

# WILD 6900 Final Project

Fitting a hierarchical linear model and evaluating impacts of model misspecification

Martin Holdrege

2020-04-26

## Introduction

Results from ecological experiments can often modeled using frequentist mixed effects models or similarly using hierarchical models in a bayesian framework. For this project, data was simulated from a simple linear hierarchical process, where there was one predictor variable, with a random slope and intercept. The objective was to repeatedly simulate data and fit models to see how well the models estimated to the true parameters. One possible mistake that can be made when modeling hierarchical data is to ignore the fact that groups exist within the data and to, for example, just fit a single slope and intercept. This ignores the fact that observations within a group are often more similar to each other, or share some other characteristic. Groupings like this may occur for example if there are sub-populations, multiple measurements of single plot (pseudo-replication), or if measurements are grouped in time. To understand the effects of ignoring the hierarchical structure of the simulated data, correctly specified hierarchical model results were compared to results from a simple linear regression with a single slope and intercept. For these comparisons Bayesian models with uninformative were used.

## Methods

I simulated a small data set that was generated by a linear hierarchical process. The simulated data was meant to reflect a hypothetical data set generated by a small precipitation manipulation experiment. I am defining  $y$  to be stem growth (continuous response variable),  $x$  to be precipitation intensity (continuous predictor variable), and  $i$  to be an index for year, and  $j$  to be plot. That is, this is a multi year experiment, and for each year growth rate is estimated. Six years were simulated, in each year there were 4 treatment levels ( $x$ , treated as a continuous variable), and 6 replicate plots per treatment level. The sample sizes and effect size used were relatively small, so the data could be somewhat representative of an under powered ecological experiment. The overall model generating process was:

$$y_{ij} = \beta_{0i} + \beta_{1i}x + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_{\epsilon}^2)$$

When simulating the data,  $\sigma_{\epsilon}$  was set at 0.3. The random intercepts ( $\beta_{0i}$ ) were simulated as follows:

$$\beta_{0i} \sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2)$$

where the hyperparameter  $\mu_{\beta_0}$  (population mean intercept) was set at 4, and  $\sigma_{\beta_0}$  (standard deviation of intercepts between years) was 0.4. Similarly random slopes ( $\beta_{1i}$ ) were simulated as follows:

$$\beta_{01i} \sim N(\mu_{\beta_1}, \sigma_{\beta_1}^2)$$

where the hyperparameter  $\mu_{\beta_1}$  (population mean slope) was set at 0.2, and  $\sigma_{\beta_1}$  (standard deviation of slopes between years) was 0.3.

A bayesian hierarchical model was fit to this data. The model was specified ‘correctly’ based on the known data generating process described above. That is the model was set up as follows (notation follows Hobbs and Hooten (2015)):

$$[\beta_{0i}, \beta_{1i}, \mu_{\beta_0}, \mu_{\beta_1}, \sigma_{\epsilon}^2, \sigma_{\beta_0}^2, \sigma_{\beta_1}^2 | y_{ij}] \propto [y_{ij} | \beta_{0i}, \beta_{1i}, \sigma_{\epsilon}^2] [\beta_{0i}, \beta_{1i} | \mu_{\beta_0}, \mu_{\beta_1}, \sigma_{\beta_0}^2, \sigma_{\beta_1}^2] [\mu_{\beta_0}] [\mu_{\beta_1}] [\sigma_{\beta_0}^2] [\sigma_{\beta_1}^2] [\sigma_{\epsilon}^2]$$

The likelihood was estimated as follows (where  $i$  is year and  $j$  is plot):

$$[y_{ij} | \beta_{0i}, \beta_{1i}, \sigma_{\epsilon}^2] = Normal(\beta_{0i} | \mu_{\beta_0}, \sigma_{\beta_0}^2) * Normal(\beta_{1i} | \mu_{\beta_1}, \sigma_{\beta_1}^2) * \prod_{i=1}^I \prod_{j=1}^J Normal(y_{ij} | \beta_{0i} + \beta_{1i} x_{ij}, \sigma_{\epsilon}^2)$$

In all cases uninformative priors were used.

The prior for the intercept hyper-parameter was:

$$[\mu_{\beta_0}] = Normal(\mu_{\beta_0} | 0, 10^2)$$

The prior for the slope hyper-parameter was:

$$[\mu_{\beta_1}] = Normal(\mu_{\beta_1} | 0, 10^2)$$

The prior for the error term was:

