Multilevel Models for Applied Social Research

Instructor: Paul Lambert
Univ. Stirling

Co-facilitator: Kate O'Hara
Univ. Stirling & Univ. Edinburgh

0930-1500hrs, 18/Jun/2025

Github resources:
https://github.com/paul-lambert/SGSSS-2025/

0930-1100	Talk 1: Classical perspectives on multilevel modelling
	Talk 2: Realistic complexity
1115-1230	Lab 1: Implementing selected popular multilevel models Talk 3: Case study on effect scores from random effects residuals
1330-1500	Talk 4: MLMs meet econometrics Lab 2: Responding to complex data and to complex analytical options

Intro on Talk 2: Realistic complexity

- *'Realistic complexity' refers to the compelling ideal of parsimony of outputs in context of an appropriately ambitious analysis*

(2a) Parameters for random intercepts and slopes

(2b) Higher level residuals

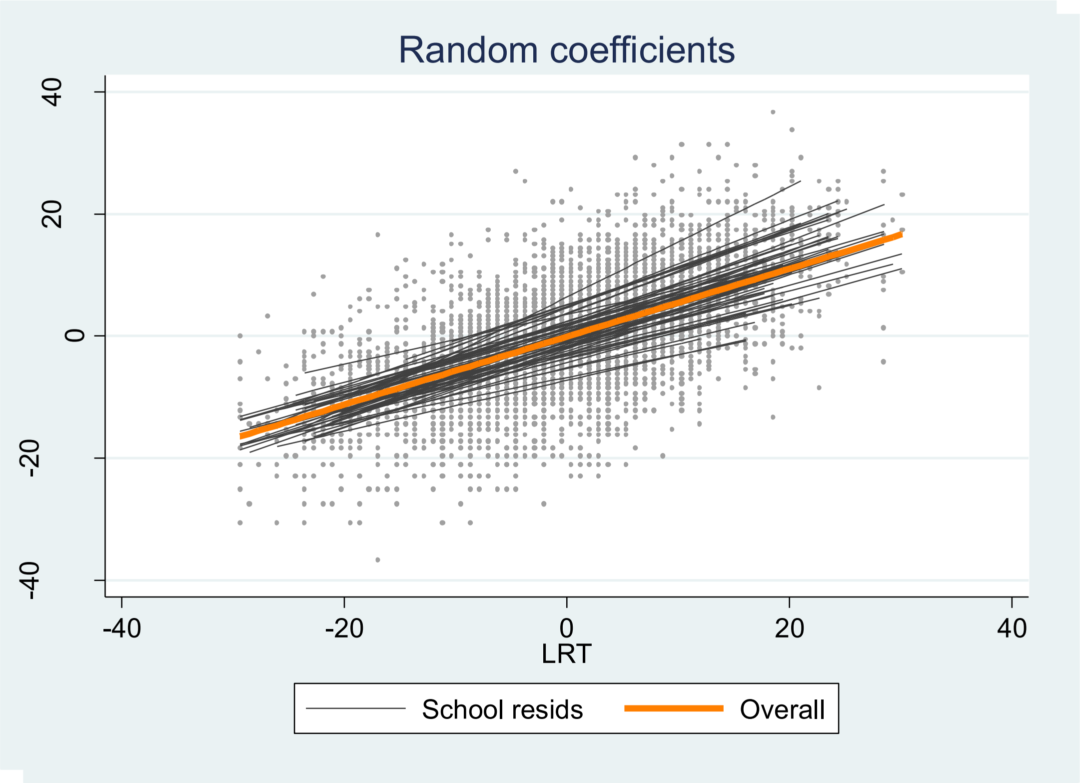
(2c) Multilevel models with more than two levels

(2d) Multilevel models with non-linear outcome variables

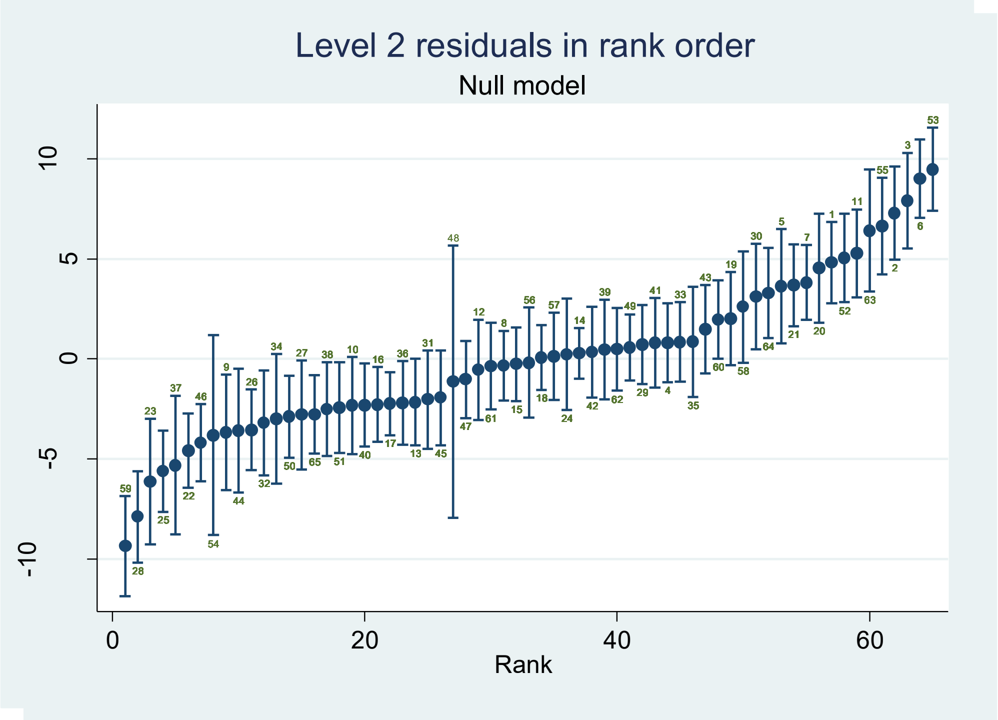
(2e) Realistic complexity in software tools

2a) Parameters for random intercepts and slopes

- *'Realistic complexity' refers to the compelling ideal of parsimony of outputs in context of an appropriately ambitious analysis*



Model-based predicted slopes: Quite a complicated model system, but summarised by only a handful of parameters
(above, that's β_0 and β_1 plus $\sigma_{\mu 0j}$ and $\sigma_{\mu 1j}$)



The 'caterpillar plot' for higher level residuals: Quite a complicated set of results, but summarised succinctly, coherently and/or instrumentally

2a) Parameters for random intercepts and slopes

…random effects models estimate extra parameters to represent decomposition of error variance into parts associated with higher and lower level units…

- Linear model:

$$Y_i = \beta X_i + \varepsilon_i$$

- Multilevel model ('random intercepts') ('mixed model')

$$Y_{ij} = \beta X_{ij} + \mu_j + \varepsilon_{ij}$$

- 'Random intercepts' models allow just a single set difference in the errors for each cluster, characterised by a variance estimate for the variation within them, i.e. $var(\mu_j)$ or $\sigma^2_{\mu j}$ or, as an SD, $\sigma_{\mu j}$

The concept ('random intercept')

The parameter (variance characterising the spread of the random intercept)

In general purpose packages, there are usually extension specifications which can be made in order to allow for and estimate parameters for random intercepts and slopes, e.g:

```
regress ghq fem age age2 cohab
mixed ghq fem age age2 cohab ||ohid:,
```

2a) Parameters for random intercepts and slopes

Stata format implementation of a random effects random intercepts multilevel model.

