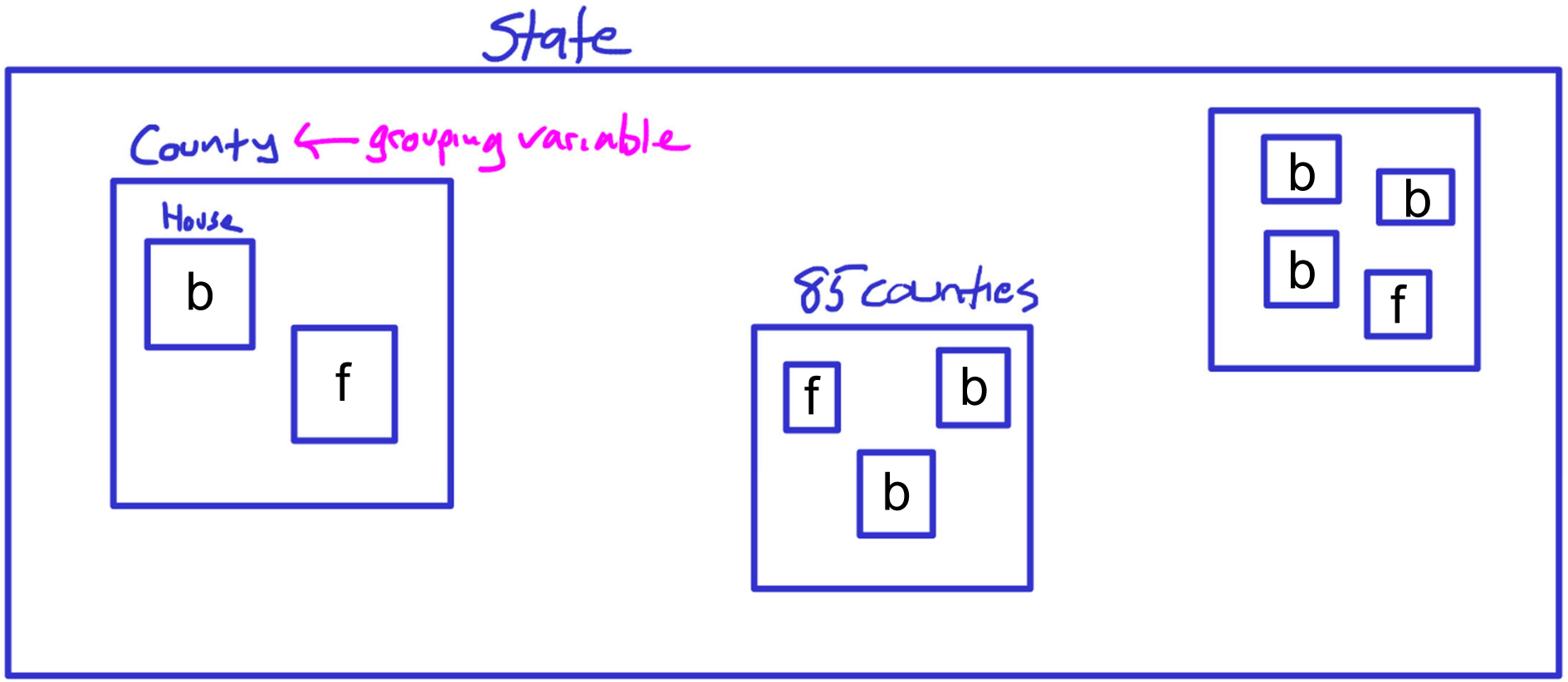
Today

- Multilevel models: radon case study
  - Model 2: 1 predictor at house scale
  - Model 3: 2 predictors at different scales

