Mixed models in R using the lme4 package
Part 7: Generalized linear mixed models

Douglas Bates

University of Wisconsin - Madison
and R Development Core Team
<Douglas.Bates@R-project.org>

University of Lausanne
July 3, 2009

Outline

Generalized Linear Mixed Models

Specific distributions and links

Data description and initial exploration

Model building

Conclusions from the example

Outline

Outline

Generalized Linear Mixed Models

Specific distributions and links

Data description and initial exploration

Model building

Conclusions from the example

Outline

Generalized Linear Mixed Models

Specific distributions and links

Data description and initial exploration

Model building

Conclusions from the example

Outline

Generalized Linear Mixed Models

Specific distributions and links

Data description and initial exploration

Model building

Conclusions from the example

Outline

Generalized Linear Mixed Models

- When using linear mixed models (LMMs) we assume that the response being modeled is on a continuous scale.
- Sometimes we can bend this assumption a bit if the response is an ordinal response with a moderate to large number of levels. For example, the Scottish secondary school test results were integer values on the scale of 1 to 10.
- However, an LMM is not suitable for modeling a binary response, an ordinal response with few levels or a response that represents a count. For these we use generalized linear mixed models (GLMMs).
- To describe GLMMs we return to the representation of the response as an n-dimensional, vector-valued, random variable, $\mathcal{Y}$, and the random effects as a q-dimensional, vector-valued, random variable, $\mathcal{B}$.

Parts of LMMs carried over to GLMMs

- Random variables
  - $\boldsymbol{\mathcal{Y}}$ the response variable
  - $\boldsymbol{\mathcal{B}}$ the (possibly correlated) random effects
  - $\boldsymbol{\mathcal{U}}$ the orthogonal random effects
- Parameters
  - $\boldsymbol{\beta}$ - fixed-effects coefficients
  - σ - the common scale parameter (not always used)
  - $\boldsymbol{\theta}$ - parameters that determine $\mathrm{Var}(\boldsymbol{\mathcal{B}}) = \sigma^2\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$
- Some matrices
  - $\boldsymbol{X}$ the $n \times p$ model matrix for $\boldsymbol{\beta}$
  - $\boldsymbol{Z}$ the $n \times q$ model matrix for $\boldsymbol{b}$
  - $\boldsymbol{P}$ fill-reducing $q \times q$ permutation (from $\boldsymbol{Z}$)
  - $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ relative covariance factor, s.t. $\mathrm{Var}(\boldsymbol{\mathcal{B}}) = \sigma^2\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$
  - $\boldsymbol{U}(\boldsymbol{\theta}) = \boldsymbol{Z}\boldsymbol{\Lambda}(\boldsymbol{\theta})$

The conditional distribution, $\mathcal{Y}|\mathcal{U}$

- For GLMMs, the marginal distribution, $\mathcal{B} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is the same as in LMMs except that σ^2 is omitted. We define $\mathcal{U} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q)$ such that $\mathcal{B} = \boldsymbol{\Lambda}(\boldsymbol{\theta})\mathcal{U}$.
- For GLMMs we retain some of the properties of the conditional distribution

  $$(\mathcal{Y}|\mathcal{U} = \boldsymbol{u}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}, \sigma^2\boldsymbol{I}\right) \text{ where } \boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}(\boldsymbol{u}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{u}$$

  Specifically
  - The distribution $\mathcal{Y}|\mathcal{U} = \boldsymbol{u}$ depends on $\boldsymbol{u}$ only through the conditional mean, $\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}(\boldsymbol{u})$.
  - Elements of $\mathcal{Y}$ are *conditionally independent*. That is, the distribution of $\mathcal{Y}|\mathcal{U} = \boldsymbol{u}$ is completely specified by the univariate, conditional distributions, $\mathcal{Y}_i|\mathcal{U}, i = 1, \ldots, n$.
  - These univariate, conditional distributions all have the same form. They differ only in their means.
- GLMMs differ from LMMs in the form of the univariate, conditional distributions and in how $\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}(\boldsymbol{u})$ depends on $\boldsymbol{u}$.

Outline

Some choices of univariate conditional distributions

- Typical choices of univariate conditional distributions are:
  - The *Bernoulli* distribution for binary (0/1) data, which has probability mass function

$$p(y|\mu) = \mu^y(1-\mu)^{1-y}, \quad 0 < \mu < 1, \quad y = 0, 1$$

  - Several independent binary responses can be represented as a *binomial* response, but only if all the Bernoulli distributions have the same mean.
  - The *Poisson* distribution for count $(0, 1, \dots)$ data, which has probability mass function

$$p(y|\mu) = e^{-\mu}\frac{\mu^y}{y!}, \quad 0 < \mu, \quad y = 0, 1, 2, \dots$$

- All of these distributions are completely specified by the conditional mean. This is different from the conditional normal (or Gaussian) distribution, which also requires the common scale parameter, σ.

The link function, g

- When the univariate conditional distributions have constraints on μ, such as $0 < \mu < 1$ (Bernoulli) or $0 < \mu$ (Poisson), we cannot define the conditional mean, $\boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{U}}}$, to be equal to the linear predictor, $\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}(\boldsymbol{\theta})\boldsymbol{u}$, which is unbounded.
- We choose an invertible, univariate *link function*, g, such that $\eta = g(\mu)$ is unconstrained. The vector-valued link function, $\boldsymbol{g}$, is defined by applying g component-wise.

$$\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\mu}) \quad \text{where} \quad \eta_i = g(\mu_i), \quad i = 1, \ldots, n$$

- We require that g be invertible so that $\mu = g^{-1}(\eta)$ is defined for $-\infty < \eta < \infty$ and is in the appropriate range ($0 < \mu < 1$ for the Bernoulli or $0 < \mu$ for the Poisson). The vector-valued inverse link, $\boldsymbol{g}^{-1}$, is defined component-wise.