$$[\sigma_{\epsilon}^2] = uniform(\sigma_{\epsilon}^2 | 0, 20)$$

The hierarchical model was fit in two ways. First, an MCMC metropolis sampler was built from scratch in R. This was done so all the steps of the model fitting process were better understood. When using this sampler chains were 10,000 iterations long, with the first 500 iterations discarded as burn in. To test for bias in this metropolis sampler, the data set was simulated 20 times and each time the parameters were estimated. Notable problems with the code developed were, 1) that it was very slow (more than an order of magnitude slower than JAGS code), and 2) for simplicity only one tuning parameter was used. The issue with only one tuning parameter being used was that while the overall acceptance rate of proposal values was around 40% the acceptance rate for individual parameters varied substantially (parameters with less variance ideally should be coded so they have a different, smaller, value of the tuning parameter). For this reason, subsequent model fitting was done using JAGS. The JAGS model was defined as described above, except chains were 20,000 iterations long. To save computing time only a single chain was built in each run (in initial tests mixing was good). Five hundred data sets were simulated with a model fit to each one. Subsequent analysis compared the parameter estimates (and their credible intervals) to the true known population parameters. Of most interest was whether the slope hyperparameter (treatment effect) estimates were about right. Clearly, a more thorough examination of the accuracy of this model would require a greater number of simulations, but this was not done here due to time and computing constraints. Additionally, for each of the 500 models only the summaries were saved so model diagnostics were not checked. However, when initially building the model, trace plots from a single run were checked.

A second, incorrect, model was also fit to this data. This second model ignores the fact that data within a given year is more similar than data between years, and treats all observations as independent (i.e. a sort of pseudo-replication is ignored). Specifically, the model ‘thinks’ the data generating process is:

$$y_j = \beta_0 + \beta_1 x + \epsilon_j$$

where the error term follows a normal distribution:

$$\epsilon_j \sim N(0, \sigma_\epsilon^2)$$

For this incorrectly specified model the posterior (left) and joint distribution (right) are:

$$[\beta_0, \beta_1, \sigma_\epsilon^2 | y_j] \propto [y_j | \beta_0, \beta_1, \sigma_\epsilon^2][\beta_0][\beta_1][\sigma_\epsilon^2]$$

Priors were:

$$[\beta_0] = Normal(\beta_0 | 0, 10^2)$$

$$[\beta_1] = Normal(\beta_1 | 0, 10^2)$$

$$[\sigma_\epsilon^2] = uniform(\sigma_\epsilon^2 | 0, 20)$$

The second model was fit to 500 simulated datasets using JAGS. For each data set only one chain was used with 5,000 iterations, the first 500 iterations were discarded as burn in. Fewer iterations were used because this is a very simple model (only 3 parameters estimated) and mixing was still good. Lastly, the 500 parameter estimates and credible intervals were compared to the known true parameters to evaluate bias in parameter estimates. All code used in this analysis is available at [https://github.com/MartinHoldrege/bayesian\\_simulation](https://github.com/MartinHoldrege/bayesian_simulation)

## Results

The data generating process leads to a small positive treatment effect (on average) that varies from year to year (Figure 1).

