Embedding the Big Bang Cosmological Model into a Bayesian Hierarchical Model

David A. van Dyk

Statistics Section, Imperial College London

Joint work with Roberto Trotta, Xiyun Jiao, and Hikmatali Shariff

Joint Statistical Meetings
August 2014

Imperial College London

Themes

The Accelerating Expansion of the Universe

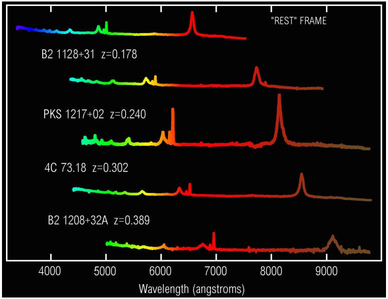
- The role of Dark Energy are Dark Matter in the evolutionary history of the Universe remain mysterious.
- Charting the expansion history is key to testing physical theories for Dark Matter and Dark Energy.
- To do this, we embed cosmological models into a Bayesian hierarchical model.
- Principled handling of data and model complexity.
- Gain better astronomical measurements along the way.

Outline

1. Measuring the Expansion of the Universe
2. A Hierarchical Statistical Model
3. Shrinkage Estimates of Absolute Magnitudes

The Expanding Universe

Redshift



http://www.noao.edu/image_gallery/html/im0566.html

For "nearby" objects,

$$z = \text{velocity}/c$$
$$\text{velocity} = H_0 \text{ distance.}$$

Hubble's Famous Diagram

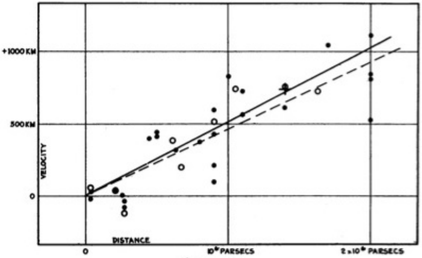


FIGURE 1

Velocity-Distance Relation among Extra-Galactic Nebulae.

Radial velocities, corrected for solar motion, are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually.

Hubble (1929)

The Big Bang!

Imperial College London

Distance Modulus in an Expanding Universe

Apparent magnitude - Absolute magnitude = Distance modulus:

$$m - M = \mu = 5\log_{10}(\text{distance}[\text{Mpc}]) + 25$$

Computing absolute magnitudes, relationship between μ and z

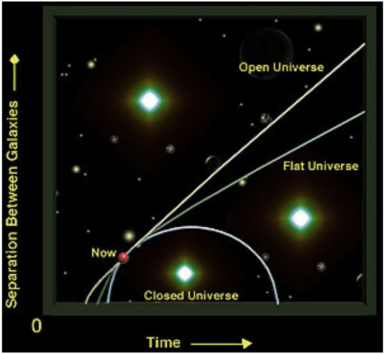
- For nearby objects,

  $$\text{distance} = zc/H_0.$$

  (Correcting for peculiar/local velocities.)

- For distant objects, involves expansion history of Universe:

  $$\mu = g(z, \Omega_\Lambda, \Omega_M, H_0)$$

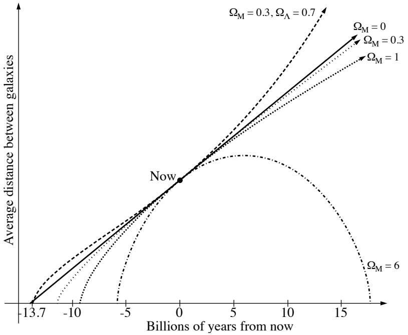


http://skyserver.sdss.org/dr1/en/astro/universe/universe.asp

Accelerating Expansion of the Universe

Recall: $m - M = \mu = g(z, \Omega_\Lambda, \Omega_M, H_0)$

- 2011 Physics Nobel Prize: discovery that expansion rate is increasing.
- Dark Energy is principle theorized explanation of acceleration.
- Ω_Λ: density of dark energy (describes acceleration).
- Ω_M: total matter.



If we observe both m and M we can infer μ and the cosmological parameters.

Type Ia Supernovae as Standardizable Candles

If mass surpasses "Chandrasekhar threshold" of $1.44 M_{\odot}$...

