conform nicely to what would be expected under the statistical model in equation (EQ::DATA). These two points also appear to be influential on the fit. When they are dropped, constant w models have approximately the same fit, but higher order model fits change substantially, with a significant increase in uncertainty for moderate to large z.

In Table 9 of Wood-Vasey et al. (*14*), supernova d083 has a chi-squared per degree of freedom of 0.25, which is far in the lower tail of the distribution (left tail area $< 3 \times 10^{-4}$). This might suggest that the error bars are too large. The distinguishing feature of supernova 04D3gt is that it has the highest value of A_V. Most of the extinction values are in the $0.1 - -0.2$ range, but $A_V = 1.127$ for this supernova. In addition, there is a discrepancy in the reported distance modulus between the Reiss et al. silver sample, where $\mu = 42.22$, and the Wood-Vasey value of $\mu = 41.35$. The data points for both these supernova are, at the very least, suspect.

The larger implication of these potential outliers for statistical inference about dark energy is that the data being fit are subject to potential selection bias. If the selected subset of observations that "pass" to the inferential stage can have a significant effect on the fit, then that initial screening and selection process should be incorporated into the statistical analysis. To treat the screening and inference as separate stages fails to account for uncertainty in the screening process and biases induced by the selection criteria.

The Need for More Data. A key question is whether current data are sufficient to resolve the differences among interesting models for w and q. We argue here that the answer is no.

First, even under strong assumptions and with essentially optimal procedures, there is not enough evidence to distinguish among interesting models. The cosmological constant model is suggestively on the boundary at the 13% level, but no conclusive differences are supported by the data.

Second, we can use as a minimal criterion for resolvability the power of the likelihood ratio test for distinguishing the cosmological constant model from a constant w model. With existing standard errors, it is straightforward to compute this power by simulation through equations (EQ::R) and (EQ::DATA). Table 3 shows the power of this test for various significance levels and alternatives, all of which are low. Table 4 shows the power for distinguishing a constant w model from a piecewise constant model with one breakpoint for a similar variety of alternatives, which are substantially lower.

Third, an even more striking demonstration of model degeneracy is given by Figure FIG::DEGENERACY. This shows two very different equations of state that give virtually indistinguishable fits to the data, with a chi-squared deviation of 0.04. This example is driven primarily by uncertainty in Ω_m, which can be reduced using other (non-SNe) data. ATTN

ATTN TABLES 3 and 4 HERE

An important question follows: what new data will make it possible to distinguish among interesting cosmological models or to establish conclusively that the cosmological constant model is the best choice. Figure INFO::PERT plots the increase in Fisher information for β_1 as a function of z for both a linear and piecewise constant ($K = 2$) model, with the measurement variance normalized out. Unsurprisingly, the greatest gains for resolving higher order structure are obtained with new data at high redshift (and low noise level).

An alternative way to characterize what is needed is to find the factor by which measurement errors must be reduced before the power to detect some higher order structure is above a specified threshold. As with Tables 3 and 4 above, we compute the power to detect specific differences in w with a likelihood ratio test comparing a constant w model to a piecewise constant model with 2 breakpoints. We compute the power as a function of the standard error scaling and convert that into a required sample size. Table 5 gives the results. We find that a factor of ATTN reduction in the standard errors of the original measurements is sufficient to achieve power 0.9 to detect a difference of 0.25 between the two constant pieces at moderate redshift. This translates to an increase in sample size of roughly a factor of 10. Such an increase should become available with the launch of NASA's Joint Dark Energy Mission (JDEM, *22*). With that amount of data, uncertainties are reduced enough that nontrivial inferences for w can be made.

ATTN TABLE 5 HERE

ATTN FIGURE

ATTN FIGURE

Conclusion

This paper offers both good news and bad news about dark energy. The good news is that we have shown how to estimate the dark energy equation of state from supernovae data precisely and without the need to estimate derivatives. This allows us to directly test various cosmological models without imposing strong assumptions about the form of w. It gives us a statistically efficient method for estimating w and q. And it offers several methods for constructing rigorous confidence bands for these functions. Moreover, our method places assumptions where they are needed – namely, directly on w rather than on r, and there is no need for an initial estimate of derivatives of r. These procedures are essentially as good as possible given the available information about the form of w. Our approach supports both parametric and nonparametric models for the equation of state, making it possible to incorporate prior information. Using this approach, we were able to simply test a variety of competing models and firmly reject at least one model.