"Canonical" link functions

- There are many choices of invertible scalar link functions, g, that we could use for a given set of constraints.
- For the Bernoulli and Poisson distributions, however, one link function arises naturally from the definition of the probability mass function. (The same is true for a few other, related but less frequently used, distributions, such as the gamma distribution.)
- To derive the canonical link, we consider the logarithm of the probability mass function (or, for continuous distributions, the probability density function).
- For distributions in this "exponential" family, the logarithm of the probability mass or density can be written as a sum of terms, some of which depend on the response, y, only and some of which depend on the mean, μ, only. However, only one term depends on **both** y and μ, and this term has the form $y \cdot g(\mu)$, where g is the canonical link.

The canonical link for the Bernoulli distribution

- The logarithm of the probability mass function is

$$\log(p(y|\mu)) = \log(1-\mu) + y \log\left(\frac{\mu}{1-\mu}\right), \ 0 < \mu < 1, \ y = 0, 1.$$

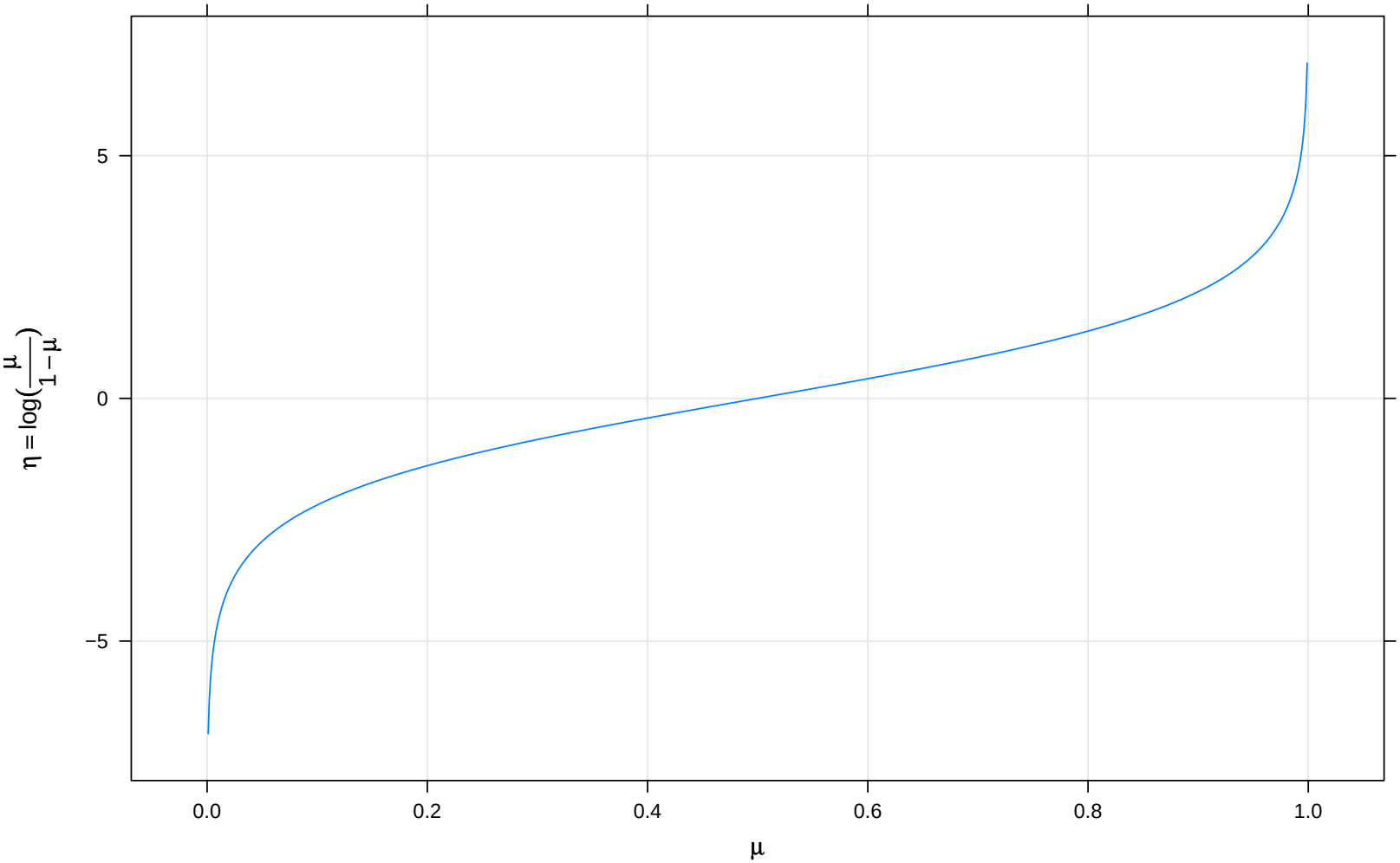
- Thus, the canonical link function is the *logit* link