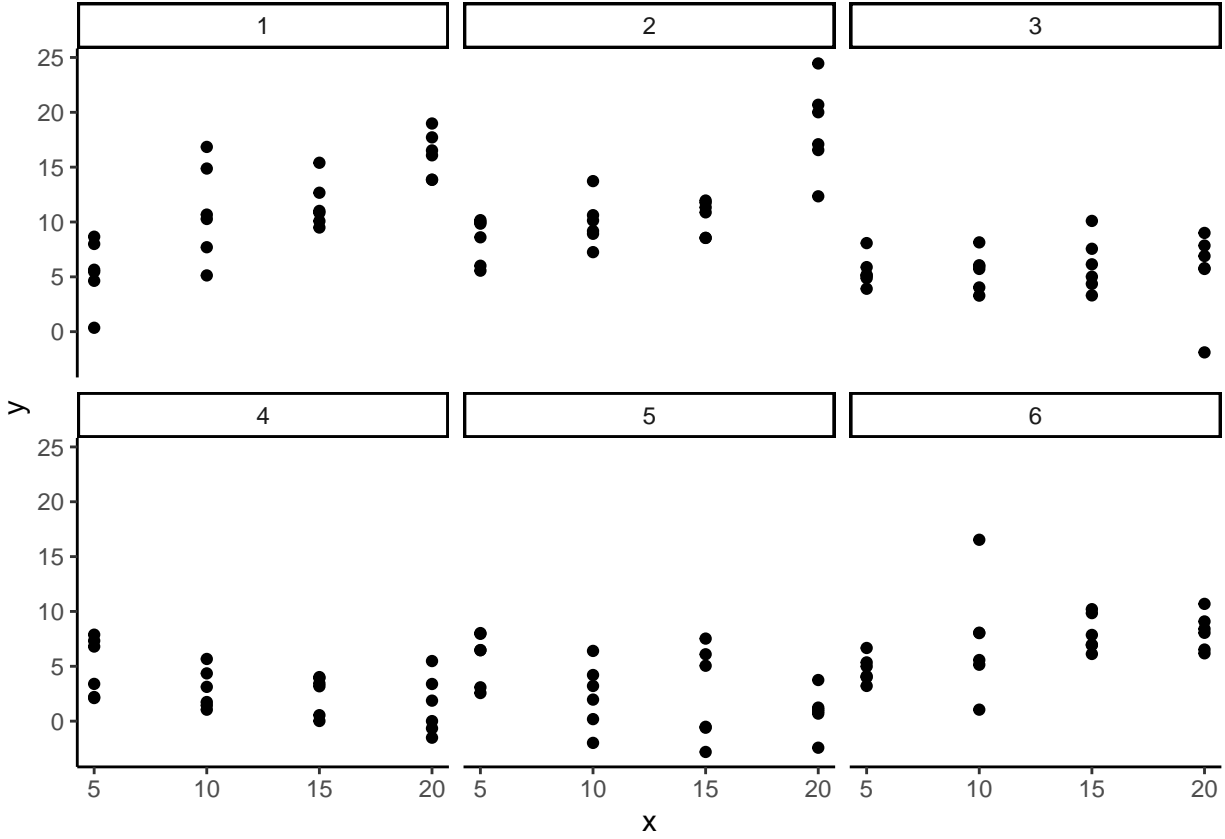


Figure 1: One realization of the data simulation. Each panel is a separate year

### Manually coded random effects model

When using the metropolis sampler to fit the model, the credible intervals mostly contained the true parameter values (Figure 2). However, not enough simulations were run to estimate whether 95% of credible intervals actually contained the true parameter values. However, note that the mean estimates of  $\sigma_{\beta_0}$  and  $\sigma_{\beta_1}$  were consistently over-estimates (Figure 2).

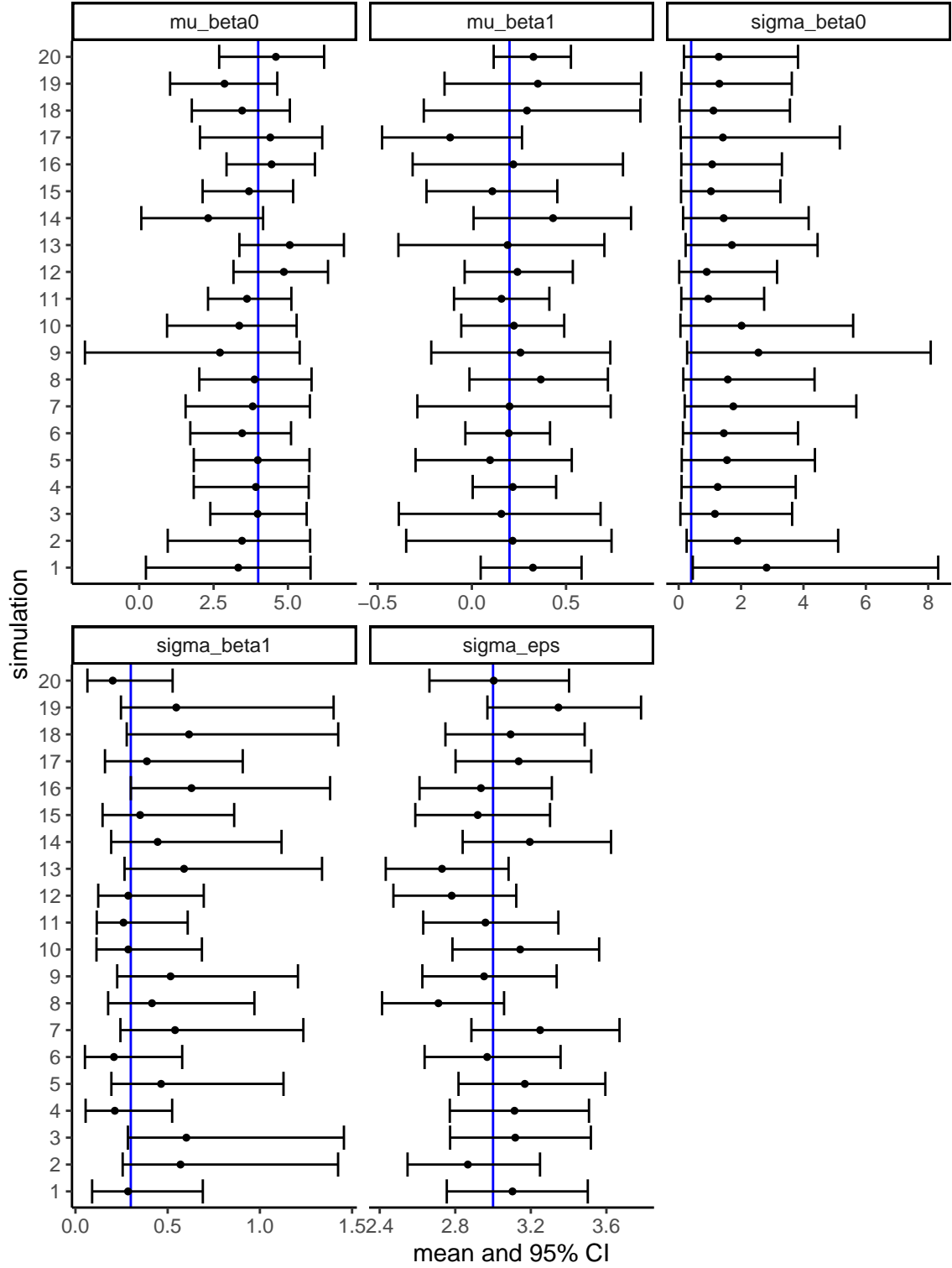


Figure 2: Mean and credible interval of intervals of parameters for models fit to 20 simulated data sets. True population values shown with blue vertical line.