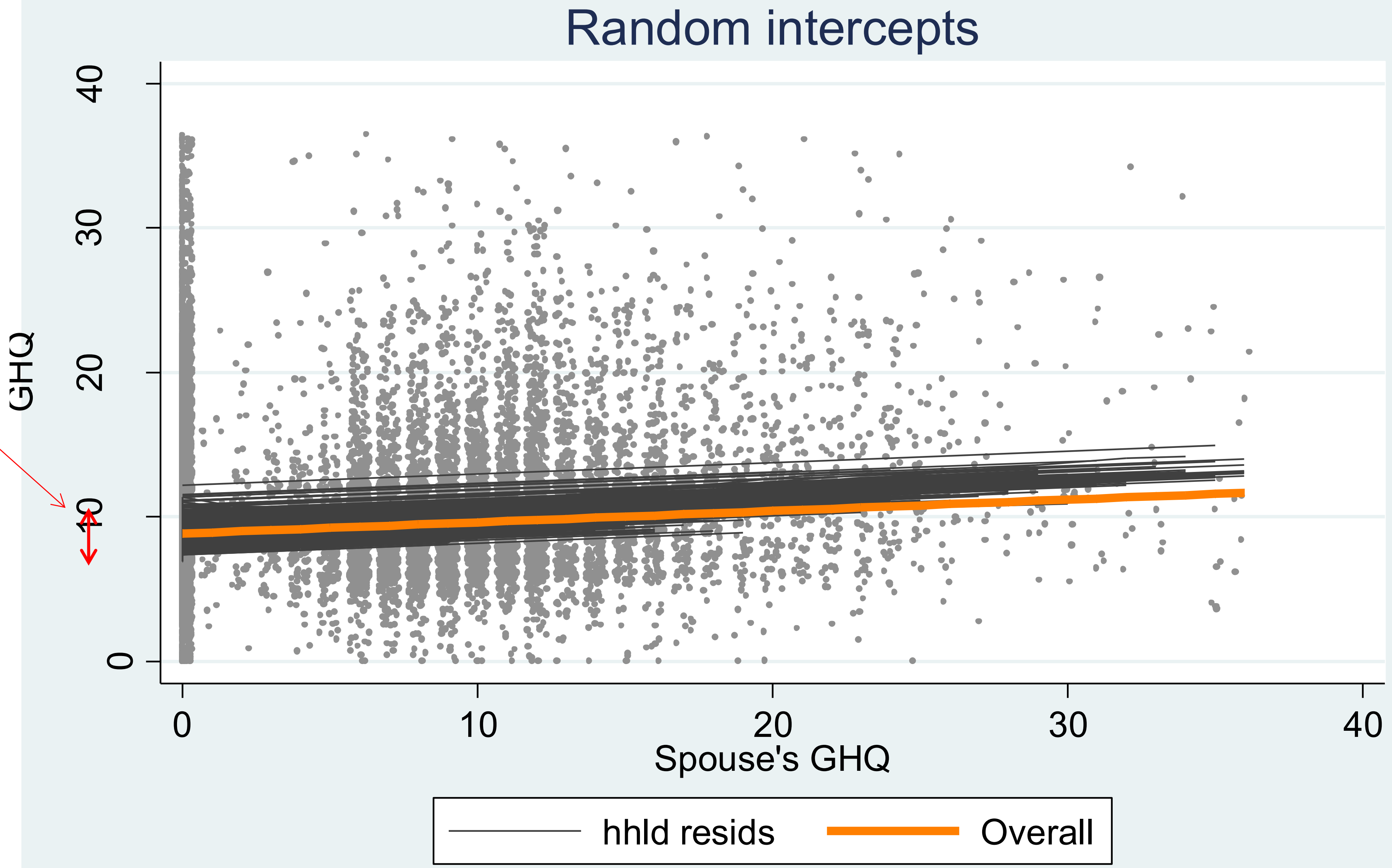
Here, the data comes from wave 15 of the UK's BHPS, the outcome is GHQ (score indicative of poor subjective well-being) with a model allowing for numerous individual and household level explanatory variables plus an additional random intercept term to estimate residual patterns in the model errors that are associated with clustering of 13.5k records within 8k households

```
. mixed ghq fem age age2 cohab emp_10hrs spghq2 gdn lnkids ||ohid: if miss1==0, mle stddev

Performing EM optimization ...

Performing gradient-based optimization:
Iteration 0:   log likelihood = -41846.795
Iteration 1:   log likelihood =  -41772.15
Iteration 2:   log likelihood = -41772.144
Iteration 3:   log likelihood = -41772.144

Computing standard errors ...

Mixed-effects ML regression                     Number of obs     =     13,496
Group variable: ohid                            Number of groups  =      7,857
                                                Obs per group:
                                                              min =          1
                                                              avg =        1.7
                                                              max =          7
                                                Wald chi2(8)      =     687.24
Log likelihood = -41772.144                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         ghq | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         fem |    1.30026   .0931852    13.95   0.000     1.117621      1.4829
         age |   .1848677   .0146967    12.58   0.000     .1560627    .2136728
        age2 |  -.0018707   .0001521   -12.30   0.000    -.0021688   -.0015726
       cohab |  -1.536452   .1461732   -10.51   0.000    -1.822947   -1.249958
   emp_10hrs |  -1.610601   .1145078   -14.07   0.000    -1.835032    -1.38617
      spghq2 |   .0778965   .0098672     7.89   0.000     .0585572    .0972358
         gdn |   -.557334   .1495786    -3.73   0.000    -.8505026   -.2641653
      lnkids |   .3914777   .1108057     3.53   0.000     .1743025    .6086529
       _cons |   8.496411   .3252431    26.12   0.000     7.858946    9.133876
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
ohid: Identity               |
                   sd(_cons) |   1.471419   .2137471      1.106842    1.956081
-----------------------------+------------------------------------------------
                sd(Residual) |   5.146973   .0644899      5.022115    5.274936
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 13.56          Prob >= chibar2 = 0.0001
```

Parameter for the random intercept: standard deviation for cluster-to-cluster variation in the intercepts ($\sigma_{\mu j}$)

2a) Parameters for random intercepts and slopes

(The RI standard deviation parameter is telling us about the distribution of parallel slopes around the intercept)



Souce: BHPS, 2005, adults with and without partners.
Regression line for male, age 38, average/modal other characteristics.