$$\eta = g(\mu) = \log\left(\frac{\mu}{1-\mu}\right).$$

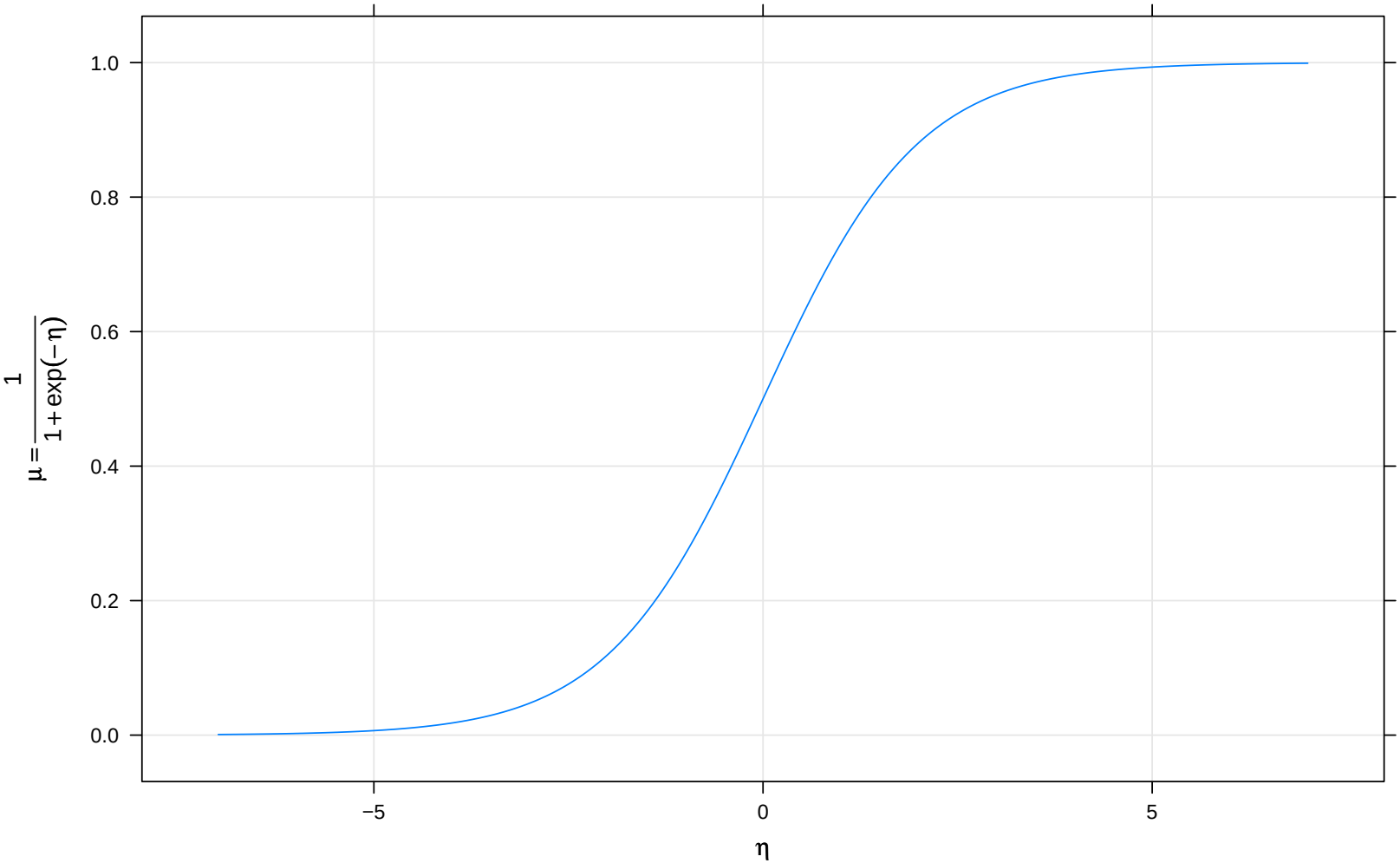
- Because $\mu = P[\mathcal{Y} = 1]$, the quantity $\mu/(1-\mu)$ is the odds ratio (in the range $(0, \infty)$) and g is the logarithm of the odds ratio, sometimes called "log odds".
- The inverse link is

$$\mu = g^{-1}(\eta) = \frac{e^{\eta}}{1+e^{\eta}} = \frac{1}{1+e^{-\eta}}$$

Plot of canonical link for the Bernoulli distribution



Plot of inverse canonical link for the Bernoulli distribution



The canonical link for the Poisson distribution

- The logarithm of the probability mass is

$$\log(p(y|\mu)) = \log(y!) - \mu + y\log(\mu)$$

- Thus, the canonical link function for the Poisson is the *log* link

$$\eta = g(\mu) = \log(\mu)$$

- The inverse link is

$$\mu = g^{-1}(\eta) = e^{\eta}$$

The canonical link related to the variance

- For the canonical link function, the derivative of its inverse is the variance of the response.
- For the Bernoulli, the canonical link is the logit and the inverse link is $\mu = g^{-1}(\eta) = 1/(1 + e^{-\eta})$. Then

$$\frac{d\mu}{d\eta} = \frac{e^{-\eta}}{(1+e^{-\eta})^2} = \frac{1}{1+e^{-\eta}} \frac{e^{-\eta}}{1+e^{-\eta}} = \mu(1-\mu) = \mathrm{Var}(\mathcal{Y})$$

- For the Poisson, the canonical link is the log and the inverse link is $\mu = g^{-1}(\eta) = e^{\eta}$. Then

$$\frac{d\mu}{d\eta} = e^{\eta} = \mu = \mathrm{Var}(\mathcal{Y})$$

The unscaled conditional density of $\mathcal{U}|\mathcal{Y}=y$

- As in LMMs we evaluate the likelihood of the parameters, given the data, as

$$L(\boldsymbol{\theta},\boldsymbol{\beta}|\boldsymbol{y}) = \int_{\mathbb{R}^q} [\mathcal{Y}|\mathcal{U}](\boldsymbol{y}|\boldsymbol{u})\,[\mathcal{U}](\boldsymbol{u})\,d\boldsymbol{u},$$

- The product $[\mathcal{Y}|\mathcal{U}](\boldsymbol{y}|\boldsymbol{u})[\mathcal{U}](\boldsymbol{u})$ is the unscaled (or *unnormalized*) density of the conditional distribution $\mathcal{U}|\mathcal{Y}$.
- The density $[\mathcal{U}](\boldsymbol{u})$ is a spherical Gaussian density $\frac{1}{(2\pi)^{q/2}}e^{-\|\boldsymbol{u}\|^2/2}$.
- The expression $[\mathcal{Y}|\mathcal{U}](\boldsymbol{y}|\boldsymbol{u})$ is the value of a probability mass function or a probability density function, depending on whether $\mathcal{Y}_i|\mathcal{U}$ is discrete or continuous.
- The linear predictor is $\boldsymbol{g}(\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}(\boldsymbol{\theta})\boldsymbol{u}$. Alternatively, we can write the conditional mean of $\mathcal{Y}$, given $\mathcal{U}$, as

$$\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}(\boldsymbol{u}) = \boldsymbol{g}^{-1}\left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}(\boldsymbol{\theta})\boldsymbol{u}\right)$$