## Random effects model fit with JAGS

For each of the 500 simulations the mean of the posteriors was extracted (Figure 3). The estimates of  $\mu_{\beta_0}$  and  $\mu_{\beta_1}$ , are very good (no bias in the means). However the estimates (means) of  $\sigma_{\beta_0}$  are all high, even when medians (not shown) of the posteriors of  $\sigma_{\beta_0}$  were examined they also tended to be too high. Similarly, distributions of the lower (Figure 4) and upper (Figure 5) bounds of the credible intervals indicate the  $\beta$  parameters were well estimated. The credible intervals also mostly contain the true values of the  $\sigma$  parameters. Credible intervals of  $\mu_{\beta_0}$  were overly conservative (Table 1), in that that  $\sim 99\%$  contained the true parameter value (95% would be expected), this is presumably because  $\sigma_{\beta_0}$  was over-estimated.

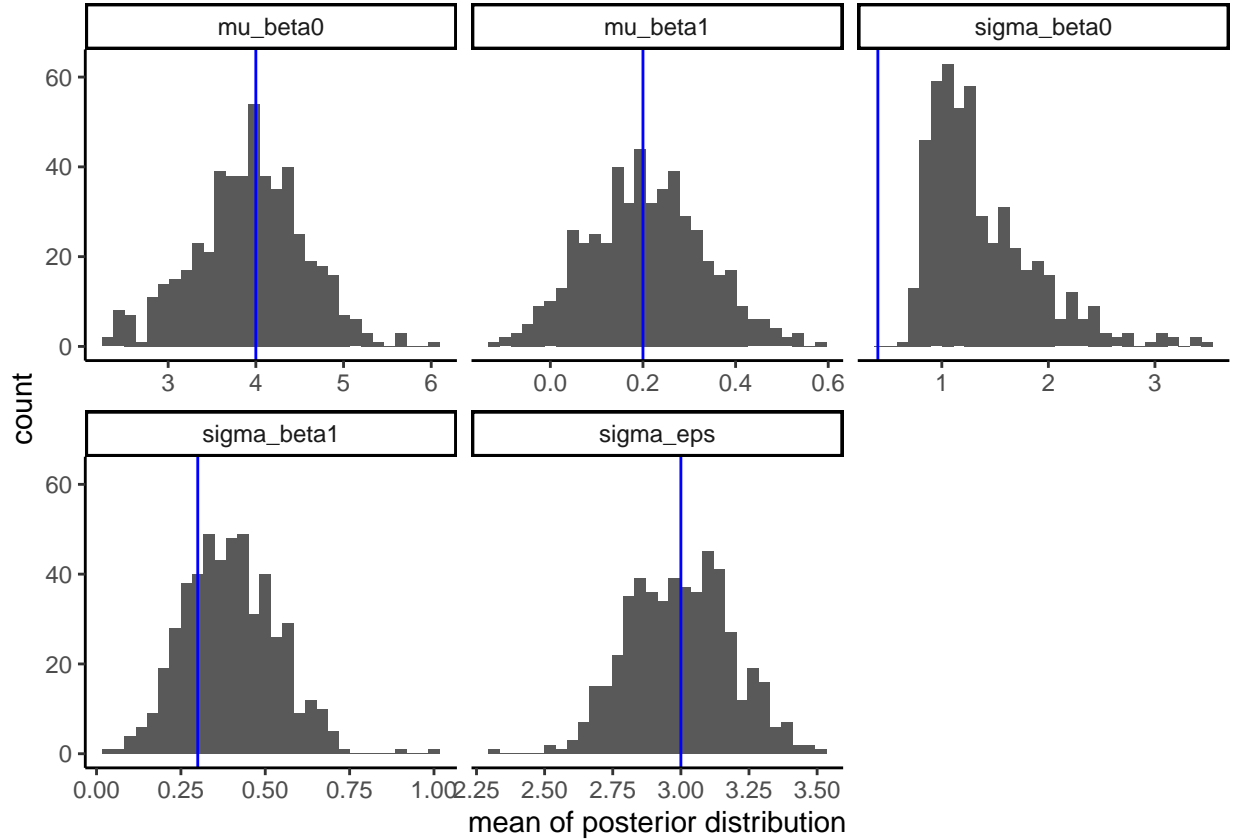


Figure 3: Mean of posterior distribution from random effects model fit with JAGS, for each of 500 simulations. Blue lines show true parameter values.