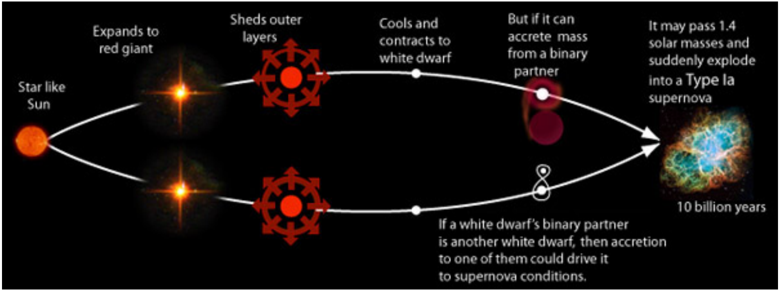


Image Credit: http://hyperphysics.phy-astr.gsu.edu/hbase/astro/snovcn.html

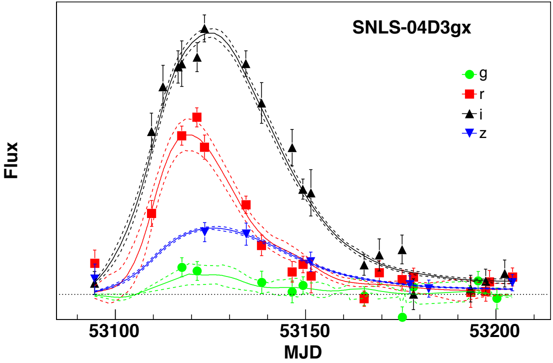
Common "flashpoint" → similar absolute magnitudes

$$M_i \sim \mathrm{N}(M_0, \sigma^2_{\text{int}}).$$

Non-linear Regression: $m_{Bi} = g(z_i, \Omega_\Lambda, \Omega_M, H_0) + M_i$

Imperial College London

Photometric Light Curves: The Raw Data



Guy et al (2007)

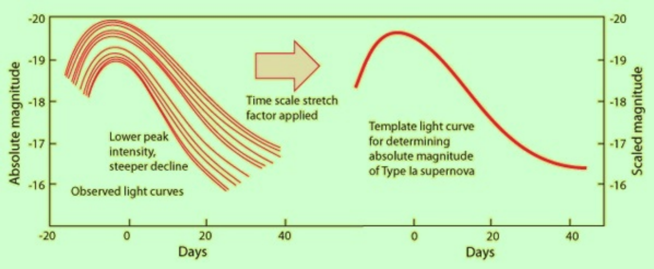
MJD = days since November 17, 1858

We use peak B band magnitudes (apparent magnitude $= -2.5\log_{10}(\text{flux})$)

Predicting Absolute Magnitude

SN1a absolute magnitudes are correlated with characteristics of the explosion / light curve:

- x_i: rescale light curve to match mean template
- c_i: describes how flux depends on color (spectrum)



Credit: http://hyperphysics.phy-astr.gsu.edu/hbase/astro/snovcn.html

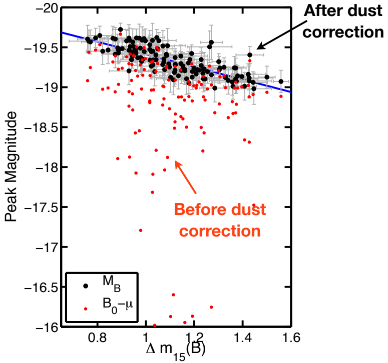
Cite: SALT and SALT II (Guy *et al.* 2005 and 2007).

Imperial College London

Phillips Corrections

- Recall: $m_{Bi} = \mu_i + M_i$
  $$M_i \sim \mathrm{N}(M_0, \sigma^2_{\mathrm{int}}).$$
- Regression:
  $$M_i = -\alpha x_i + \beta c_i + M_i^\epsilon,$$
  $$m_{Bi} = \mu_i - \alpha x_i + \beta c_i + M_i^\epsilon,$$
  $$M_i^\epsilon \sim \mathrm{N}(M_0, \sigma^2_\epsilon).$$
- $\sigma^2_\epsilon \leqslant \sigma^2_{\mathrm{int}}$
- Including x_i and c_i reduces variance and increases precision of estimates.