2a) Parameters for random intercepts and slopes

Total variance $= \sigma_u^2 + \sigma_\varepsilon^2 = (\text{sd\_cons})^2 + (\text{sdResidual})^2$
$= (1.471)^2 + (5.147)^2$
$= 2.16 + 26.49 = 28.65$

Intra-cluster correlation $= \rho$
$= \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2)$
$= 0.076$

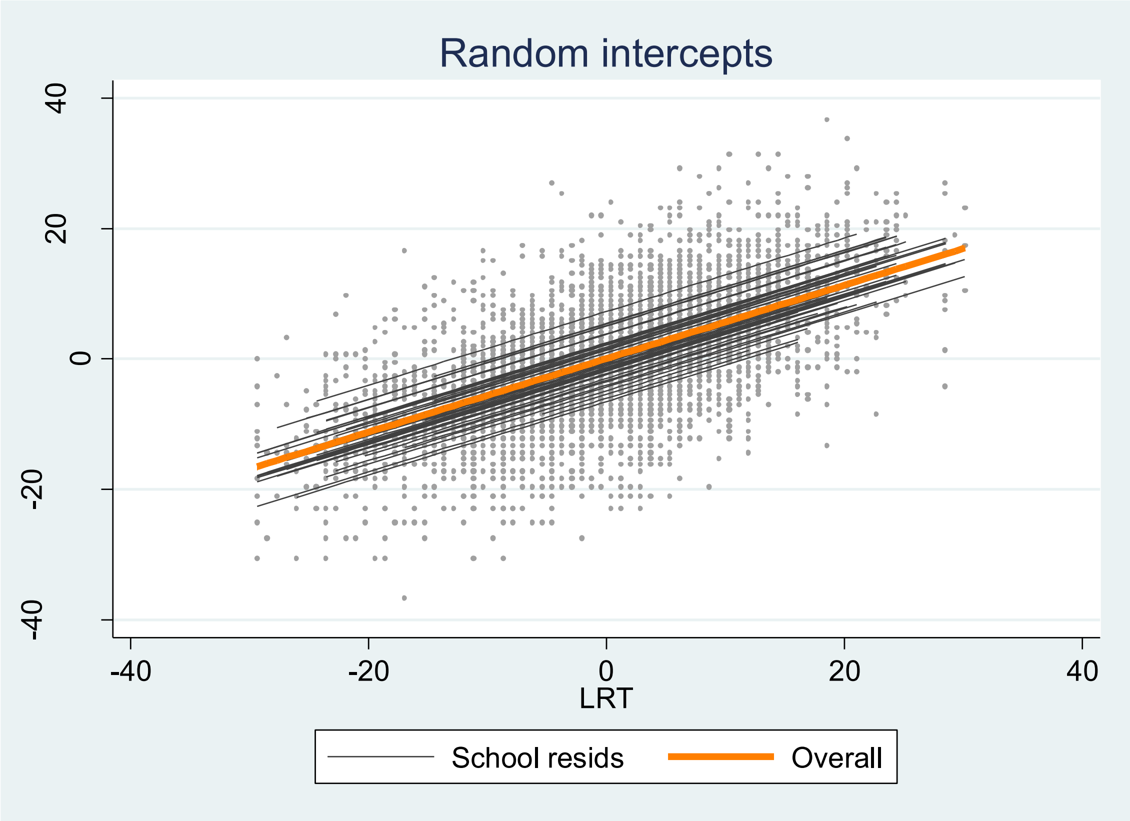
(ICC=proportion of variance associated with level 2)

Intra-cluster correlation as a summarising parameter (a Variance Partition Coefficient)

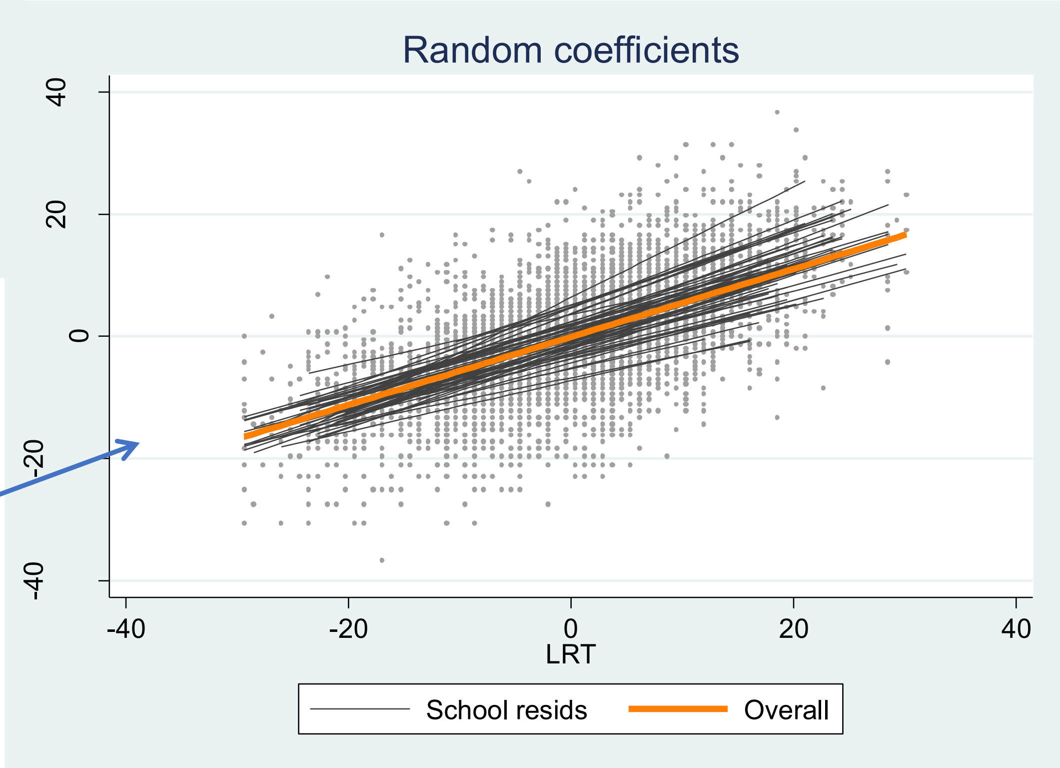
	null	+ coeff.s
Deviance	84185	83544
Level 2 var	4.51	2.16
Level 1 var	25.84	26.49
Total var	30.29	28.65
ICC	0.147	0.076

- ICC conventionally used to tell us about the source of the effects (e.g. is pupil attainment mainly influenced by the family, class or school?)
- Often, the ICC will decline when we add fixed part terms related to the higher level (since relatively less remains unexplained about the higher level)
- *Correlation interpretation:* ICC is the average correlation between two cases from the same cluster
- *{Design effect interpretation: ICC is a function of the difference between actual and effective sample size}*
- *{ICC applies to the 'random intercepts' model, and doesn't have a simple interpretation for the 'random coefficients' model}*

2a) Parameters for random intercepts and slopes



‘random intercepts’ (the line for each cluster can shift up and down but not change angle)



‘random slopes’ (the slope of the regression line for each cluster can differ)

The ‘Random Coefficients’ or ‘Random slopes’ multilevel model

Fixed part *Random part*

$$Y_{ij} = \beta_p X_{pij} + \mu_{pj} X_{pij} + \varepsilon_{pij} X_{pij}$$ [hypothetical]

$$Y_{ij} = \beta_0 X_0 + \beta_1 X_{1ij} + \beta_2 X_{2j} + \mu_{1j} X_{1ij} + \mu_{0j} X_0 + \varepsilon_{0ij} X_0$$ [typical]

‘Random slopes’ = Error is conceived to be a function of one or more X variables, other than the intercept, then parameterized by a variance estimate at cluster and/or lower levels.