The conditional mode of $\mathcal{U}|\mathcal{Y} = y$

- In general the likelihood, $L(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{y})$ does not have a closed form. To approximate this value, we first determine the *conditional mode*

$$\tilde{\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \arg\max_{\boldsymbol{u}} [\mathcal{Y}|\mathcal{U}](\boldsymbol{y}|\boldsymbol{u})\,[\mathcal{U}](\boldsymbol{u})$$

  using a quadratic approximation to the logarithm of the unscaled conditional density.
- This optimization problem is (relatively) easy because the quadratic approximation to the logarithm of the unscaled conditional density can be written as a penalized, weighted residual sum of squares,

$$\tilde{\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \arg\min_{\boldsymbol{u}} \left\| \begin{bmatrix} \boldsymbol{W}^{1/2}(\boldsymbol{\mu}) \left(\boldsymbol{y} - \boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}(\boldsymbol{u})\right) \\ -\boldsymbol{u} \end{bmatrix} \right\|^2$$

  where $\boldsymbol{W}(\boldsymbol{\mu})$ is the diagonal weights matrix. The weights are the inverses of the variances of the $\mathcal{Y}_i$.

The PIRLS algorithm

- Parameter estimates for generalized linear models (without random effects) are usually determined by iteratively reweighted least squares (IRLS), an incredibly efficient algorithm. PIRLS is the penalized version. It is iteratively reweighted in the sense that parameter estimates are determined for a fixed weights matrix $\boldsymbol{W}$ then the weights are updated to the current estimates and the process repeated.
- For fixed weights we solve

$$\min_{\boldsymbol{u}} \left\| \begin{bmatrix} \boldsymbol{W}^{1/2}\left(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{U}}}(\boldsymbol{u})\right) \\ -\boldsymbol{u} \end{bmatrix} \right\|^2$$

as a nonlinear least squares problem with update, $\boldsymbol{\delta}_{\boldsymbol{u}}$, given by

$$\boldsymbol{P}\left(\boldsymbol{U}\boldsymbol{M}\boldsymbol{W}\boldsymbol{M}\boldsymbol{U}'+\boldsymbol{I}\right)\boldsymbol{P}'\boldsymbol{\delta}_{\boldsymbol{u}} = \boldsymbol{U}\boldsymbol{M}\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{\mu})-\boldsymbol{u}$$

where $\boldsymbol{M} = d\boldsymbol{\mu}/d\boldsymbol{\eta}$ is the (diagonal) Jacobian matrix. Recall that for the canonical link, $\boldsymbol{M} = \mathrm{Var}(\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{U}}) = \boldsymbol{W}^{-1}$.

The Laplace approximation to the deviance

- At convergence, the sparse Cholesky factor, $\boldsymbol{L}$, used to evaluate the update is

$$\boldsymbol{L}\boldsymbol{L}' = \boldsymbol{P}\left(\boldsymbol{U}\boldsymbol{M}\boldsymbol{W}\boldsymbol{M}\boldsymbol{U}' + \boldsymbol{I}\right)\boldsymbol{P}'$$

  or

$$\boldsymbol{L}\boldsymbol{L}' = \boldsymbol{P}\left(\boldsymbol{U}\boldsymbol{M}\boldsymbol{U}' + \boldsymbol{I}\right)\boldsymbol{P}'$$

  if we are using the canonical link.
- The integrand of the likelihood is approximately a constant times the density of the $\mathcal{N}(\tilde{\boldsymbol{u}}, \boldsymbol{L}\boldsymbol{L}')$ distribution.
- On the deviance scale (negative twice the log-likelihood) this corresponds to

$$d(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{y}) = d_g(\boldsymbol{y}, \boldsymbol{\mu}(\tilde{\boldsymbol{u}})) + \|\tilde{\boldsymbol{u}}\|^2 + \log(|\boldsymbol{L}|^2)$$

  where $d_g(\boldsymbol{y}, \boldsymbol{\mu}(\tilde{\boldsymbol{u}}))$ is the GLM deviance for $\boldsymbol{y}$ and $\boldsymbol{\mu}$.