The bad news, however, is that current supernova data do not yet provide tight enough inferences to make nontrivial claims about w. We have shown that even with strong additional assumptions the uncertainties are too large to reliably distinguish among most competing cosmological models. Although the results are suggestive, no conclusive differences from the cosmological constant model are supported by the data. Models with higher order structure for w exhibit large standard errors for the model parameters, and in more complicated models, these variances would only grow.

On balance, the good news outweighs the bad. A sample size increase by roughly a factor of ATTN would rule would likely be sufficient to distinguish among interesting hypotheses about the equation of state. The upcoming JDEM mission and the proliferation of automated surveys makes this a likely prospect in the near future, and when the data become available, the methods in this paper give the tools to produce sharp inferences.

References and Notes

1. A. G. Reiss, et al., *Astron. J.* **116**, 1009 (1998).

2. S. Perlmutter, et al., *Nature* **391**, 51 (1998).

3. e.g., R. Scranton, et al., preprint (astro-ph/0307335); P. Fosalba, et al., *Astrophys. J.* **597**, L89 (2003); S. Boughn, R. Crittenden, *Nature* **427**, 45 (2004); M. R. Nolta, et al., *Astrophys. J.* **608**, 10 (2004); S. W. Allen, et al., *Mon. Not. R. Astron. Soc.* **353**, 457 (2004).

4. D. Huterer, E. Linder, *Phys. Rev. D* **75**, 3519 (2007).

5. D. Huterer, M. S. Turner, *Phys. Rev. D* **60**, 081301 (1999).

6. M. Bucher, D. Spergel, *Phys. Rev. D* **60**, 043505 (1999).

7. R. R. Caldwell, E. V. Linder, *Phys. Rev. Lett.* **95**, 141301 (2005).

8. I. Zlatev, L. Wang, P.J., Steinhardt *Phys. Rev. Let.* **82**, 896 (1999); P. J. Steinhardt, L. Wang, I. Zlatev, *Phys. Rev. D* **59**, 123504 (1999).

9. K. Freese, M. Lewis, *Phys. Let. B* **540**, 1 (2002). P. Gondolo, K. Freese, *Phys. Rev. D* **68**, 063509 (2003).

10. R. R. Caldwell, M. Kamionkowski, N. N. Weinberg, *Phys. Rev. Lett.* **91**, 071301 (2003).

11. e.g., Y. Wang, G. Lovelace, *Astrophys. J.* **562**, 115 (2001); M. Tegmark, *Phys. Rev. D* **66**, 103507 (2002); D. Huterer, G. Starkman, *Phys. Rev. Lett.* **90**, 031301 (2003); T. D. Saini, *Mon. Not. R. Aston. Soc.* **344**, 129 (2003); Y. Wang, P. Mukherjee, *Astrophys. J.* **606**, 654 (2004); D. Huterer, A. Cooray, *Phys. Rev. D* **71**, 023506 (2005); Y. Wang, M. Tegmark, *Phys. Rev. D* **71**, 103513 (2005); C. Shapiro, M. Turner, *Astrophys. J.* **649**, 563 (2006); C. Stephan-Otto, *Phys. Rev. D* **74**, 023507 (2006); Y. Wang, K. Freese, *Phys. Lett. B* **632**, 449 (2006); L. M. Krauss, K. Jones-Smith, D. Huterer, preprint (astro-ph/0701692).

12. e.g., R. A. Daly, S. G. Djorgovski, *Astrophys. J.* **597**, 9 (2003); R. A. Daly, S. G. Djorgovski, *Astrophys. J.* **612**, 652 (2004); R. A. Daly, S. G. Djorgovski, preprint (astro-ph/0512576); R. A. Daly, S. G. Djorgovski, preprint (astro-ph/0609791); A. Shafieloo, et al., *Mon. Not. R. Astron. Soc.* **366**, 1081 (2006).

13. D.L. Donoho, *Appl. Comput. Harmon. Anal.* **2**, 101 (1995).

14. **ATTN need cites** Davis et al. (astro-ph/0701510). These data combine data from Wood-Vasey et al. 2007 (astro-ph/0701041), [Table 9 passed] and Riess et al. 2007 (astro-ph/0611572), [Table 6 gold].

15. P. Astier, et al., *Astron. Astrophys.* **447**, 31 (2006).

16. E.L. Lehmann and J.P. Romano, *Testing Statistical Hypotheses*, 3rd edition. Springer, NY, 590 (2006).

17. Y. Baraud, *Annals of Statistics*, **32**, 528 (2004). Davies, Kovac, Meise (2007) Confidence Regions, Nonparametric Regression. Technical Report, University of Duisburg-Essen. Y. Ingster and I.A. Suslina, *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, Springer, NY (2006).