- Random slopes at higher levels: Allowing for possibility that the impact of explanatory variables on the outcome varies from cluster to cluster
- Realistically, only helpful for lower level explanatory variables (e.g. Goldstein 2003: 65)
- Hard to estimate/disentangle numerous random slopes [most applications only use a few]
- Can have level 1 random slopes (modelling heteroscedasticity) [but relatively uncommon]
- Sometimes useful to parameterize the correlation between different random slopes and intercepts

2a) Parameters for random intercepts and slopes

- This model has both a random intercept and a random slope (on 'lnkids')

- There could be household-to-household variation in both
  - Slope: 'lnkids' coefficient ranges approx. 0.0-0.7 from household to household
  - Intercept: '_cons' ranges approx. 10-16 from household to household

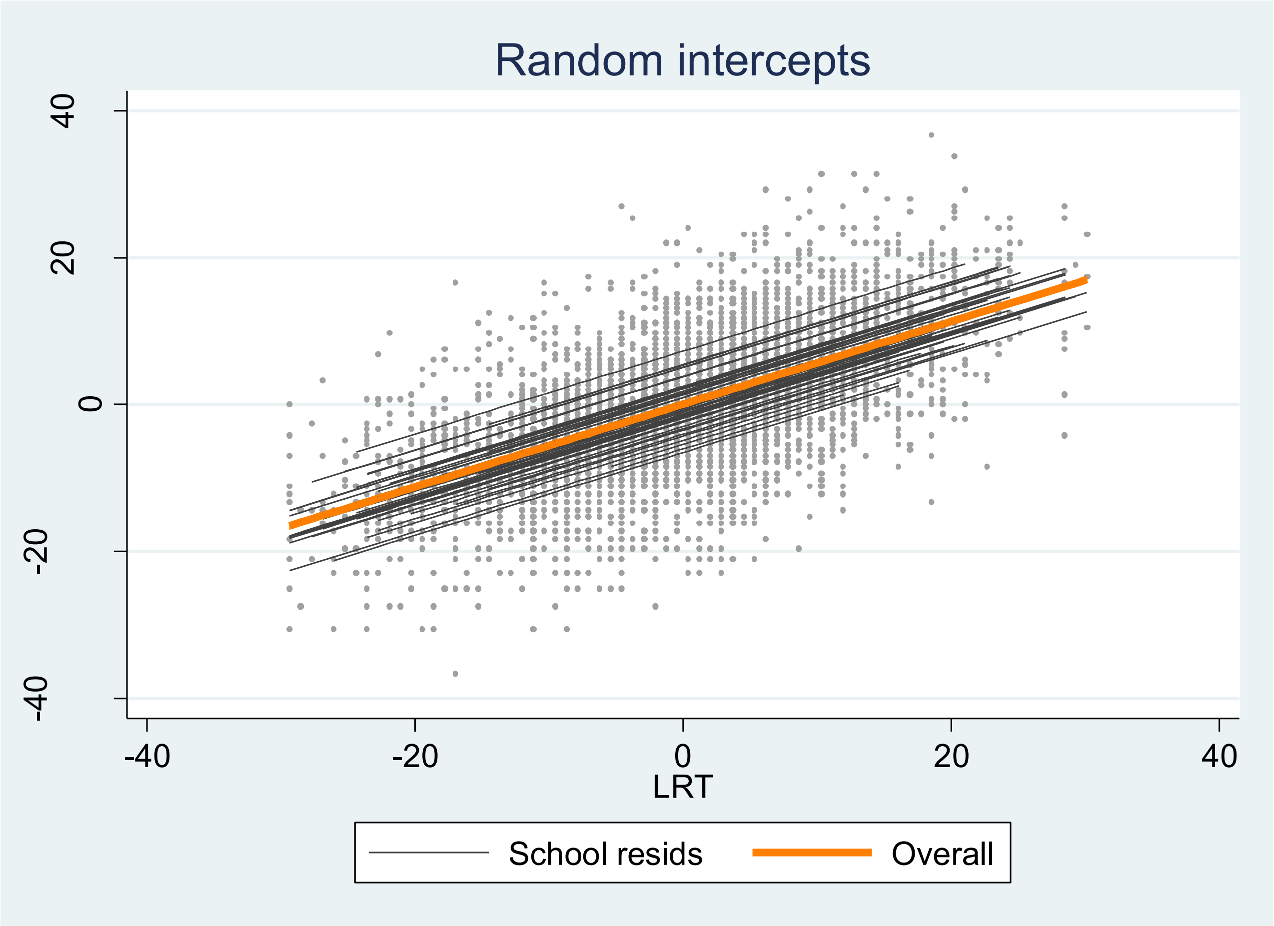
- Subtle complications
  - Here, the random slopes parameter hasn't improved model fit compared to earlier random intercepts (negligible change in Log likelihood -> we'd usually favour RI model)
  - I also imposed a non-standard constraint (cov(u_1, u_0)=0) to make the model identification stronger (with 'cov(ind)', evident in 'ohid: independent')

```
. mixed ghq fem c_age c_age2 cohab emp_10hrs spghq2 gdn lnkids ///
>         ||ohid:lnkids if miss1==0, mle stddev cov(ind)

Performing EM optimization ...

Performing gradient-based optimization:
Iteration 0:   log likelihood = -42111.623
Iteration 1:   log likelihood = -41776.667
Iteration 2:   log likelihood = -41772.144
Iteration 3:   log likelihood = -41772.142

Computing standard errors ...

Mixed-effects ML regression                     Number of obs     =     13,496
Group variable: ohid                            Number of groups  =      7,857
                                                Obs per group:
                                                              min =          1
                                                              avg =        1.7
                                                              max =          7
                                                Wald chi2(8)      =     686.92
Log likelihood = -41772.142                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
         ghq | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         fem |   1.300032   .0931782    13.95   0.000     1.117406    1.482658
       c_age |   .1848736   .0146956    12.58   0.000     .1560706    .2136765
      c_age2 |  -.0018707   .0001521   -12.30   0.000    -.0021688   -.0015726
       cohab |  -1.534827   .1461712   -10.50   0.000    -1.821317   -1.248337
   emp_10hrs |   -1.61033   .1145099   -14.06   0.000    -1.834766   -1.385895
      spghq2 |   .0777487   .0098676     7.88   0.000     .0584085    .0970889
         gdn |  -.5573813   .1495716    -3.73   0.000    -.8505363   -.2642264
      lnkids |   .3917099   .1108789     3.53   0.000     .1743913    .6090284
       _cons |   12.89781    .194322    66.37   0.000     12.51694    13.27867
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
ohid: Independent            |
                  sd(lnkids) |   .1743521   .0226385      .1351774    .2248796
                   sd(_cons) |   1.469373   .2141252      1.104308    1.955122
-----------------------------+------------------------------------------------
               sd(Residual) |   5.146578   .0645077      5.021686    5.274577
------------------------------------------------------------------------------
```

2b) Higher level residuals

The group level lines are calculated from the group level residuals, they are not estimated in the model outputs themselves.

