# Statistical Models for Dark Energy

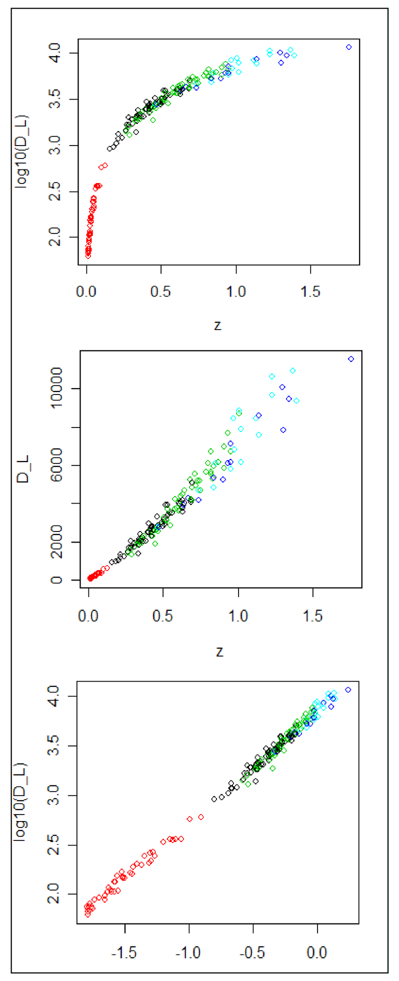
TRACY HOLSCLAW, RISHI GRAHAM, BRUNO SANSO, AND HERBIE LEE (UCSC ), KATRIN HEITMANN AND SALMAN HABIB (LANL)

Applied Math and Statistics, University of California, Santa Cruz, CA 95064, USA

## Abstract

The fact that the Universe is expanding has been known since the 1920's. If the Universe was filled with ordinary matter, the expansion should be decelerating. Beginning in 1998, however, observational evidence has been accumulating in favor of an accelerating expansion of the Universe. The unknown driver of the acceleration has been termed dark energy. The nature of dark energy can be investigated by studying its equation of state, that is the relationship of its pressure to its density. The equation of state can be measured via a study of the luminosity distance-redshift relation for supernovae. In this study we employ supernovae data, including measurement errors, to determine whether the equation of state is constant or not. Our method is based on fitting a Bayesian regression model and taking multiple derivatives of the predictive mean. The data is first transformed and then a Bayesian model is fit to it. The first derivative of the resulting model is the Hubble parameter, and the second derivative, w(z), is the equation of state parameter. The estimation of w(z) is the term of interest. This work stems from collaboration between UCSC and Los Alamos National Laboratory (LANL) in the context of the Institute for Scalable Scientific Data Management (ISSDM) project.

## Original Data



This is the raw data plot (z vs. μ). This data is obviously not linear but is the data we wish to fit. So we need to use some transforms to make the data linear.

First, we use the fact that $D_L = 10^{\frac{\mu-25}{5}}$ transform the data in terms of so $\mu = 5\log_{10} D_L + 25$. A regression equation: will have problems with non-constant variance. To correct for this problem we will instead transform the z variable.

Here the z variable is transformed by $\log_{10} z$ and we use $\log_{10} D_L$. The regression equation fit was: $\log_{10} D_L = \beta_0 + \beta_1 \log_{10} z$. This equation showed some sign of slight curvature so the squared term was added for a new regression equation:

$$\log_{10} D_L = \beta_0 + \beta_1 \log_{10} z + \beta_2 (\log_{10} z)^2$$

(You could leave this equation in terms of μ instead of and it would still be linear.)

## Where we want to go … w(z)

We start with our regression equation and take some derivatives:

$$\log_{10} D_L = \beta_0 + \beta_1 \log_{10} z + \beta_2 (\log_{10} z)^2$$

$$D_L = 10^{\beta_0 + \beta_1 \log_{10} z + \beta_2 (\log_{10} z)^2}$$

$$D_L' = z^{-1}(\beta_1 + 2\beta_2 \log_{10} z)\, 10^{\beta_0 + \beta_1 \log_{10} z + \beta_2 (\log_{10} z)^2}$$

$$D_L'' = z^{-1}(\beta_1 + 2\beta_2 \log_{10} z)\, D_L' + \frac{(-\beta_1 \ln 10 + 2\beta_2 - 2\beta_2 \ln z)}{z^2 \ln 10} D_L$$