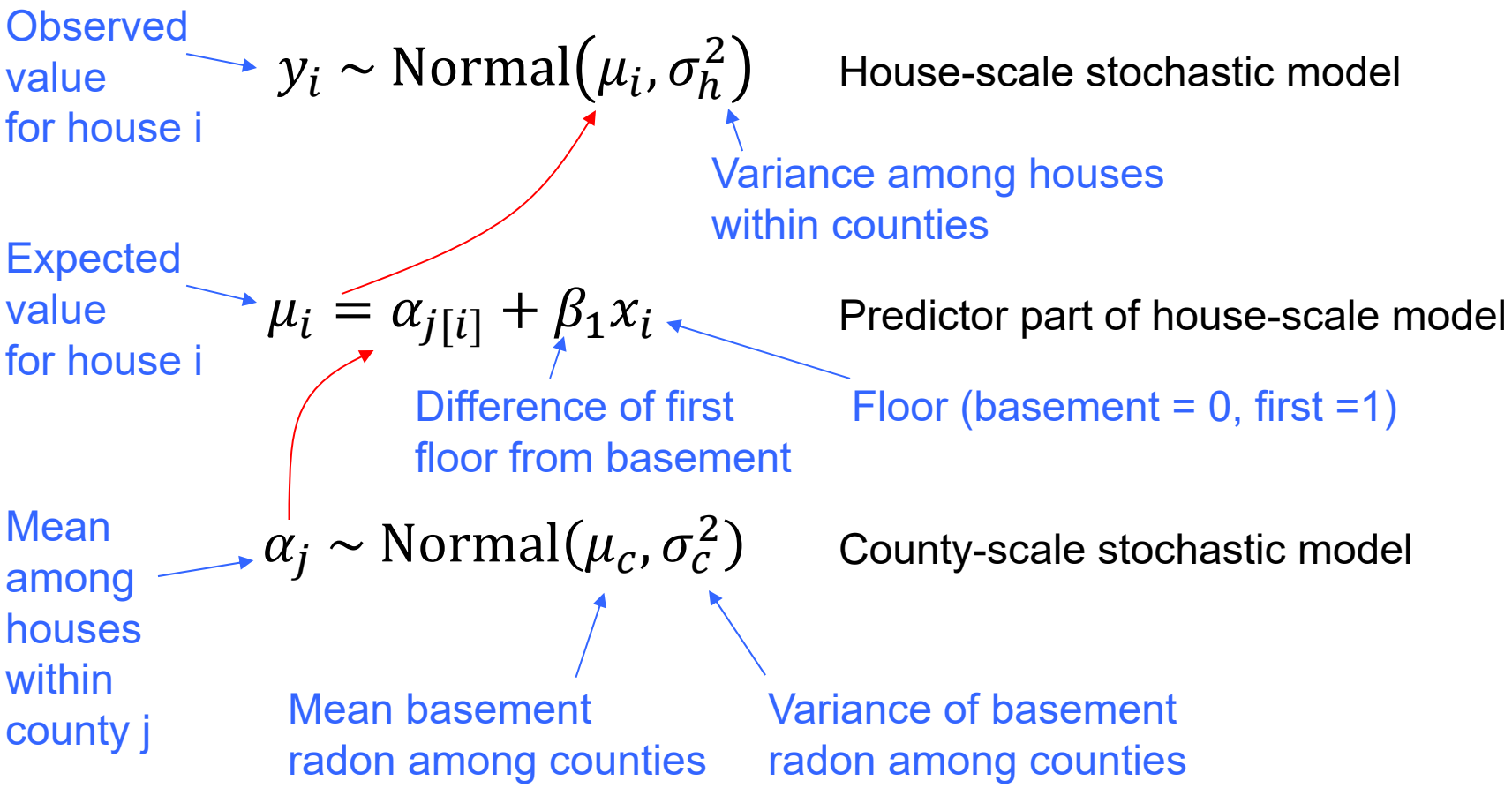
Model 2: 1 predictor



Predictor (fixed effect) is at house scale

Writing model 2

Multilevel model, with 1 predictor at house scale



Linear model syntax

```
log_radon ~ floor + (1|county)
```

Equivalent:

```
log_radon ~ 1 + floor + (1|county)
```

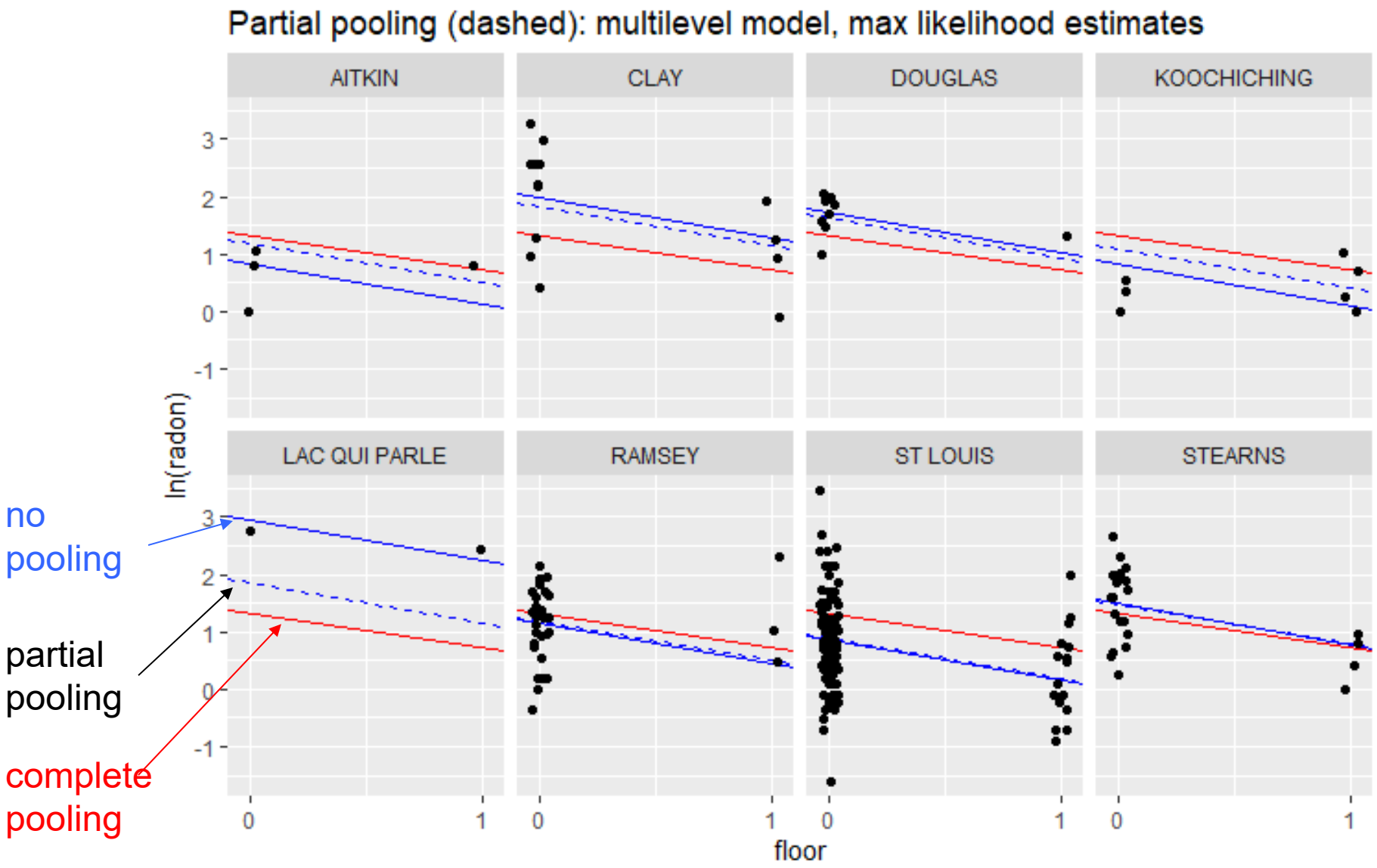
Model with one house-scale predictor

See radon_multilevel_M2.R
radon_multilevel_M2.md

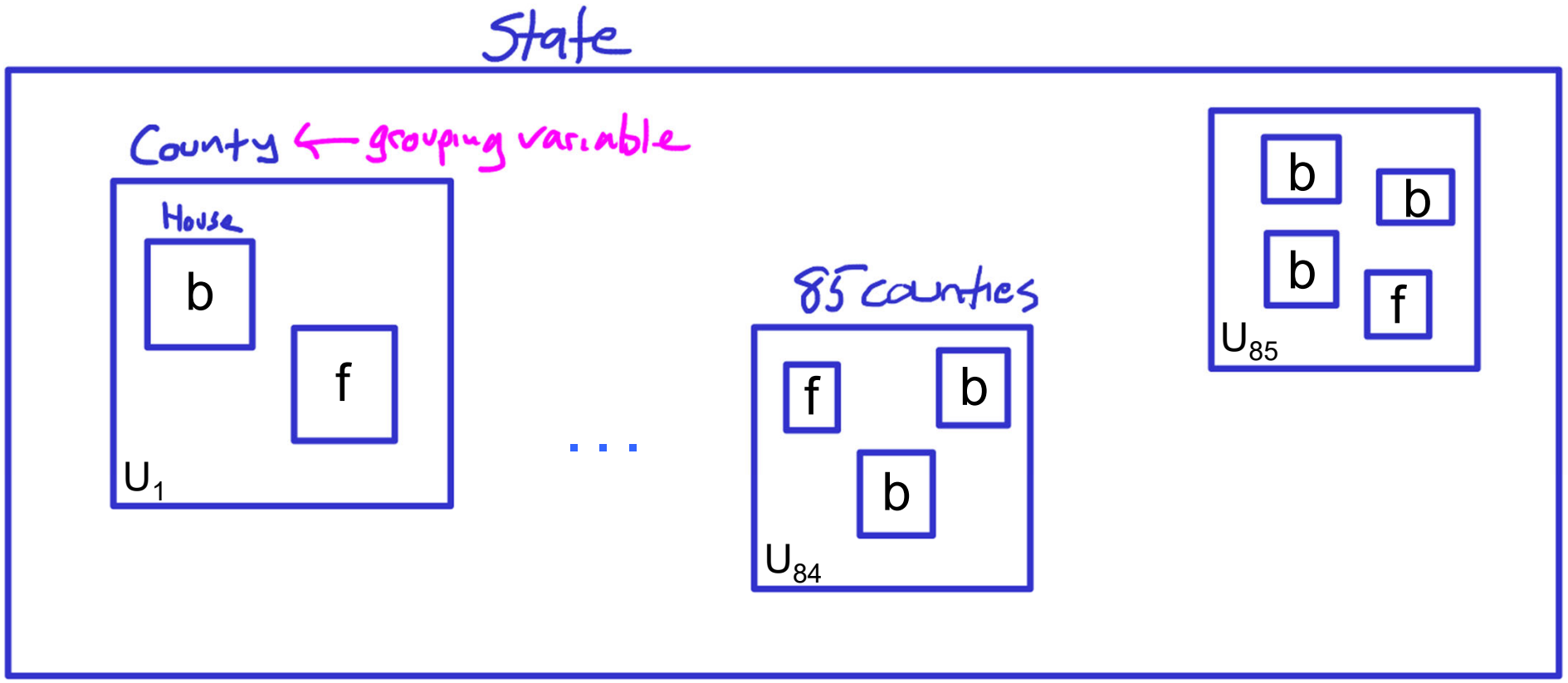
```
#` Examine the parameter estimates:
summary(ppfit)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: log_radon ~ floor_x + (1 | county)
##    Data: radon_dat
##
##      AIC      BIC   logLik deviance df.resid
##   2171.7   2190.9  -1081.8   2163.7      915
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4071 -0.6164  0.0056  0.6398  3.4288
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  county   (Intercept) 0.1053   0.3245
##  Residual             0.5703   0.7552
## Number of obs: 919, groups:  county, 85
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.46116    0.05124  28.516
## floor_x     -0.69264    0.07036  -9.844
##
## Correlation of Fixed Effects:
##         (Intr)
## floor_x -0.290
```

= sqrt(Variance)
i.e. this is not uncertainty of the Variance estimate


Partial pooling (dashed): multilevel model, max likelihood estimates
AITKIN
CLAY
DOUGLAS
KOOCHICHING
LAC QUI PARLE
RAMSEY
ST LOUIS
STEARNS
ln(radon)
floor
no pooling
partial pooling
complete pooling

Model 3: 2 predictors



Predictors (fixed effects):

floor: house scale, categorical (b, f)

