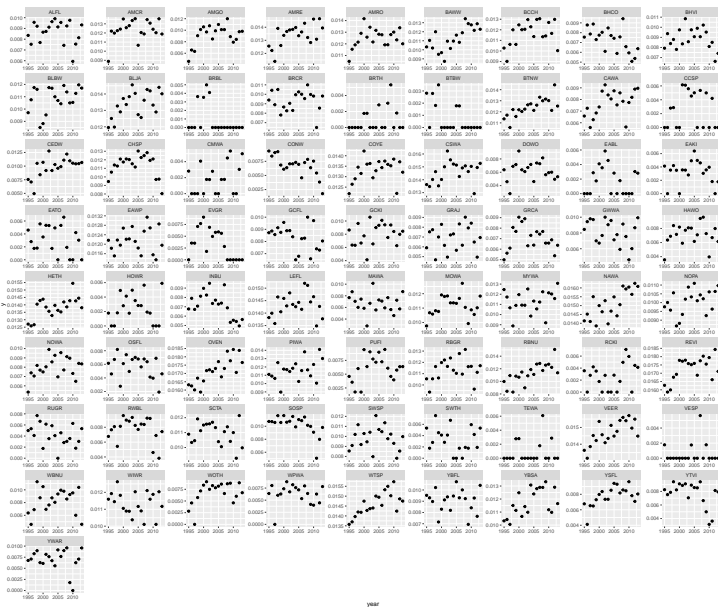# Hierarchical linear models (cont.)
Random intercept, random slope

Dr. Jarad Niemi

STAT 544 - Iowa State University
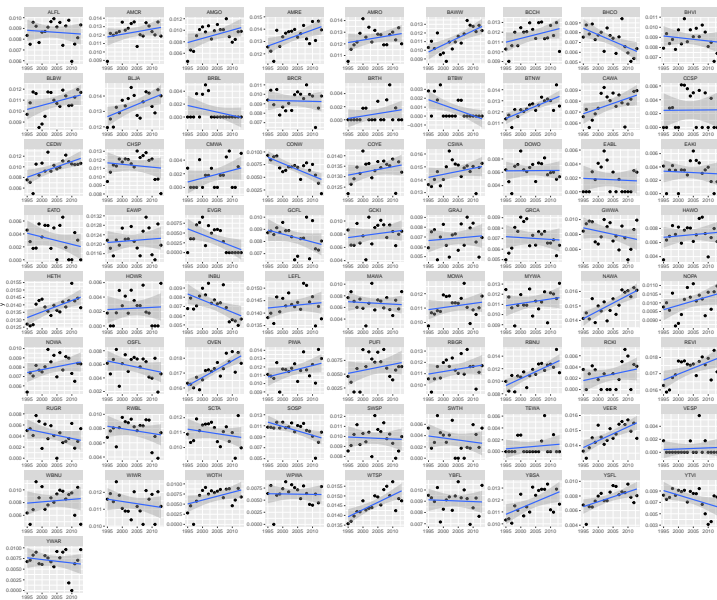
April 25, 2016

year

Independent regressions

Initially, we could consider the model

$$y_{st} \stackrel{ind}{\sim} N(\beta_{s,0} + x_{st}\beta_{s,1}, \sigma_s^2)$$

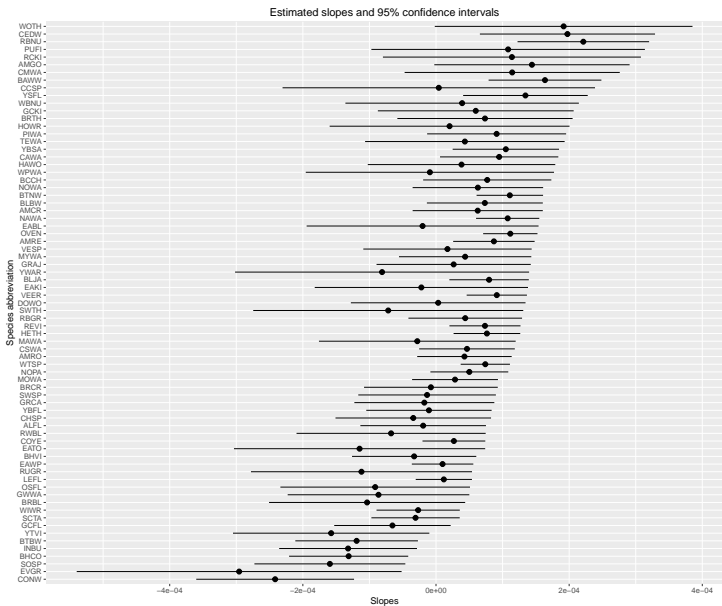where

- y_{st} is the mean log count $(+1)$ for species s at time t
- x_{st} is the year (minus 2005) for species s at time t

This model treats each species completely independently.

year

Estimated slopes and 95% confidence intervals

Random intercept, random slope model

A reasonable assumption is to treat these species exchangeably and put a distribution on the intercept and slope.