18. For current cosmological parameter estimates, see http://lambda.gsfc.nasa.gov/product/map/current/; we adopt the value of $H_o^2\Omega_M$ derived from "all" data, for the LCDM model.

19. e.g., D. Huterer, M. S. Turner, *Phys. Rev. D* **64**, 123527 (2001).

20. C.R. Genovese and L. Wasserman, *Ann. Stat.* to appear (2007).

21. B., Efron, *Ann. Stat.*, **7**, 1 (1979) and B. Efron. and R.J. Tibshirani *An Introduction to the Bootstrap*, Chapman & Hall, NY (1994).

22. http://universe.nasa.gov/program/probes/jdem.html

23. At its simplest, a statistical model for a data vector $X = (X_1, \dots, X_n)$ is a collection of possible probability distributions for X indexed by a parameter θ. A statistical model is *parametric* if θ is a finite-dimensional object; the model is *nonparametric* if θ is an infinite-dimensional object, such as a function or an infinite sequence of coefficients in some basis expansion. Because there are effectively an infinite number of parameters in a nonparametric model, a priori constraints (usually smoothness or sparsity) are necessary for meaningful inferences, and inferences become sharp more slowly (as n grows) than with a parametric model, when both models are correct. The advantage of a nonparametric approach is greater flexibility in functional form; with large data sets, misfit from an incorrect parametric model can be significant.

24. Schwarz ... ATTB

Supplementary Material

Why Derivative Estimation Is Hard. We represent derivative estimation as an ill-posed inverse problem. Suppose we have data of the form

$$Y_i = f(z_i) + \sigma_i \epsilon_i,$$

where $\mathrm{E}(\epsilon_i) = 0$ and f is an unknown function in a nonparametric model. Suppose also we want to make inferences about f'. Then we can write

$$Y = Kf' + \Sigma^{1/2}\epsilon,$$

where $Y = (Y_1, \ldots, Y_n)$, $K = (K_1, \ldots, K_n)$, $K_i = \int_0^{z_i}$ is an integral operator, and $\Sigma^{1/2} = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$. We would like to recover f' from $K^{-1}Y$, but because K is a smoothing operator, it obscures high-frequency structure in f. (Put another way, the singular values of K decrease rapidly, making K an ill-conditioned operator.) This leads to huge variance inflation.

More formally, the eigenfunctions of the operator K^*K produce an orthonormal basis $\phi_1, \ldots, \phi_n$ with associated eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. Here, K^* is the adjoint of K given by

$$K^* u = \sum_{i=1}^n u_i 1_{[0, z_i]},$$

for any vector $u = (u_1, \ldots, u_n)$. Then,

$$f' = \sum_{j=1}^n \beta_j \phi_j + f'_\perp = \sum_{j=1}^n \lambda_j^{-1} \langle K\phi_j, Kf \rangle \phi_j \ + \ f'_\perp = \sum_{j=1}^n \lambda_j^{-1/2} \langle a_j, Kf \rangle \phi_j \ + \ f'_\perp,$$

where $a_j = K\phi_j / ||K\phi_j||$ and where $rf_\perp$ is orthogonal to the ϕ_js. Notice that the component $f_\perp$ cannot be estimated from the data: because $Kf'_\perp = 0$, its contribution is not observed.

Using an optimal estimation procedure for f', the mean squared error is approximately

$$\sum_{j=1}^n \min(\beta_j^2, \lambda_j^{-1}\tau_j^2),$$

where each τ_j^2 is a linear combination of the σ_k^2s (*13*). Because the λ_js decrease, large components at high order have a substantial contribution to the mean squared error. Put another way, derivative estimation is hard. And this problem becomes more severe for higher-order derivatives.

Deriving Equation (EQ::R). In this subsection, we show how to derive the expression for co-moving distance r as a function of w. The specification of w in terms of r' and r'' yields a differential equation whose solutions form a one-dimensional family in $r'(0)$. We have

$$\begin{aligned} w(z) &= \frac{\frac{3}{2}H_0^2\Omega_m(1+z)^3 + (1+z)\frac{r''(z)}{(r'(z))^3}}{\frac{3}{2}H_0^2\Omega_m(1+z)^3 - \frac{3}{2}(r'(z))^{-2}} - 1 && (1) \\ &= \frac{\frac{3}{2}H_0^2\Omega_m(1+z)^3(r'(z))^3 + r''(z)(1+z)}{\frac{3}{2}H_0^2\Omega_m(1+z)^3(r'(z))^3 - \frac{3}{2}r'(z)} - 1 && (2) \\ &= \frac{(1+z)r''(z) + \frac{3}{2}r'(z)}{\frac{3}{2}H_0^2\Omega_m(1+z)^3(r'(z))^3 - \frac{3}{2}r'(z)}. && (3) \end{aligned}$$