uranium: county scale, continuous +

Writing model 2

Multilevel model, with 1 predictor at house scale

$$y_i \sim \text{Normal}(\mu_i, \sigma_h^2)$$ House-scale stochastic model

$$\mu_i = \alpha_{j[i]} + \beta_1 x_i$$ Predictor part of house-scale model

$$\alpha_j \sim \text{Normal}(\mu_c, \sigma_c^2)$$ County-scale stochastic model

Writing model 3

Multilevel model, with 2 predictors at different scales

$$y_i \sim \text{Normal}(\mu_i, \sigma_h^2)$$ House-scale stochastic model

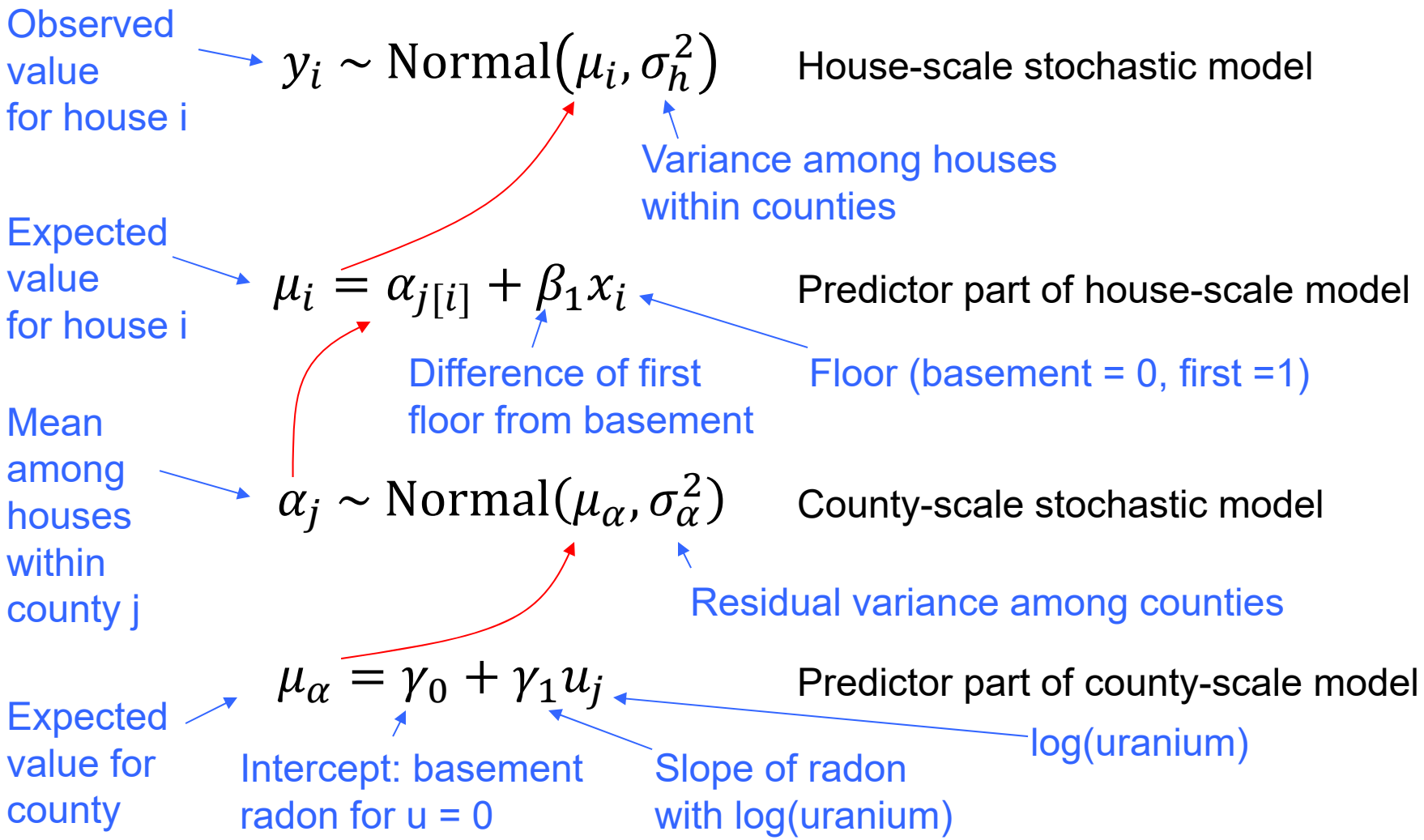
$$\mu_i = \alpha_{j[i]} + \beta_1 x_i$$ Predictor part of house-scale model

$$\alpha_j \sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2)$$ County-scale stochastic model