Then we need Hubble's parameter and its derivative (where x = z + 1):

$$H(z) = \left[\frac{d}{dz}\left(\frac{D_L}{1+z}\right)\right]^{-1} = \frac{(1+z)^2}{(1+z)D_L' - D_L} \quad \text{and} \quad H'(x) = \frac{-D_L'' x^3 + 2D_L' x^2 - 2D_L x}{(xD_L' - D_L)^2}$$

Finally, we can solve for w(x) and we use the constants $H_0 = 71.0 \pm 2.6$ and $\Omega_{m0} = 0.265 \pm 0.03$:

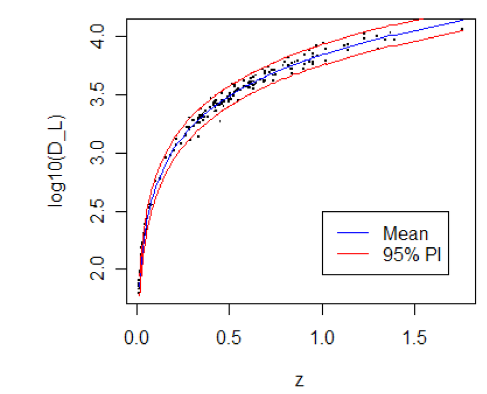
$$w(x) = \frac{2q(x) - 1}{3(1 - \Omega_m(x))} \quad \text{where} \quad q(x) = x\frac{H'(x)}{H(x)} - 1 \quad \text{and} \quad \Omega_m(x) = \left(\frac{H_0}{H(x)}\right)^2 x^3 \Omega_{m0}$$

## Bayesian Normal Regression

**Normal Model-** 95% Probability Intervals for a run of 10,000 samples

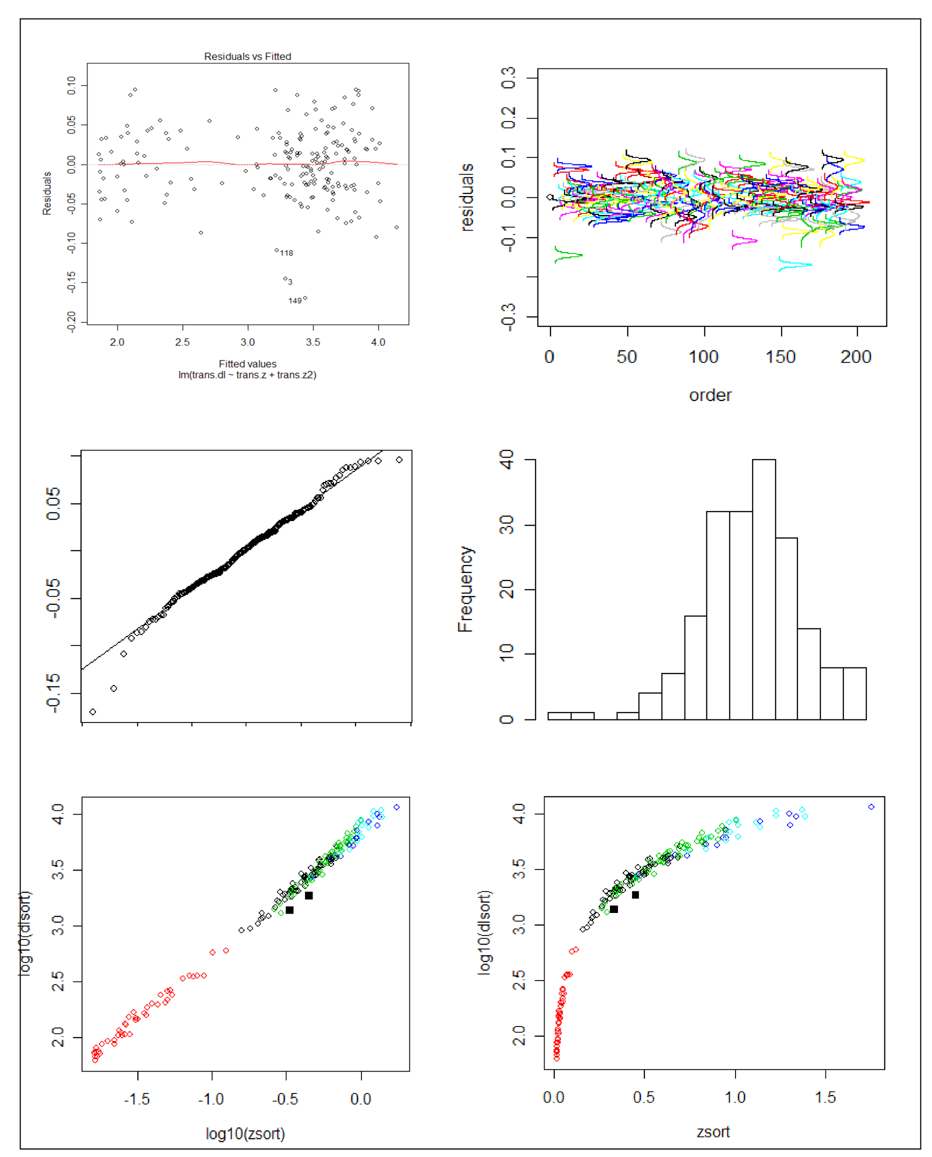
| Parameter | Lower bound | Estimated value | Upper bound |
|---|---|---|---|
| $\beta_0$ | 3.83568909 | 3.84734588 | 3.85898462 |
| $\beta_1$ | 1.15207258 | 1.19561118 | 1.23907095 |
| $\beta_2$ | 0.02351677 | 0.04881526 | 0.07349912 |