It follows that

$$r''(z) + \frac{3}{2}\frac{1+w(z)}{1+z}r'(z) = \frac{3}{2}H_0^2\Omega_m(1+z)^2 w(z)(r'(z))^3. \tag{4}$$

Write $g \equiv r'$ and $\theta = g(0)$. Note that $r(0) = 0$. The above differential equation becomes

$$g'(z) - \frac{1}{2}U(z)g(z) = -\frac{1}{2}V(z)g^3(z) \tag{5}$$

where $U(z) = -3\frac{1+w(z)}{1+z}$ and $V(z) = -3H_0^2\Omega_m(1+z)^2 w(z)$. Define $h = g^{-2}$, then $h' = -2g'/g^3$. In terms of h, the differential equation becomes

$$h'(z) + U(z)h(z) = V(z). \tag{6}$$

If we define $T(z) = \int_0^z U(s)\,ds$, then

$$\left(e^{T(z)}h(z)\right)' = e^{T(z)}U(z)h(z) + e^{T(z)}h'(z) = e^{T(z)}V(z); \tag{7}$$

hence, using $h(0) = 1/\theta^2$, we have

$$h(z) = e^{-T(z)} \int_0^z e^{T(s)} V(s)\, ds + \frac{1}{\theta^2}\, e^{-T(z)} \tag{8}$$

Using the boundary condition $\theta = 1/H_0$ gives

$$\begin{aligned}
r'(z) &= \left[e^{-T(z)} \int_0^z e^{T(s)} V(s)\, ds + H_0^2\, e^{-T(z)}\right]^{-\frac{1}{2}} &(9)\\
&= \left[3H_0^2\Omega_m e^{3\int_0^z \frac{1+w(s)}{1+s}\, ds} \int_0^z e^{-3\int_0^s \frac{1+w(t)}{1+t}\, dt}(1+s)^2(-w(s))\, ds \;+\; H_0^2 e^{3\int_0^z \frac{1+w(s)}{1+s}\, ds}\right]^{-\frac{1}{2}} &(10)\\
&= H_0^{-1}\left[\Omega_m(1+z)^3 e^{-3\int_0^z \frac{-w(s)}{1+s}\, ds} \int_0^z e^{3\int_0^s \frac{-w(t)}{1+t}\, dt} 3\left(\frac{-w(s)}{1+s}\right)\, ds \;+\; (1+z)^3 e^{-3\int_0^z \frac{-w(s)}{1+s}\, ds}\right]^{-\frac{1}{2}} &(11)\\
&= H_0^{-1}\left[\Omega_m(1+z)^3(1 - e^{-3\int_0^z \frac{-w(s)}{1+s}\, ds}) \;+\; (1+z)^3 e^{-3\int_0^z \frac{-w(s)}{1+s}\, ds}\right]^{-\frac{1}{2}} &(12)\\
&= H_0^{-1}\left[\Omega_m(1+z)^3 + (1-\Omega_m)(1+z)^3 e^{-3\int_0^z \frac{-w(s)}{1+s}\, ds}\right]^{-\frac{1}{2}} &(13)
\end{aligned}$$

Integrating gives the r in equation (EQ::R).

Because $r' \geq 0$, r is necessarily monotone. In fact, r' is bounded above and below by the cases where the exponential term is zero or one, respectively. This gives the stated bounds on the derivative. Concavity of r (and consequently $\log r$) can be a powerful tool for sharpening inferences. Conditions for concavity follow directly from the above expression for r'. Let $h = 1/(r')^2$; then, $r''(z) = -(1/2)h^{-3/2}(z)h'(z) = -(1/2)(r'(z))^3 h'(z)$, so $r'' \leq 0$ if and only if $h' \geq 0$. But

$$h'(z) = 3(1+z)^2\left[\Omega_M + (1-\Omega_m)(1+w(z))e^{-3\int_0^z \frac{-w(s)}{1+s}\, ds}\right].$$

It follows that r is concave if, for all $z \geq 0$,

$$w(z) \geq -\frac{1}{1-\Omega_m}\left[(1-\Omega_m) + \Omega_m e^{3\int_0^z \frac{-w(s)}{1+s}\, ds}\right].$$

This holds, for example, if $w \geq -1/(1-\Omega_m)$, as claimed. This is the best one can do near zero, but away from zero, sufficient smoothness conditions on w can lead to concavity as well.