Modifications to the algorithm

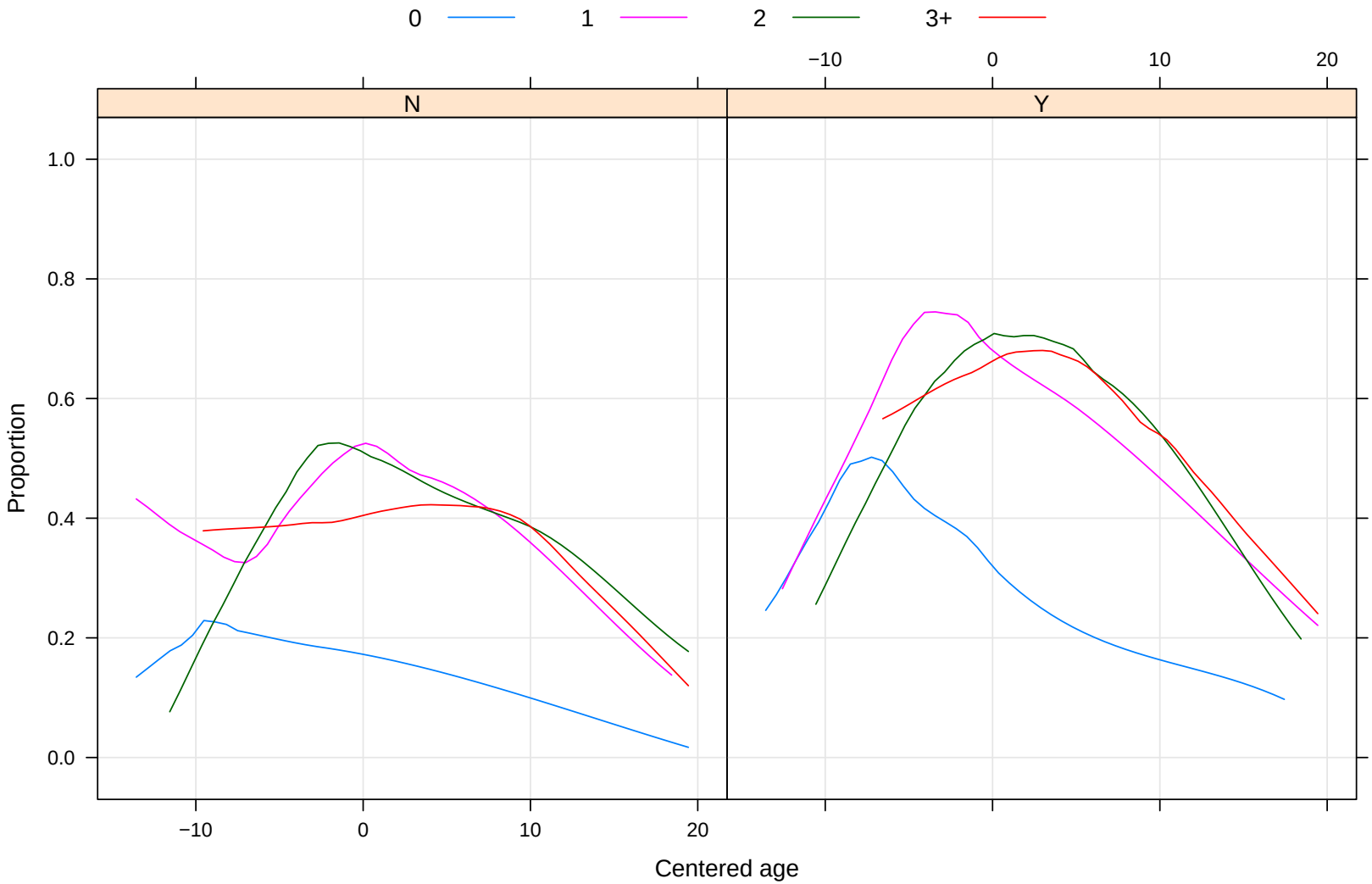
- Notice that this deviance depends on the fixed-effects parameters, $\boldsymbol{\beta}$, as well as the variance-component parameters, $\boldsymbol{\theta}$. This is because $\log(|\boldsymbol{L}|^2)$ depends on $\boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{U}}}$ and, hence, on $\boldsymbol{\beta}$. For LMMs $\log(|\boldsymbol{L}|^2)$ depends only on $\boldsymbol{\theta}$.
- It is likely that modifying the PIRLS algorithm to optimize simultaneously on $\boldsymbol{u}$ and $\boldsymbol{\beta}$ would result in a value that is very close to the deviance profiled over $\boldsymbol{\beta}$.
- Another approach, which is being implemented as a Google Summer of Code project, is adaptive Gauss-Hermite quadrature (AGQ). This has a similar structure to the Laplace approximation but is based on more evaluations of the unscaled conditional density near the conditional modes. It is only appropriate for models in which the random effects are associated with only one grouping factor

Outline

Contraception data

- One of the data sets in the `"mlmRev"` package, derived from data files available on the multilevel modelling web site, is from a fertility survey of women in Bangladesh.
- One of the responses is whether or not the woman currently uses artificial contraception (i.e. a binary response)
- Covariates included the woman's age (on a centered scale), the number of live children she had, whether she lived in an urban or rural setting, and the district in which she lived.
- Instead of plotting such data as points, we use the 0/1 response to generate scatterplot smoother curves versus age for the different groups.

Contraception use versus age by urban and livch



Comments on the data plot

- These observational data are unbalanced (some districts have only 2 observations, some have nearly 120). They are not longitudinal (no "time" variable).
- Binary responses have low per-observation information content (exactly one bit per observation). Districts with few observations will not contribute strongly to estimates of random effects.
- Within-district plots will be too imprecise so we only examine the global effects in plots.
- The comparisons on the multilevel modelling site are for fits of a model that is linear in `age`, which is clearly inappropriate.
- The form of the curves suggests at least a quadratic in `age`.
- The urban versus rural differences may be additive.
- It appears that the `livch` factor could be dichotomized into "0" versus "1 or more".

Outline

Preliminary model using Laplacian approximation

```
Generalized linear mixed model fit by the Laplace approximation
Formula: use ~ age + I(age^2) + urban + livch + (1 | district)
   Data: Contraception
  AIC  BIC logLik deviance
 2389 2433  -1186     2373
Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.22586  0.47524
Number of obs: 1934, groups: district, 60
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0350725  0.1743606  -5.936 2.91e-09
age          0.0035327  0.0092311   0.383    0.702
I(age^2)    -0.0045623  0.0007252  -6.291 3.15e-10
urbanY       0.6972694  0.1198788   5.816 6.01e-09
livch1       0.8150439  0.1621898   5.025 5.03e-07
livch2       0.9165123  0.1850995   4.951 7.37e-07
livch3+      0.9150213  0.1857689   4.926 8.41e-07
```

Comments on the model fit

- This model was fit using the Laplacian approximation to the deviance.
- There is a highly significant quadratic term in `age`.
- The linear term in `age` is not significant but we retain it because the `age` scale has been centered at an arbitrary (and unknown) value.
- The `urban` factor is highly significant (as indicated by the plot).
- Levels of `livch` greater than 0 are significantly different from 0 but may not be different from each other.