Low-z calibration sample



Mandel et al (2011)

Light curve stretch

Brighter SNIa are slower decliners over time.

Data

Two samples:

1. An Sloan Digital Sky Survey (2009) sample of 288 SNIa.[1]
2. A larger JLA sample[2] of 740 SNIa observed with SDSS, Hubble Space Telescope, SNLS (Canada-France-Hawaii Telescope), and several other telescopes for low z SNIa.

[1] Kessler et al., 2009, arXiv:0908.4274

[2] Betoule, et al., 2014, arXiv:1401.4064v1

Outline

The Baseline Hierarchical Model

Level 1: c_i, x_i, and m_{Bi} are observed with error.

$$\begin{pmatrix} \hat{c}_i \\ \hat{x}_i \\ \hat{m}_{Bi} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} c_i \\ x_i \\ m_{Bi} \end{pmatrix}, \Sigma_i \right\}$$

with $m_{Bi} = \mu_i + M_i^\epsilon - \alpha x_i + \beta c_i$ and $\mu_i = g(z_i, \Omega_\Lambda, \Omega_M, H_0)$

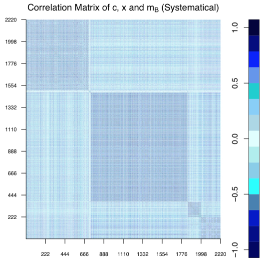
Level 2:

1. $c_i \sim \mathrm{N}(c_0, R_c^2)$
2. $x_i \sim \mathrm{N}(x_0, R_x^2)$
3. $M_i^\epsilon \sim \mathrm{N}(M_0, \sigma_\epsilon^2)$

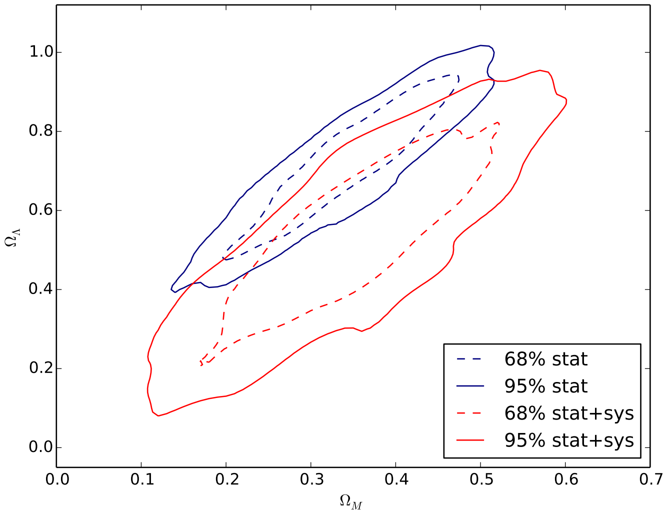
Level 3: Priors on α, β, Ω_Λ, Ω_M, H_0, c_0, R_c^2, x_0, R_x^2, M_0, σ_ϵ^2

Accounting for Systematic/Instrumental Effects

- Systematic errors: differences among telescopes, their components, and observational conditions.
- Total Variance: $\Sigma = \Sigma_{\text{stat}} + \Sigma_{\text{sys}}$
- Blocks: SNLS, HST, SDSS, low z.
- Similar to random effect for telescope.

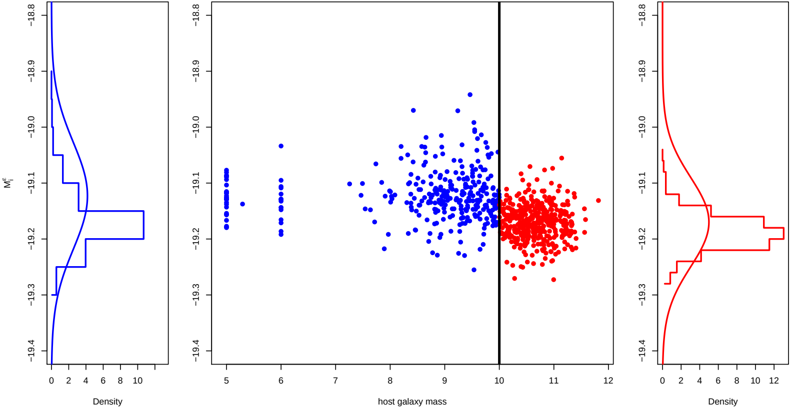


Effect on Fitted Cosmological Parameters



Adjusting for Galactic Masses

Can we further reduce the residual error by adjusting for the mass of the host galaxy?