Then a random intercept, random slope model is

$$
\begin{aligned}
y_{st} &\stackrel{ind}{\sim} N(\beta_{s,0} + x_{st}\beta_{s,1}, \sigma^2) \\
\beta_s &\stackrel{ind}{\sim} N(\mu_\beta, \Sigma_\beta)
\end{aligned}
$$

where $\beta_s = (\beta_{s,0}, \beta_{s,1})'$ and σ^2, μ_β, and Σ_β are parameters to be estimated.

Notice that there is now a common variance for all species.

Random intercept and random slope model in R

```
m2 = lmer(y~I(year-2005) + (I(year-2005)|abbrev), d)
summary(m2)


Linear mixed model fit by REML ['lmerMod']
Formula: y ~ I(year - 2005) + (I(year - 2005) | abbrev)
   Data: d

REML criterion at convergence: -13786.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.6888 -0.5157  0.0381  0.5437  3.6365

Random effects:
 Groups   Name          Variance  Std.Dev. Corr
 abbrev   (Intercept)   1.799e-05 4.242e-03
          I(year - 2005) 5.998e-09 7.745e-05 0.49
 Residual               2.015e-06 1.419e-03
Number of obs: 1387, groups:  abbrev, 73

Fixed effects:
               Estimate Std. Error t value
(Intercept)   8.543e-03  4.980e-04  17.156
I(year - 2005) 1.502e-05  1.143e-05   1.314

Correlation of Fixed Effects:
            (Intr)
I(yer-2005) 0.393
```
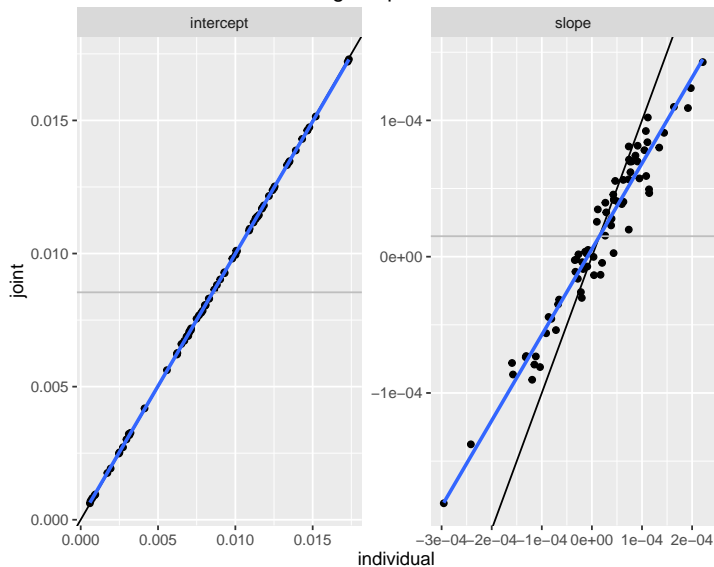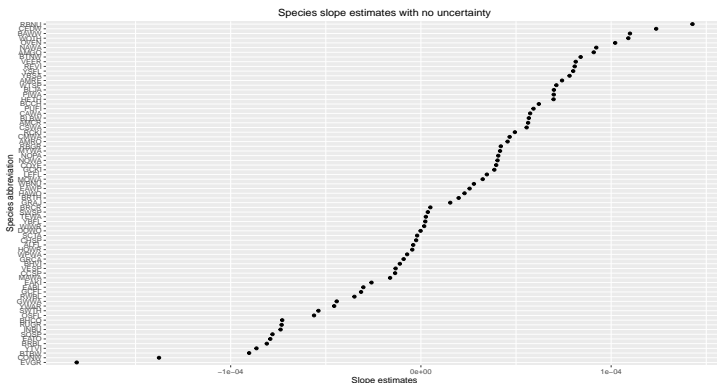
Shrinkage of point estimates

Which species have significant decline?