Outline

Reduced model with dichotomized livch

```
Generalized linear mixed model fit by the Laplace approximation
Formula: use ~ age + I(age^2) + urban + ch + (1 | district)
   Data: Contraception
  AIC  BIC logLik deviance
 2385 2419  -1187     2373
Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.22470  0.47402
Number of obs: 1934, groups: district, 60
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0064262  0.1678949  -5.994 2.04e-09
age          0.0062563  0.0078404   0.798    0.425
I(age^2)    -0.0046354  0.0007163  -6.471 9.73e-11
urbanY       0.6929504  0.1196687   5.791 7.01e-09
chY          0.8603757  0.1473539   5.839 5.26e-09
```

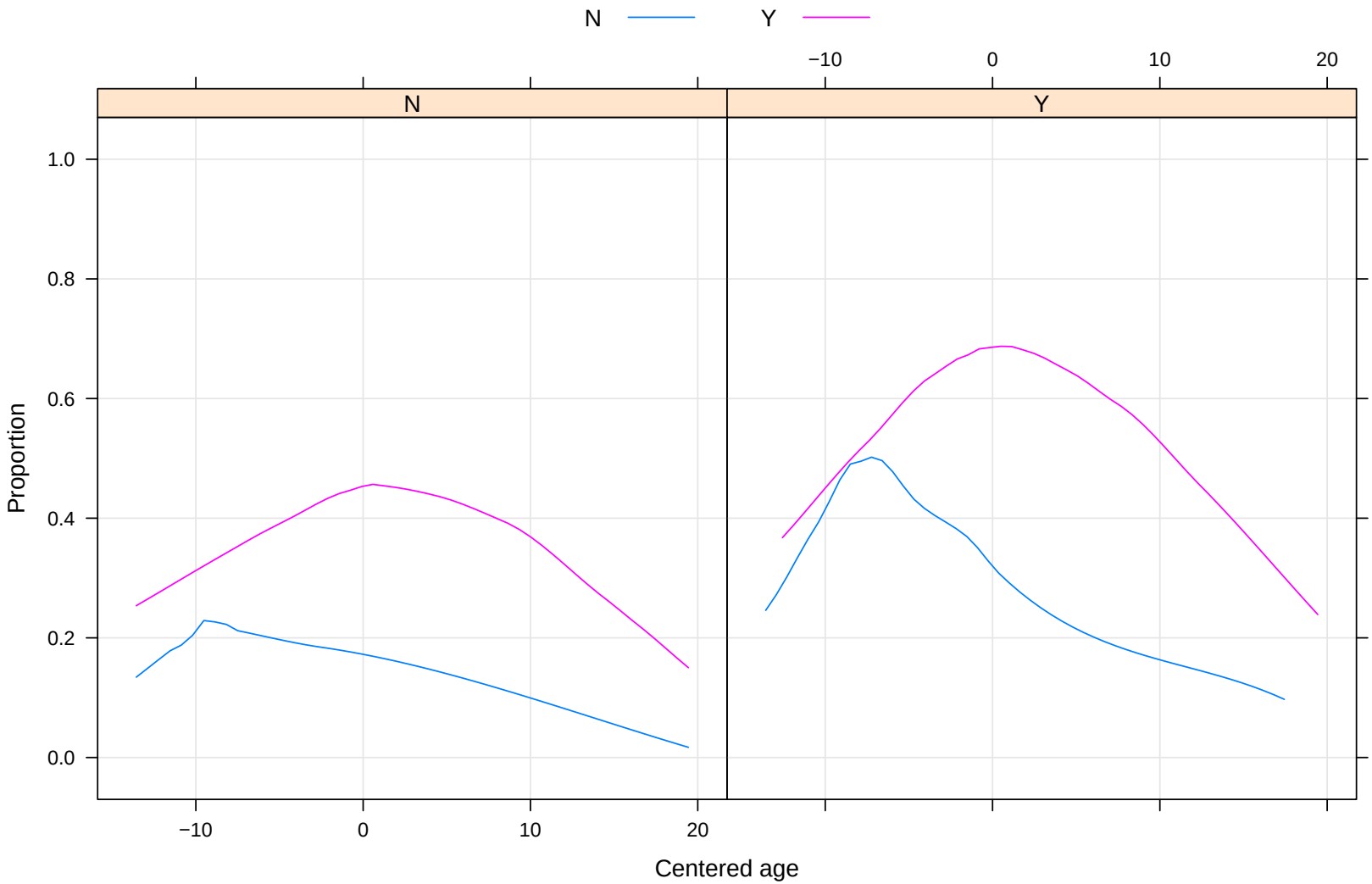
Comparing the model fits

- A likelihood ratio test can be used to compare these nested models.

```
> anova(cm2, cm1)

Data: Contraception
Models:
cm2: use ~ age + I(age^2) + urban + ch + (1 | district)
cm1: use ~ age + I(age^2) + urban + livch + (1 | district)
    Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
cm2  6 2385.2 2418.6 -1186.6
cm1  8 2388.7 2433.3 -1186.4 0.4571      2     0.7957
```

- The large p-value indicates that we would not reject `cm2` in favor of `cm1` hence we prefer the more parsimonious `cm2`.
- The plot of the scatterplot smoothers according to live children or none indicates that there may be a difference in the age pattern between these two groups.