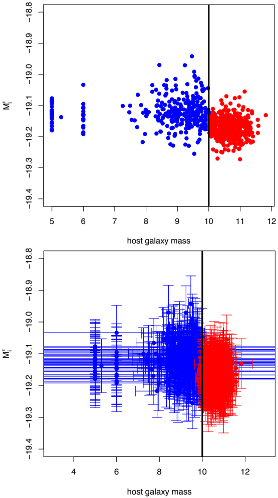
Comparing $\mathrm{E}(M_i^\epsilon \mid Y)$ with estimated mass [$\log_{10} M_\odot$] of host galaxy.

Adjusting for Galactic Masses

- The distribution of M_i^ϵ appears to depend on host galaxy mass $= w$.
- Only observe $\hat{w}_i \sim N(w_i, \sigma_{wi})$.
- We separate the population:

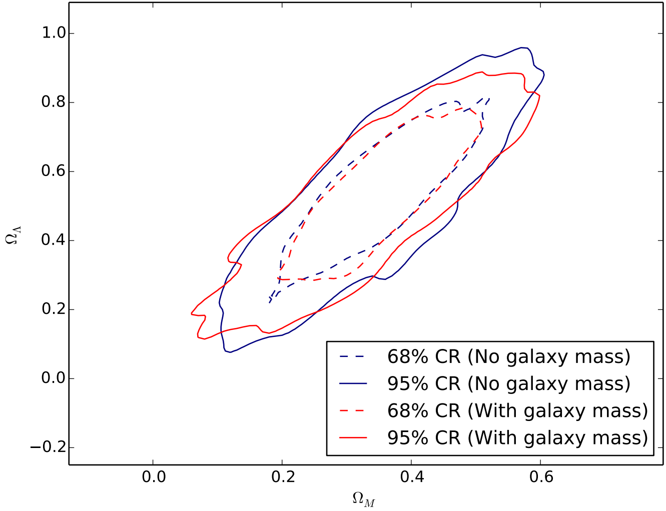
$$M_i^\epsilon \sim \mathrm{N}(M_{01}, \sigma^2_{\epsilon 1}) \text{ if galaxy mass } = w_i < 10$$
$$M_i^\epsilon \sim \mathrm{N}(M_{02}, \sigma^2_{\epsilon 2}) \text{ if galaxy mass } = w_i \geqslant 10.$$

- This reduces residual variance.
- Better strategies:
  - ⋆ $M_i^\epsilon \sim \mathrm{N}(M_0 + \psi w_i, \sigma^2_\epsilon)$
  - ⋆ $m_{Bi} = \mu_i + M_i^\epsilon - \alpha x_i + \beta c_i + \boldsymbol{\psi w_i}$ with $w_i \sim \mathrm{N}(w_0, R_w^2)$
- Non-linearity / interaction?



Imperial College London

Effect on Fitted Cosmological Parameters



Checking the Cosmological Model

We model:

$$m_{Bi} = g(z_i, \Omega_\Lambda, \Omega_M, H_0) - \alpha x_i + \beta c_i + M_i^\epsilon$$

How good of a fit is the cosmological model,
$g(z_i, \Omega_\Lambda, \Omega_M, H_0)$?