Equation (EQ::R) also gives an expression for the deceleration q in terms of w.

$$
\begin{aligned}
q(z) &= -1-(1+z)\frac{r''(z)}{r'(z)} & (14)\\
&= -1-(1+z)[\log r'(z)]' & (15)\\
&= -1+\frac{1+z}{2}\left[\log\left(H_0^2\Omega_m(1+z)^3\left(1-e^{-3\int_0^z \frac{-w(s)}{1+s}\,ds}\right)+\frac{1}{\theta^2}\right)\right]' & (16)
\end{aligned}
$$

which becomes

$$
q(z)=\frac{1}{2}+\frac{3}{2}\frac{(1-\Omega_m)w(z)}{(1-\Omega_m)+\Omega_m e^{3\int_0^z \frac{-w(s)}{1+s}\,ds}} \qquad (EQ::Q::W)
$$

When $w\equiv -1$, equation (EQ::Q::W) reduces to

$$
q(z)=-1+\frac{3}{2}\frac{H_0^2\Omega_m\theta^2(1+z)^3}{H_0^2\Omega_m\theta^2\left((1+z)^3-1\right)+1},
$$

and the crossing point occurs at $z_0=(2(1-H_0^2\Omega_m\theta^2)/H_0^2\Omega_m\theta^2)^{1/3}-1$. We can solve for r in terms of q using a simpler method than that given above for w:

$$
r(z)=\lambda_1 z+\lambda_2\int_0^z \mathrm{d}s\, e^{-\int_0^s \mathrm{d}u\,\frac{1+q(u)}{1+u}}.
$$

The extra dimension in the solution space cannot be eliminated in general because q is insensitive to a linear term in the co-moving distance. However, comparison with equation (13) shows that either $\lambda_1=0$ or $r''\equiv 0$. This gives two families of solutions

$$
r(z)=\theta\int_0^z \mathrm{d}s\, e^{-\int_0^s \mathrm{d}u\,\frac{1+q(u)}{1+u}} \qquad \text{or} \qquad r(z)=\theta z, \qquad (EQ::R::Q)
$$

where $r'(0)=\theta>0$. For example, the case $q\equiv 0$, the boundary between accelerating and decelerating corresponds to $r(z)=\theta\log(1+z)$ because the linear solution is ruled out for purely decelerating universes.

By considering various forms for w, equation (EQ::R) produces various models for r. Suppose that we expand w in a basis as $w(z)=-\sum_{j=1}^{d}\beta_j\psi_j(z)$. Let $\tilde{\psi}_j(z)=3\int_0^z \psi_j(u)/(1+u)\,du$.

Equation (EQ::R::Basis) follows immediately. When the ψ_js are polynomials, this reduces to the form of equation (EQ::R::POLY). For integer $n \geq 0$,

$$\int_0^z \frac{u^n}{1+u}\,du = \frac{z^{n+1}}{n+1}\, F\left(\begin{array}{c} 1 \quad n+1 \\ n+2 \end{array} \middle| -z \right) = (-1)^n \left[\log(1+z) + \sum_{k=1}^{n} (-1)^k \frac{z^k}{k} \right],$$

where F is the generalized hypergeometric function F_1^2. For instance, if $\psi_j(z) = -z^{j-1}$, for $j = 1, \ldots, d$, let $\alpha_k = \sum_{m=k}^{d-1} (-1)^m \beta_{m+1}$, for $k = 0, \ldots, d-1$. Then,

$$\begin{aligned} \int_0^z \frac{-w(u)}{1+u} &= \left(\sum_{k=0}^{d-1} (-1)^k \beta_{k+1} \right) \log(1+z) + \sum_{k=1}^{d-1} (-1)^k \beta_{k+1} \sum_{\ell=1}^{k} (-1)^\ell \frac{z^\ell}{\ell} && (17) \\ &= \left(\sum_{k=0}^{d-1} (-1)^k \beta_{k+1} \right) \log(1+z) + \sum_{\ell=1}^{d-1} \sum_{k=\ell}^{d-1} (-1)^{k+\ell} \frac{\beta_{k+1}}{\ell}\, z^\ell && (18) \\ &= \alpha_0 \log(1+z) + \sum_{\ell=1}^{d-1} (-1)^\ell \alpha_\ell\, \frac{z^\ell}{\ell}. && (19) \end{aligned}$$

Equation (EQ::R::POLY) follows directly. Equation (EQ::R::APOLY) is derived similarly, using simple integration of $(1+z)^{-(j+1)}$ for each j.