The quantities of interest here are $\beta_{s,1}$ and whether these quantities are negative, i.e. indicating an average decease in counts over time. But how can we calculate pvalues or confidence intervals for the random effects themselves?

Species slope estimates with no uncertainty

Bayesian random intercept, random slope model

The model

$$
\begin{aligned}
y_{st} &\stackrel{ind}{\sim} N(\beta_{s,0} + x_{st}\beta_{s,1}, \sigma^2) \\
\beta_s &\stackrel{ind}{\sim} N(\mu_\beta, \Sigma_\beta)
\end{aligned}
$$

and a prior

$$
p(\sigma, \mu_\beta, \Sigma_\beta) \propto p(\sigma)p(\mu_\beta)p(\Sigma_\beta)
$$

and

- $\sigma \sim Ca^+(0,1)$,
- $p(\mu_\beta) \propto 1$, and
- $\Sigma_\beta \sim ?$

Conjugate prior for a covariance matrix

The natural conjugate prior for a covariance matrix is the inverse-Wishart distribution, which has density

$$p(\Sigma) \propto |\Sigma|^{-(\nu+d+1)/2} \exp\left(-\frac{1}{2}\text{tr}\left(S\Sigma^{-1}\right)\right)$$

with $\nu > d - 1$ and S is a positive definite matrix. The expected value is

$$E[\Sigma] = \frac{S}{\nu - d - 1}$$

for $\nu > d + 1$. We write $\Sigma \sim IW(\nu, S^{-1})$.

Special cases:

- If $\nu = d + 1$ and S is diagonal, then each of the correlations in Σ has a marginal uniform prior.
- Jeffreys prior

$$p(\Sigma) = |\Sigma|^{-(d+1)/2}$$

Issues with the inverse-Wishart distribution

If $\Sigma \sim IW(\nu, S)$, then $\Sigma_{ii} \sim IG([\nu - (d-1)]/2, S_{ii}/2)$. In particular, if $\nu = d + 1$ and $S = \mathrm{I}$ (to ensure marginally uniform priors on the correlations), then $\Sigma_{ii} \sim IG(1, 1/2)$.

The problems

- although the correlations are marginally uniform, they are not independent *a priori* of the variances (diagonal elements of Σ),
- the inverse gamma distribution has a region near zero of extremely low density that can cause extreme bias toward larger values for truly small variances,
- this in turn causes the correlation to be shrunk toward zero.

Deconstructing the covariance matrix

Let

$$\Sigma = \text{diag}(\sigma)\Omega\text{diag}(\sigma)$$

where

- σ is a vector of standard deviations
- Ω is a correlation matrix

which results in the standard deviations and correlations being independent *a priori*.

Now we can put whatever prior we want on σ and Ω, e.g. $\sigma_i \overset{ind}{\sim} Ca^+(0, ?)$.

LKJ correlation matrix prior

The LKJ (Lewandowski, Kurowicka, and Joe 2009) distribution is

$$p(\Omega) = |\Omega|^{\eta-1}$$

where Ω is a correlation matrix with implicit dimension d and $\eta > 0$ is the shape parameter.