$$\mu_\alpha = \gamma_0 + \gamma_1 u_j$$ Predictor part of county-scale model

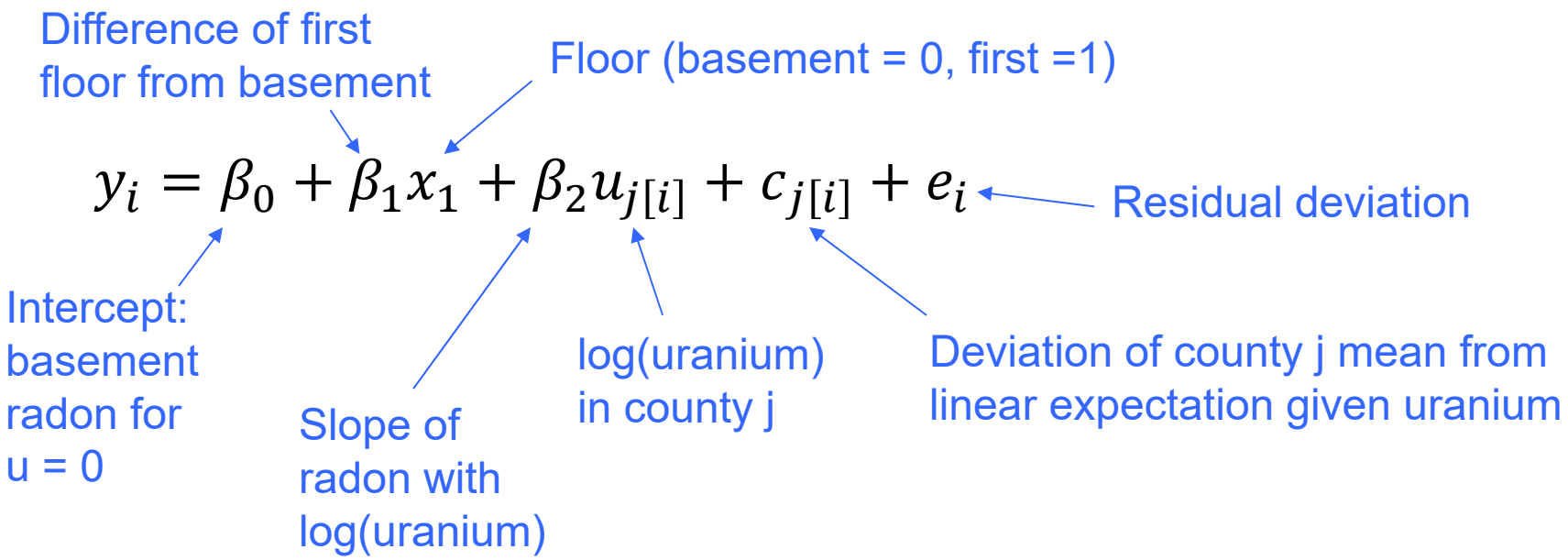
Writing model 3

Multilevel model, with 2 predictors at different scales



Alternative parameterization

Multilevel model, with 2 predictors at different levels



$c_j \sim \text{Normal}(0, \sigma_c^2)$ — Residual variance among counties within state — County-scale stochastic model

$e_i \sim \text{Normal}(0, \sigma_e^2)$ — Residual variance among houses within counties — House-scale stochastic model

Linear model syntax

```
log_radon ~ floor + logu + (1|county)
```

Model with two predictors, each at different levels

House level predictor: floor (basement, first floor)
County level predictor: uranium concentration in substrate

See radon_multilevel_M3.R
radon_multilevel_M3.md

```
summary(ppfit)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: log_radon ~ floor_x + logu + (1 | county)
##    Data: radon_dat
##
##      AIC      BIC   logLik deviance df.resid
##   2132.8   2156.9  -1061.4   2122.8      914
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.9976 -0.6163  0.0307  0.6561  3.3794
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  county   (Intercept) 0.02127  0.1458
##  Residual             0.57499  0.7583
## Number of obs: 919, groups:  county, 85
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.46427    0.03714  39.421
## floor_x     -0.66644    0.06865  -9.708
## logu         0.72320    0.08965   8.067
##
## Correlation of Fixed Effects:
##         (Intr) flor_x
## floor_x -0.361
## logu     0.154 -0.011
```

Now a lot less because uranium accounts for it

```
ppfit_bayes <- stan_lmer(log_radon ~ floor_x + logu + (1|county), data=radon_dat)
```

```
print(summary(ppfit_bayes)[,c("mean","sd","n_eff","Rhat")], digits=3)
```

```
##                                              mean     sd n_eff  Rhat
## (Intercept)                                 1.46e+00 0.0390  2510 1.000
## floor_x                                    -6.68e-01 0.0700  5342 0.999
## logu                                        7.22e-01 0.0942  2140 1.000
## b[(Intercept) county:AITKIN]               -2.46e-02 0.1465  4445 1.001
## b[(Intercept) county:ANOKA]                 1.00e-02 0.1027  2583 1.000
## b[(Intercept) county:BECKER]                1.33e-02 0.1519  5190 0.999
## b[(Intercept) county:BELTRAMI]              1.12e-01 0.1477  3595 0.999
## b[(Intercept) county:BENTON]                7.96e-03 0.1459  5771 1.000
```

Random effects

```
## b[(Intercept) county:WATONWAN]              1.28e-01 0.1619  2822 1.000
## b[(Intercept) county:WILKIN]                2.57e-02 0.1635  5435 1.000
## b[(Intercept) county:WINONA]               -6.49e-02 0.1304  4673 0.999
## b[(Intercept) county:WRIGHT]                8.30e-02 0.1253  4844 1.000
## b[(Intercept) county:YELLOW_MEDICINE]      -4.40e-02 0.1599  4798 1.000
## sigma                                       7.59e-01 0.0183  3662 1.000
## Sigma[county:(Intercept),(Intercept)]       2.71e-02 0.0165  1062 1.002
## mean_PPD                                    1.22e+00 0.0348  4574 1.000
## log-posterior                              -1.18e+03 9.4548   787 1.004
```

σ = s.d.