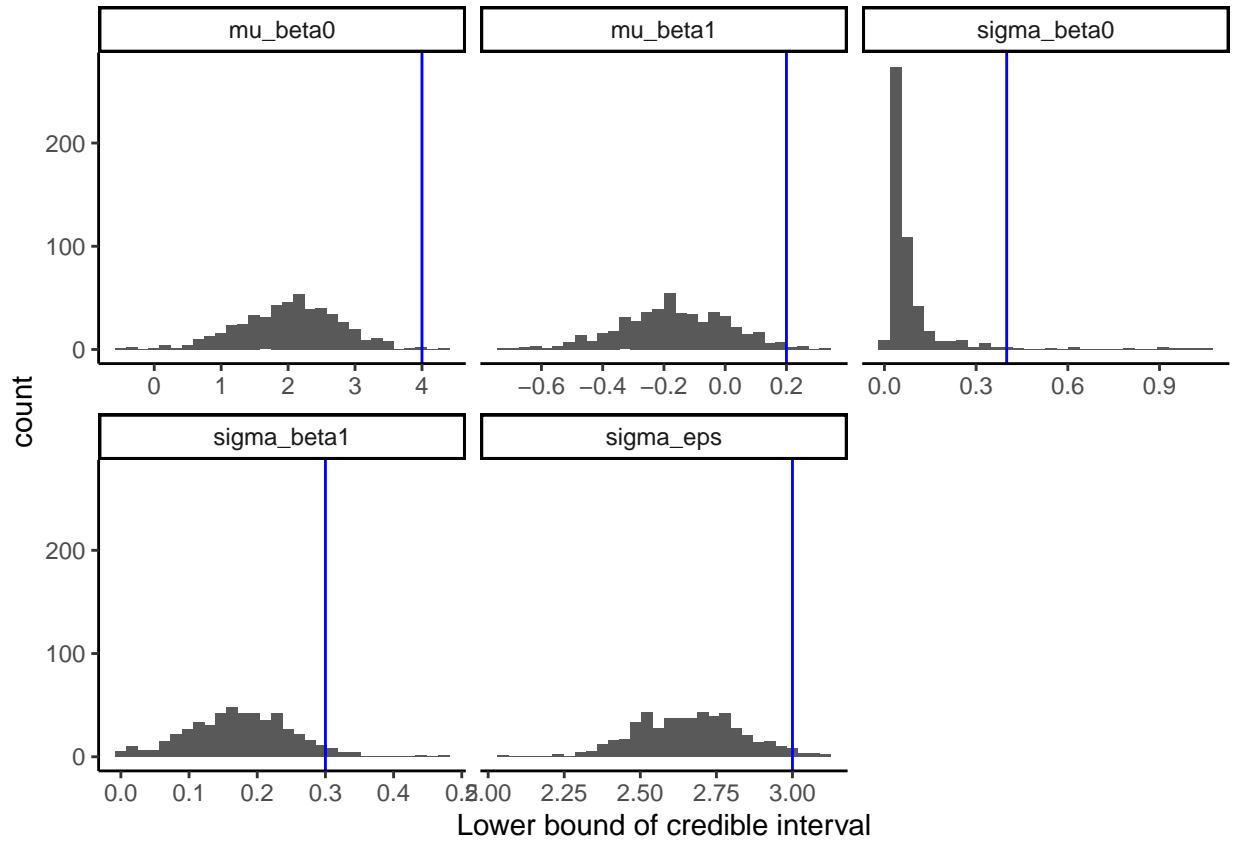


Figure 4: Lower bound of 95 percent credible intervals from random effects model fit with JAGS, for each of 500 simulations. Blue lines show true parameter values.