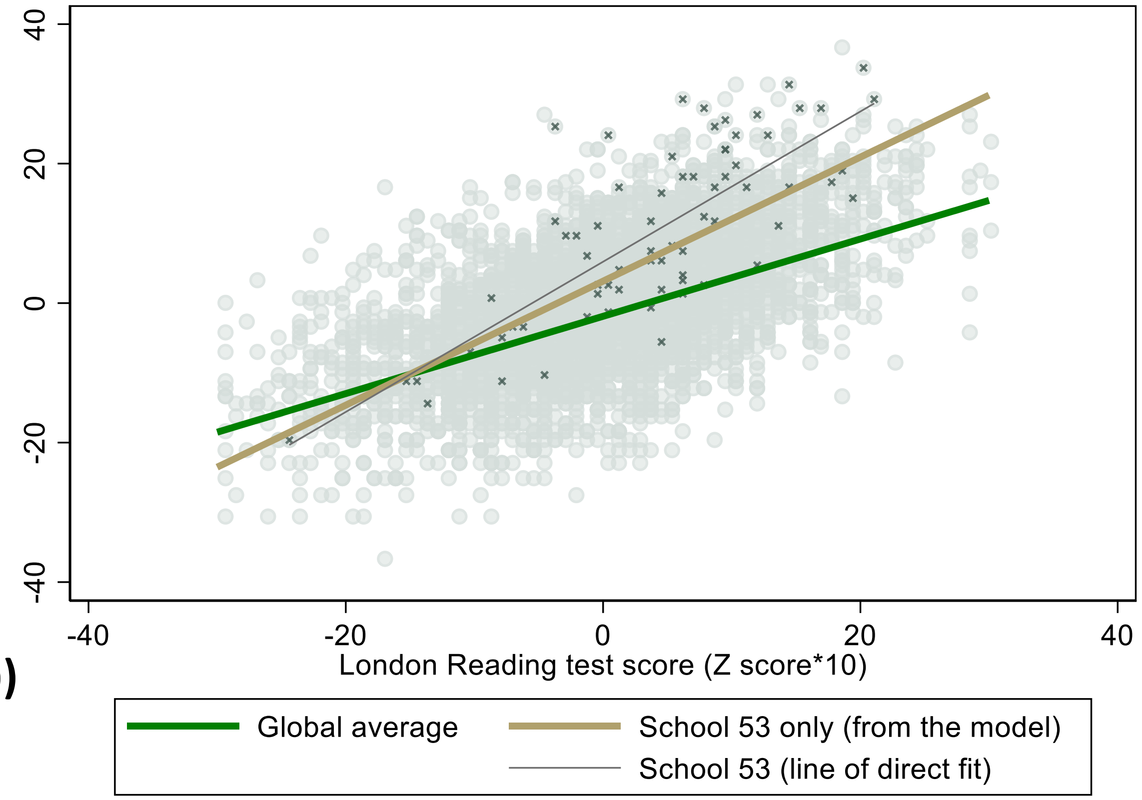
..importantly, they are not wholly unconstrained, but are 'shrunk' according to the overall pattern..



- Level 2 residuals are error patterns at the group level and are substantively informative
- Level 1 residuals are the individual error terms and are usually only used to check model assumptions

2b) Higher level residuals

'Empirical Bayes' residuals..?



The **'Empirical Bayes' residuals** for cluster-based random effects are a 'shrunk' adjustment to the arithmetic residuals for the cluster, with 'shrinkage' towards the population level pattern

- $\mathbf{EB(\mu_j) = \lambda_j \mu_j}$, where $\boldsymbol{\lambda_j = (\sigma^2_\mu / (\sigma^2_\mu + (\sigma^2_\varepsilon/n_j)))}$
- $\mathbf{EB(\beta_{pj}) = \lambda_j \beta_{pj} + (1 - \lambda_j)\beta_{p.}}$ (where 'β_{pj}' represents $\beta_p + \mu_{pj}$)

- 'Shrinkage factor' or 'reliability' $\boldsymbol{\lambda_j}$ deflates the impact of cluster specific μ_j when n_j (cluster size) is smaller, and impacts much less when n_j gets larger
- 'Shrinkage' means that EB(μ_j) is generally a more compelling estimate for the net distinctiveness of cluster j than is its arithmetic value μ_j
- *Research in practice*

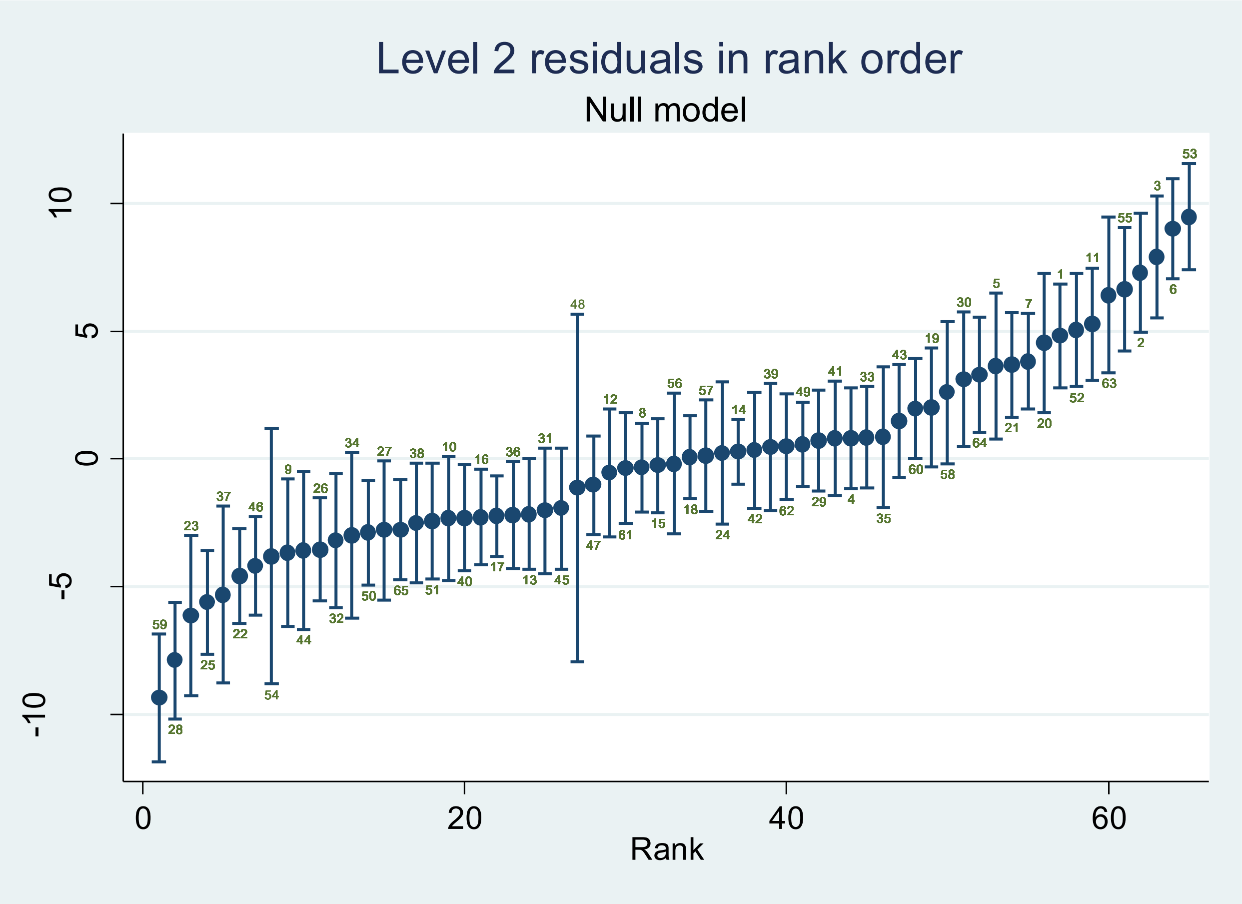
'Fixed effects residuals' (the arithmetic average perturbations based on the cases representing the cluster) are often reported when Empirical Bayes random effects residuals would arguably be more appropriate