Σ = variance

Two predictors
Floor (house scale)
Uranium (county scale)

Each point below
is the intercept above

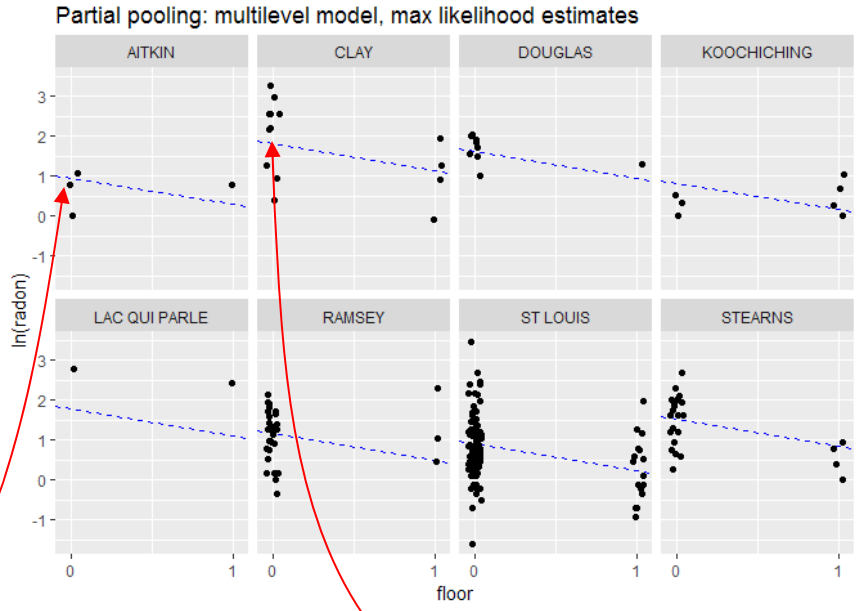


Fig. 12.5

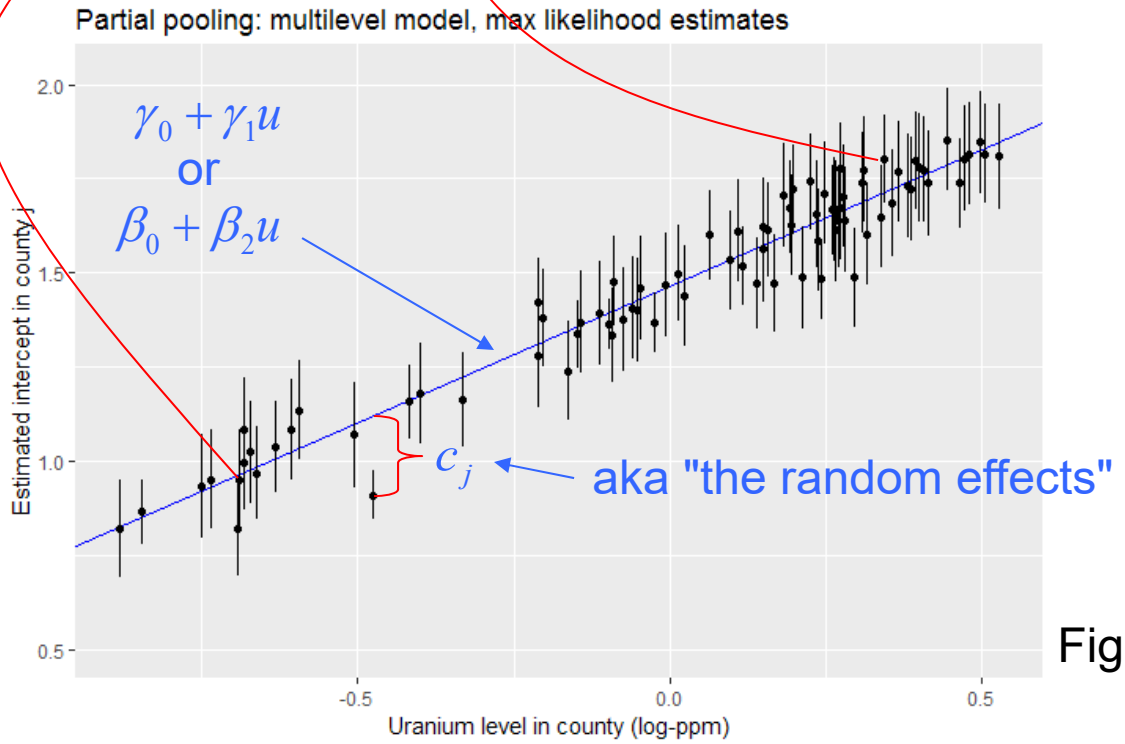