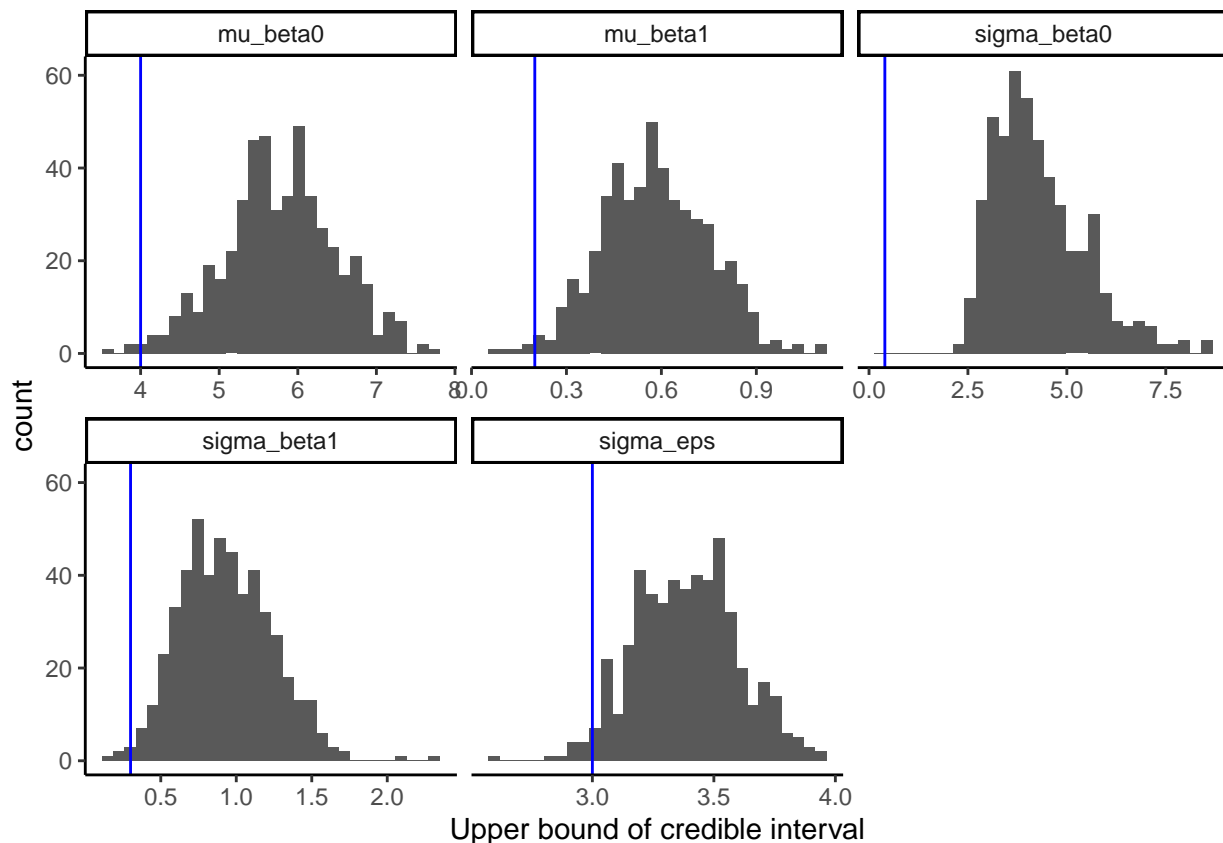


Figure 5: Upper bound of 95 percent credible intervals from random effects model fit with JAGS, for each of 500 simulations. Blue lines show true parameter values.

Table 1: Table 1. Percent of credible intervals that contain the true parameter value. Based on hierarchical models fit with JAGS to 500 simulated datasets.

parameter	% Correct
$\mu_{\beta_0}$	99.2 %
$\mu_{\beta_1}$	97.6 %
$\sigma_{\beta_0}$	97.6 %
$\sigma_{\beta_1}$	95 %
$\sigma_{\epsilon}$	95.8 %

## Incorrect model fit with JAGS

The ‘incorrect’ model (i.e. just a single slope/intercept) still estimated the intercept  $\mu_{\beta_0}$  well (Figures 6, 7, 8). Estimates of  $\mu_{\beta_1}$  don’t appear biased (Figure 5), but the credible intervals are too narrow (Figures 7 & 8), and only 69.8 % of the credible intervals contained the true parameter (Table 2). The error term estimated by the model is substantially higher than the true  $\sigma_{\epsilon}$  (Figures 6, 7, 8).