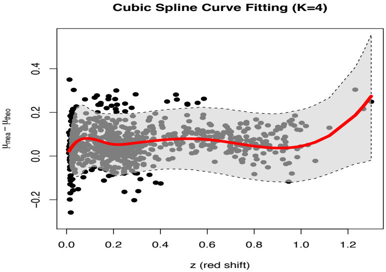
We can check the model by adding a cubic spline term:

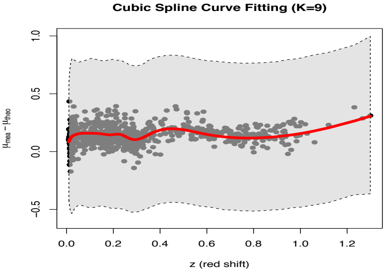
$$m_{Bi} = g(z_i, \Omega_\Lambda, \Omega_M, H_0) + h(z_i) + M_i^\epsilon - \alpha x_i + \beta c_i + M_i^\epsilon$$

where, $h(z_i)$ is cubic spline term with K knots.

Checking the Cosmological Model

Fitted cubic spline, $h(z)$, and its errors:





Can use similar methods to compare with competing cosmological models.

Statistical Computation

MH within Gibbs:	Decondition:	Permute:	Trim:
1: $p(\mu, X \mid \Omega, \beta, T)$	1: $p(\mu, X \mid \Omega, \beta, T)$	1: $\mathcal{M}(\Omega, \mu, X \mid \beta, T)$	1: $\mathcal{M}(\Omega \mid \beta, T)$
2: $\mathcal{M}(\Omega \mid \mu, X, \beta, T)$	2: $\mathcal{M}(\Omega, \boldsymbol{\mu}, \boldsymbol{X} \mid \beta, T)$	2: $\mathcal{M}(\beta, \mu, X \mid \Omega, T)$	2: $\mathcal{M}(\beta \mid \Omega, T)$
3: $p(\beta \mid \mu, X, \Omega, T)$	3: $\mathcal{M}(\beta, \boldsymbol{\mu}, \boldsymbol{X} \mid \Omega, T)$	3: $p(\mu, X \mid \Omega, \beta, T)$	3: $p(\mu, X \mid \Omega, \beta, T)$
4: $p(T \mid \mu, X, \Omega, \beta)$	4: $p(T \mid \mu, X, \Omega, \beta)$	4: $p(T \mid \mu, X, \Omega, \beta)$	4: $p(T \mid \mu, X, \Omega, \beta)$

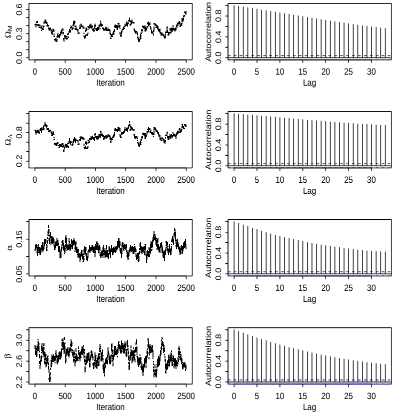
Baseline Hierarchical Model:

- ⋆ Let X represent the random effects
  ⋆ μ and T their means and variances, respectively
  ⋆ β the regression coefficients
  ⋆ Ω the cosmological parameters
- Final sampler is an MH with Partially Collapsed Sampler.[3]
- Steps 1-2 analytically marginalize out X and μ.
- Construct with care: permuting steps may change the stationary distribution of the chain.

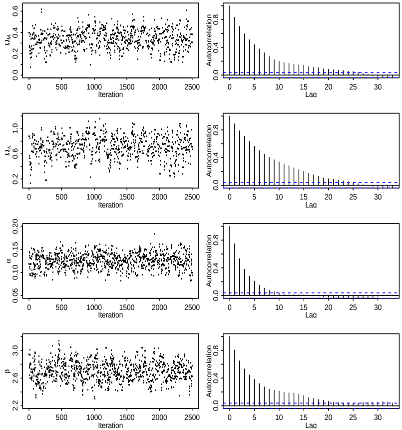
[3] van Dyk and Jiao (2014). The MH within PCG Sampler, JCGS, to appear.

Improved Mixing

MH within Gibbs Sampler



MH within PCG Sampler



New ASA Interest Group!

New! Astrostatistics Interest Group New!