2b) Higher level residuals

Caterpillars

- The caterpillar plot works out the higher level residuals and their standard errors, and ranks them by means (usually with influential ‘shrinkage’)
- Can identify general patterns and/or extreme cases (often model extreme cases explicitly by a single dummy variable)

This graph is the GCSE attainment by schools dataset, and is generated in lab



2c) Multilevel models with more than two levels

E.g. of three level hierarchical systems:

Regions (Level 3)	k	**1**										**2**			**3**	**.**
Households (Level 2)	j	**1**			**2**		**3**	**4**				**5**	**6**		**7**	**.**
Individuals (Level 1)	i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**.**

A random intercepts model could be…

$$Y_{ijk} = \beta_p X_{pijk} + \beta_q Z_{qjk} + \beta_r A_{rk} + \quad v_{0k} + \mu_{0jk} + \varepsilon_{0ijk}$$

A random slopes example could be….

$$Y_{ijk} = \beta_p X_{pijk} + \beta_q Z_{qjk} + \beta_r A_{rk} + \quad v_{1k} X_{1ijk} + \mu_{1jk} X_{1ijk} + v_{0k} + \mu_{0jk} + \varepsilon_{0ijk}$$

Extensions to 3 or more levels are 'easy' when the levels are hierarchical (at least in terms of conceptualisation, technical formulation and interpretation)

However, model identification is often difficult…

…Identification challenges, in practice, mean that most applied studies with multilevel models using random effects still only use two levels

2c) Multilevel models with more than two levels

Example for the 3-level model…

- Random part of model can now have components at three levels (**v**, **u** and **ε**)

Total error variance (random intercepts) $= \sigma_v^2 + \sigma_u^2 + \sigma_\varepsilon^2$
$= (\text{sd\_cons3})^2 + (\text{sd\_cons2})^2 + (\text{sd\_cons1})^2$

Variance Partition Coefficients/Intra-Cluster Correlations:

Level 3 VPC/ICC $= \sigma_v^2 \,/\, (\sigma_v^2 + \sigma_u^2 + \sigma_\varepsilon^2)$

Level 2 VPC $= \sigma_u^2 \,/\, (\sigma_v^2 + \sigma_u^2 + \sigma_\varepsilon^2)$
Level 2 ICC $= (\sigma_u^2 + \sigma_v^2) \,/\, (\sigma_v^2 + \sigma_u^2 + \sigma_\varepsilon^2)$

- depends upon desired interpretation, e.g RH&S (2022, 8.5)
(VPC=proportion of variance at the level; ICC = similarity of units within clusters)

Modelling 3+ levels works best if the levels relate to different things, have plenty of units/cases, and when only relatively few 'random slopes' are allowed for

More complex level structures include **'cross-classified'** models and **'multiple membership models'**,

- ➢ *random effects can be modelled but need specialist routines and are harder to estimate and to interpret*

In practice, studies often model fewer levels than could exist in theory, and often substitute random with fixed effects for supplementary levels

2d) Multilevel models with non-linear outcome variables

(e.g. Rabe-Hesketh and Skrondal 2022, vol 2; Heck et al. 2012; Hox et al. 2017)

Non-linear outcomes include

- **Binary outcomes**
- **Multinomial outcomes**
- **Ordered categories**
- **Counts**
- **Durations**

A popular take-home is that we (now) have modelling tools that…

- …let us explore how explanatory variables link to non-linear outcomes in a comparable way to a linear outcomes model….
- …can (now) adjust to add multilevel random effects….
- …give basic interpretations that work the same way as for linear outcome models (e.g. the sign and significance of coefficients, and relative explanatory power of a model)

However…

Non-linear outcomes + Random effects	=	*Trouble, lots of the time…*

- Estimation procedures are much more impactful and unstable
- Detecting the trace of empirical patterns is harder
- Interpretation of coefficients is harder
- Estimates about the multilevel structure are harder to summarise

2d) Multilevel models with non-linear outcome variables

- *For non-linear outcomes, the statistical trick is the 'Generalised Linear Model' (e.g. McCullagh and Nelder 1989) which allows us to model something that represents the outcome in a format comparable to a linear response*

e.g….