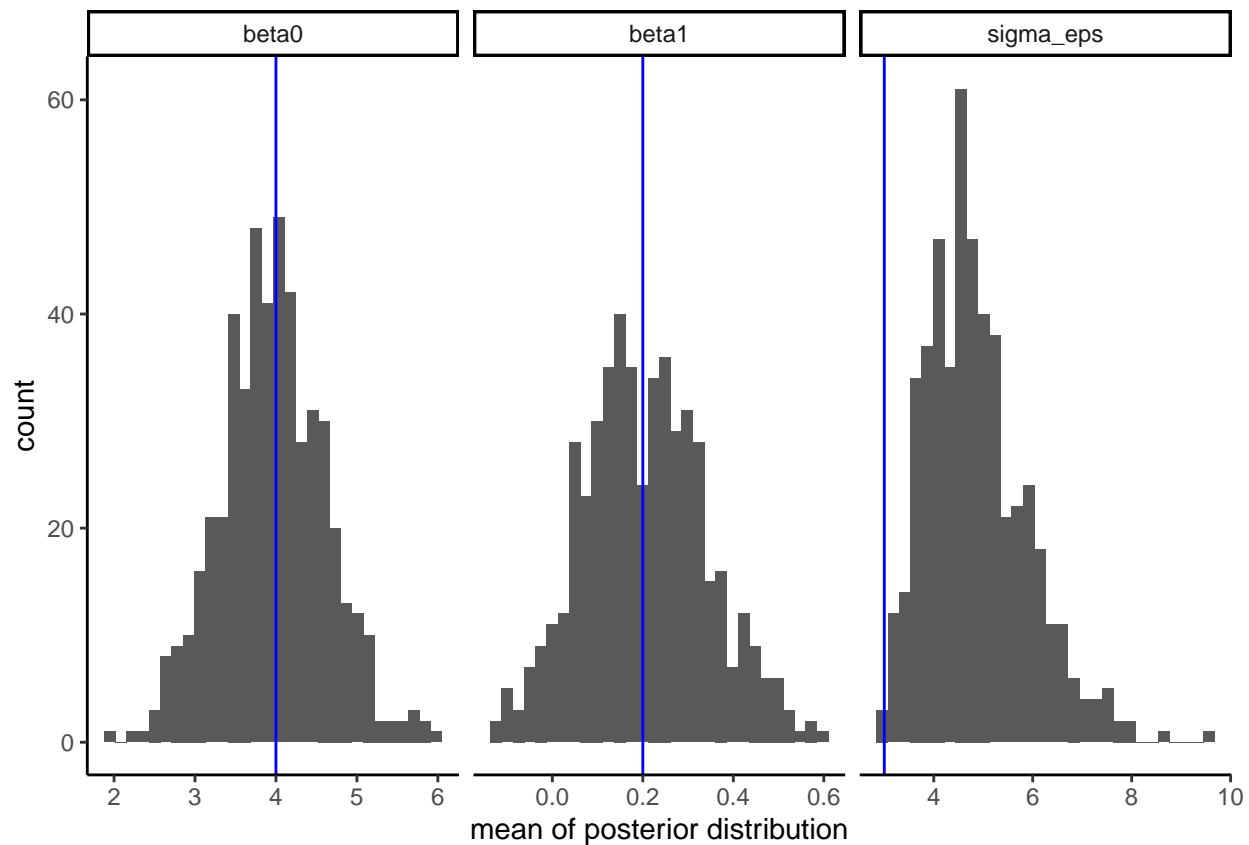


Figure 6: Mean of posterior distribution from the incorrect model, for each of 500 simulations. Blue lines show true parameter values.

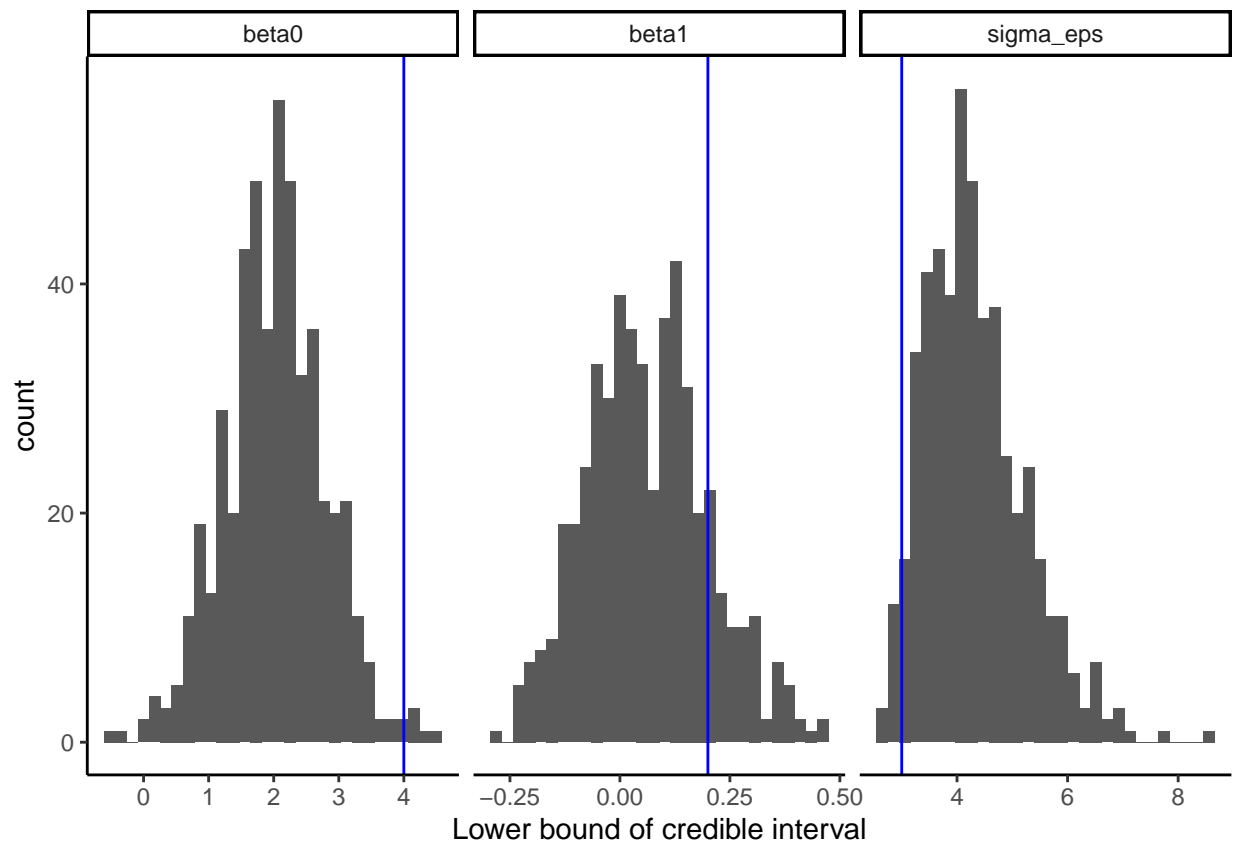


Figure 7: Lower bound of 95 percent credible intervals from the incorrect model, for each of 500 simulations. Blue lines show true parameter values.