**Fitted line and 95% Predictive Intervals**




The analysis was done using a Bayesian regression model with non-informative prior. 10,000 draws were taken for each Beta parameter being estimated, as well as, the estimated variance being drawn from an Inverse Chi-square distribution. There was no problem with burn in, mixing, or autocorrelation of the posteriors because Gibbs steps were used.

## Residual Analysis

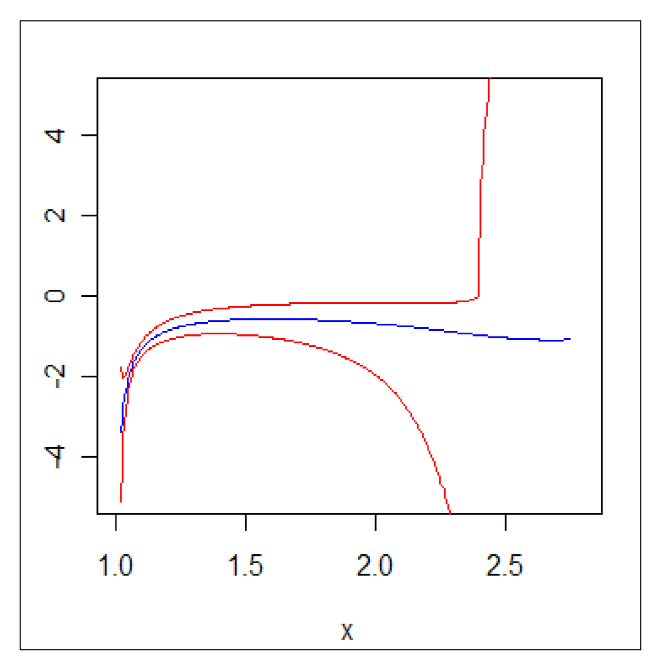



The residual plot on the left we can use this to examine the assumptions of constant variance. The plot of Bayesian residuals on the right show the distribution of each residual.

The plot on the left is a QQ plot for the Normal distribution which is showing a few outliers and possibly heavier tails than a Normal distribution. And the histogram on the right looks somewhat Normal but may be skewed.

These last two plots are of the original data but with the two outliers of interest highlighted. We wanted to see where these points fell with respect to z. (The multi-colors mark different data collectors.)

## Fit of w(z)




The estimate of w(z) is around the value negative one. The blue line is the mean fit and the red lines are the 95% probability intervals. There is a large amount of variation for the larger values of z because of a few points in this region.