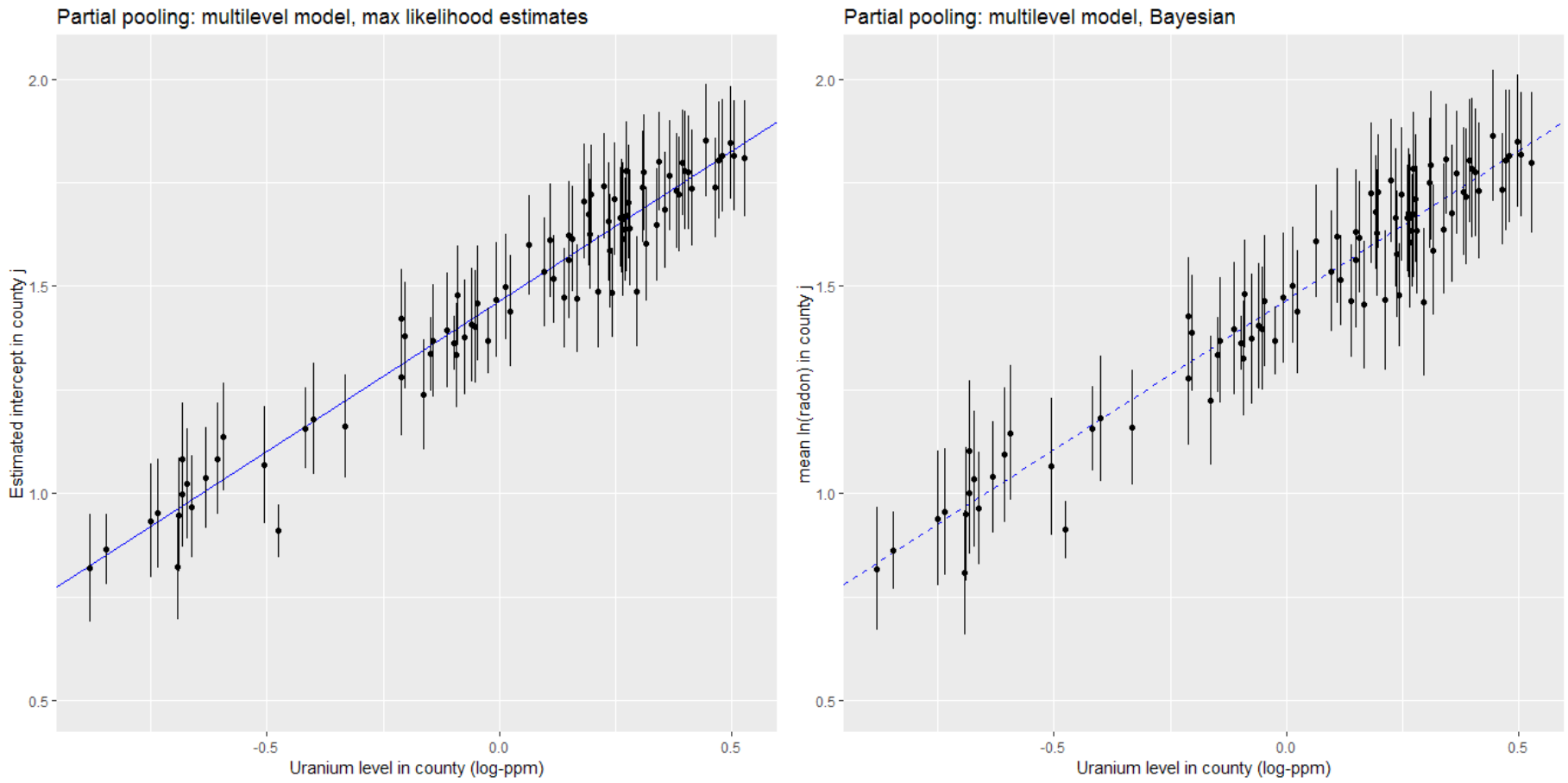


Fig. 12.6

Each point is:

$$\beta_0 + \beta_2 u_j + c_j$$

Max lik vs Bayesian



Some minor differences (e.g. Bayesian intervals a bit wider)