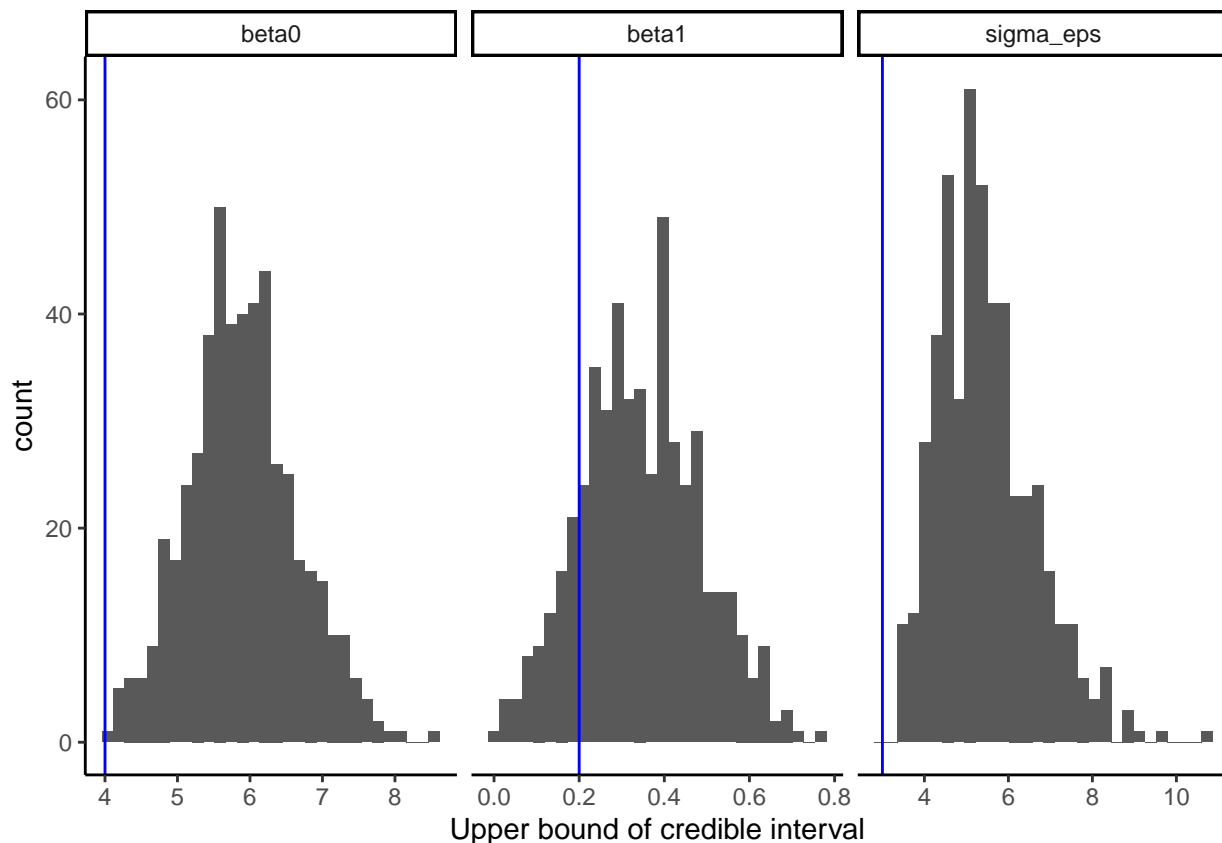


Figure 8: Upper bound of 95 percent credible intervals from the incorrect model, for each of 500 simulations. Blue lines show true parameter values.

Table 2: Table 1. Percent of credible intervals that contain the true parameter value. Based on ‘incorrect’ linear models fit with JAGS to 500 simulated datasets.

parameter	% Correct
beta0	98.8 %
beta1	69.8 %
sigma_eps	3.6 %

## Discussion

The hierarchical models fit the data quite well. It is encouraging that results from the manually coded MCMC and JAGS are similar. One issue with the model fit, both when building a metropolis sampler and when JAGS was used, is that  $\sigma_{\beta_0}$  was over-estimated. I’m unsure why this is the case, the effect of this problem is that credible intervals of estimates of  $\mu_{\beta_0}$  were overly conservative. Note, that while the point estimates of  $\sigma_{\beta_0}$  were biased, the 95% credible intervals contained the true value at about the expected rate (Table 1).

Hierarchical models are often used when there are sub-groups, and observations within a sub-group are more similar to each other. On an intuitive level I think the issue with this is that estimates of uncertainty

can be too low because the data doesn't have as many completely independent data points as the model 'thinks'. In this simulation the problem of miss-specifying the model by not taking year to year variability into account was clearly evident. Credible intervals of the slope parameter were too narrow (although not seemingly biased in one direction). Therefore, ignoring the true hierarchical nature of the data leads to over-confident estimations (and potentially a greater number of 'significant' findings than is truly warranted). The error term was over-estimated by the miss-specified model. However, that isn't terribly interesting or surprising because it had to soak up all variability that the hierarchical model ascribed to the random slope and intercept.

A more thorough simulation of hypothetical experimental data should also simulate the fact that repeated measurements of the same plot over years are likely correlated. A model taking this into account would likely require a random intercept that allowed for a different intercept to be fit for each plot.