- Logistic regression
- Probit regression
- Ordered logistic regression
- Multinomial logistic regression
- Poisson regression
- Parametric survival models

Linear functional form: can model the exact value of Y:

$\mathbf{Y_i = \beta X_i + \varepsilon_i}$ (ε_i has a normal distribution)

Non-linear functional form: model for a linear response that can be connected to Y (e.g. the probability of categories of Y):

$\boldsymbol{\eta_i = \beta X_i + \varepsilon_i}$ (η_i is connected to probability of Y_i by a 'link function', and a probability model to describe the 'response distribution' $Y | \eta$)

Link function: $\eta_i = g(Y_i) = g(p(Y_i))$

Inverse link function: $p(Y_i) = Y_i = h(\eta_i) = g^{-1}(\eta_i)$

2d) Multilevel models with non-linear outcome variables

- *To then add multilevel models with random effects to non-linear outcomes models involved debates and new estimation tools, but ultimately comparable procedures are usually available…*

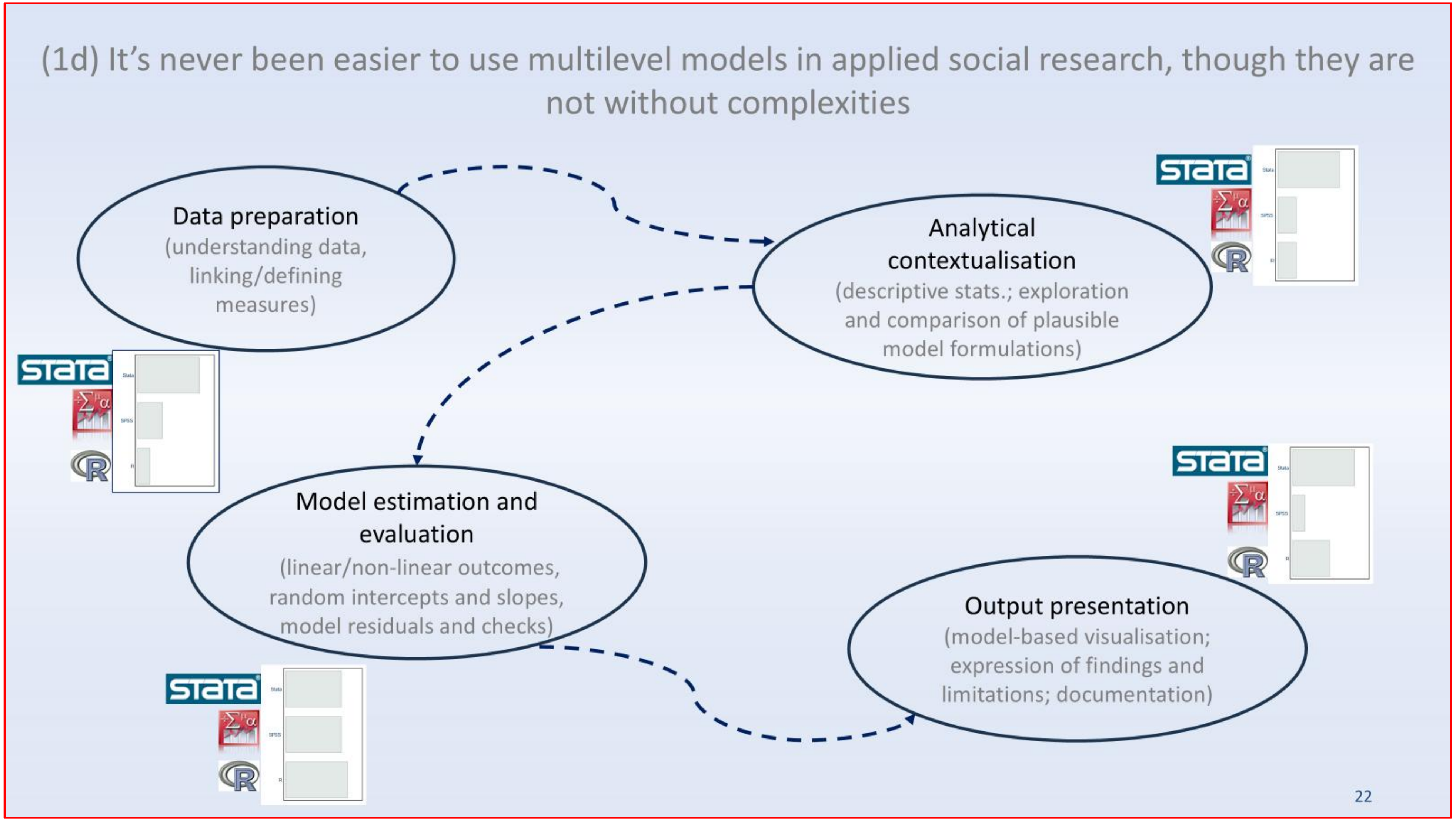
This model is a random slopes multilevel logistic regression (outcome: binary indicator of having poor health on Scottish census 2011 teaching dataset). It:

- Struggled to converge
- Gives comparable broad results to a linear model
- Gives beta parameters that are hard to interpret
- Gives random part parameters that are vulnerable to estimation (and compare to a fixed-by-design lower level variance component)

```
. melogit poor_health female c_age age_fem ||occ_ind_gp:agegpt,

Fitting fixed-effects model:

Iteration 0:   log likelihood = -14114.453
Iteration 1:   log likelihood = -12350.178
Iteration 2:   log likelihood = -12332.949
Iteration 3:   log likelihood = -12332.913
Iteration 4:   log likelihood = -12332.913

Refining starting values:

Grid node 0:   log likelihood = -12179.879

Fitting full model:

Iteration 0:   log likelihood = -12179.879  (not concave)
Iteration 1:   log likelihood =  -12164.17  (not concave)
Iteration 2:   log likelihood = -12119.908  (not concave)
Iteration 3:   log likelihood = -12063.372  (not concave)
Iteration 4:   log likelihood = -12034.437  (not concave)
Iteration 5:   log likelihood = -12004.586  (not concave)
Iteration 6:   log likelihood = -11991.311  (not concave)
Iteration 7:   log likelihood =  -11979.98
Iteration 8:   log likelihood = -11974.064  (backed up)
Iteration 9:   log likelihood = -11970.948
Iteration 10:  log likelihood = -11969.194
Iteration 11:  log likelihood = -11969.091
Iteration 12:  log likelihood = -11969.091

Mixed-effects logistic regression               Number of obs     =     63,388
Group variable: occ_ind_gp                      Number of groups  =        118

                                                Obs per group:
                                                              min =          3
                                                              avg =      537.2
                                                              max =     14,435

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(3)      =     967.11
Log likelihood = -11969.091                     Prob > chi2       =     0.0000

  poor_health | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]

       female |   .1282234     .05366     2.39   0.017     .0230517    .2333951
        c_age |   .0484283   .0017969    26.95   0.000     .0449064    .0519503
      age_fem |  -.0066825   .0019679    -3.40   0.001    -.0105395   -.0028254
        _cons |  -3.543554   .0693899   -51.07   0.000    -3.679556   -3.407552

occ_ind_gp
  var(agegpt) |   .0027651   .0009485                      .0014117    .0054162
   var(_cons) |   .2134169   .0551743                      .1285787    .3542326
```

2e) Realistic complexity in software tools

- *General purpose packages often support features that make adding 'realistic complexity' via multilevel models quite easy*