At the JSM:

- Sunday at 4 PM: Bayesian Astrostatistics
- Wednesday at 8:30 AM: Big Data in Astrostatistics
- Wednesday at 10:30 AM: Informal Meeting outside the "Big Data in Astrostatistics" session room
- Wednesday at 2:00 PM: Analysis of Kepler Data at SAMSI
- Thursday at 8:30 AM: IOL: Astrostatistics

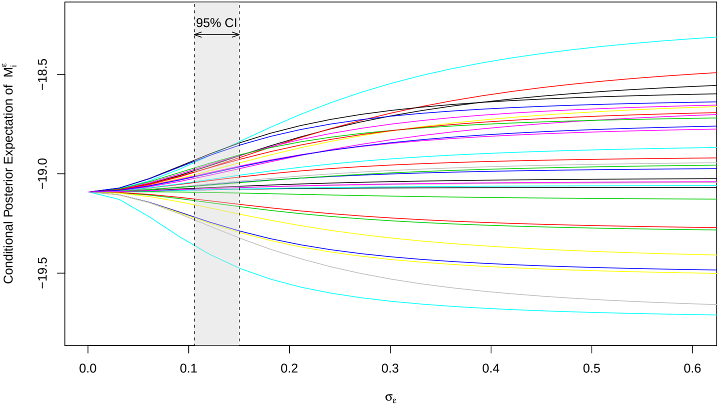
For more information:

`http://community.amstat.org/astrostats/home`

Outline

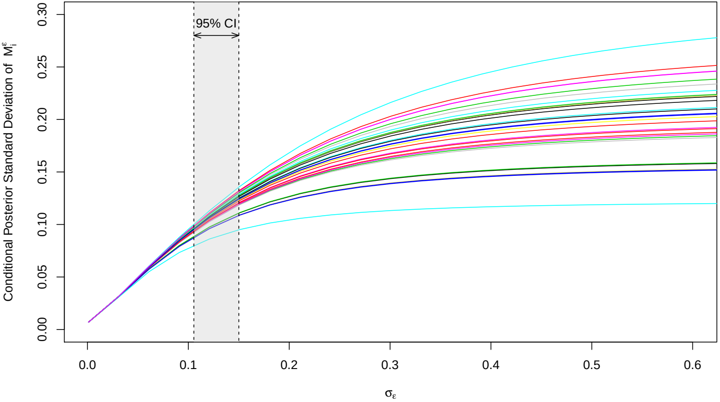
Shrinkage Estimates in Hierarchical Model

A statistical byproduct: low MSE estimates of M_i^ϵ.



Shrinkage Errors in Hierarchical Model

Reduced standard errors



Fitting Absolute Magnitudes Without Shrinkage

Under the model, absolute magnitudes are given by

$$M_i^{\epsilon} = m_{Bi} - \mu_i + \alpha x_i - \beta c_i \text{ with } \mu_i = g(z_i, \Omega_\Lambda, \Omega_M, H_0)$$

Setting

1. $\alpha, \beta, \Omega_\Lambda$, and Ω_M to their minimum χ^2 estimates,
2. $H_0 = 72 km/s/Mpc$, and
3. m_{Bi}, x_i, and c_i to their observed values

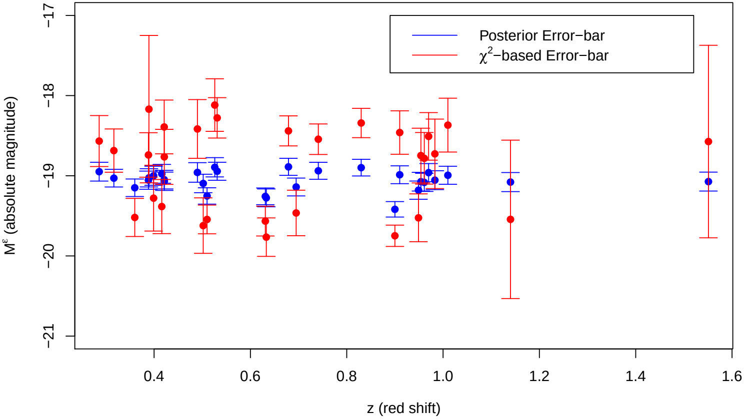
we have

$$\hat{M}_i^{\epsilon} = \hat{m}_{Bi} - g(\hat{z}_i, \hat{\Omega}_\Lambda, \hat{\Omega}_M, \hat{H}_0) + \hat{\alpha}\hat{x}_i - \hat{\beta}\hat{c}_i$$

