# How lmer Works

Joint dist:

$$p(y, u; \theta, \beta, \sigma^2) \propto (\sigma^2)^{-(N+Q)/2} \exp\left\{ -\frac{1}{2\sigma^2} \left\| \begin{bmatrix} y \\ 0 \end{bmatrix} - \begin{bmatrix} Z\Lambda(\theta) & X \\ I & 0 \end{bmatrix} \begin{bmatrix} u \\ \beta \end{bmatrix} \right\|^2 \right\}$$

1. Calculate joint mode of $u, \beta$: $\tilde{u}, \tilde{\beta}$
   - 1.1 Compute decomposition of augmented design, $L, R_{ZX}, R_X$
   - 1.2 $(A^\top A)^{-1} A^\top y'$
2. Integrate $u$ - marginal likelihood
3. Profile $\beta$
4. Profile $\sigma^2$
5. Loop over deviance

# Marginal likelihood

$$p(y;\theta,\beta,\sigma^2) \propto (\sigma^2)^{-N/2}|L(\theta)|^{-1} \exp\left\{-\frac{1}{2\sigma^2}\left\|R_X(\beta-\tilde{\beta})\right\|^2\right\} \times$$
$$\exp\left\{-\frac{1}{2\sigma^2}\left\|\begin{bmatrix} y \\ 0 \end{bmatrix} - \begin{bmatrix} Z\Lambda(\theta) & X \\ I & 0 \end{bmatrix}\begin{bmatrix} \tilde{u} \\ \tilde{\beta} \end{bmatrix}\right\|^2\right\}$$

- Joint mode in $\beta$ maximizes likelihood
- Joint mode gives PWRSS, maximizes $\sigma^2$

# Priors "For Free"

Add $\beta \sim \mathrm{N}(0, \sigma^2 \Sigma_\beta)$, joint dist:

$$p(y, u; \theta, \beta, \sigma^2) \propto (\sigma^2)^{-(N+Q+P)/2} |\Sigma_\beta|^{-1/2}$$
$$\exp\left\{ -\frac{1}{2\sigma^2} \left\| \begin{bmatrix} y \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} Z\Lambda(\theta) & X \\ 0 & \Sigma_\beta^{-1/2} \\ I & 0 \end{bmatrix} \begin{bmatrix} u \\ \beta \end{bmatrix} \right\|^2 \right\}$$

Changes:

1. PWRSS includes $\tilde{\beta}^\top \Sigma_\beta^{-1} \tilde{\beta}$ (falls out of joint mode)
2. $R_X^\top R_X = X^\top X - R_{ZX}^\top R_{ZX} + \Sigma_\beta^{-1}$

# Priors "For Free"

Add $\beta \sim \mathrm{N}(0, \sigma^2 \Sigma_\beta)$, joint dist:

$$p(y, u; \theta, \beta, \sigma^2) \propto (\sigma^2)^{-(N+Q+P)/2} |\Sigma_\beta|^{-1/2}$$
$$\exp\left\{ -\frac{1}{2\sigma^2} \left\| \begin{bmatrix} y \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} Z\Lambda(\theta) & X \\ 0 & \Sigma_\beta^{-1/2} \\ I & 0 \end{bmatrix} \begin{bmatrix} u \\ \beta \end{bmatrix} \right\|^2 \right\}$$

Changes:

1. PWRSS includes $\tilde{\beta}^\top \Sigma_\beta^{-1} \tilde{\beta}$ (falls out of joint mode)
2. $R_X^\top R_X = X^\top X - R_{ZX}^\top R_{ZX} + \Sigma_\beta^{-1}$

$\beta \sim \mathrm{N}(0, \Sigma_\beta)$ much, much harder

# More "For Free"

- $\sigma^2 \sim \Gamma^{-1}$ - adjust degrees of freedom, add constant to PWRSS
- $\sigma^2 \sim \Gamma$, $\sigma \sim \Gamma^{-1}$ yield quadratic maximizer for $\sigma^2$
- $\sigma^2 = \sigma_0^2$, skip profiling step and plug in

PWRSS less useful/concrete

# $\theta$ Priors

Any prior on $\Sigma(\theta)$ can be tacked on as "penalty"

What of $\tilde{\Sigma} = \sigma^2\Sigma(\theta)$?

- Inv-Wishart yields functional form equivalent to Inv-Gamma on $\sigma^2$
- Wishart and $\tilde{\Sigma}^{1/2} \sim$ Inv-Wishart yield quadratic in $\sigma^2$ as before

# GLMMs

- $\sigma^2$ not (yet) relevant
- $\beta$ not profiled, numerically optimized

All priors easy

# Wishlist

Hooks into/pause optimization at:

- For given $\theta$, just do decomp & let me tweak $R_X$
- For given $\theta$, decomp, just do joint mode
- Less ambitiously, way to penalize deviance
- Something like an interface

For last point, can already wrap deviance

Or, just throw everything into numeric optimizer