- if $\eta = 1$, then the density is uniform over correlation matrices of dimension d
- if $\eta > 1$, the identity matrix is the modal correlation matrix with a sharper peak in the density for larger values of η
- if $\eta < 1$, the density has a trough at the identity matrix.

```
model = "
data {
  int<lower=1> n_species;
  int<lower=1> n_years;
  vector[n_years] y[n_species];
  matrix[n_years,2] X;
}
parameters {
  real<lower=0> sigma;
  vector[2] beta[n_species];
  vector[2] mu_beta;
  vector<lower=0>[2] sigma_beta;
  corr_matrix[2] L;
}
model {
  sigma ~ cauchy(0,1);
  sigma_beta ~ cauchy(0,1);
  L ~ lkj_corr(1.0);
  beta ~ multi_normal(mu_beta, diag_matrix(sigma_beta) * L * diag_matrix(sigma_beta));
  for (s in 1:n_species) y[s] ~ normal(X*beta[s], sigma);
}
"
```

```
SAMPLING FOR MODEL 'bec085ae9301179f3728ab3b0612c7d6' NOW (CHAIN 1).

Chain 1, Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 1, Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 1, Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 1, Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 1, Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 1, Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 1, Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 1, Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 1, Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 1, Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 1, Iteration: 1800 / 2000 [ 90%]  (Sampling)
Chain 1, Iteration: 2000 / 2000 [100%]  (Sampling)#
#  Elapsed Time: 13.86 seconds (Warm-up)
#                35.45 seconds (Sampling)
#                49.31 seconds (Total)
#

SAMPLING FOR MODEL 'bec085ae9301179f3728ab3b0612c7d6' NOW (CHAIN 2).

Chain 2, Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 2, Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 2, Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 2, Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 2, Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 2, Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 2, Iteration: 1001 / 2000 [ 50%]  (Sampling)
Chain 2, Iteration: 1200 / 2000 [ 60%]  (Sampling)
Chain 2, Iteration: 1400 / 2000 [ 70%]  (Sampling)
Chain 2, Iteration: 1600 / 2000 [ 80%]  (Sampling)
Chain 2, Iteration: 1800 / 2000 [ 90%]  (Sampling)
Chain 2, Iteration: 2000 / 2000 [100%]  (Sampling)#
#  Elapsed Time: 14.89 seconds (Warm-up)
#                46.14 seconds (Sampling)
```
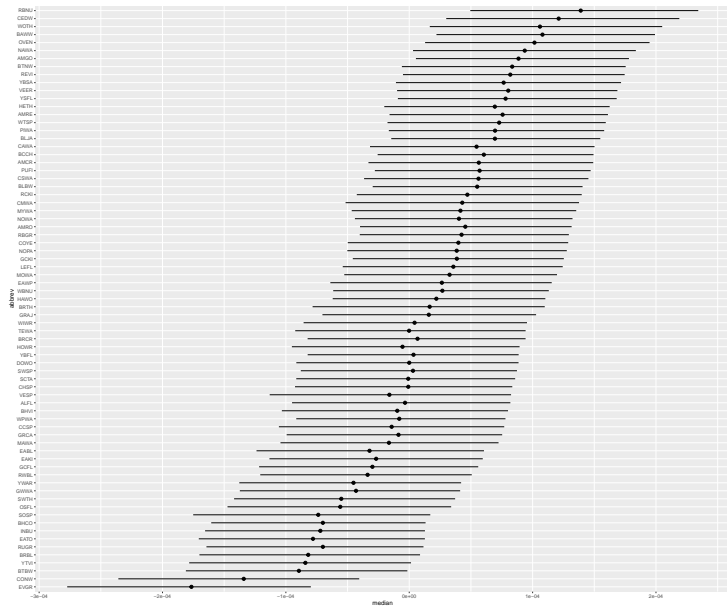
Hierarchical model for the variances

The model

$$
\begin{aligned}
y_{st} &\stackrel{ind}{\sim} N(\beta_{s,0} + x_{st}\beta_{s,1}, \sigma_s^2) \\
\beta_s &\stackrel{ind}{\sim} N(\mu_\beta, \Sigma_\beta) \\
\sigma_s &\stackrel{ind}{\sim} LN(\mu_\sigma, \tau_\sigma)
\end{aligned}
$$

and a prior

$$
p(\mu_\sigma, \tau_\sigma, \mu_\beta, \Sigma_\beta) \propto p(\mu_\sigma)p(\tau_\sigma)p(\mu_\beta)p(\Sigma_\beta)
$$

and

- $p(\mu_\sigma) \propto 1$,
- $\tau_\sigma \sim Ca^+(0,1)$,
- $p(\mu_\beta) \propto 1$, and
- Σ_β as before

```
model2 = "
data {
  int<lower=1> n_species;
  int<lower=1> n_years;
  vector[n_years] y[n_species];
  matrix[n_years,2] X;
}
parameters {
  real<lower=0> sigma[n_species];
  real mu;
  real<lower=0> tau;
  vector[2] beta[n_species];
  vector[2] mu_beta;
  vector<lower=0>[2] sigma_beta;
  corr_matrix[2] L;
}
model {
  tau ~ cauchy(0,1);
  sigma ~ lognormal(mu,tau);
  sigma_beta ~ cauchy(0,1);
  L ~ lkj_corr(1.0);
  beta ~ multi_normal(mu_beta, diag_matrix(sigma_beta) * L * diag_matrix(sigma_beta));
  for (s in 1:n_species) y[s] ~ normal(X*beta[s], sigma[s]);
}
"
```