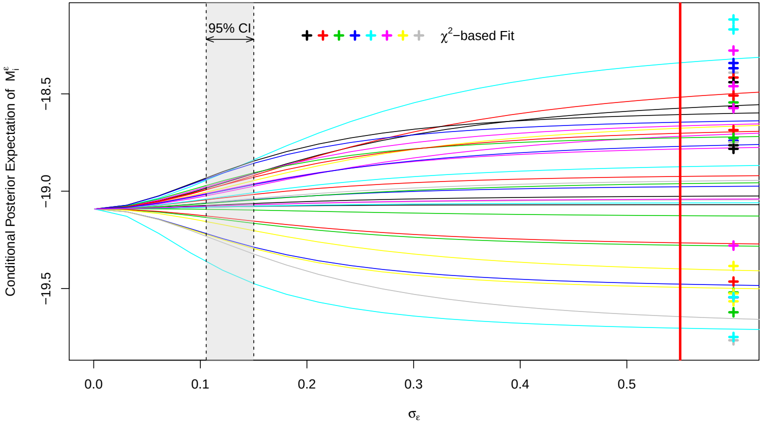
with error

$$\approx \sqrt{\mathrm{Var}(\hat{m}_{Bi}) + \hat{\alpha}^2 \mathrm{Var}(\hat{x}_i) + \hat{\beta}^2 \mathrm{Var}(\hat{c}_i)}$$

Comparing the Estimates



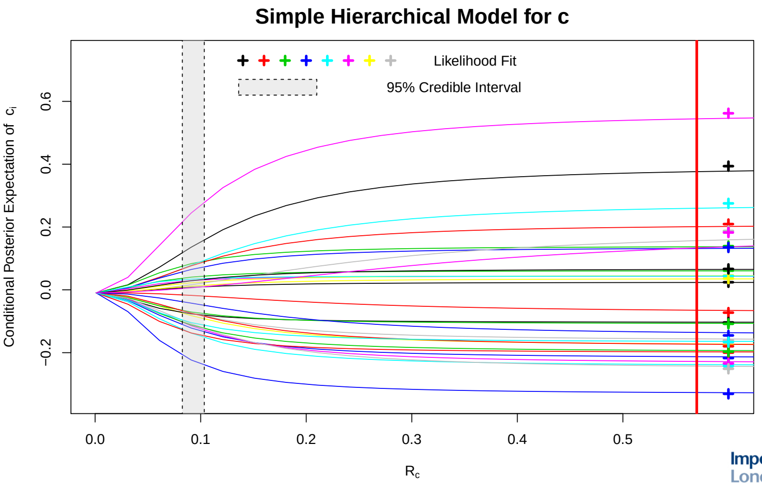
Comparing the Estimates



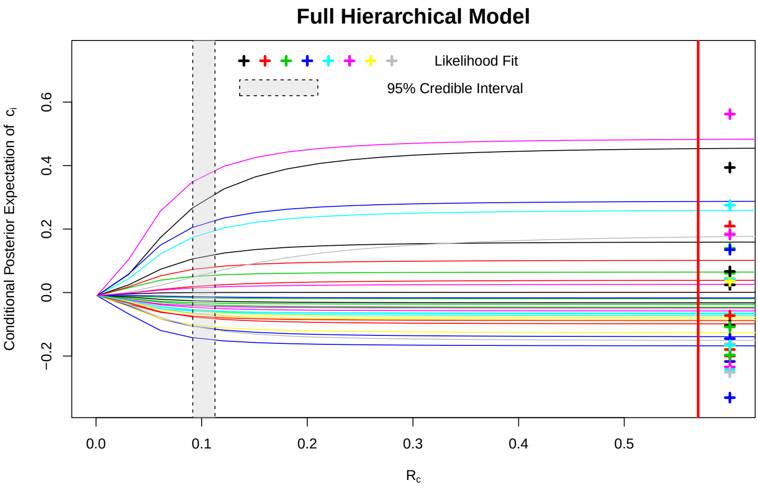
Bayes estimates are offset even without shrinkage.