Remark. For modeling purposes it might be worth re-expressing w as

$$w(z) = -\frac{1+z}{3} v'(z),$$

where $v(0) = 0$. This gives a one-to-one correspondence between w and v and r is more simply expressed as

$$r(z) = H_0^{-1} \int_0^z ds\, \left[\Omega_m (1+s)^3 + (1 - \Omega_m)(1+s)^3 e^{-v(s)} \right]^{-\frac{1}{2}}.$$

It is also straightforward to compute the Fisher Information matrix for the model. Let ∇r denote the gradient of r with respect to β, θ as a $d + 1 \times 1$ vector. Then, the log-likelihood $\ell(\beta, \theta) = -(1/2) \sum_{i=1}^n (Y_i - r(z_i; \beta, \theta))^2 / \sigma_i^2$ has a Fisher information matrix given by

$$I(\beta, \theta) = \sum_{i=1}^n \frac{1}{\sigma_i^2} \frac{\nabla r(z_i; \beta, \theta)^T \nabla r(z_i; \beta, \theta)}{r^2(z_i; \beta, \theta)}. \qquad (EQ:FISHER)$$

If G is the $n \times (d+1)$ matrix with G_{ij} equal to the jth component of $\nabla r(z_i;\beta,\theta)/r(z_i;\beta,\theta)$ and $\Sigma = \mathrm{diag}(\sigma_i^2)$, then the information matrix is given by $I(\beta,\theta) = G^T\Sigma^{-1}G$. When the ψ_js are polynomials, r can be computed in closed form, allowing a simple computation of the Fisher information matrix in β,θ. The observed Fisher information also has a relatively simple form that is easy to compute numerically. Under mild regularity conditions on the ψ_js in a basis expansion of w, we have that

$$\frac{\partial r}{\partial \beta_j}(z) = \frac{1}{2}\int_0^z ds\, (r'(s))^3(1-\Omega_m)(1+s)^3 e^{-\sum_j \beta_j \tilde{\psi}_j(s)}\tilde{\psi}_j(s),$$

and that

$$\frac{\partial r}{\partial H_0}(z) = -H_0^{-1} r(z),$$

and that

$$\frac{\partial r}{\partial \Omega_m}(z) = -\frac{1}{2}\int_0^z ds\, (r'(s))^3(1+s)^3(1-e^{-\sum_j \beta_j \tilde{\psi}_j(s)}).$$

Null Hypotheses for Testing Cosmological Models Here we provide further details about the hypothesis testing procedure.

1. Equation of state bounds for the quintessence models.

To derive equations (EQ::THAW) and (EQ::FREEZE), we begin by transforming equations (EQ::THAW::CON::lna) and (EQ::FREEZE::CON::lna) from the scale factor a to redshift z. Replacing $dw/d\ln a$ by $-(1+z)w'(z)$ and reversing the inequality because of the negative sign yields the corresponding equations

$$-3\frac{1+w(z)}{1+z} \le w'(z) \le -\frac{1+w(z)}{1+z}, \qquad (EQ :: THAW :: CON :: z)$$

for thawing solutions, and

$$-0.2w(z)\frac{1+w(z)}{1+z} \le w'(z) \le -3w(z)\frac{1+w(z)}{1+z}, \qquad (EQ :: FREEZE :: CON :: z)$$

for freezing solutions.

Begin with the assumption that $w > -1$, which we will weaken below. For the thawing equalities, divide through by $1 + w$ to get $w'/(1 + w) = (\log(1 + w))'$ and thus

$$-3\frac{1}{1+z} \leq (\log(1 + w(z)))' \leq -\frac{1}{1+z}.$$

Integrating through from 0 to z yields

$$-3\log(1 + z) \leq \log(1 + w(z)) - \log(1 + w(0)) \leq -\log(1 + z),$$

and taking exponents

$$(1 + z)^{-3} \leq \frac{1 + w(z)}{1 + w(0)} \leq (1 + z)^{-1},$$

which leads directly to equation (EQ::THAW).

Similarly for the freezing solutions, $w'/w(1 + w) = (\log(w/(1 + w)))'$. Dividing through and integrating as before gives

$$-0.2\log(1 + z) \leq \log\left(\frac{w(z)}{1 + w(z)}\right) - \log\left(\frac{w(0)}{1 + w(0)}\right) \leq -3\log(1 + z).$$

Taking exponents and simplifying gives equation (EQ::FREEZE).