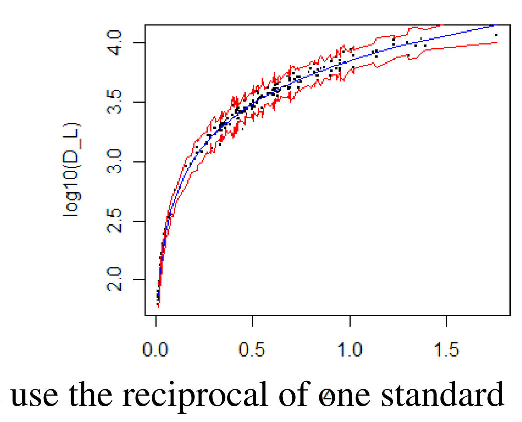
Other papers have also estimated w(z) around negative one, so these results are consistent with other sources. Which means our model is consistent with others' findings.

## Weighted Bayesian Normal Regression

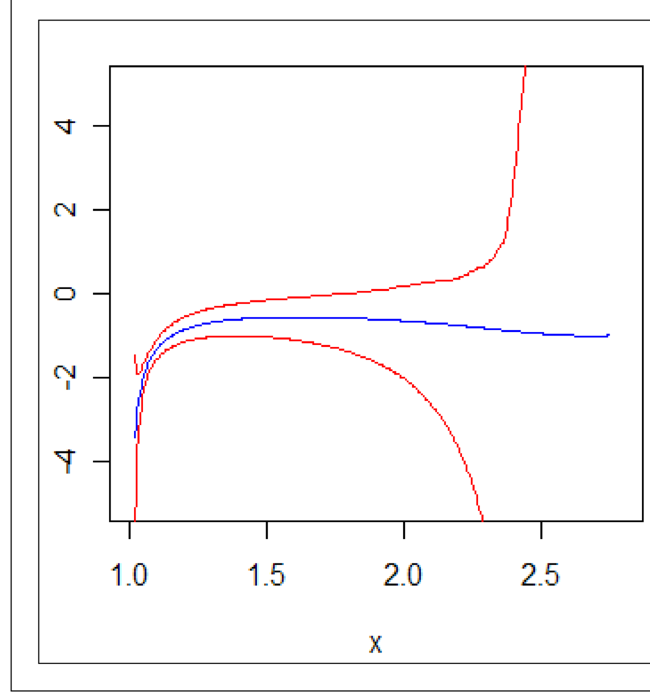
**Normal Model with weights-** 95% Probability Intervals

| Parameter | Lower bound | Estimated value | Upper bound |
|---|---|---|---|
| $\beta_0$ | 3.837621 | 3.84935690 | 3.86093713 |
| $\beta_1$ | 1.165519 | 1.20938644 | 1.25306955 |
| $\beta_2$ | 0.032682 | 0.05757346 | 0.08279315 |

**Fitted line and 95% Predictive Intervals**




Each supernova has its own error measurement. We use the reciprocal of one standard deviation of the estimated error as the weight for each supernova. This means that supernovas with more measurement error will carry less weight in our model. We transform the weights to match the transforms we put on the data to make it linear. After analyzing this model we see there is not a large difference in the parameter estimate of the fit of the line compared to the regression without weights.




The weights still leave our Q-Q plot with heavy tails and the same outliers that need to be analyzed as the previous model. A heavier tailed distribution may fit the data better.

These measurement weights do affect w(z). These cause our 95% probability bands to be wider, which is expected since there is more uncertainty in the model.

## Conclusions

- We have shown that simple Bayesian regression models can be used to obtain valid information about w(z) when using transformations on the data
- Both models here allow for estimation of uncertainty bands in a straightforward way
- Incorporating the measurement error via weights affects the estimation of w(z)
- For larger values of z there is a large amount of uncertainty

## Future Work

- We will fit a t-distribution with heavy tails to account for the non-Normality in the Q-Q plot
- A neural net will be fit and w(z) approximated from that fit
- We will look at other similar data sets
- We can examine the few outliers or larger z values more carefully
- We can analyze more of the measurement error and also the error in some of the constant terms used to approximate w(z) ie. the Hubble constant

## References

Heitmann, Katrin. Cosmic Calibration: Reconstructing Dark Energy from Supernovae. Los Alamos National Laboratory.

Sahni, V. & Starobinsky, A. (2006). *Reconstructing Dark Energy*. Int.J.Mod. Phys. D15, 2105.

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