Fitting a simple hierarchical model for c_i

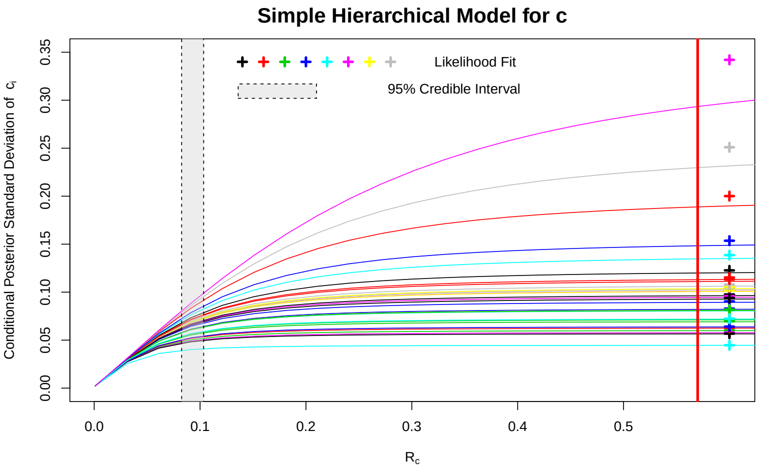
Model: $\hat{c}_i \sim N(c_i, \sigma_{ci})$ with $c_i \sim \mathrm{N}(c_0, R_c^2)$.



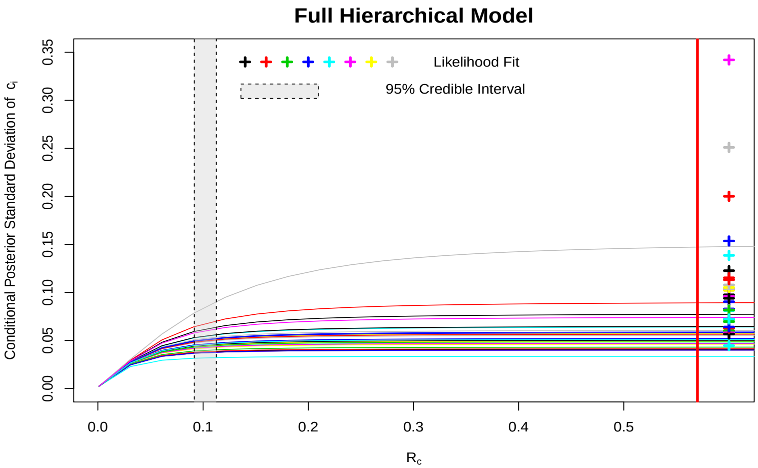
Additional shrinkage due to regression



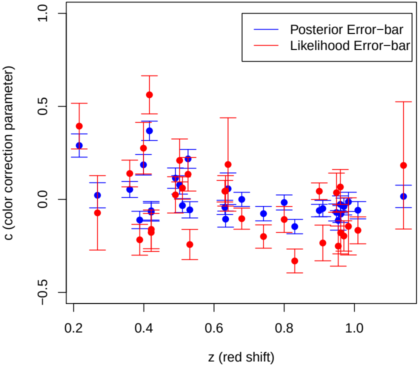
Errors under simple hierarchical model for c_i

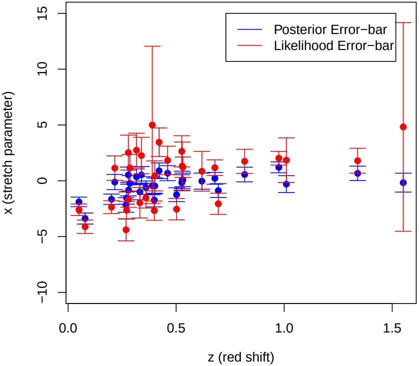


Reduced errors due to regression



Comparing the Estimates of c_i and x_i





Discussion

- *Bayesian science-driven hierarchical model* provides a platform for honest handling of model & data complexity.
- Sophisticated computation allows for effecient model fitting.
- Estimation of groups of parameters describing populations of sources not uncommon in astronomy.
- These parameters may or may not be of primary interest.
- Modeling the distribution of object-specific parameters can dramatically *reduce both error bars and MSE* ...
- ... especially with noisy observations of similar objects.