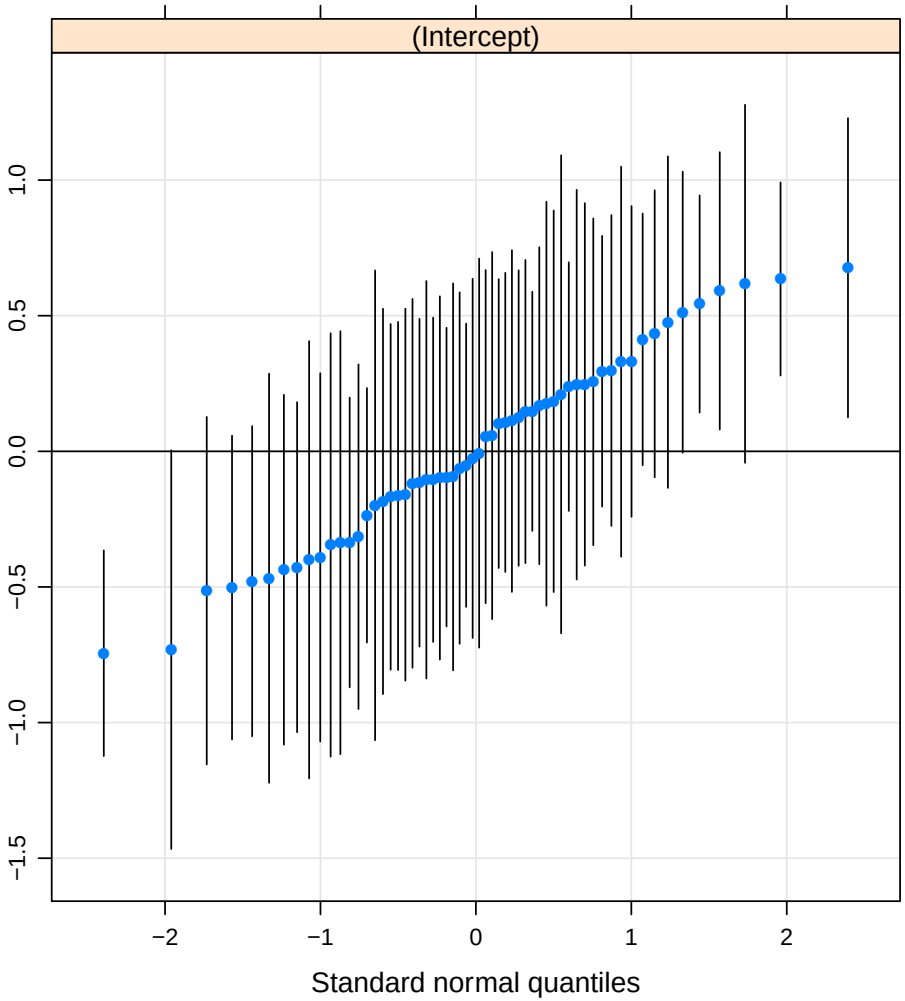
Contraception use versus age by urban and ch



Allowing age pattern to vary with ch

```
Generalized linear mixed model fit by the Laplace approximation
Formula: use ~ age * ch + I(age^2) + urban + (1 | district)
   Data: Contraception
  AIC  BIC logLik deviance
 2379 2418  -1183     2365
Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.22306  0.4723
Number of obs: 1934, groups: district, 60
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.3233178  0.2144470  -6.171 6.79e-10
age         -0.0472956  0.0218394  -2.166   0.0303
chY          1.2107859  0.2069938   5.849 4.93e-09
I(age^2)    -0.0057572  0.0008358  -6.888 5.64e-12
urbanY       0.7140327  0.1202579   5.938 2.89e-09
age:chY      0.0683522  0.0254347   2.687   0.0072
```

Prediction intervals on the random effects



Extending the random effects

- We may want to consider allowing a random effect for urban/rural by district. This is complicated by the fact the many districts only have rural women in the study

	district															
urban	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
N	54	20	0	19	37	58	18	35	20	13	21	23	16	17	14	18
Y	63	0	2	11	2	7	0	2	3	0	0	6	8	101	8	2

	district															
urban	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
N	24	33	22	15	10	20	15	14	49	13	39	45	25	45	27	24

Including a random effect for urban by district

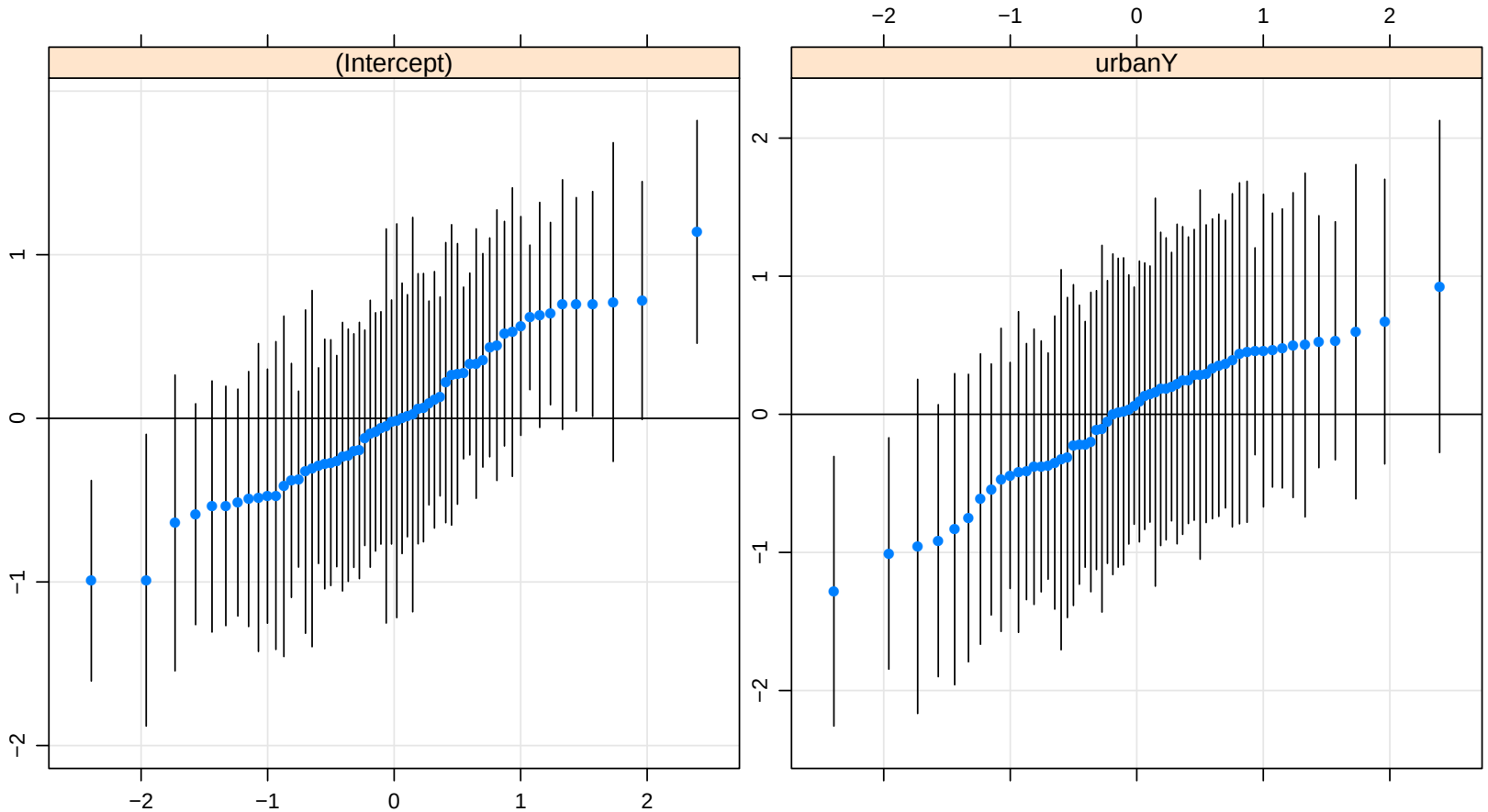
```
Generalized linear mixed model fit by the Laplace approximation
Formula: use ~ age * ch + I(age^2) + urban + (urban | district)
   Data: Contraception
  AIC  BIC logLik deviance
 2372 2422  -1177     2354
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 district (Intercept) 0.37830  0.61506
          urbanY      0.52613  0.72535  -0.793
Number of obs: 1934, groups: district, 60
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.3442631  0.2227667  -6.034 1.60e-09
age         -0.0461836  0.0219446  -2.105  0.03533
chY          1.2116527  0.2082372   5.819 5.93e-09
I(age^2)    -0.0056514  0.0008431  -6.703 2.04e-11
urbanY       0.7902096  0.1600484   4.937 7.92e-07
age:chY      0.0664682  0.0255674   2.600  0.00933

Correlation of Fixed Effects:
         (Intr) age    chY    I(g^2) urbanY
age      0.696
```