2e) Realistic complexity in software tools

Some examples in Stata, SPSS and R… [Referring to the GCSE dataset that features in lab examples] [in all packages, there is more than one way to specify the same models]	
[Stata]: regress gcse lrt girl sch_2 sch_3 [SPSS]: regression var=gcse lrt girl sch_2 sch_3 /dependent=gcse /method=enter . [R]: m1 <- with(gcse_data, lm(formula = gcse ~ lrt + girl + sch_2 + sch_3))	**A single level linear outcomes regression model**
[Stata]: mixed gcse lrt girl sch_2 sch_3 ||school:, stddev [SPSS]: mixed gcse with lrt girl sch_2 sch_3 /fixed lrt girl sch_2 sch_3 |sstype(3) /method=reml /print=corb solution r /random=intercept | subject(school) covtype(vc) . [R]: m2 <- lmer(formula = gcse ~ lrt + girl+ sch_2 + sch_3 + (1 | school), data = gcse_data)	**Adding 'random intercepts' for clustering by 'school'**
[Stata]: mixed gcse lrt girl sch_2 sch_3 ||school:lrt, stddev cov(un) [SPSS]: mixed gcse with lrt girl sch_2 sch_3 /fixed lrt girl sch_2 sch_3 |sstype(3) /method=reml /print=corb solution r /random=intercept lrt | subject(school) covtype(un) . [R]: m3 <- lmer(formula = gcse ~ lrt + girl + sch_2 + sch_3 + (1 + lrt | school), data = gcse_data)	**Adding one random slope, for the role of the 'lrt' explanatory variable**
[Stata]: melogit hi_gcse lrt girl sch_2 sch_3 ||school:lrt, intp(7) [SPSS]: genlinmixed /data_structure subjects=school /fields target=hi_gcse (reference=0) /target_options distribution=binomial link=logit /fixed effects= lrt girl sch_2 sch_3 use_intercept=true /random effects=lrt use_intercept=true subjects=school covariance_type=unstructured. [R]: m4 <- glmer(gcse ~ lrt + girl + scg_2 + sch_3 + (1 + lrt | school), family = binomial(link="logit"), data = gcse_data)	**A binary outcomes version of the random slopes model**

2e) Realistic complexity in software tools

<table>
<tr><th colspan="2">Some examples in Stata, SPSS and R… [Referring to the GCSE dataset that features in lab examples]
[in all packages, there is more than one way to specify the same models]</th></tr>
<tr><td>[Stata]: regress gcse lrt girl sch_2 sch_3
[SPSS]: regression var=gcse lrt girl sch_2 sch_3 /dependent=gcse /method=enter .
[R]: m1 <- with(gcse_data, lm(formula = gcse ~ lrt + girl + sch_2 + sch_3))</td><td>A single level linear outcomes regression model</td></tr>
<tr><td>[Stata]: mixed gcse lrt girl sch_2 sch_3 ||school:, stddev
[SPSS]: mixed gcse with lrt girl sch_2 sch_3 /fixed lrt girl sch_2 sch_3 |sstype(3)
/method=reml /print=corb solution r /random=intercept | subject(school) covtype(vc) .
[R]: m2 <- lmer(formula = gcse ~ lrt + girl+ sch_2 + sch_3 + (1 | school), data = gcse_data)</td><td>Adding ‘random intercepts’ for clustering by ‘school’</td></tr>
<tr><td>[Stata]: mixed gcse lrt girl sch_2 sch_3 ||school:lrt, stddev cov(un)
[SPSS]: mixed gcse with lrt girl sch_2 sch_3 /fixed lrt girl sch_2 sch_3 |sstype(3)
/method=reml /print=corb solution r /random=intercept lrt |subject(school) covtype(un).
[R]: m3 <- lmer(formula = gcse ~ lrt + girl + sch_2 + sch_3 + (1 + lrt | school), data = gcse_data)</td><td>Adding one random slope, for the role of the ‘lrt’ explanatory variable</td></tr>
<tr><td>[Stata]: melogit hi_gcse lrt girl sch_2 sch_3 ||school:lrt, intp(7)
[SPSS]: genlinmixed /data_structure subjects=school /fields target=hi_gcse (reference=0)
/target_options distribution=binomial link=logit /fixed effects= lrt girl sch_2 sch_3
use_intercept=true /random effects=lrt use_intercept=true subjects=school
covariance_type=unstructured .
[R]: m4 <- glmer(gcse ~ lrt + girl + scg_2 + sch_3 + (1 + lrt | school),
family = binomial(link="logit"), data = gcse_data)</td><td>A binary outcomes version of the random slopes model</td></tr>
</table>

2e) Realistic complexity in software tools

Some examples in Stata, SPSS and R… [Referring to the GCSE dataset that features in lab examples] [in all packages, there is more than one way to specify the same models]	
[Stata]: regress gcse lrt girl sch_2 sch_3 [SPSS]: regression var=gcse lrt girl sch_2 sch_3 /dependent=gcse /method=enter . **[R]: m1 <- with(gcse_data, lm(formula = gcse ~ lrt + girl + sch_2 + sch_3))**	**A single level linear outcomes regression model**
[Stata]: mixed gcse lrt girl sch_2 sch_3 ||school:, stddev [SPSS]: mixed gcse with lrt girl sch_2 sch_3 /fixed lrt girl sch_2 sch_3 |sstype(3) /method=reml /print=corb solution r /random=intercept | subject(school) covtype(vc) . **[R]: m2 <- lmer(formula = gcse ~ lrt + girl+ sch_2 + sch_3 + (1 | school), data = gcse_data)**	**Adding 'random intercepts' for clustering by 'school'**
[Stata]: mixed gcse lrt girl sch_2 sch_3 ||school:lrt, stddev cov(un) [SPSS]: mixed gcse with lrt girl sch_2 sch_3 /fixed lrt girl sch_2 sch_3 |sstype(3) /method=reml /print=corb solution r /random=intercept lrt | subject(school) covtype(un) . **[R]: m3 <- lmer(formula = gcse ~ lrt + girl + sch_2 + sch_3 + (1 + lrt | school), data = gcse_data)**	**Adding one random slope, for the role of the 'lrt' explanatory variable**
[Stata]: melogit hi_gcse lrt girl sch_2 sch_3 ||school:lrt, intp(7) [SPSS]: genlinmixed /data_structure subjects=school /fields target=hi_gcse (reference=0) /target_options distribution=binomial link=logit /fixed effects= lrt girl sch_2 sch_3 use_intercept=true /random effects=lrt use_intercept=true subjects=school covariance_type=unstructured. **[R]: m4 <- glmer(gcse ~ lrt + girl + scg_2 + sch_3 + (1 + lrt | school), family = binomial(link="logit"), data = gcse_data)**	**A binary outcomes version of the random slopes model**

Summary: Realistic complexity

Parameters for random slopes and intercepts

Mulitlevel models with more than two levels

Higher level residuals

Multilevel models with non-linear outcome variables

Software tools

References cited

- Goldstein, H. (2003). *Multilevel Statistical Models, Third Edition*. Arnold.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2012). *Multilevel Modeling of Categorical Outcomes Using IBM SPSS* Routledge.
- Hox, J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications (Third Edition)*. Routledge.
- McCullagh, P., & Nelder, J. (1989). *Generalised Linear Models*. Chapman and Hall.
- Rabe-Hesketh, S., & Skrondal, A. (2022). *Multilevel and Longitudinal Modeling Using Stata, Fourth edition (two-volume set)*. Stata Press.