For freezing solutions, $w' \geq 0$, and if $w(0) = -1$, then $w \equiv -1$. For thawing solutions, either $w(0) = -1$ and $w \equiv -1$, or $w(z) = -1$ for some $z > 0$. The latter case leads to a contradiction given the bounds on w' and continuity of w. Hence, the bounds hold for $-1 \leq w \leq 0.8$.

2. Constructing the Confidence set $\mathcal{C}$

Suppose that for $i = 1, \ldots, n$, $Y_i = f_i + \sigma_i\epsilon_i$, where σ_is are known numbers and the ϵ_is are independent Gaussian variables. This corresponds to equation (EQ::DATA) with $f_i = \log_1 0r(z_i)$. If the vector $f = (f_1, \ldots, f_n)$ denotes the true but unknown values of the function at the observed points, then a $1 - \alpha$ confidence set $\mathcal{C}$ for f is a random set, constructed from the data, that satisfies

$$P\{\mathcal{C} \ni f\} \geq 1 - \alpha.$$

We want $\mathcal{C}$ to be as small as possible.

There are several ways to construct such confidence sets. One way is to invert a goodness of fit test, such as the chi-squared test. The chi-squared statistic for the null hypothesis $f = f^0$ equals $T^2(f_0) = (1/n)\sum_{i=1}^{n}(Y_i - f_i^0)^2/\sigma_i^2$ on n degrees of freedom. The set of f^0 that are not rejected by this test at level α form a $1-\alpha$ confidence set. That is, we can define $\mathcal{C} = \{f^0 : T^2(f_0) \leq \chi^2_{n,\alpha}/n\}$, where $\chi^2_{n,\alpha}$ is the upper-tail α quantile of the corresponding chi-squared distribution. This defines a confidence set because, for the true f, $P\{\chi^2(f) \leq \chi^2_{n,\alpha}\} = 1-\alpha$.

The chi-squared confidence set is simple to use, but it has several major drawbacks. The confidence set is relatively large; the radius of the set $\chi n, \alpha/\sqrt{n}$ is $O(1)$ no matter how large n is. The set is constructed from a rough estimator of f, namely the data. The size of the set is independent of the data and thus cannot adjust to evidence of smoothness. And some prior information, such as shape restrictions, is difficult to incorporate in practice.

There are practical confidence set procedures that address all these drawbacks. We consider two, adaptive chi-squared confidence sets from Baraud and shape-restricted confidence bands from Davies et al. (*17*). Both achieve asymptotically optimal size and adjust their size based on the data. Both are computationally practical, though somewhat more work than the naive chi-square confidence set. And both allow us to incorporate prior information about the co-moving distance to produce a smaller confidence set.

The procedure of Baraud takes advantage of the fact that a smaller confidence set is possible when f is smoothly varying. One specifies a nested sequence of subspaces $V_1 \subset V_2 \subset \cdots \subset V_m$, where V_m is the set of all vectors and the rest of the subspaces have dimension less than $n/2$. Lower dimensional subspaces correspond to spaces of smoother f; the final subspace is included to ensure that all true vectors can be covered. For each subspace V_i, we do a chi-squared goodness of fit test of the null hypothesis $f \in V_i$. For those subspaces for which the null is not rejected, we form a ball of specified radius (depending on the subspace's dimension) around the projection of the data on that subspace. The confidence set is the intersection of these balls.

The procedure of Davis et al. generates confidence bands under the assumption that f is monotone and concave. Let I_{ik} be the set of $1 \le j \le n$ such that $z_i \le z_j \le z_k$ for some $i \le k$. Consider the following statistics, modified from the original paper to account for the different standard errors of the measurements:

$$T_{ik} = \frac{1}{\sqrt{\#(I_{jk})}} \sum_{j \in I_{jk}} \frac{y_j - f_j}{\sigma_j}.$$

When f is the true vector, the T_{ik}s each are mean zero Gaussian variables. The procedure begins with a confidence set for the T_{ik}s. In the original paper, this is a confidence cube with edge length equal to twice the $1-\alpha$ quantile of $\max_{i,k} |T_{ik}|$. The key is that the initial confidence set have linear boundaries. Constraints for concavity and monotonicity are also linear. The confidence bands are constructed by maximizing and minimizing f_j subject to f lying in the initial confidence set and satisfying the shape restrictions. This can be expressed as two linear programs for each j.