Significance of the additional random effect

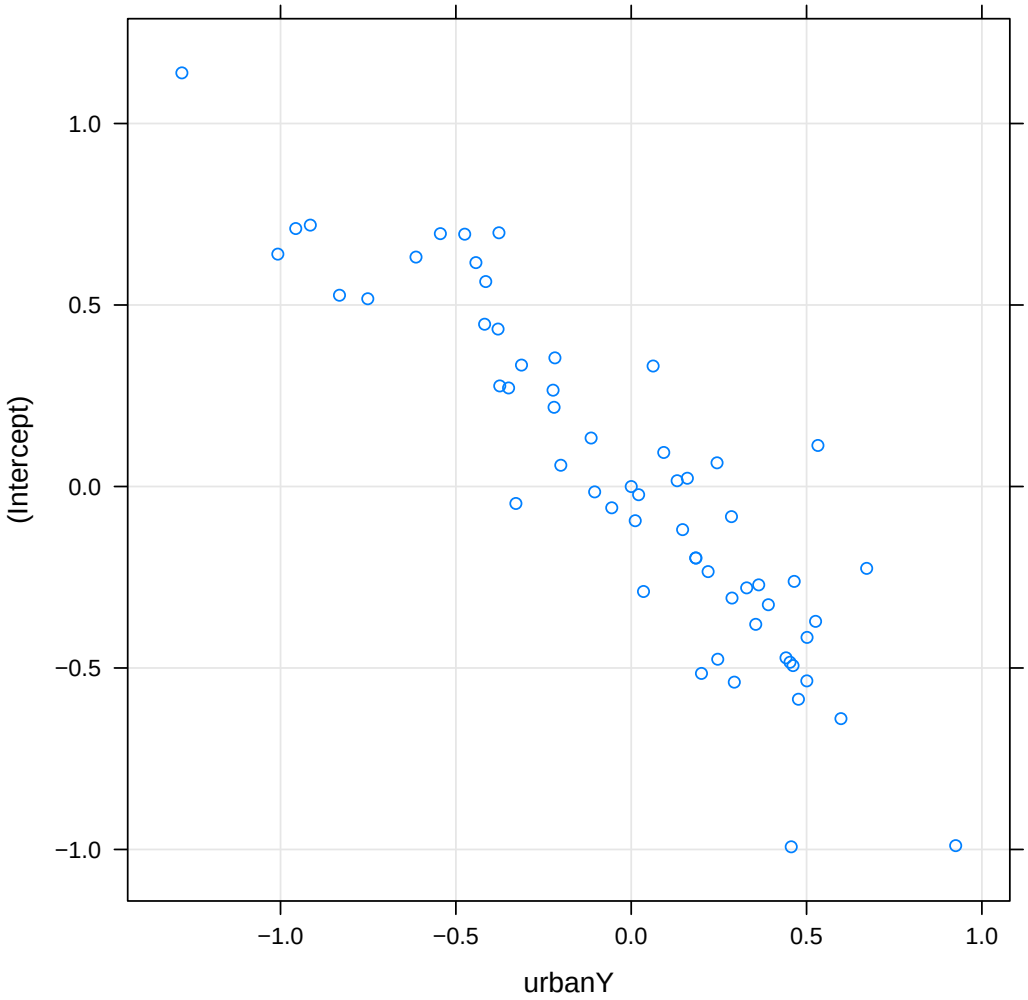
```
> anova(cm4, cm3)

Data: Contraception
Models:
cm3: use ~ age * ch + I(age^2) + urban + (1 | district)
cm4: use ~ age * ch + I(age^2) + urban + (urban | district)
    Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
cm3  7 2379.2 2418.2 -1182.6
cm4  9 2371.5 2421.6 -1176.8 11.651      2   0.002951
```

- The additional random effect is highly significant in this test.
- Most of the prediction intervals still overlap zero.
- A scatterplot of the random effects shows several random effects vectors falling along a straight line. These are the districts with all rural women or all urban women.

Prediction intervals for the bivariate random effects



Scatter plot of the BLUPs



Outline

Conclusions from the example

- Again, carefully plotting the data is enormously helpful in formulating the model.
- Observational data tend to be unbalanced and have many more covariates than data from a designed experiment. Formulating a model is typically more difficult than in a designed experiment.
- A generalized linear model is fit by adding a value, typically `binomial` or `poisson`, for the optional argument `family` in the call to `lmer`.
- MCMC sampling is not provided for GLMMs at present but will be added.
- We use likelihood-ratio tests and z-tests in the model building.