We modify the Davies et al. procedure in several ways. First, because monotonicity and concavity are used in the procedure in log space, the confidence bands need not be concave in r space. We adjust for this by optimizing the bands in r space, finding the smallest bands consistent with the shape restrictions there, including the additional constraint that $r(0) = 0$. This involves two additional linear programs for each z_i. Second, we use a smaller initial confidence set for the T_{ik}s, a degenerate ellipsoid, which is drastically smaller than the hyper-cube of Davies et al. but has the proper coverage. This requires second order cone programming, which **ATTN: This last part is not yet reflected in the results.**

Finally, we can search the chi-squared ball imposing the shape constraints using quadratic programming. This reduces the effective size of the confidence set substantially.

With the current data, all of these methods give comparable results, but the new methods give higher power, with the advantage increasing with n.

3. Computational details

The Baraud procedure has a several "tuning parameters," including the selection of subspaces, the allocation of confidence level to each of the subspaces, the significance levels of the chi-squared goodness of fit tests, and the overall confidence level. We use spaces of B-splines in $\log_{10}(z)$ of dimension 2, 3, 4, 5, 10, and 20 along with the full subspace of dimension n. In this parameterization, $\log_{10}(r)$ is nearly linear and is well smit by a low dimensional space of splines. Of total probability α allowed for misfit, we allocate $\alpha/2$ to the full subspace and divide the remaining $\alpha/2$ evenly among the spaces. We give aach of the individual goodness of fit tests significance level 0.2. For large data sets, the Baraud procedure will eventually outperform the chi-squared confidence ball, but with our current data, the results are essentially the same.

The Davies procedure applies to $\log_{10}(r)$ which is monotone and concave because r is. The result are concave and monotone lower confidence bands for $\log_{10}(r)$ that can be transferred to r space. The procedure requires solving two linear programs for each z_i for the initial bands and that much again for the improvement in r space. Searching the resulting bands in r space is straightforward. In some cases, one of the vectors in the one or two parameter family generated by our testing procedure lies within the bands. Otherwise, such a vector can be found by finding a feasible solution for a simple linear program in with monotonicity and concavity constraints and bounds dictated by the bands.

In all of the hypotheses we test, there are one or two free parameters that must be varied. A simple grid search is practical and straightforward. For any value of the free parameters, we can tell whether the corresponding r lies in the Davies et al. confidence bands by direct comparison and with the Baraud confidence set by computing the distance to each of the ball centers. The search can be made faster in the Baraud case through finding the best match with quadratic programming, but that did not prove necessary in practice.

ATTN This whole bit needs polishing and clarifying. Very likely should drop Baraud because it does nothing for us, but I'm holding on to it for the moment.

Bootstrap and Confidence Bands The bootstrap is a technique for approximating the sampling distribution of a statistic by resampling from the data set (*21*). Pseudo-data are drawn from the empiricial distribution of the data and the statistic computed; this is repeated many times to approximate a desired feature of the statistics' sampling distribution. The bootstrap is most commonly used to generate standard errors and confidence intervals for complicated statistics.

In a regression problem such as this, resampling naively from the empirical distribution would obscure the relationship with redshift. Instead, we resample from the residuals. The basic procedure is as follows:

1. Compute the maximum likelihood estimator $(\hat{\beta}, \hat{\theta})$, where β is the vector of parameters for w and $\theta \equiv r'(0)$.

2. Compute residuals $e_i = Y_i - \hat{r}(z_i)$.

3. Using the standard errors of Y_i and the the linear approximation at the maximum likelihood estimator, standardize the residuals to unit variance. Call these standardized residuals ϵ_i.

4. For $b = 1, \ldots, B$, for some large B, draw pseudo-noise from the empirical distribution of the ϵ_i. Call these $\epsilon_i^{*(b)}$ for $i = 1, \ldots, n$.

5. Generate pseudo-data

$$Y_i^{*(b)} = \hat{r}(z_i) + \sigma(z_i)\epsilon_i^{*(b)}.$$

6. Compute the maximum likelihood estimates $(\hat{\beta}^{*(b)}, \hat{\theta}^{*(b)})$ from each pseudo-data set.

7. Compute standard errors and confidence intervals for these parameters from the $(\hat{\beta}^{*(b)}, \hat{\theta}^{*(b)})$'s as in (*21*).

We uses bootstrap confidence intervals to compute confidence bands for w and q by computing the largest and smallest values of the functions at each redshift that are consistent with the confidence intervals on the parameters.